CFA Institute®
CFA Program

2022 CFA® PROGRAM CURRICULUM LEVEL II VOLUMES 1–6

© 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006 by CFA Institute. All rights reserved.

This copyright covers material written expressly for this volume by the editor/s as well as the compilation itself. It does not cover the individual selections herein that first appeared elsewhere. Permission to reprint these has been obtained by CFA Institute for this edition only. Further reproductions by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval systems, must be arranged with the individual copyright holders noted.

CFA®, Chartered Financial Analyst®, AIMR-PPS®, and GIPS® are just a few of the trademarks owned by CFA Institute. To view a list of CFA Institute trademarks and the Guide for Use of CFA Institute Marks, please visit our website at www.cfainstitute.org.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the services of a competent professional should be sought.

All trademarks, service marks, registered trademarks, and registered service marks are the property of their respective owners and are used herein for identification purposes only.

ISBN 978-1-950157-61-7 (paper)
ISBN 978-1-119705-03-1 (ebk)

10 9 8 7 6 5 4 3 2 1

Please visit our website at
www.WileyGlobalFinance.com.

WILEY

CFA Institute®
CFA Program

QUANTITATIVE METHODS AND ECONOMICS

CFA® Program Curriculum
2022 • LEVEL II • VOLUME 1

WILEY

CONTENTS

O indicates an optional segment

How to Use the CFA Program Curriculum

Congratulations on your decision to enter the Chartered Financial Analyst (CFA®) Program. This exciting and rewarding program of study reflects your desire to become a serious investment professional. You are embarking on a program noted for its high ethical standards and the breadth of knowledge, skills, and abilities (competencies) it develops. Your commitment should be educationally and professionally rewarding.

The credential you seek is respected around the world as a mark of accomplishment and dedication. Each level of the program represents a distinct achievement in professional development. Successful completion of the program is rewarded with membership in a prestigious global community of investment professionals. CFA charterholders are dedicated to life-long learning and maintaining currency with the ever-changing dynamics of a challenging profession. CFA Program enrollment represents the first step toward a career-long commitment to professional education.

The CFA exam measures your mastery of the core knowledge, skills, and abilities required to succeed as an investment professional. These core competencies are the basis for the Candidate Body of Knowledge (CBOK™). The CBOK consists of four components:

- A broad outline that lists the major CFA Program topic areas (www.cfainstitute.org/programs/cfa/curriculum/cbok);
- Topic area weights that indicate the relative exam weightings of the top-level topic areas (www.cfainstitute.org/programs/cfa/curriculum);
- Learning outcome statements (LOS) that advise candidates about the specific knowledge, skills, and abilities they should acquire from readings covering a topic area (LOS are provided in candidate study sessions and at the beginning of each reading); and
- CFA Program curriculum that candidates receive upon exam registration.

Therefore, the key to your success on the CFA exams is studying and understanding the CBOK. The following sections provide background on the CBOK, the organization of the curriculum, features of the curriculum, and tips for designing an effective personal study program.

BACKGROUND ON THE CBOK

CFA Program is grounded in the practice of the investment profession. CFA Institute performs a continuous practice analysis with investment professionals around the world to determine the competencies that are relevant to the profession, beginning with the Global Body of Investment Knowledge (GBIK®). Regional expert panels and targeted surveys are conducted annually to verify and reinforce the continuous feedback about the GBIK. The practice analysis process ultimately defines the CBOK. The CBOK reflects the competencies that are generally accepted and applied by investment professionals. These competencies are used in practice in a generalist context and are expected to be demonstrated by a recently qualified CFA charterholder.

© 2021 CFA Institute. All rights reserved.

The CFA Institute staff—in conjunction with the Education Advisory Committee and Curriculum Level Advisors, who consist of practicing CFA charterholders—designs the CFA Program curriculum in order to deliver the CBOK to candidates. The exams, also written by CFA charterholders, are designed to allow you to demonstrate your mastery of the CBOK as set forth in the CFA Program curriculum. As you structure your personal study program, you should emphasize mastery of the CBOK and the practical application of that knowledge. For more information on the practice analysis, CBOK, and development of the CFA Program curriculum, please visit www.cfainstitute.org.

ORGANIZATION OF THE CURRICULUM

The Level II CFA Program curriculum is organized into 10 topic areas. Each topic area begins with a brief statement of the material and the depth of knowledge expected. It is then divided into one or more study sessions. These study sessions should form the basic structure of your reading and preparation. Each study session includes a statement of its structure and objective and is further divided into assigned readings. An outline illustrating the organization of these study sessions can be found at the front of each volume of the curriculum.

The readings are commissioned by CFA Institute and written by content experts, including investment professionals and university professors. Each reading includes LOS and the core material to be studied, often a combination of text, exhibits, and in-text examples and questions. End of Reading Questions (EORQs) followed by solutions help you understand and master the material. The LOS indicate what you should be able to accomplish after studying the material. The LOS, the core material, and the EORQs are dependent on each other, with the core material and EORQs providing context for understanding the scope of the LOS and enabling you to apply a principle or concept in a variety of scenarios.

The entire readings, including the EORQs, are the basis for all exam questions and are selected or developed specifically to teach the knowledge, skills, and abilities reflected in the CBOK.

You should use the LOS to guide and focus your study because each exam question is based on one or more LOS and the core material and practice problems associated with the LOS. As a candidate, you are responsible for the entirety of the required material in a study session.

We encourage you to review the information about the LOS on our website (www.cfainstitute.org/programs/cfa/curriculum/study-sessions), including the descriptions of LOS "command words" on the candidate resources page at www.cfainstitute.org.

FEATURES OF THE CURRICULUM

End of Reading Questions/Solutions *All End of Reading Questions (EORQs) as well as their solutions are part of the curriculum and are required material for the exam.* In addition to the in-text examples and questions, these EORQs help demonstrate practical applications and reinforce your understanding of the concepts presented. Some of these EORQs are adapted from past CFA exams and/or may serve as a basis for exam questions.

Glossary For your convenience, each volume includes a comprehensive Glossary. Throughout the curriculum, a **bolded** word in a reading denotes a term defined in the Glossary.

Note that the digital curriculum that is included in your exam registration fee is searchable for key words, including Glossary terms.

LOS Self-Check We have inserted checkboxes next to each LOS that you can use to track your progress in mastering the concepts in each reading.

Source Material The CFA Institute curriculum cites textbooks, journal articles, and other publications that provide additional context or information about topics covered in the readings. As a candidate, you are not responsible for familiarity with the original source materials cited in the curriculum.

Note that some readings may contain a web address or URL. The referenced sites were live at the time the reading was written or updated but may have been deactivated since then.

Some readings in the curriculum cite articles published in the *Financial Analysts Journal®*, which is the flagship publication of CFA Institute. Since its launch in 1945, the *Financial Analysts Journal* has established itself as the leading practitioner-oriented journal in the investment management community. Over the years, it has advanced the knowledge and understanding of the practice of investment management through the publication of peer-reviewed practitioner-relevant research from leading academics and practitioners. It has also featured thought-provoking opinion pieces that advance the common level of discourse within the investment management profession. Some of the most influential research in the area of investment management has appeared in the pages of the *Financial Analysts Journal*, and several Nobel laureates have contributed articles.

Candidates are not responsible for familiarity with *Financial Analysts Journal* articles that are cited in the curriculum. But, as your time and studies allow, we strongly encourage you to begin supplementing your understanding of key investment management issues by reading this, and other, CFA Institute practice-oriented publications through the Research & Analysis webpage (www.cfainstitute.org/en/research).

Errata The curriculum development process is rigorous and includes multiple rounds of reviews by content experts. Despite our efforts to produce a curriculum that is free of errors, there are times when we must make corrections. Curriculum errata are periodically updated and posted by exam level and test date online (www.cfainstitute.org/en/programs/submit-errata). If you believe you have found an error in the curriculum, you can submit your concerns through our curriculum errata reporting process found at the bottom of the Curriculum Errata webpage.

DESIGNING YOUR PERSONAL STUDY PROGRAM

Create a Schedule An orderly, systematic approach to exam preparation is critical. You should dedicate a consistent block of time every week to reading and studying. Complete all assigned readings and the associated problems and solutions in each study session. Review the LOS both before and after you study each reading to ensure that

you have mastered the applicable content and can demonstrate the knowledge, skills, and abilities described by the LOS and the assigned reading. Use the LOS self-check to track your progress and highlight areas of weakness for later review.

Successful candidates report an average of more than 300 hours preparing for each exam. Your preparation time will vary based on your prior education and experience, and you will probably spend more time on some study sessions than on others.

You should allow ample time for both in-depth study of all topic areas and additional concentration on those topic areas for which you feel the least prepared.

CFA INSTITUTE LEARNING ECOSYSTEM (LES)

As you prepare for your exam, we will email you important exam updates, testing policies, and study tips. Be sure to read these carefully.

Your exam registration fee includes access to the CFA Program Learning Ecosystem (LES). This digital learning platform provides access, even offline, to all of the readings and End of Reading Questions found in the print curriculum organized as a series of shorter online lessons with associated EORQs. This tool is your one-stop location for all study materials, including practice questions and mock exams.

The LES provides the following supplemental study tools:

Structured and Adaptive Study Plans The LES offers two ways to plan your study through the curriculum. The first is a structured plan that allows you to move through the material in the way that you feel best suits your learning. The second is an adaptive study plan based on the results of an assessment test that uses actual practice questions.

Regardless of your chosen study path, the LES tracks your level of proficiency in each topic area and presents you with a dashboard of where you stand in terms of proficiency so that you can allocate your study time efficiently.

Flashcards and Game Center The LES offers all the Glossary terms as Flashcards and tracks correct and incorrect answers. Flashcards can be filtered both by curriculum topic area and by action taken—for example, answered correctly, unanswered, and so on. These Flashcards provide a flexible way to study Glossary item definitions.

The Game Center provides several engaging ways to interact with the Flashcards in a game context. Each game tests your knowledge of the Glossary terms a in different way. Your results are scored and presented, along with a summary of candidates with high scores on the game, on your Dashboard.

Discussion Board The Discussion Board within the LES provides a way for you to interact with other candidates as you pursue your study plan. Discussions can happen at the level of individual lessons to raise questions about material in those lessons that you or other candidates can clarify or comment on. Discussions can also be posted at the level of topics or in the initial Welcome section to connect with other candidates in your area.

Practice Question Bank The LES offers access to a question bank of hundreds of practice questions that are in addition to the End of Reading Questions. These practice questions, only available on the LES, are intended to help you assess your mastery of individual topic areas as you progress through your studies. After each practice question, you will receive immediate feedback noting the correct response and indicating the relevant assigned reading so you can identify areas of weakness for further study.

Mock Exams The LES also includes access to three-hour Mock Exams that simulate the morning and afternoon sessions of the actual CFA exam. These Mock Exams are intended to be taken after you complete your study of the full curriculum and take practice questions so you can test your understanding of the curriculum and your readiness for the exam. If you take these Mock Exams within the LES, you will receive feedback afterward that notes the correct responses and indicates the relevant assigned readings so you can assess areas of weakness for further study. We recommend that you take Mock Exams during the final stages of your preparation for the actual CFA exam. For more information on the Mock Exams, please visit www.cfainstitute.org.

PREP PROVIDERS

You may choose to seek study support outside CFA Institute in the form of exam prep providers. After your CFA Program enrollment, you may receive numerous solicitations for exam prep courses and review materials. When considering a prep course, make sure the provider is committed to following the CFA Institute guidelines and high standards in its offerings.

Remember, however, that there are no shortcuts to success on the CFA exams; reading and studying the CFA Program curriculum *is* the key to success on the exam. The CFA Program exams reference only the CFA Institute assigned curriculum; no prep course or review course materials are consulted or referenced.

SUMMARY

Every question on the CFA exam is based on the content contained in the required readings and on one or more LOS. Frequently, an exam question is based on a specific example highlighted within a reading or on a specific practice problem and its solution. To make effective use of the CFA Program curriculum, please remember these key points:

1. All pages of the curriculum are required reading for the exam.
2. All questions, problems, and their solutions are part of the curriculum and are required study material for the exam. These questions are found at the end of the readings in the print versions of the curriculum. In the LES, these questions appear directly after the lesson with which they are associated. The LES provides immediate feedback on your answers and tracks your performance on these questions throughout your study.
3. We strongly encourage you to use the CFA Program Learning Ecosystem. In addition to providing access to all the curriculum material, including EORQs, in the form of shorter, focused lessons, the LES offers structured and adaptive study planning, a Discussion Board to communicate with other candidates, Flashcards, a Game Center for study activities, a test bank of practice questions, and online Mock Exams. Other supplemental study tools, such as eBook and PDF versions of the print curriculum, and additional candidate resources are available at www.cfainstitute.org.
4. Using the study planner, create a schedule and commit sufficient study time to cover the study sessions. You should also plan to review the materials, answer practice questions, and take Mock Exams.
5. Some of the concepts in the study sessions may be superseded by updated rulings and/or pronouncements issued after a reading was published. Candidates are expected to be familiar with the overall analytical framework contained in the assigned readings. Candidates are not responsible for changes that occur after the material was written.

FEEDBACK

At CFA Institute, we are committed to delivering a comprehensive and rigorous curriculum for the development of competent, ethically grounded investment professionals. We rely on candidate and investment professional comments and feedback as we work to improve the curriculum, supplemental study tools, and candidate resources.

Please send any comments or feedback to info@cfainstitute.org. You can be assured that we will review your suggestions carefully. Ongoing improvements in the curriculum will help you prepare for success on the upcoming exams and for a lifetime of learning as a serious investment professional.

Quantitative Methods

STUDY SESSION

Study Session 1	Quantitative Methods (1)
Study Session 2	Quantitative Methods (2)

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to explain regression and time series analysis and their uses in investment decision-making. The candidate should also be able to interpret the results and implications of a regression and time-series analysis in an investment context.

Quantitative methods such as regression and time series provide the means to identify and assess the relationships that exist between variables. Measuring the direction and strength of these relationships, with some level of confidence, can provide valuable insights for many investment-related activities. Moreover, large structured and unstructured datasets are now prevalent in investment management. Quantitative techniques for analyzing such datasets, via classification, simplification and clustering, based on machine learning, are presented with applications to investment management.

© 2021 CFA Institute. All rights reserved.

QUANTITATIVE METHODS

STUDY SESSION 1

Quantitative Methods (1)

This study session provides coverage on how linear regression and time-series analysis are used as tools in financial analysis for identifying relationships among variables. The session begins by examining simple linear regression with a single (independent) variable to explain or predict the value of another (dependent) variable. Multiple regression, using more than one independent variable to explain or predict a dependent variable, is explored next. Time-series analysis, in which the dependent variable's past values are included as independent variables, concludes the session.

READING ASSIGNMENTS

© 2021 CFA Institute. All rights reserved.

READING

1

Introduction to Linear Regression

by Pamela Peterson Drake, PhD, CFA

Pamela Peterson Drake, PhD, CFA, is at James Madison University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe a simple linear regression model and the roles of the dependent and independent variables in the model;
☐	**b.** describe the least squares criterion, how it is used to estimate regression coefficients, and their interpretation;
☐	**c.** explain the assumptions underlying the simple linear regression model, and describe how residuals and residual plots indicate if these assumptions may have been violated;
☐	**d.** calculate and interpret the coefficient of determination and the *F*-statistic in a simple linear regression;
☐	**e.** describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression;
☐	**f.** formulate a null and an alternative hypothesis about a population value of a regression coefficient, and determine whether the null hypothesis is rejected at a given level of significance;
☐	**g.** calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable;
☐	**h.** describe different functional forms of simple linear regressions.

SIMPLE LINEAR REGRESSION 1

a describe a simple linear regression model and the roles of the dependent and independent variables in the model

© 2021 CFA Institute. All rights reserved.

Financial analysts often need to examine whether a variable is useful for explaining another variable. For example, the analyst may want to know whether earnings growth, or perhaps cash flow growth, helps explain the company's value in the marketplace. **Regression analysis** is a tool for examining this type of issue.

Suppose an analyst is examining the return on assets (ROA) for an industry and observes the ROA for the six companies shown in Exhibit 1. The average of these ROAs is 12.5%, but the range is from 4% to 20%.

Exhibit 1 Return on Assets of Selected Companies

Company	ROA (%)
A	6
B	4
C	15
D	20
E	10
F	20

In trying to understand why the ROAs differ among these companies, we could look at why the ROA of Company A differs from that of Company B, why the ROA of Company A differs from that of Company D, why the ROA of Company F differs from that of Company C, and so on, comparing each pair of ROAs. A way to make this a simpler exploration is to try to understand why each company's ROA differs from the mean ROA of 12.5%. We look at the sum of squared deviations of the observations from the mean to capture variations in ROA from their mean. Let Y represent the variable that we would like to explain, which in this case is the return on assets. Let Y_i represent an observation of a company's ROA, and let $\overline{Y}$ represent the mean ROA for the sample of size n. We can describe the variation of the ROAs as

$$\text{Variation of } Y = \sum_{i=1}^{n}(Y_i - \overline{Y})^2. \quad (1)$$

Our goal is to understand what drives these returns on assets or, in other words, what explains the variation of Y. The variation of Y is often referred to as the **sum of squares total (SST)**, or the total sum of squares.

We now ask whether it is possible to explain the variation of the ROA using another variable that also varies among the companies; note that if this other variable is constant or random, it would not serve to explain why the ROAs differ from one another. Suppose the analyst believes that the capital expenditures in the previous period, scaled by the prior period's beginning property, plant, and equipment, are a driver for the ROA variable. Let us represent this scaled capital expenditures variable as CAPEX, as we show in Exhibit 2.

Exhibit 2 Return on Assets and Scaled Capital Expenditures

Company	ROA (%)	CAPEX (%)
A	6.0	0.7
B	4.0	0.4
C	15.0	5.0

Exhibit 2 (Continued)

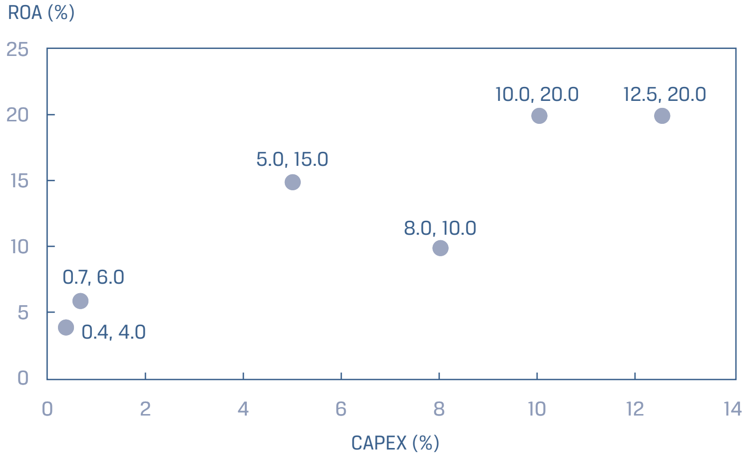
Company	ROA (%)	CAPEX (%)
D	20.0	10.0
E	10.0	8.0
F	20.0	12.5
Arithmetic mean	12.50	6.10

The variation of X, in this case CAPEX, is calculated as

$$\text{Variation of } X = \sum_{i=1}^{n}(X_i - \bar{X})^2. \quad (2)$$

We can see the relation between ROA and CAPEX in the **scatter plot** (or scattergram) in Exhibit 3, which represents the two variables in two dimensions. Typically, we present the variable whose variation we want to explain along the vertical axis and the variable whose variation we want to use to explain that variation along the horizontal axis. Each point in this scatter plot represents a paired observation that consists of CAPEX and ROA. From a casual visual inspection, there appears to be a positive relation between ROA and CAPEX: Companies with higher CAPEX tend to have a higher ROA.

Exhibit 3 Scatter Plot of ROA and CAPEX



In the ROA example, we use the capital expenditures to explain the returns on assets. We refer to the variable whose variation is being explained as the **dependent variable**, or the explained variable; it is typically denoted by Y. We refer to the variable whose variation is being used to explain the variation of the dependent variable as the **independent variable**, or the explanatory variable; it is typically denoted by X. Therefore, in our example, the ROA is the dependent variable (Y) and CAPEX is the independent variable (X).

A common method for relating the dependent and independent variables is through the estimation of a linear relationship, which implies describing the relation between the two variables as represented by a straight line. If we have only one independent variable, we refer to the method as **simple linear regression (SLR)**; if we have more than one independent variable, we refer to the method as multiple regression.

Linear regression allows us to test hypotheses about the relationship between two variables, by quantifying the strength of the relationship between the two variables, and to use one variable to make predictions about the other variable. Our focus is on linear regression with a single independent variable—that is, simple linear regression.

EXAMPLE 1

Identifying the Dependent and Independent Variables in a Regression

An analyst is researching the relationship between corporate earnings growth and stock returns. Specifically, she is interested in whether earnings revisions affect stock price returns in the same period. She collects five years of monthly data on "Wall Street" EPS revisions for a sample of 100 companies and on their monthly stock price returns over the five-year period.

What are the dependent and independent variables in her model?

Solution

The dependent variable is monthly stock price returns, and the independent variable is Wall Street EPS revisions, since in the analyst's model, the variation in monthly stock price returns is being explained by the variation in EPS revisions.

2 ESTIMATING THE PARAMETERS OF A SIMPLE LINEAR REGRESSION

b describe the least squares criterion, how it is used to estimate regression coefficients, and their interpretation

2.1 The Basics of Simple Linear Regression

Regression analysis begins with the dependent variable, the variable whose variation you are seeking to explain. The independent variable is the variable whose variation you are using to explain changes in the dependent variable. For example, you might try to explain small-stock returns (the dependent variable) using returns to the S&P 500 Index (the independent variable). Or you might try to explain a country's inflation rate (the dependent variable) as a function of growth in its money supply (the independent variable).

As the name implies, **linear regression** assumes a linear relationship between the dependent and the independent variables. The goal is to fit a line to the observations on Y and X to minimize the squared deviations from the line; this is the least squares criterion—hence, the name least squares regression. Because of its common use, linear regression is often referred to as ordinary least squares (OLS) regression.

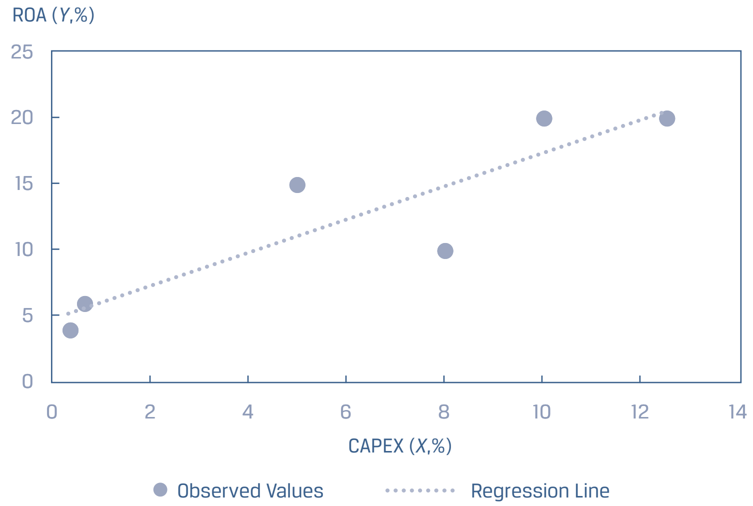
Using notation, the linear relation between the dependent and independent variables is described as

$$Y_i = b_0 + b_1 X_i + \varepsilon_i,\ i = 1, \ldots, n. \quad (3)$$

Equation 3 is a model that does not require that every (Y, X) pair for an observation fall on the regression line. This equation states that the dependent variable, Y, is equal to the **intercept**, b_0, plus a **slope coefficient**, b_1, multiplied by the independent variable, X, plus an **error term**, ε. The error term, or simply the error, represents the difference between the observed value of Y and that expected from the true underlying population relation between Y and X. We refer to the intercept, b_0, and the slope coefficient, b_1, as the **regression coefficients**. A way that we often describe this simple linear regression relation is that Y is regressed on X.

Consider the ROA and CAPEX scatter diagram from Exhibit 3, which we elaborate on in Exhibit 4 by including the fitted regression line. This line represents the average relationship between ROA and CAPEX; not every observation falls on the line, but the line describes the mean relation between ROA and CAPEX.

Exhibit 4 Fitted Regression Line of ROA and CAPEX



2.2 Estimating the Regression Line

We cannot observe the population parameter values b_0 and b_1 in a regression model. Instead, we observe only $\hat{b}_0$ and $\hat{b}_1$, which are estimates (as indicated by the "hats" above the coefficients) of the population parameters based on the sample. Thus, predictions must be based on the parameters' estimated values, and testing is based on estimated values in relation to the hypothesized population values.

We estimate the regression line as the line that best fits the observations. *In simple linear regression, the estimated intercept, $\hat{b}_0$, and slope, $\hat{b}_1$, are such that the sum of the squared vertical distances from the observations to the fitted line is minimized.* The focus is on the sum of the squared differences between the observations on Y_i and the corresponding estimated value, $\hat{Y}_i$, on the regression line.

We represent the value of the dependent variable for the ith observation that falls on the line as $\hat{Y}_i$, which is equal to $\hat{b}_0 + \hat{b}_1 X_i$. The $\hat{Y}_i$ is what the estimated value of the Y variable would be for the ith observation based on the mean relationship between

Y and X. The **residual** for the ith observation, e_i, is how much the observed value of Y_i differs from the $\hat{Y}_i$ estimated using the regression line: $e_i = Y_i - \hat{Y}_i$. Note the subtle difference between the error term and the residual: The error term refers to the true underlying population relationship, whereas the residual refers to the fitted linear relation based on the sample.

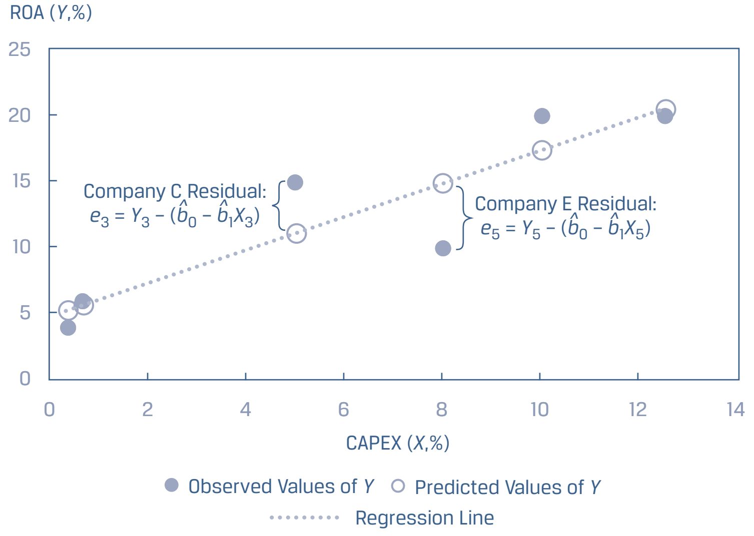
Fitting the line requires minimizing the sum of the squared residuals, the **sum of squares error (SSE)**, also known as the residual sum of squares:

$$\text{Sum of squares error} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$
$$= \sum_{i=1}^{n}\left[Y_i - \left(\hat{b}_0 + \hat{b}_1 X_i\right)\right]^2 \quad (4)$$
$$= \sum_{i=1}^{n} e_i^2.$$

Using least squares regression to estimate the values of the population parameters of b_0 and b_1, we can fit a line through the observations of X and Y that explains the value that Y takes for any particular value of X.

As seen in Exhibit 5, the residuals are represented by the vertical distances from the fitted line (see the third and fifth observations, Companies C and E, respectively) and are, therefore, in the units of measurement represented by the dependent variable. The residual is in the same unit of measurement as the dependent variable: If the dependent variable is in euros, the error term is in euros, and if the dependent variable is in growth rates, the error term is in growth rates.

Exhibit 5 Residuals of the Linear Regression



How do we calculate the intercept ($\hat{b}_0$) and the slope ($\hat{b}_1$) for a given sample of (Y, X) pairs of observations? The slope is the ratio of the covariance between Y and X to the variance of X, where $\overline{Y}$ is the mean of the Y variable and $\overline{X}$ is the mean of X variable:

$$\hat{b}_1 = \frac{\text{Covariance of } Y \text{ and } X}{\text{Variance of } X} = \frac{\dfrac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{n-1}}{\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}.$$

Simplifying,

$$\hat{b}_1 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}. \quad (5)$$

Once we estimate the slope, we can then estimate the intercept using the mean of Y and the mean of X:

$$\hat{b}_0 = \overline{Y} - \hat{b}_1\overline{X}. \quad (6)$$

We show the calculation of the slope and the intercept in Exhibit 6.

Exhibit 6 Estimating Slope and Intercept for the ROA Model

Company	ROA (Y_i)	CAPEX (X_i)	$(Y_i - \overline{Y})^2$	$(X_i - \overline{X})^2$	$(Y_i - \overline{Y})(X_i - \overline{X})$
A	6.0	0.7	42.25	29.16	35.10
B	4.0	0.4	72.25	32.49	48.45
C	15.0	5.0	6.25	1.21	−2.75
D	20.0	10.0	56.25	15.21	29.25
E	10.0	8.0	6.25	3.61	−4.75
F	20.0	12.5	56.25	40.96	48.00
Sum	75.0	36.6	239.50	122.64	153.30
Arithmetic mean	12.5	6.100			

Slope coefficient: $\hat{b}_1 = \dfrac{153.30}{122.64} = 1.25.$

Intercept: $\hat{b}_0 = 12.5 - (1.25 \times 6.10) = 4.875$

ROA regression model: $\hat{Y}_i = 4.875 + 1.25X_i + \varepsilon_i.$

Notice the similarity of the formula for the slope coefficient and that of the pairwise correlation. The sample correlation, r, is the ratio of the covariance to the product of the standard deviations:

$$r = \frac{\text{Covariance of } Y \text{ and } X}{\left(\begin{array}{c}\text{Standard deviation}\\ \text{of } Y\end{array}\right)\left(\begin{array}{c}\text{Standard deviation}\\ \text{of } X\end{array}\right)}.$$

The subtle difference between the slope and the correlation formulas is in the denominator: For the slope, this is the variance of the independent variable, but for the correlation, the denominator is the product of the standard deviations. For our ROA and CAPEX analysis,

Covariance of Y and X: $$\text{cov}_{XY} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{n-1} = \frac{153.30}{5} = 30.66. \quad \text{(7a)}$$

Standard deviation of Y and X:

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{239.50}{5}} = 6.9210;$$
$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{122.64}{5}} = 4.9526. \quad \text{(7b)}$$

$$r = \frac{30.66}{(6.9210)(4.9526)} = 0.89458945.$$

Because the denominators of both the slope and the correlation are positive, the sign of the slope and the correlation are driven by the numerator: If the covariance is positive, both the slope and the correlation are positive, and if the covariance is negative, both the slope and the correlation are negative.

HOW DO ANALYSTS PERFORM SIMPLE LINEAR REGRESSION?

Typically, an analyst will use the data analysis functions on a spreadsheet, such as Microsoft Excel, or a statistical package in the R or Python programming languages to perform linear regression analysis. The following are some of the more common choices in practice.

Simple Linear Regression: Intercept and Slope

- *Excel*: Use the INTERCEPT, SLOPE functions.
- *R*: Use the lm function.
- *Python*: Use the sm.OLS function in the statsmodels package.

Correlations

- *Excel*: Use the CORREL function.
- *R*: Use the cor function in the stats library.
- *Python*: Use the corrcoef function in the numpy library.

Note that in R and Python, there are many choices for regression and correlation analysis.

2.3 Interpreting the Regression Coefficients

What is the meaning of the regression coefficients? The intercept is the value of the dependent variable if the value of the independent variable is zero. Importantly, this does not make sense in some contexts, especially if it is unrealistic that the independent variable would be zero. For example, if we have a model where money supply explains GDP growth, the intercept has no meaning because, practically speaking, zero money supply is not possible. If the independent variable were money supply growth, however, the intercept is meaningful. The slope is the change in the dependent variable for a one-unit change in the independent variable. If the slope is positive, then the change in the independent variable and that of the dependent variable will be in the same direction; if the slope is negative, the change in the independent variable and that of the dependent variable will be in opposite directions.

INTERPRETING POSITIVE AND NEGATIVE SLOPES

Suppose the dependent variable (Y) is in millions of euros and the independent variable (*X*) is in millions of US dollars.

If the slope is positive 1.2, then

↑ USD1 million → ↑ EUR1.2 million

↓ USD1 million → ↓ EUR1.2 million

If the slope is negative 1.2, then

↑ USD1 million → ↓ EUR1.2 million

↓ USD1 million → ↑ EUR1.2 million

Using the ROA regression model from Exhibit 6, we would interpret the estimated coefficients as follows:

- The return on assets for a company is 4.875% if the company makes no capital expenditures.
- If CAPEX increases by one unit—say, from 4% to 5%—ROA increases by 1.25%.

Using the estimated regression coefficients, we can determine the values of the dependent variable if they follow the average relationship between the dependent and independent variables. A result of the mathematics of the least squares fitting of the regression line is that the expected value of the residual term is zero: E(ε) = 0.

We show the calculation of the predicted dependent variable and residual term for each observation in the ROA example in Exhibit 7. Note that the sum and average of Y_i and $\hat{Y}_1$ are the same, and the sum of the residuals is zero.

Exhibit 7 Calculation of the Dependent Variable and Residuals for the ROA and CAPEX Model

	(1)	(2)	(3)	(4)
Company	ROA (Y_i)	CAPEX (X_i)	Predicted ROA ($\hat{Y}_i$)	(1) – (3) Residual (e_i)
A	6.0	0.7	5.750	0.250
B	4.0	0.4	5.375	–1.375
C	15.0	5.0	11.125	3.875
D	20.0	10.0	17.375	2.625
E	10.0	8.0	14.875	–4.875
F	20.0	12.5	20.500	–0.500

(continued)

Exhibit 7 (Continued)

Company	(1) ROA (Y_i)	(2) CAPEX (X_i)	(3) Predicted ROA ($\hat{Y}_i$)	(4) (1) – (3) Residual (e_i)
Sum	75.0	36.6	75.000	0.000
Average	12.5	6.1	12.5	0.000

For Company C (i = 3),

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i + \varepsilon_i = 4.875 + 1.25X_i + \varepsilon_i$$

$$\hat{Y}_i = 4.875 + (1.25 \times 5.0) = 4.875 + 6.25 = 11.125$$

$Y_i - \hat{Y}_i = e_i = 15.0 - 11.125 = 3.875$, the vertical distance in Exhibit 5.

Whereas the sum of the residuals must equal zero by design, the focus of fitting the regression line in a simple linear regression is minimizing the sum of the squared residual terms.

2.4 Cross-Sectional vs. Time-Series Regressions

Regression analysis uses two principal types of data: cross sectional and time series. A cross-sectional regression involves many observations of X and Y for the same time period. These observations could come from different companies, asset classes, investment funds, countries, or other entities, depending on the regression model. For example, a cross-sectional model might use data from many companies to test whether predicted EPS growth explains differences in price-to-earnings ratios during a specific time period. Note that if we use cross-sectional observations in a regression, we usually denote the observations as $i = 1, 2, \ldots, n$.

Time-series data use many observations from different time periods for the same company, asset class, investment fund, country, or other entity, depending on the regression model. For example, a time-series model might use monthly data from many years to test whether a country's inflation rate determines its short-term interest rates. If we use time-series data in a regression, we usually denote the observations as $t = 1, 2, \ldots, T$. Note that in the sections that follow, we primarily use the notation $i = 1, 2, \ldots, n$, even for time series.

EXAMPLE 2

Estimating a Simple Linear Regression Model

An analyst is exploring the relationship between a company's net profit margin and research and development expenditures. He collects data for an industry and calculates the ratio of research and development expenditures to revenues (RDR) and the net profit margin (NPM) for eight companies. Specifically, he wants to explain the variation that he observes in the net profit margin by using the variation he observes in the companies' research and development spending. He reports the data in Exhibit 8.

Exhibit 8 Observations on NPM and RDR for Eight Companies

Company	NPM (%)	RDR (%)
1	4	8
2	5	10
3	10	6
4	9	5
5	5	7
6	6	9
7	12	5
8	3	10

1 What is the slope coefficient for this simple linear regression model?

2 What is the intercept for this regression model?

3 How is this estimated linear regression model represented?

4 What is the pairwise correlation between NPM and RDR?

Solutions

1 The slope coefficient for the regression model is −1.3, and the details for the inputs to this calculation are in Exhibit 9.

Exhibit 9 Details of Calculation of Slope of NPM Regressed on RDR

Company	NPM (%) (Y_i)	RDR (%) (X_i)	$Y_i - \bar{Y}$	$X_i - \bar{X}$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})(X_i - \bar{X})$
1	4	8	−2.8	0.5	7.5625	0.25	−1.375
2	5	10	−1.8	2.5	3.0625	6.25	−4.375
3	10	6	3.3	−1.5	10.5625	2.25	−4.875
4	9	5	2.3	−2.5	5.0625	6.25	−5.625
5	5	7	−1.8	−0.5	3.0625	0.25	0.875
6	6	9	−0.8	1.5	0.5625	2.25	−1.125
7	12	5	5.3	−2.5	27.5625	6.25	−13.125
8	3	10	−3.8	2.5	14.0625	6.25	−9.375
Sum	54.0	60.0	0.0	0.0	71.5000	30.00	−39.0
Average	6.75	7.5					

Slope coefficient: $\hat{b}_1 = \frac{-39}{30} = -1.3.$

2 The intercept of the regression model is 16.5:

$$\text{Intercept: } \hat{b}_0 = 6.75 - (-1.3 \times 7.5) = 6.75 + 9.75 = 16.5$$

3 The regression model is represented by $\hat{Y}_i = 16.5 - 1.3X_i + \varepsilon_i$.

4 The pairwise correlation is –0.8421:

$$r = \frac{-39/7}{\sqrt{71.5/7}\sqrt{30/7}} = \frac{-5.5714}{(3.1960)(2.0702)} = -0.8421.$$

EXAMPLE 3

Interpreting Regression Coefficients

An analyst has estimated a model that regresses a company's return on equity (ROE) against its growth opportunities (GO), defined as the company's three-year compounded annual growth rate in sales, over 20 years and produces the following estimated simple linear regression:

$$\text{ROE}_i = 4 + 1.8\ \text{GO}_i + \varepsilon_i.$$

Both variables are stated in percentages, so a GO observation of 5% is included as 5.

1 The predicted value of the company's ROE if its GO is 10% is closest to:

A 1.8%.

B 15.8%.

C 22.0%.

2 The change in ROE for a change in GO from 5% to 6% is closest to:

A 1.8%.

B 4.0%.

C 5.8%.

3 The residual in the case of a GO of 8% and an observed ROE of 21% is closest to:

A –1.8%.

B 2.6%.

C 12.0%.

Solutions

1 C is correct. The predicted value of ROE = 4 + (1.8 × 10) = 22.

2 A is correct. The slope coefficient of 1.8 is the expected change in the dependent variable (ROE) for a one-unit change in the independent variable (GO).

3 B is correct. The predicted value is ROE = 4 + (1.8 × 8) = 18.4. The observed value of ROE is 21, so the residual is 2.6 = 21.0 – 18.4.

ASSUMPTIONS OF THE SIMPLE LINEAR REGRESSION MODEL

3

c explain the assumptions underlying the simple linear regression model, and describe how residuals and residual plots indicate if these assumptions may have been violated

We have discussed how to interpret the coefficients in a simple linear regression model. Now we turn to the statistical assumptions underlying this model. Suppose that we have n observations of both the dependent variable, Y, and the independent variable, X, and we want to estimate the simple linear regression of Y regressed on X. We need to make the following four key assumptions to be able to draw valid conclusions from a simple linear regression model:

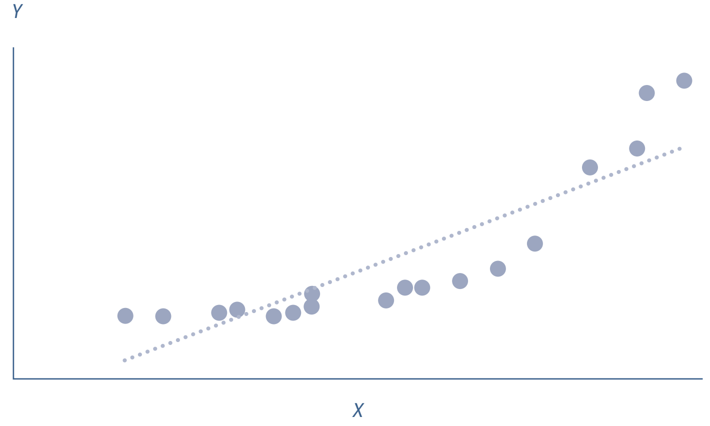
1. Linearity: The relationship between the dependent variable, Y, and the independent variable, X, is linear.
2. Homoskedasticity: The variance of the regression residuals is the same for all observations.
3. Independence: The observations, pairs of Ys and Xs, are independent of one another. This implies the regression residuals are uncorrelated across observations.
4. Normality: The regression residuals are normally distributed.

Now we take a closer look at each of these assumptions and introduce the "best practice" of examining residual plots of regression results to identify potential violations of these key assumptions.

3.1 Assumption 1: Linearity

We are fitting a linear model, so we must assume that the true underlying relationship between the dependent and independent variables is linear. If the relationship between the independent and dependent variables is nonlinear in the parameters, estimating that relation with a simple linear regression model will produce invalid results: The model will be biased, because it will under- and overestimate the dependent variable at certain points. For example, $Y_i = b_0 e^{b_1 X_i} + \varepsilon_i$ is nonlinear in b_1, so we should not apply the linear regression model to it. Exhibit 10 shows an example of this exponential model, with a regression line indicated. You can see that this line does not fit this relationship well: For lower and higher values of X, the linear model underestimates the Y, whereas for the middle values, the linear model overestimates Y.

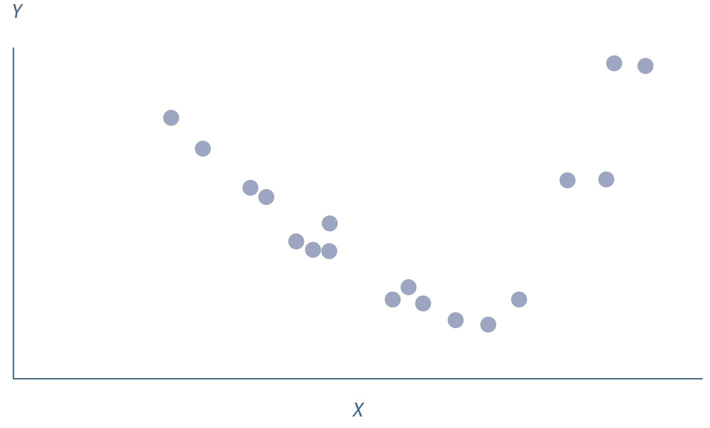
Exhibit 10 Illustration of Nonlinear Relationship Estimated as a Linear Relationship



Another implication of this assumption is that the independent variable, *X*, must not be random; that is, it is non-stochastic. If the independent variable is random, there would be no linear relation between the dependent and independent variables. Although we may initially assume that the independent variable in the regression model is not random, that assumption may not always be true.

When we look at the residuals of a model, what we would like to see is that the residuals are random. The residuals should not exhibit a pattern when plotted against the independent variable. As we show in Exhibit 11, the residuals from the Exhibit 10 linear regression do not appear to be random but, rather, exhibit a relationship with the independent variable, *X*, falling for some range of *X* and rising in another.

Exhibit 11 Illustration of Residuals in a Nonlinear Relationship Estimated as a Linear Relationship



3.2 Assumption 2: Homoskedasticity

Assumption 2, that the variance of the residuals is the same for all observations, is known as the **homoskedasticity** assumption. In terms of notation, this assumption relates to the squared residuals:

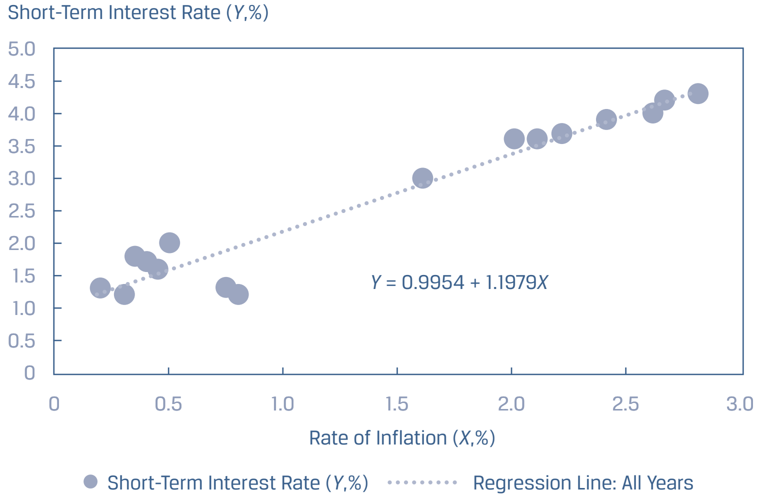
$$E\left(\varepsilon_i^2\right) = \sigma_\varepsilon^2,\ i = 1, \ldots, n. \quad (8)$$

If the residuals are not homoscedastic, that is, if the variance of residuals differs across observations, then we refer to this as **heteroskedasticity**.

Suppose you are examining a time series of short-term interest rates as the dependent variable and inflation rates as the independent variable over 16 years. We may believe that short-term interest rates (Y) and inflation rates (X) should be related (that is, interest rates are higher with higher rates of inflation. If this time series spans many years, with different central bank actions that force short-term interest rates to be (artificially) low for the last eight years of the series, then it is likely that the residuals in this estimated model will appear to come from two different models. We will refer to the first eight years as Regime 1 (normal rates) and the second eight years as Regime 2 (low rates). If the model fits differently in the two regimes, the residuals and their variances will be different.

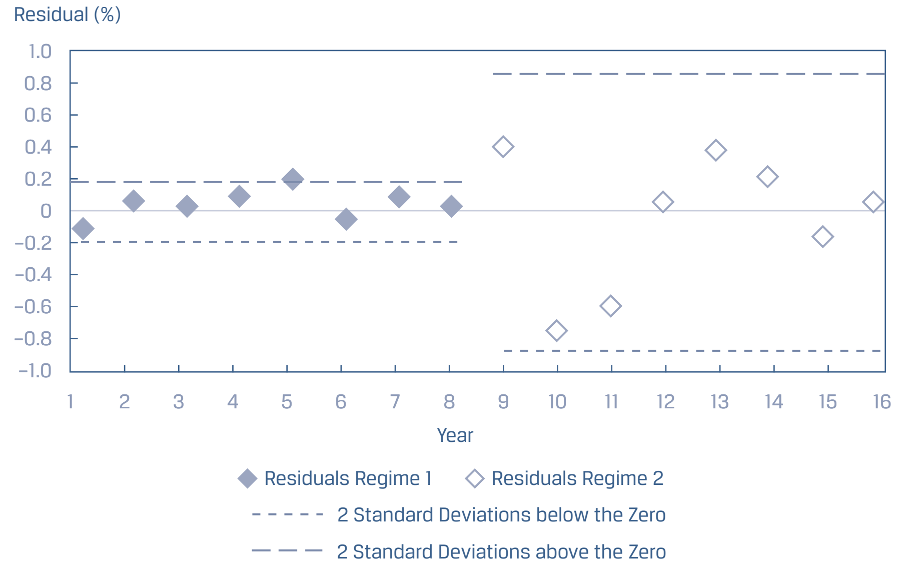
You can see this situation in Exhibit 12, which shows a scatter plot with an estimated regression line. The slope of the regression line over all 16 years is 1.1979.

Exhibit 12 Scatter Plot of Interest Rates (*Y*) and Inflation Rates (*X*)



We plot the residuals of this model in Exhibit 13 against the years. In this plot, we indicate the distance that is two standard deviations from zero (the mean of the residuals) for the first eight years' residuals and then do the same for the second eight years. As you can see, the residuals appear different for the two regimes: the variation in the residuals for the first eight years is much smaller than the variation for the second eight years.

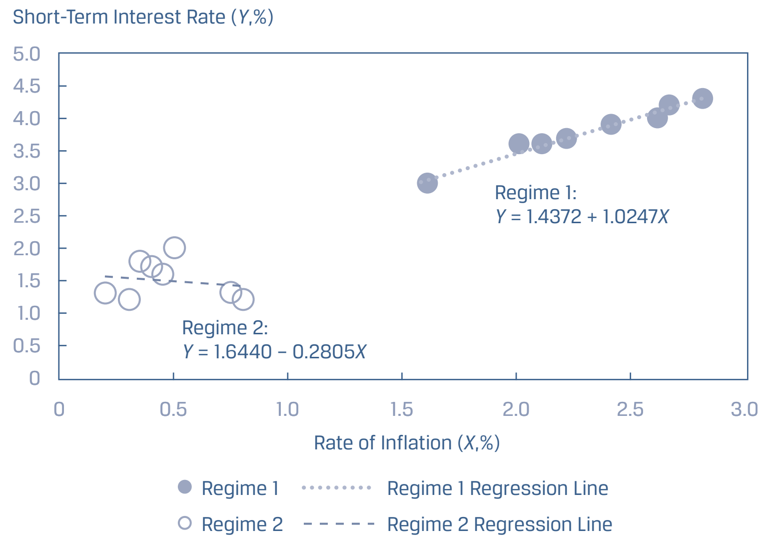
Exhibit 13 Residual Plot for Interest Rates (*Y*) vs. Inflation Rates (*X*) Model



Why does this happen? The model seems appropriate, but when we examine the residuals (Exhibit 13), an important step in assessing the model fit, we see that the model fits better in some years compared with others. The difference in variance of residuals between the two regimes is apparent from the much wider band around residuals for Regime 2 (the low-rate period). This indicates a clear violation of the homoskedasticity assumption.

If we estimate a regression line for each regime, we can see that the model for the two regimes is quite different, as we show in Exhibit 14. In the case of Regime 1 (normal rates), the slope is 1.0247, whereas in Regime 2 (low rates) the slope is −0.2805. In sum, the clustering of residuals in two groups with much different variances clearly indicates the existence of distinct regimes for the relationship between short-term interest rates and the inflation rate.

Exhibit 14 Fitted Regression Lines for the Two Regimes

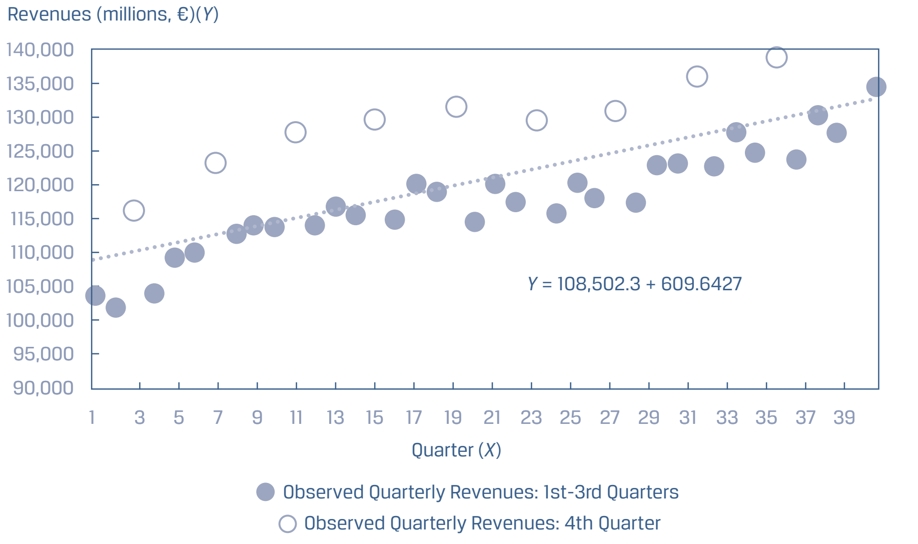


3.3 Assumption 3: Independence

We assume that the observations (Y and X pairs) are uncorrelated with one another, meaning they are independent. If there is correlation between observations (that is, autocorrelation), they are not independent and the residuals will be correlated. The assumption that the residuals are uncorrelated across observations is also necessary for correctly estimating the variances of the **estimated parameters** of b_0 and b_1 (i.e., $\hat{b}_0$ and $\hat{b}_1$) that we use in hypothesis tests of the intercept and slope, respectively. It is important to examine whether the residuals exhibit a pattern, suggesting a violation of this assumption. Therefore, we need to visually and statistically examine the residuals for a regression model.

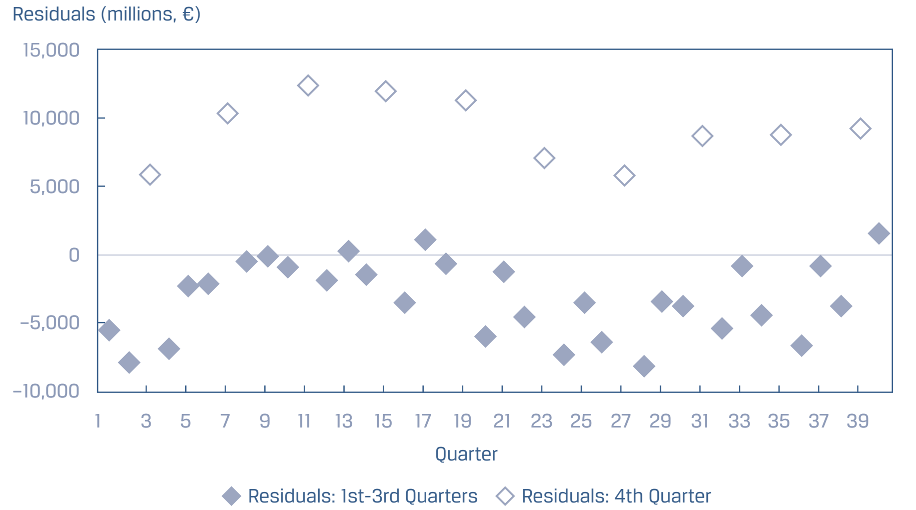
Consider the quarterly revenues of a company regressed over 40 quarters, as shown in Exhibit 15, with the regression line included. It is clear that these revenues display a seasonal pattern, an indicator of autocorrelation.

Exhibit 15 Regression of Quarterly Revenues vs. Time (40 Quarters)



In Exhibit 16, we plot the residuals from this model and see that there is a pattern. These residuals are correlated, specifically jumping up in Quarter 4 and then falling back the subsequent quarter. In sum, the patterns in both Exhibits 15 and 16 indicate a violation of the assumption of independence.

Exhibit 16 Residual Plot for Quarterly Revenues vs. Time Model



3.4 Assumption 4: Normality

The assumption of normality requires that the residuals be normally distributed. This does not mean that the dependent and independent variables must be normally distributed; it only means that the residuals from the model are normally distributed. However, in estimating any model, it is good practice to understand the distribution of the dependent and independent variables to explore for outliers. An outlier in either or both variables can substantially influence the fitted line such that the estimated model will not fit well for most of the other observations.

With normally distributed residuals, we can test a particular hypothesis about a linear regression model. For large sample sizes, we may be able to drop the assumption of normality by appealing to the central limit theorem; asymptotic theory (which deals with large samples) shows that in many cases, the test statistics produced by standard regression programs are valid even if the model's residuals are not normally distributed.

EXAMPLE 4

Assumptions of Simple Linear Regression

An analyst is investigating a company's revenues and estimates a simple linear time-series model by regressing revenues against time, where time—1, 2, . . . , 15—is measured in years. She plots the company's observed revenues and the estimated regression line, as shown in Exhibit 17. She also plots the residuals from this regression model, as shown in Exhibit 18.

Exhibit 17 Revenues vs. Time Using Simple Linear Regression

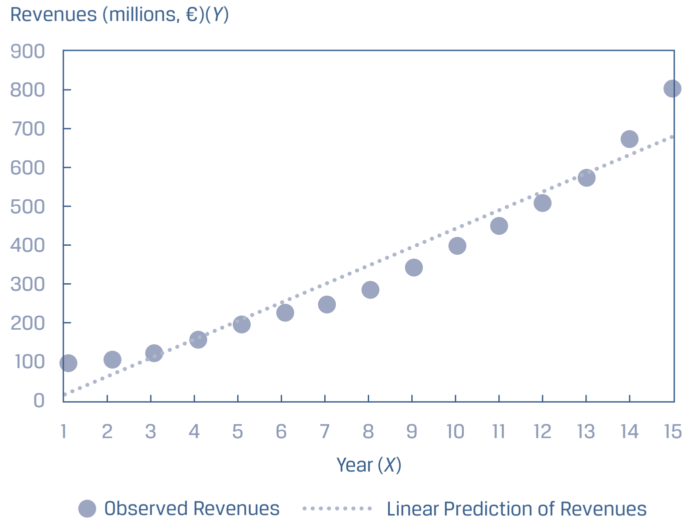
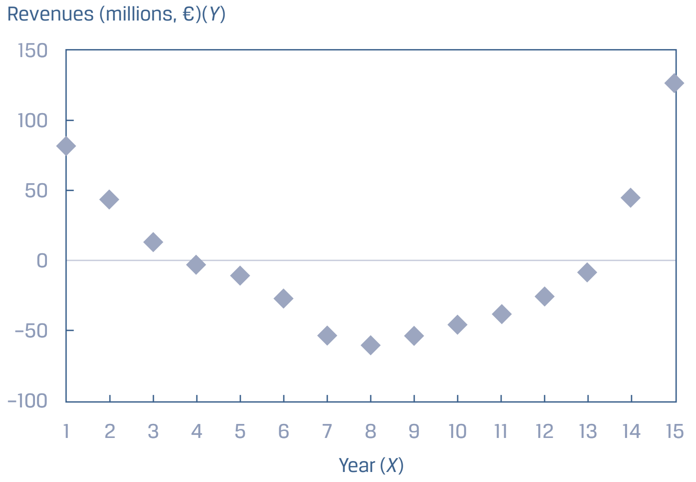


Exhibit 18 Residual Plot for Revenues vs. Time



Based on Exhibits 17 and 18, describe which assumption(s) of simple linear regression the analyst's model may be violating.

Solution

The correct model is not linear, as evident from the pattern of the revenues in Exhibit 17. In the earlier years (i.e., 1 and 2) and later years (i.e., 14 and 15), the linear model underestimates revenues, whereas for the middle years (i.e., 7–11), the linear model overestimates revenues. Moreover, the curved pattern of residuals in Exhibit 18 indicates potential heteroskedasticity (residuals have unequal variances), lack of independence of observations, and non-normality (a concern given the small sample size of n = 15). In sum, the analyst should be concerned that her model violates all the assumptions governing simple linear regression (linearity, homoskedasticity, independence, and normality).

4 ANALYSIS OF VARIANCE

d calculate and interpret the coefficient of determination and the *F*-statistic in a simple linear regression

The simple linear regression model sometimes describes the relationship between two variables quite well, but sometimes it does not. We must be able to distinguish between these two cases to use regression analysis effectively. Remember our goal is to explain the variation of the dependent variable. So, how well has this goal been achieved, given our choice of independent variable?

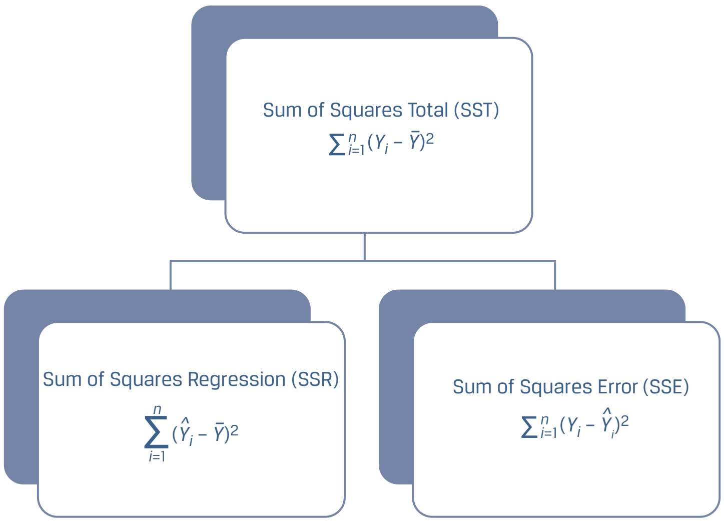
4.1 Breaking down the Sum of Squares Total into Its Components

We begin with the sum of squares total and then break it down into two parts: the sum of squares error and the **sum of squares regression (SSR)**. The sum of squares regression is the sum of the squared differences between the predicted value of the dependent variable, $\hat{Y}_i$, based on the estimated regression line, and the mean of the dependent variable, $\bar{Y}$:

$$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2. \tag{9}$$

We have already defined the sum of squares total, which is the total variation in *Y*, and the sum of squares error, the unexplained variation in *Y*. Note that the sum of squares regression is the explained variation in *Y*. So, as illustrated in Exhibit 19, SST = SSR + SSE, meaning total variation in *Y* equals explained variation in *Y* plus unexplained variation in *Y*.

Exhibit 19 Breakdown of Variation of Dependent Variable



We show the breakdown of the sum of squares total formula for our ROA regression example in Exhibit 20. The total variation of ROA that we want to explain (SST) is 239.50. This number comprises the variation unexplained (SSE), 47.88, and the variation explained (SSR), 191.63. These sum of squares values are important inputs into measures of the fit of the regression line.

Exhibit 20 Breakdown of Sum of Squares Total for ROA Model

Company	ROA (Y_i)	CAPEX (X_i)	Predicted ROA ($\hat{Y}$)	Variation to Be Explained $(Y_i - \bar{Y})^2$	Variation Unexplained $(Y_i - \hat{Y}_i)^2$	Variation Explained $(\hat{Y}_i - \bar{Y})^2$
A	6.0	0.7	5.750	42.25	0.063	45.563
B	4.0	0.4	5.375	72.25	1.891	50.766
C	15.0	5.0	11.125	6.25	15.016	1.891
D	20.0	10.0	17.375	56.25	6.891	23.766
E	10.0	8.0	14.875	6.25	23.766	5.641
F	20.0	12.5	20.500	56.25	0.250	64.000
				239.50	47.88	191.625
Mean	12.50					

Sum of squares total = 239.50.
Sum of squares error = 47.88.
Sum of squares regression = 191.63.

4.2 Measures of Goodness of Fit

There are several measures that we can use to evaluate goodness of fit—that is, how well the regression model fits the data. These include the coefficient of determination, the *F*-statistic for the test of fit, and the standard error of the regression.

The **coefficient of determination**, also referred to as the *R*-squared or R^2, is the percentage of the variation of the dependent variable that is explained by the independent variable:

$$\text{Coefficient of determination} = \frac{\text{Sum of squares regression}}{\text{Sum of squares total}}$$

$$\text{Coefficient of determination} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}. \tag{10}$$

By construction, the coefficient of determination ranges from 0% to 100%. In our ROA example, the coefficient of determination is 191.625 ÷ 239.50, or 0.8001, so 80.01% of the variation in ROA is explained by CAPEX. In a simple linear regression, the square of the pairwise correlation is equal to the coefficient of determination:

$$r^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} = R^2.$$

In our earlier ROA regression analysis, $r = 0.8945$, so we now see that r^2 is indeed equal to the coefficient of determination (R^2), since $(0.8945)^2 = 0.8001$.

Whereas the coefficient of determination—the portion of the variation of the dependent variable explained by the independent variable—is descriptive, it is not a statistical test. To see if our regression model is likely to be statistically meaningful, we will need to construct an *F*-distributed test statistic.

In general, we use an F-distributed test statistic to compare two variances. In regression analysis, we can use an F-distributed test statistic to test whether the slopes in a regression are equal to zero, with the slopes designated as b_i, against the alternative hypothesis that at least one slope is not equal to zero:

H_0: $b_1 = b_2 = b_3 = \ldots = b_k = 0$.

H_a: At least one b_k is not equal to zero.

For simple linear regression, these hypotheses simplify to

H_0: $b_1 = 0$.

H_a: $b_1 \neq 0$.

The F-distributed test statistic is constructed by using the sum of squares regression and the sum of squares error, each adjusted for degrees of freedom; in other words, it is the ratio of two variances. We divide the sum of squares regression by the number of independent variables, represented by k. In the case of a simple linear regression, k = 1, so we arrive at the **mean square regression (MSR)**, which is the same as the sum of squares regression:

$$\text{MSR} = \frac{\text{Sum of squares regression}}{k} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{1}.$$

So, for simple linear regression,

$$\text{MSR} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2. \tag{11}$$

Next, we calculate the **mean square error (MSE)**, which is the sum of squares error divided by the degrees of freedom, which are $n - k - 1$. In simple linear regression, $n - k - 1$ becomes $n - 2$:

$$\text{MSE} = \frac{\text{Sum of squares error}}{n - k - 1}.$$

$$\text{MSE} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n - 2}. \tag{12}$$

Therefore, the F-distributed test statistic (MSR/MSE) is

$$F = \frac{\dfrac{\text{Sum of squares regression}}{k}}{\dfrac{\text{Sum of squares error}}{n - k - 1}} = \frac{\text{MSR}}{\text{MSE}}$$

$$F = \frac{\dfrac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{1}}{\dfrac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n - 2}}, \tag{13}$$

which is distributed with 1 and $n - 2$ degrees of freedom in simple linear regression. The F-statistic in regression analysis is one sided, with the rejection region on the right side, because we are interested in whether the variation in Y explained (the numerator) is larger than the variation in Y unexplained (the denominator).

4.3 ANOVA and Standard Error of Estimate in Simple Linear Regression

e describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression

We often represent the sums of squares from a regression model in an **analysis of variance (ANOVA)** table, as shown in Exhibit 21, which presents the sums of squares, the degrees of freedom, the mean squares, and the *F*-statistic. Notice that the variance of the dependent variable is the ratio of the sum of squares total to $n - 1$.

Exhibit 21 Analysis of Variance Table for Simple Linear Regression

Source	Sum of Squares	Degrees of Freedom	Mean Square	*F*-Statistic
Regression	$\text{SSR} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	1	$\text{MSR} = \dfrac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{1}$	$F = \dfrac{\text{MSR}}{\text{MSE}} = \dfrac{\dfrac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{1}}{\dfrac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}}$
Error	$\text{SSE} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	$n - 2$	$\text{MSE} = \dfrac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}$	
Total	$\text{SST} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$	$n - 1$		

From the ANOVA table, we can also calculate the **standard error of the estimate** (s_e), which is also known as the standard error of the regression or the root mean square error. The s_e is a measure of the distance between the observed values of the dependent variable and those predicted from the estimated regression; the smaller the s_e, the better the fit of the model. The s_e, along with the coefficient of determination and the *F*-statistic, is a measure of the goodness of the fit of the estimated regression line. Unlike the coefficient of determination and the *F*-statistic, which are relative measures of fit, the standard error of the estimate is an absolute measure of the distance of the observed dependent variable from the regression line. Thus, the s_e is an important statistic used to evaluate a regression model and is used in calculating prediction intervals and performing tests on the coefficients. The calculation of s_e is straightforward once we have the ANOVA table because it is the square root of the MSE:

$$\text{Standard error of the estimate } (s_e) = \sqrt{\text{MSE}} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}}. \quad (14)$$

We show the ANOVA table for our ROA regression example in Exhibit 22, using the information from Exhibit 20. For a 5% level of significance, the critical *F*-value for the test of whether the model is a good fit (that is, whether the slope coefficient is different from zero) is 7.71. We can get this critical value in the following ways:

- *Excel* [F.INV(0.95,1,4)]
- *R* [qf(.95,1,4)]
- *Python* [from scipy.stats import f *and* f.ppf(.95,1,4)]

With a calculated F-statistic of 16.0104 and a critical F-value of 7.71, we reject the null hypothesis and conclude that the slope of our simple linear regression model for ROA is different from zero.

Exhibit 22 ANOVA Table for ROA Regression Model

Source	Sum of Squares	Degrees of Freedom	Mean Square	*F*-Statistic
Regression	191.625	1	191.625	16.0104
Error	47.875	4	11.96875	
Total	239.50	5		

The calculations to derive the ANOVA table and ultimately to test the goodness of fit of the regression model can be time consuming, especially for samples with many observations. However, statistical packages, such as SAS, SPSS Statistics, and Stata, as well as software, such as Excel, R, and Python, produce the ANOVA table as part of the output for regression analysis.

EXAMPLE 5

Using ANOVA Table Results to Evaluate a Simple Linear Regression

Suppose you run a cross-sectional regression for 100 companies, where the dependent variable is the annual return on stock and the independent variable is the lagged percentage of institutional ownership (INST). The results of this simple linear regression estimation are shown in Exhibit 23. Evaluate the model by answering the questions below.

Exhibit 23 ANOVA Table for Annual Stock Return Regressed on Institutional Ownership

Source	Sum of Squares	Degrees of Freedom	Mean Square
Regression	576.1485	1	576.1485
Error	1,873.5615	98	19.1180
Total	2,449.7100		

1. What is the coefficient of determination for this regression model?
2. What is the standard error of the estimate for this regression model?
3. At a 5% level of significance, do we reject the null hypothesis of the slope coefficient equal to zero if the critical F-value is 3.938?
4. Based on your answers to the preceding questions, evaluate this simple linear regression model.

Solutions

1 The coefficient of determination is sum of squares regression/sum of squares total: 576.148 ÷ 2,449.71 = 0.2352, or 23.52%.

2 The standard error of the estimate is the square root of the mean square error: $\sqrt{19.1180} = 4.3724$.

3 Using a six-step process for testing hypotheses, we get the following:

Step 1	State the hypotheses.	H_0: $b_1 = 0$ versus H_a: $b_1 \neq 0$
Step 2	Identify the appropriate test statistic.	$F = \frac{\text{MSR}}{\text{MSE}}$ with 1 and 98 degrees of freedom.
Step 3	Specify the level of significance.	α = 5% (one tail, right side).
Step 4	State the decision rule.	Critical F-value = 3.938. Reject the null hypothesis if the calculated F-statistic is greater than 3.938.
Step 5	Calculate the test statistic.	$F = \frac{576.1485}{19.1180} = 30.1364$
Step 6	Make a decision.	Reject the null hypothesis because the calculated F-statistic is greater than the critical F-value. There is sufficient evidence to indicate that the slope coefficient is different from 0.0.

4 The coefficient of determination indicates that variation in the independent variable explains 23.52% of the variation in the dependent variable. Also, the F-statistic test confirms that the model's slope coefficient is different from 0 at the 5% level of significance. In sum, the model seems to fit the data reasonably well.

HYPOTHESIS TESTING OF LINEAR REGRESSION COEFFICIENTS 5

f formulate a null and an alternative hypothesis about a population value of a regression coefficient, and determine whether the null hypothesis is rejected at a given level of significance

5.1 Hypothesis Tests of the Slope Coefficient

We can use the F-statistic to test for the significance of the slope coefficient (that is, whether it is significantly different from zero), but we also may want to perform other hypothesis tests for the slope coefficient—for example, testing whether the population slope is different from a specific value or whether the slope is positive. We can use a t-distributed test statistic to test such hypotheses about a regression coefficient.

Suppose we want to check a stock's valuation using the market model; we hypothesize that the stock has an average systematic risk (i.e., risk similar to that of the market), as represented by the coefficient on the market returns variable. Or we may want to test the hypothesis that economists' forecasts of the inflation rate are unbiased (that is, on average, not overestimating or underestimating actual inflation rates). In each case, does the evidence support the hypothesis? Such questions as these can be addressed with hypothesis tests on the regression slope. To test a hypothesis about a slope, we calculate the test statistic by subtracting the hypothesized population slope (B_1) from the estimated slope coefficient ($\hat{b}_1$) and then dividing this difference by the standard error of the slope coefficient, $s_{\hat{b}_1}$:

$$t = \frac{\hat{b}_1 - B_1}{s_{\hat{b}_1}}. \tag{15}$$

This test statistic is t-distributed with $n - k - 1$ or $n - 2$ degrees of freedom because two parameters (an intercept and a slope) were estimated in the regression.

The **standard error of the slope coefficient** ($s_{\hat{b}_1}$) for a simple linear regression is the ratio of the model's standard error of the estimate (s_e) to the square root of the variation of the independent variable:

$$s_{\hat{b}_1} = \frac{s_e}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}. \tag{16}$$

We compare the calculated t-statistic with the critical values to test hypotheses. Note that the greater the variability of the independent variable, the lower the standard error of the slope (Equation 16) and hence the greater the calculated t-statistic (Equation 15). If the calculated t-statistic is outside the bounds of the critical t-values, we reject the null hypothesis, but if the calculated t-statistic is within the bounds of the critical values, we fail to reject the null hypothesis. Similar to tests of the mean, the alternative hypothesis can be two sided or one sided.

Consider our previous simple linear regression example with ROA as the dependent variable and CAPEX as the independent variable. Suppose we want to test whether the slope coefficient of CAPEX is different from zero to confirm our intuition of a significant relationship between ROA and CAPEX. We can test the hypothesis concerning the slope using the six-step process, as we show in Exhibit 24. As a result of this test, we conclude that the slope is different from zero; that is, CAPEX is a significant explanatory variable of ROA.

Exhibit 24 Test of the Slope for the Regression of ROA on CAPEX

Step 1	State the hypotheses.	H_0: $b_1 = 0$ versus H_a: $b_1 \neq 0$
Step 2	Identify the appropriate test statistic.	$t = \frac{\hat{b}_1 - B_1}{s_{\hat{b}_1}}$ with 6 − 2 = 4 degrees of freedom.
Step 3	Specify the level of significance.	α = 5%.

Exhibit 24 (Continued)

Step 4	State the decision rule.	Critical t-values = ±2.776. We can determine this from **Excel** *Lower:* T.INV(0.025,4) *Upper:* T.INV(0.975,4) **R** qt(c(.025,.975),4) **Python** from scipy.stats import t *Lower:* t.ppf(.025,4) *Upper:* t.ppf(.975,4) We reject the null hypothesis if the calculated t-statistic is less than −2.776 or greater than +2.776.
Step 5	Calculate the test statistic.	The slope coefficient is 1.25 (Exhibit 6). The mean square error is 11.96875 (Exhibit 22). The variation of CAPEX is 122.640 (Exhibit 6). $s_e = \sqrt{11.96875} = 3.459588.$ $s_{\hat{b}_1} = \dfrac{3.459588}{\sqrt{122.640}} = 0.312398.$ $t = \dfrac{1.25 - 0}{0.312398} = 4.00131.$
Step 6	Make a decision.	Reject the null hypothesis of a zero slope. There is sufficient evidence to indicate that the slope is different from zero.

A feature of simple linear regression is that the t-statistic used to test whether the slope coefficient is equal to zero and the t-statistic to test whether the pairwise correlation is zero (that is, H_0: ρ = 0 versus H_a: ρ ≠ 0) are the same value. Just as with a test of a slope, both two-sided and one-sided alternatives are possible for a test of a correlation—for example, H_0: ρ ≤ 0 versus H_a: ρ > 0. The test-statistic to test whether the correlation is equal to zero is

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

In our example of ROA regressed on CAPEX, the correlation (r) is 0.8945. To test whether this correlation is different from zero, we perform a test of hypothesis, shown in Exhibit 25. As you can see, we draw a conclusion similar to that for our test of the slope, but it is phrased in terms of the correlation between ROA and CAPEX: There is a significant correlation between ROA and CAPEX.

Exhibit 25 Test of the Correlation between ROA and CAPEX

Step 1	State the hypotheses.	H_0: ρ = 0 versus H_a: ρ ≠ 0
Step 2	Identify the appropriate test statistic.	$t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$ with 6 − 2 = 4 degrees of freedom.
Step 3	Specify the level of significance.	α = 5%.
Step 4	State the decision rule.	Critical t-values = ±2.776. Reject the null if the calculated t-statistic is less than −2.776 or greater than +2.776.

(continued)

Exhibit 25 (Continued)

Step 5	Calculate the test statistic.	$t = \dfrac{0.8945\sqrt{4}}{\sqrt{1 - 0.8001}} = 4.00131.$
Step 6	Make a decision.	Reject the null hypothesis of no correlation. There is sufficient evidence to indicate that the correlation between ROA and CAPEX is different from zero.

Another interesting feature of simple linear regression is that the test-statistic used to test the fit of the model (that is, the F-distributed test statistic) is related to the calculated t-statistic used to test whether the slope coefficient is equal to zero: $t^2 = F$; therefore, $4.00131^2 = 16.0104$.

What if instead we want to test whether there is a one-to-one relationship between ROA and CAPEX, implying a slope coefficient of 1.0. The hypotheses become H_0: $b_1 = 1$ and H_a: $b_1 \neq 1$. The calculated t-statistic is

$$t = \frac{1.25 - 1}{0.312398} = 0.80026.$$

This calculated test statistic falls within the bounds of the critical values, ±2.776, so we fail to reject the null hypothesis: There is not sufficient evidence to indicate that the slope is different from 1.0.

What if instead we want to test whether there is a positive slope or positive correlation, as our intuition suggests? In this case, all the steps are the same as in Exhibits 24 and 25 except the critical values because the tests are one sided. For a test of a positive slope or positive correlation, the critical value for a 5% level of significance is +2.132. We show the test of hypotheses for a positive slope and a positive correlation in Exhibit 26. Our conclusion is that there is sufficient evidence supporting both a positive slope and a positive correlation.

Exhibit 26 One-Sided Tests for the Slope and Correlation

		Test of the Slope	Test of the Correlation
Step 1	State the hypotheses.	H_0: $b_1 \leq 0$ versus H_a: $b_1 > 0$	H_0: $\rho \leq 0$ versus H_a: $\rho > 0$
Step 2	Identify the appropriate test statistic.	$t = \dfrac{\hat{b}_1 - B_1}{s_{\hat{b}_1}}$ with 6 − 2 = 4 degrees of freedom.	$t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$ with 6 − 2 = 4 degrees of freedom.
Step 3	Specify the level of significance.	α = 5%.	α = 5%.
Step 4	State the decision rule.	Critical t-value = 2.132. Reject the null if the calculated t-statistic is greater than 2.132.	Critical t-value = 2.132. Reject the null if the calculated t-statistic is greater than 2.132.
Step 5	Calculate the test statistic.	$t = \dfrac{1.25 - 0}{0.312398} = 4.00131$	$t = \dfrac{0.8945\sqrt{4}}{\sqrt{1 - 0.8001}} = 4.00131$
Step 6	Make a decision.	Reject the null hypothesis. There is sufficient evidence to indicate that the slope is greater than zero.	Reject the null hypothesis. There is sufficient evidence to indicate that the correlation is greater than zero.

5.2 Hypothesis Tests of the Intercept

There are occasions when we want to test whether the population intercept is a specific value. As a reminder on how to interpret the intercept, consider the simple linear regression with a company's revenue growth rate as the dependent variable (Y) and the GDP growth rate of its home country as the independent variable (X). The intercept is the company's revenue growth rate if the GDP growth rate is 0%.

The equation for the standard error of the intercept, $s_{\hat{b}_0}$, is

$$s_{\hat{b}_0} = \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}. \qquad (17)$$

We can test whether the intercept is different from the hypothesized value, B_0, by comparing the estimated intercept $(\hat{b}_0)$ with the hypothesized intercept and then dividing the difference by the standard error of the intercept:

$$t_{intercept} = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}} = \frac{\hat{b}_0 - B_0}{\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}.$$

In the ROA regression example, the intercept is 4.875%. Suppose we want to test whether the intercept is greater than 3%. The one-sided hypothesis test is shown in Exhibit 27. As you can see, we reject the null hypothesis. In other words, there is sufficient evidence that if there are no capital expenditures (CAPEX = 0), ROA is greater than 3%.

Exhibit 27 Test of Hypothesis for Intercept for Regression of ROA on CAPEX

Step 1	State the hypotheses.	H_0: $b_0 \le 3\%$ versus H_a: $b_0 > 3\%$
Step 2	Identify the appropriate test statistic.	$t_{intercept} = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}}$ with 6 − 2 = 4 degrees of freedom.
Step 3	Specify the level of significance.	α = 5%.
Step 4	State the decision rule.	Critical t-value = 2.132. Reject the null if the calculated t-statistic is greater than 2.132.
Step 5	Calculate the test statistic.	$t_{intercept} = \frac{4.875 - 3.0}{\sqrt{\frac{1}{6} + \frac{6.1^2}{122.64}}} = \frac{1.875}{0.68562} = 2.73475$
Step 6	Make a decision.	Reject the null hypothesis. There is sufficient evidence to indicate that the intercept is greater than 3%.

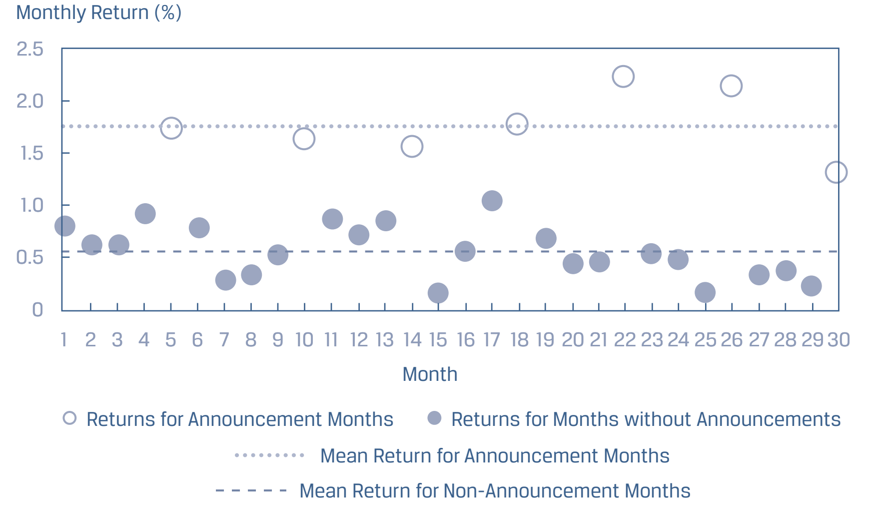
5.3 Hypothesis Tests of Slope When Independent Variable Is an Indicator Variable

Suppose we want to examine whether a company's quarterly earnings announcements influence its monthly stock returns. In this case, we could use an **indicator variable**, or dummy variable, that takes on only the values 0 or 1 as the independent variable. Consider the case of a company's monthly stock returns over a 30-month period. A simple linear regression model for investigating this question would be monthly returns, RET, regressed on the indicator variable, EARN, that takes on a value of 0 if there is no earnings announcement that month and 1 if there is an earnings announcement:

$$\text{RET}_i = b_0 + b_1\text{EARN}_i + \varepsilon_i.$$

This regression setup allows us to test whether there are different returns for earnings-announcement months versus non-earnings-announcement months. The observations and regression results are shown graphically in Exhibit 28.

Exhibit 28 Earnings Announcements, Dummy Variable, and Stock Returns



Clearly there are some months in which the returns are different from other months, and these correspond to months in which there was an earnings announcement. We estimate the simple linear regression model and perform hypothesis testing in the same manner as if the independent variable were a continuous variable. In a simple linear regression, the interpretation of the intercept is the predicted value of the dependent variable if the indicator variable is zero. Moreover, the slope, when the indicator variable is 1, is the difference in the means if we grouped the observations by the indicator variable. The results of the regression are given in Panel A of Exhibit 29.

Exhibit 29 Regression and Test of Differences Using an Indicator Variable

A. Regression Estimation Results

	Estimated Coefficients	Standard Error of Coefficients	Calculated Test Statistic
Intercept	0.5629	0.0560	10.0596
EARN	1.2098	0.1158	10.4435

Degrees of freedom = 28.
Critical t-values = +2.0484 (5% significance).

B. Test of Differences in Means

	RET for Earnings-Announcement Months	RET for Non-Earnings-Announcement Months	Difference in Means
Mean	1.7727	0.5629	1.2098
Variance	0.1052	0.0630	
Observations	7	23	
Pooled variance			0.07202
Calculated test statistic			10.4435

Degrees of freedom = 28.
Critical t-values = +2.0484 (5% significance).

We can see the following from Panel A of Exhibit 29:

- The intercept (0.5629) is the mean of the returns for non-earnings-announcement months.
- The slope coefficient (1.2098) is the difference in means of returns between earnings-announcement and non-announcement months.
- We reject the null hypothesis that the slope coefficient on EARN is equal to zero. We also reject the null hypothesis that the intercept is zero. The reason is that in both cases, the calculated test statistic exceeds the critical t-value.

We could also test whether the mean monthly return is the same for both the non-earnings-announcement months and the earnings-announcement months by testing the following:

$$H_0 : \mu_{RETearnings} = \mu_{RETNon-earnings} \text{ and } H_a : \mu_{RETearnings} \neq \mu_{RETNon-earnings}$$

The results of this hypothesis test are gleaned from Panel B of Exhibit 29. As you can see, we reject the null hypothesis that there is no difference in the mean RET for the earnings-announcement and non-earnings-announcements months at the 5% level of significance, since the calculated test statistic (10.4435) exceeds the critical value (2.0484).

5.4 Test of Hypotheses: Level of Significance and *p*-Values

The choice of significance level in hypothesis testing is always a matter of judgment. Analysts often choose the 0.05 level of significance, which indicates a 5% chance of rejecting the null hypothesis when, in fact, it is true (a Type I error, or false positive). Of course, decreasing the level of significance from 0.05 to 0.01 decreases the probability of Type I error, but it also increases the probability of Type II error—failing to reject the null hypothesis when, in fact, it is false (that is, a false negative).

The p-value is the smallest level of significance at which the null hypothesis can be rejected. The smaller the p-value, the smaller the chance of making a Type I error (i.e., rejecting a true null hypothesis), so the greater the likelihood the regression model is valid. For example, if the p-value is 0.005, we reject the null hypothesis that the true parameter is equal to zero at the 0.5% significance level (99.5% confidence). In most software packages, the p-values provided for regression coefficients are for a test of null hypothesis that the true parameter is equal to zero against the alternative that the parameter is not equal to zero.

In our ROA regression example, the calculated t-statistic for the test of whether the slope coefficient is zero is 4.00131. The p-value corresponding to this test statistic is 0.008, which means there is just a 0.8% chance of rejecting the null hypotheses when it is true. Comparing this p-value with the level of significance of 5% (and critical values of ±2.776) leads us to easily reject the null hypothesis of H_0: $b_1 = 0$.

How do we determine the p-values? Since this is the area in the distribution outside the calculated test statistic, we need to resort to software tools. For the p-value corresponding to the t = 4.00131 from the ROA regression example, we could use the following:

- **Excel** 1-T.DIST(4.00131,4,TRUE))*2
- **R** (1-pt(4.00131,4))*2
- **Python** from scipy.stats import t *and* (1 - t.cdf(4.00131,4))*2

EXAMPLE 6

Hypothesis Testing of Simple Linear Regression Results

An analyst is interested in interpreting the results of and performing tests of hypotheses for the market model estimation that regresses the daily return on ABC stock on the daily return on the fictitious Europe–Asia–Africa (EAA) Equity Index, his proxy for the stock market. He has generated the regression results presented in Exhibit 30.

Exhibit 30 Selected Results of Estimation of Market Model for ABC Stock

Standard error of the estimate (s_e)	1.26
Standard deviation of ABC stock returns	0.80
Standard deviation of EAA Equity Index returns	0.70
Number of observations	1,200
	Coefficients
Intercept	0.010
Slope of EAA Equity Index returns	0.982

1 If the critical t-values are ±1.96 (at the 5% significance level), is the slope coefficient different from zero?

2 If the critical t-values are ±1.96 (at the 5% significance level), is the slope coefficient different from 1.0?

Solutions

1 First, we calculate the variation of the independent variable using the standard deviation of the independent variable:

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} \times (n-1).$$

So,

$$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 0.70^2 \times 1{,}199 = 587.51.$$

Next, the standard error of the estimated slope coefficient is

$$s_{\hat{b}_1} = \frac{s_e}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}} = \frac{1.26}{\sqrt{587.51}} = 0.051983,$$

and the test statistic is

$$t = \frac{\hat{b}_1 - B_1}{s_{\hat{b}_1}} = \frac{0.982 - 0}{0.051983} = 18.89079$$

The calculated test statistic is outside the bounds of ±1.96, so we reject the null hypothesis of a slope coefficient equal to zero.

2 The calculated test statistic for the test of whether the slope coefficient is equal to 1.0 is

$$t = \frac{0.982 - 1}{0.051983} = -0.3463.$$

The calculated test statistic is within the bounds of ±1.96, so we fail to reject the null hypothesis of a slope coefficient equal to 1.0, which is evidence that the true population slope may be 1.0.

6 PREDICTION USING SIMPLE LINEAR REGRESSION AND PREDICTION INTERVALS

g calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable

Financial analysts often want to use regression results to make predictions about a dependent variable. For example, we might ask, "How fast will the sales of XYZ Corporation grow this year if real GDP grows by 4%?" But we are not merely interested in making these forecasts; we also want to know how certain we can be about the forecasts' results. A forecasted value of the dependent variable, $\hat{Y}_f$, is determined using the estimated intercept and slope, as well as the expected or forecasted independent variable, X_f:

$$\hat{Y}_f = \hat{b}_0 + \hat{b}_1 X_f \quad (18)$$

In our ROA regression model, if we forecast a company's CAPEX to be 6%, the forecasted ROA based on our estimated equation is 12.375%:

$$\hat{Y}_f = 4.875 + \left(1.25 \times 6\right) = 12.375$$

However, we need to consider that the estimated regression line does not describe the relation between the dependent and independent variables perfectly; it is an average of the relation between the two variables. This is evident because the residuals are not all zero.

Therefore, an interval estimate of the forecast is needed to reflect this uncertainty. The estimated variance of the prediction error, s_f^2, of Y, given X, is

$$s_f^2 = s_e^2\left[1 + \frac{1}{n} + \frac{\left(X_f - \bar{X}\right)^2}{(n-1)s_X^2}\right] = s_e^2\left[1 + \frac{1}{n} + \frac{\left(X_f - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right],$$

and the **standard error of the forecast** is

$$s_f = s_e\sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}. \qquad (19)$$

The standard error of the forecast depends on

- the standard error of the estimate, s_e;
- the number of observations, n;
- the forecasted value of the independent variable, X_f, used to predict the dependent variable and its deviation from the estimated mean, $\bar{X}$; and
- the variation of the independent variable.

We can see the following from the equation for the standard error of the forecast:

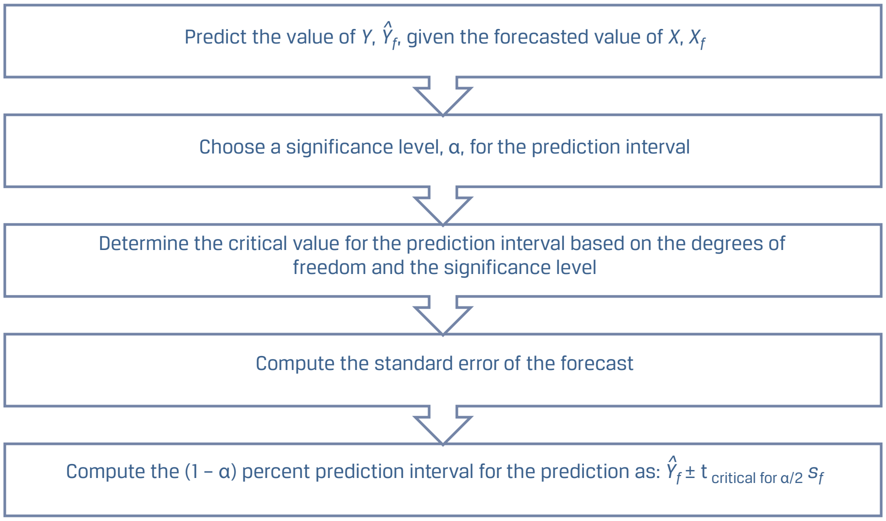
1. The better the fit of the regression model, the smaller the standard error of the estimate (s_e) and, therefore, the smaller standard error of the forecast.
2. The larger the sample size (n) in the regression estimation, the smaller the standard error of the forecast.
3. The closer the forecasted independent variable (X_f) is to the mean of the independent variable ($\bar{X}$) used in the regression estimation, the smaller the standard error of the forecast.

Once we have this estimate of the standard error of the forecast, determining a prediction interval around the predicted value of the dependent variable ($\hat{Y}_f$) is very similar to estimating a confidence interval around an estimated parameter. The prediction interval is

$$\hat{Y}_f \pm t_{critical\ for\ \alpha/2}s_f. \qquad (20)$$

We outline the steps for developing the prediction interval in Exhibit 31.

Exhibit 31 Creating a Prediction Interval around the Predicted Dependent Variable



For our ROA regression model, given that the forecasted value of CAPEX is 6.0, the predicted value of Y is 12.375:

$$\hat{Y}_f = 4.875 + 1.25X_f = 4.875 + (1.25 \times 6.0) = 12.375.$$

Assuming a 5% significance level (α), two sided, with n – 2 degrees of freedom (so, df = 4), the critical values for the prediction interval are ±2.776.

The standard error of the forecast is

$$s_f = 3.459588\sqrt{1 + \frac{1}{6} + \frac{(6 - 6.1)^2}{122.640}} = 3.459588\sqrt{1.166748} = 3.736912.$$

The 95% prediction interval then becomes

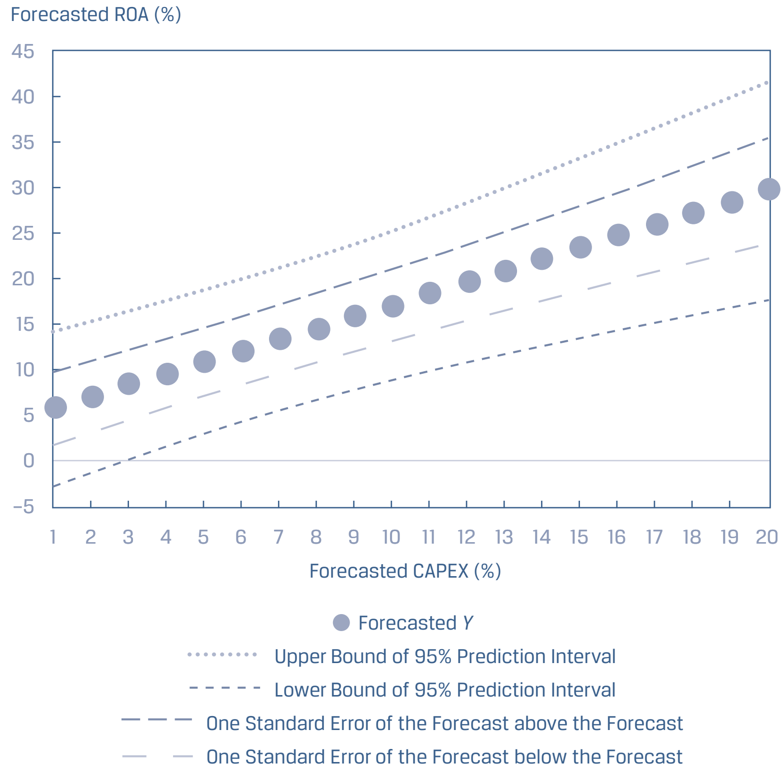
$$12.375 \pm 2.776(3.736912)$$

$$12.375 \pm 10.3737$$

$$\{2.0013 < \hat{Y}_f < 22.7487\}$$

For our ROA regression example, we can see how the standard error of the forecast (s_f) changes as our forecasted value of the independent variable gets farther from the mean of the independent variable $(X_f - \bar{X})$ in Exhibit 32. The mean of CAPEX is 6.1%, and the band that represents one standard error of the forecast, above and below the forecast, is minimized at that point and increases as the independent variable gets farther from $\bar{X}$.

Exhibit 32 ROA Forecasts and Standard Error of the Forecast



EXAMPLE 7

Predicting Net Profit Margin Using R&D Spending

Suppose we want to forecast a company's net profit margin (NPM) based on its research and development expenditures scaled by revenues (RDR), using the model estimated in Example 2 and the details provided in Exhibit 8. The regression model was estimated using data on eight companies as

$$\hat{Y}_f = 16.5 - 1.3X_f,$$

with a standard error of the estimate (s_e) of 1.8618987 and variance of RDR, $\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}$, of 4.285714, as given.

1 What is the predicted value of NPM if the forecasted value of RDR is 5?

2 What is the standard error of the forecast (s_f) if the forecasted value of RDR is 5?

3 What is the 95% prediction interval for the predicted value of NPM using critical t-values (df = 6) of ±2.447?

4 What is the predicted value of NPM if the forecasted value of RDR is 15?

5 What is the standard error of the forecast if the forecasted value of RDR is 15?

6 What is the 95% prediction interval for the predicted value of NPM using critical t-values (df = 6) of ±2.447?

Solutions

1 The predicted value of NPM is 10: 16.5 − (1.3 × 5) = 10.

2 To derive the standard error of the forecast (s_f), we first have to calculate the variation of RDR. Then, we have the all the pieces to calculate s_f:

$$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 4.285714 \times 7 = 30.$$

$$s_f = 1.8618987\sqrt{1 + \frac{1}{8} + \frac{(5 - 7.5)^2}{30}} = 2.1499.$$

3 The 95% prediction interval for the predicted value of NPM is

$$\left\{10 \pm 2.447(2.1499)\right\}$$

$$\left\{4.7392 < \hat{Y}_f < 15.2608\right\}$$

4 The predicted value of NPM is −3: 16.5 − (1.3 × 15) = −3.

5 To derive the standard error of the forecast, we first must calculate the variation of RDR. Then, we can calculate s_f:

$$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 4.285714 \times 7 = 30.$$

$$s_f = 1.8618987\sqrt{1 + \frac{1}{8} + \frac{(15 - 7.5)^2}{30}} = 3.2249.$$

6 The 95% prediction interval for the predicted value of NPM is

$$\left\{-3 \pm 2.447(3.2249)\right\}$$

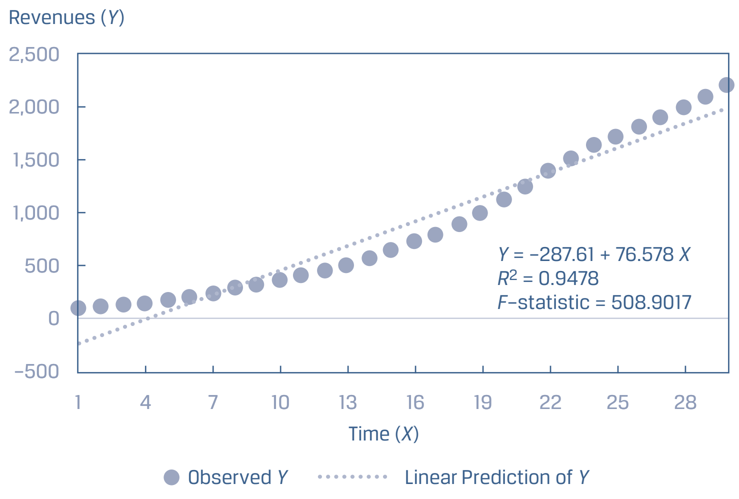
$$\left\{-10.8913 < \hat{Y}_f < 4.8913\right\}$$

FUNCTIONAL FORMS FOR SIMPLE LINEAR REGRESSION 7

h describe different functional forms of simple linear regressions

Not every set of independent and dependent variables has a linear relation. In fact, we often see non-linear relationships in economic and financial data. Consider the revenues of a company over time illustrated in Exhibit 33, with revenues as the dependent (Y) variable and time as the independent (X) variable. Revenues grow at a rate of 15% per year for several years, but then the growth rate eventually declines to just 5% per year. Estimating this relationship as a simple linear model would understate the dependent variable, revenues, for some ranges of the independent variable, time, and would overstate it for other ranges of the independent variable.

Exhibit 33 Company Revenues over Time



We can still use the simple linear regression model, but we need to modify either the dependent or the independent variables to make it work well. This is the case with many different financial or economic data that you might use as dependent and independent variables in your regression analysis.

There are several different functional forms that can be used to potentially transform the data to enable their use in linear regression. These transformations include using the log (i.e., natural logarithm) of the dependent variable, the log of the independent variable, the reciprocal of the independent variable, the square of the independent variable, or the differencing of the independent variable. We illustrate and discuss three often-used functional forms, each of which involves log transformation:

1 the **log-lin model**, in which the dependent variable is logarithmic but the independent variable is linear;

2 the **lin-log model**, in which the dependent variable is linear but the independent variable is logarithmic; and

3 the **log-log model**, where both the dependent and independent variables are in logarithmic form.

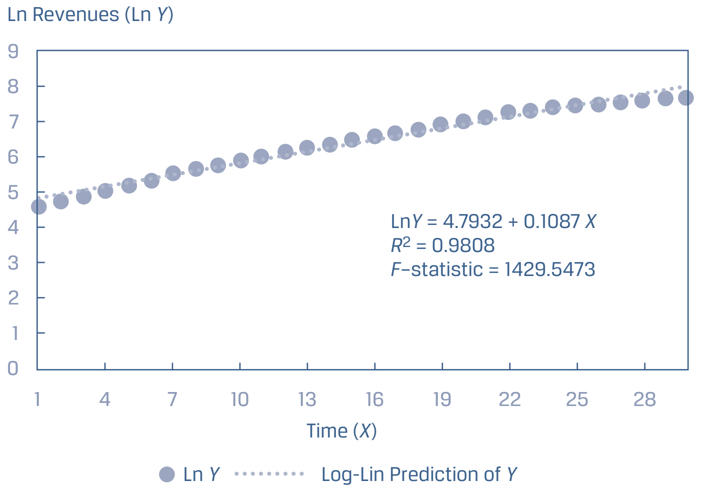
7.1 The Log-Lin Model

In the log-lin model, the dependent variable is in logarithmic form and the independent variable is not, as follows:

$$\ln Y_i = b_0 + b_1 X_i. \quad (21)$$

The slope coefficient in this model is the relative change in the dependent variable for an absolute change in the independent variable. We can transform the Y variable (revenues) in Exhibit 33 into its natural log (ln) and then fit the regression line, as we show in Exhibit 34. From this chart, we see that the log-lin model is a better fitting model than the simple linear regression model.

Exhibit 34 Log-Lin Model Applied to Company Revenues over Time



It is important to note that in working with a log-lin model, you must take care when making a forecast. For example, suppose the estimated regression model is ln $Y = -7 + 2X$. If X is 2.5%, then the forecasted value of ln Y is −2. In this case, the predicted value of Y is the antilog of −2, or $e^{-2} = 0.135335$. Another caution is that you cannot directly compare a log-lin model with a lin-lin model (that is, the regression of Y on X without any transformation) because the dependent variables are not in the same form – we would have to transform the R^2 and F-statistic to enable a comparison. However, looking at the residuals is helpful.

7.2 The Lin-Log Model

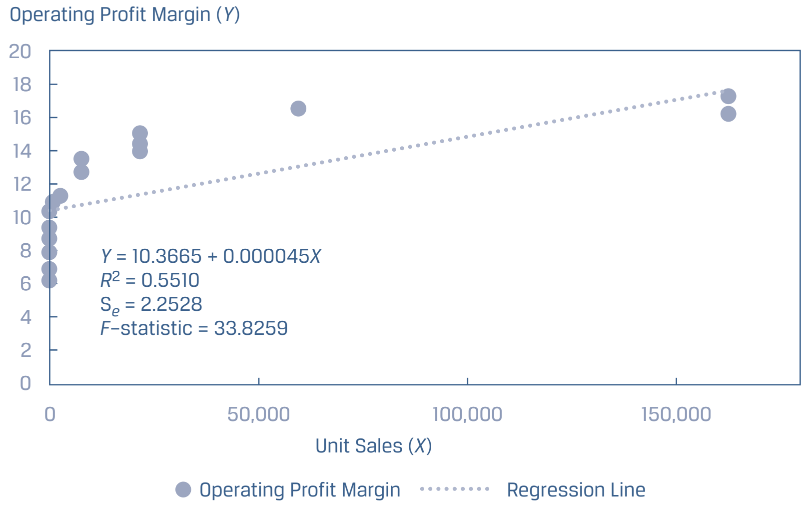
The lin-log model is similar to the log-lin model, but only the independent variable is in logarithmic form:

$$Y_i = b_0 + b_1 \ln X_i, \quad \textbf{(22)}$$

The slope coefficient in this regression model provides the absolute change in the dependent variable for a relative change in the independent variable.

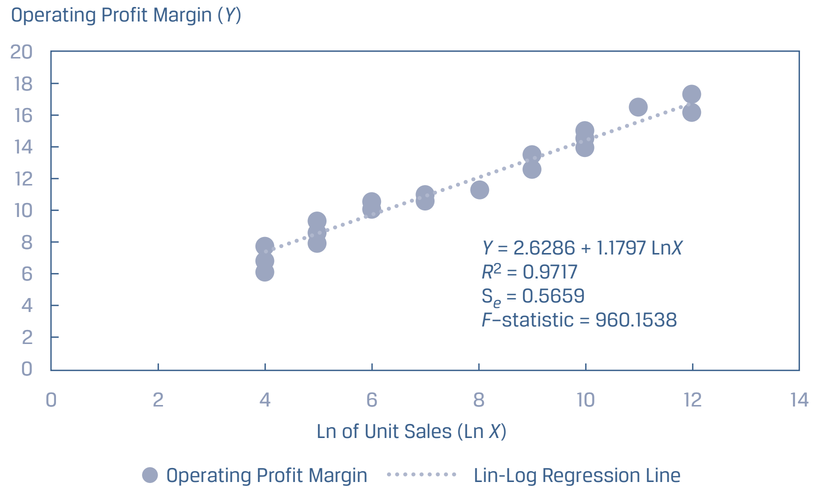
Suppose an analyst is examining the cross-sectional relationship between operating profit margin, the dependent variable (Y), and unit sales, the independent variable (X), and gathers data on a sample of 30 companies. The scatter plot and regression line for these observations are shown in Exhibit 35. Although the slope is different from zero at the 5% level (the calculated t-statistic on the slope is 5.8616, compared with critical t-values of ±2.048), given the R^2 of 55.10%, the issue is whether we can get a better fit by using a different functional form.

Exhibit 35 Relationship between Operating Profit Margin and Unit Sales



If instead we use the natural log of the unit sales as the independent variable in our model, we get a very different picture, as shown in Exhibit 36. The R^2 for the model of operating profit margin regressed on the natural log of unit sales jumps to 97.17%. Since the dependent variable is the same in both the original and transformed models, we can compare the standard error of the estimate: 2.2528 with the original independent variable and a much lower 0.5629 with the transformed independent variable. Clearly the log-transformed explanatory variable has resulted in a better fitting model.

Exhibit 36 Relationship Between Operating Profit Margin and Natural Logarithm of Unit Sales



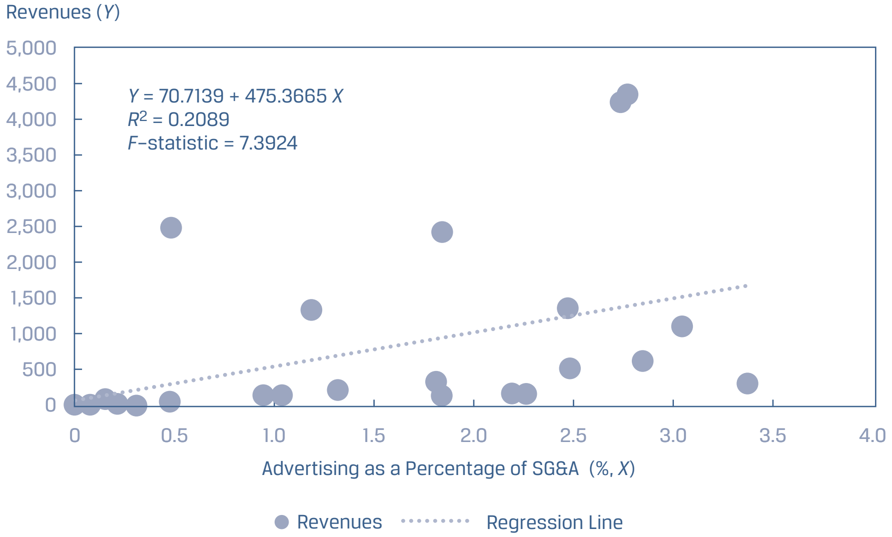
7.3 The Log-Log Model

The log-log model, in which both the dependent variable and the independent variable are linear in their logarithmic forms, is also referred to as the double-log model.

$$\ln Y_i = b_0 + b_1 \ln X_i. \tag{23}$$

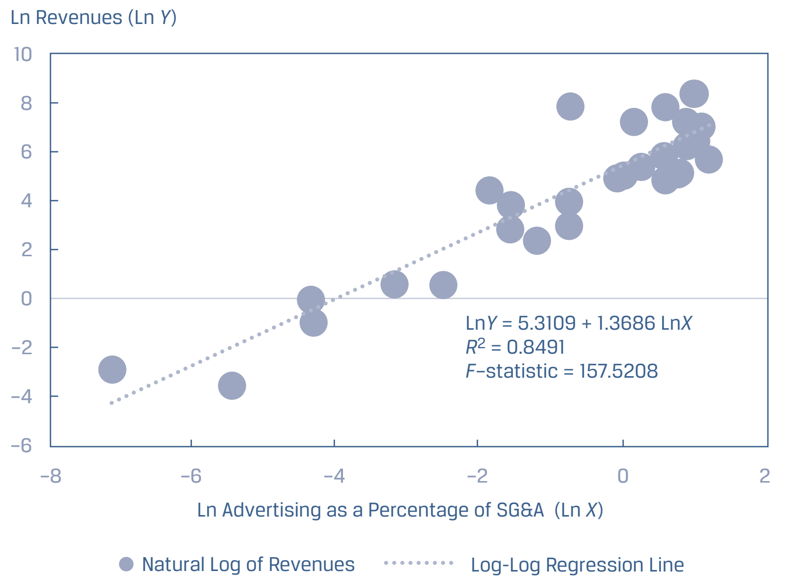
This model is useful in calculating elasticities because the slope coefficient is the relative change in the dependent variable for a relative change in the independent variable. Consider a cross-sectional model of company revenues (the Y variable) regressed on advertising spending as a percentage of selling, general, and administrative expenses, ADVERT (the X variable). As shown in Exhibit 37, a simple linear regression model results in a shallow regression line, with a coefficient of determination of just 20.89%.

Exhibit 37 Fitting a Linear Relation Between Revenues and Advertising Spending



However, if instead we use the natural logarithms of both the revenues and ADVERT, we get a much different picture of this relationship. As shown in Exhibit 38, the estimated regression line has a significant positive slope; the log-log model's R^2 increases by more than four times, from 20.89% to 84.91%; and the F-statistic jumps from 7.39 to 157.52. So, using the log-log transformation dramatically improves the regression model fit relative to our data.

Exhibit 38 Fitting a Log-Log Model of Revenues and Advertising Spending



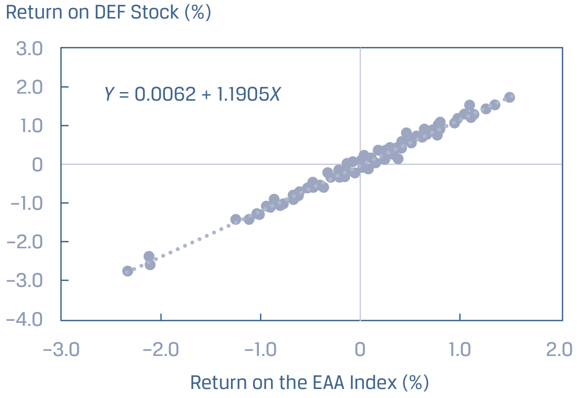
7.4 Selecting the Correct Functional Form

The key to fitting the appropriate functional form of a simple linear regression is examining the goodness of fit measures—the coefficient of determination (R^2), the F-statistic, and the standard error of the estimate (s_e)—as well as examining whether there are patterns in the residuals. In addition to fit statistics, most statistical packages provide plots of residuals as part of the regression output, which enables you to visually inspect the residuals. To reiterate an important point, what you want to see in these plots is random residuals.

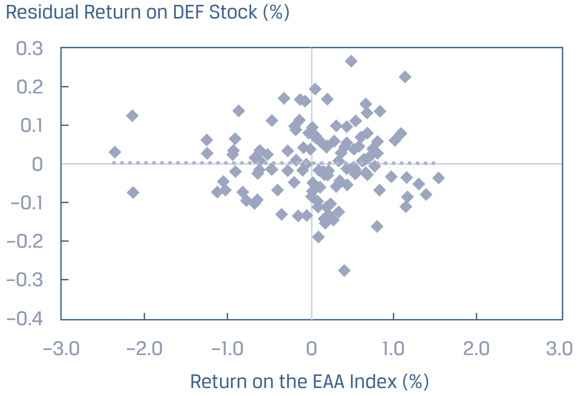
As an example, consider the relationship between the monthly returns on DEF stock and the monthly returns of the EAA Equity Index, as depicted in Panel A of Exhibit 39, with the regression line indicated. Using the equation for this regression line, we calculate the residuals and plot them against the EAA Equity Index, as shown in Panel B of Exhibit 39. The residuals appear to be random, bearing no relation to the independent variable. The distribution of the residuals, shown in Panel C of Exhibit 39, shows that the residuals are approximately normal. Using statistical software, we can investigate further by examining the distribution of the residuals, including using a normal probability plot or statistics to test for normality of the residuals.

Exhibit 39 Monthly Returns on DEF Stock Regressed on Returns on the EAA Index

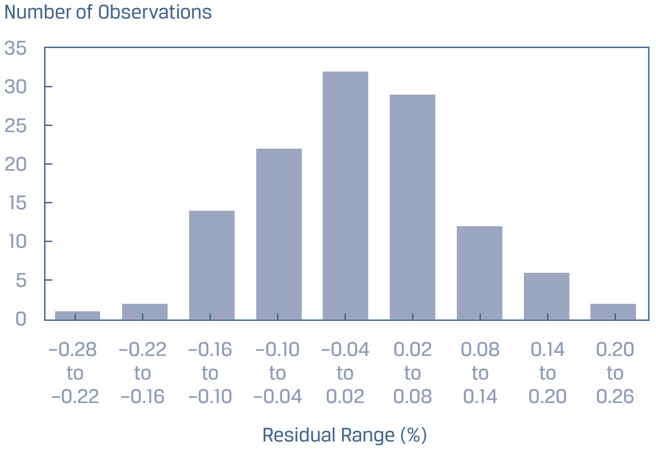
A. Scatterplot of Returns on DEF Stock and Return on the EAA Index



B. Scatterplot of Residuals and the Returns on the EAA Index



C. Histogram of Residuals



EXAMPLE 8

Comparing Functional Forms

An analyst is investigating the relationship between the annual growth in consumer spending (CONS) in a country and the annual growth in the country's GDP (GGDP). The analyst estimates the following two models:

	Model 1	Model 2
	$GGDP_i = b_0 + b_1 CONS_i + \varepsilon_i$	$GGDP_i = b_0 + b_1 \ln(CONS_i) + \varepsilon_i$
Intercept	1.040	1.006
Slope	0.669	1.994
R^2	0.788	0.867
Standard error of the estimate	0.404	0.320
F-statistic	141.558	247.040

1 Identify the functional form used in these models.

2 Explain which model has better goodness-of-fit with the sample data.

Solution

1 Model 1 is the simple linear regression with no variable transformation, whereas Model 2 is a lin-log model with the natural log of the variable CONS as the independent variable.

2 The lin-log model, Model 2, fits the data better. Since the dependent variable is the same for the two models, we can compare the fit of the models using either the relative measures (R^2 or F-statistic) or the absolute measure of fit, the standard error of the estimate. The standard error of the estimate is lower for Model 2, whereas the R^2 and F-statistic are higher for Model 2 compared with Model 1.

SUMMARY

- The dependent variable in a linear regression is the variable whose variability the regression model tries to explain. The independent variable is the variable whose variation the researcher uses to explain the variation of the dependent variable.
- If there is one independent variable in a linear regression and there are n observations of the dependent and independent variables, the regression model is $Y_i = b_0 + b_1X_i + \varepsilon_i$, $i = 1, \ldots, n$, where Y_i is the dependent variable, X_i is the independent variable, and ε_i is the error term. In this model, the coefficients b_0 and b_1 are the population intercept and slope, respectively.

- The intercept is the expected value of the dependent variable when the independent variable has a value of zero. The slope coefficient is the estimate of the population slope of the regression line and is the expected change in the dependent variable for a one-unit change in the independent variable.
- The assumptions of the classic simple linear regression model are as follows:
  - Linearity: A linear relation exists between the dependent variable and the independent variable.
  - Homoskedasticity: The variance of the error term is the same for all observations.
  - Independence: The error term is uncorrelated across observations.
  - Normality: The error term is normally distributed.
- The estimated parameters in a simple linear regression model minimize the sum of the squared errors.
- The coefficient of determination, or R^2, measures the percentage of the total variation in the dependent variable explained by the independent variable.
- To test the fit of the simple linear regression, we can calculate an F-distributed test statistic and test the hypotheses H_0: $b_1 = 0$ versus H_a: $b_1 \neq 0$, with 1 and $n - 2$ degrees of freedom.
- The standard error of the estimate is an absolute measure of the fit of the model calculated as the square root of the mean square error.
- We can evaluate a regression model by testing whether the population value of a regression coefficient is equal to a particular hypothesized value. We do this by calculating a t-distributed test statistic that compares the estimated parameter with the hypothesized parameter, dividing this difference by the standard error of the coefficient.
- An indicator (or dummy) variable takes on only the values 0 or 1 and can be used as the independent variable in a simple linear regression. In such a model, the interpretation of the intercept is the predicted value of the dependent variable if the indicator variable is 0, and when the indicator variable is 1, the slope is the difference in the means if we grouped the observations by the indicator variable.
- We calculate a prediction interval for a regression coefficient using the estimated coefficient, the standard error of the estimated coefficient, and the critical value for the t-distributed test statistic based on the level of significance and the appropriate degrees of freedom, which are $n - 2$ for simple regression.
- We can make predictions for the dependent variable using an estimated linear regression by inserting the forecasted value of the independent variable into the estimated model.
- The standard error of the forecast is the product of the standard error of the estimate and a term that reflects the sample size of the regression, the variation of the independent variable, and the deviation between the forecasted value of the independent variable and the mean of the independent variable in the regression.

- The prediction interval for a particular forecasted value of the dependent variable is formed by using the forecasted value of the dependent variable and extending above and below this value a quantity that reflects the critical *t*-value corresponding to the degrees of freedom, the level of significance, and the standard error of the forecast.
- If the relationship between the independent variable and the dependent variable is not linear, we can often transform one or both of these variables to convert this relation to a linear form, which then allows the use of simple linear regression.

PRACTICE PROBLEMS

1 Julie Moon is an energy analyst examining electricity, oil, and natural gas consumption in different regions over different seasons. She ran a simple regression explaining the variation in energy consumption as a function of temperature. The total variation of the dependent variable was 140.58, and the explained variation was 60.16. She had 60 monthly observations.

A Calculate the coefficient of determination.

B Calculate the *F*-statistic to test the fit of the model.

C Calculate the standard error of the estimate of the regression estimation.

D Calculate the sample standard deviation of monthly energy consumption.

2 Homoskedasticity is best described as the situation in which the variance of the residuals of a regression is:

A zero.

B normally distributed.

C constant across observations.

The following information relates to Questions 3–6

An analyst is examining the annual growth of the money supply for a country over the past 30 years. This country experienced a central bank policy shift 15 years ago, which altered the approach to the management of the money supply. The analyst estimated a model using the annual growth rate in the money supply regressed on the variable (SHIFT) that takes on a value of 0 before the policy shift and 1 after. She estimated the following:

	Coefficients	Standard Error	*t*-Stat.
Intercept	5.767264	0.445229	12.95348
SHIFT	−5.13912	0.629649	−8.16188

Critical *t*-values, level of significance of 0.05:

One-sided, left side: −1.701

One-sided, right side: +1.701

Two-sided: ±2.048

3 The variable SHIFT is best described as:

A an indicator variable.

B a dependent variable.

C a continuous variable.

4 The interpretation of the intercept is the mean of the annual growth rate of the money supply:

A over the enter entire period.

© 2021 CFA Institute. All rights reserved.

B after the shift in policy.

C before the shift in policy.

5 The interpretation of the slope is the:

A change in the annual growth rate of the money supply per year.

B average annual growth rate of the money supply after the shift in policy.

C difference in the average annual growth rate of the money supply from before to after the shift in policy.

6 Testing whether there is a change in the money supply growth after the shift in policy, using a 0.05 level of significance, we conclude that there is:

A sufficient evidence that the money supply growth changed.

B not enough evidence that the money supply growth is different from zero.

C not enough evidence to indicate that the money supply growth changed.

---

7 You are examining the results of a regression estimation that attempts to explain the unit sales growth of a business you are researching. The analysis of variance output for the regression is given in the following table. The regression was based on five observations (n = 5).

Source	df	Sum of Squares	Mean Square	*F*	*p*-Value
Regression	1	88.0	88.0	36.667	0.00904
Residual	3	7.2	2.4		
Total	4	95.2			

A Calculate the sample variance of the dependent variable using information in the table.

B Calculate the coefficient of determination for this estimated model.

C What hypothesis does the F-statistic test?

D Is the F-test significant at the 0.05 significance level?

E Calculate the standard error of the estimate.

8 An economist collected the monthly returns for KDL's portfolio and a diversified stock index. The data collected are shown in the following table:

Month	Portfolio Return (%)	Index Return (%)
1	1.11	−0.59
2	72.10	64.90
3	5.12	4.81
4	1.01	1.68
5	−1.72	−4.97
6	4.06	−2.06

The economist calculated the correlation between the two returns and found it to be 0.996. The regression results with the KDL return as the dependent variable and the index return as the independent variable are given as follows:

Regression Statistics	
R^2	0.9921
Standard error	2.8619
Observations	6

Source	df	Sum of Squares	Mean Square	F	p-Value
Regression	1	4,101.6205	4,101.6205	500.7921	0.0000
Residual	4	32.7611	8.1903		
Total	5	4,134.3815			

	Coefficients	Standard Error	t-Statistic	p-Value
Intercept	2.2521	1.2739	1.7679	0.1518
Index return (%)	1.0690	0.0478	22.3784	0.0000

When reviewing the results, Andrea Fusilier suspected that they were unreliable. She found that the returns for Month 2 should have been 7.21% and 6.49%, instead of the large values shown in the first table. Correcting these values resulted in a revised correlation of 0.824 and the following revised regression results:

Regression Statistics	
R^2	0.6784
Standard error	2.0624
Observations	6

Source	df	Sum of Squares	Mean Square	F	p-Value
Regression	1	35.8950	35.8950	8.4391	0.044
Residual	4	17.0137	4.2534		
Total	5	52.91			

	Coefficients	Standard Error	t-Statistic	p-Value
Intercept	2.2421	0.8635	2.5966	0.060
Slope	0.6217	0.2143	2.9050	0.044

Explain how the bad data affected the results.

The following information relates to Questions 9–12

Kenneth McCoin, CFA, is a challenging interviewer. Last year, he handed each job applicant a sheet of paper with the information in the following table, and he then asked several questions about regression analysis. Some of McCoin's questions, along with a sample of the answers he received to each, are given below. McCoin told the applicants that the independent variable is the ratio of net income to sales for restaurants with a market cap of more than $100 million and the dependent variable is the ratio of cash flow from operations to sales for those restaurants. Which of the choices provided is the best answer to each of McCoin's questions?

Regression Statistics	
R^2	0.7436
Standard error	0.0213
Observations	24

Source	**df**	**Sum of Squares**	**Mean Square**	***F***	***p*-Value**
Regression	1	0.029	0.029000	63.81	0
Residual	22	0.010	0.000455		
Total	23	0.040			

	Coefficients	**Standard Error**	***t*-Statistic**	***p*-Value**
Intercept	0.077	0.007	11.328	0
Net income to sales (%)	0.826	0.103	7.988	0

9 The coefficient of determination is *closest* to:

A 0.7436.

B 0.8261.

C 0.8623.

10 The correlation between X and Y is *closest* to:

A −0.7436.

B 0.7436.

C 0.8623.

11 If the ratio of net income to sales for a restaurant is 5%, the predicted ratio of cash flow from operations (CFO) to sales is *closest* to:

A −4.054.

B 0.524.

C 4.207.

12 Is the relationship between the ratio of cash flow to operations and the ratio of net income to sales significant at the 0.05 level?

A No, because the R^2 is greater than 0.05

B No, because the p-values of the intercept and slope are less than 0.05

C Yes, because the p-values for F and t for the slope coefficient are less than 0.05

The following information relates to Questions 13–17

Howard Golub, CFA, is preparing to write a research report on Stellar Energy Corp. common stock. One of the world's largest companies, Stellar is in the business of refining and marketing oil. As part of his analysis, Golub wants to evaluate the sensitivity of the stock's returns to various economic factors. For example, a client recently asked Golub whether the price of Stellar Energy Corp. stock has tended to rise following increases in retail energy prices. Golub believes the association between the two variables is negative, but he does not know the strength of the association.

Golub directs his assistant, Jill Batten, to study the relationships between (1) Stellar monthly common stock returns and the previous month's percentage change in the US Consumer Price Index for Energy (CPIENG) and (2) Stellar monthly common stock returns and the previous month's percentage change in the US Producer Price Index for Crude Energy Materials (PPICEM). Golub wants Batten to run both a correlation and a linear regression analysis. In response, Batten compiles the summary statistics shown in Exhibit 1 for 248 months. All the data are in decimal form, where 0.01 indicates a 1% return. Batten also runs a regression analysis using Stellar monthly returns as the dependent variable and the monthly change in CPIENG as the independent variable. Exhibit 2 displays the results of this regression model.

Exhibit 1 Descriptive Statistics

	Stellar Common Stock Monthly Return	Lagged Monthly Change	
		CPIENG	PPICEM
Mean	0.0123	0.0023	0.0042
Standard deviation	0.0717	0.0160	0.0534
Covariance, Stellar vs. CPIENG	−0.00017		
Covariance, Stellar vs. PPICEM	−0.00048		
Covariance, CPIENG vs. PPICEM	0.00044		
Correlation, Stellar vs. CPIENG	−0.1452		

Exhibit 2 Regression Analysis with CPIENG

Regression Statistics	
R^2	0.0211
Standard error of the estimate	0.0710
Observations	248

(continued)

Exhibit 2 (Continued)

Regression Statistics

	Coefficients	Standard Error	*t*-Statistic
Intercept	0.0138	0.0046	3.0275
CPIENG (%)	−0.6486	0.2818	−2.3014

Critical *t*-values

One-sided, left side: −1.651

One-sided, right side: +1.651

Two-sided: ±1.967

13 Which of the following best describes Batten's regression?

A Time-series regression

B Cross-sectional regression

C Time-series and cross-sectional regression

14 Based on the regression, if the CPIENG *decreases* by 1.0%, the expected return on Stellar common stock during the next period is *closest* to:

A 0.0073 (0.73%).

B 0.0138 (1.38%).

C 0.0203 (2.03%).

15 Based on Batten's regression model, the coefficient of determination indicates that:

A Stellar's returns explain 2.11% of the variability in CPIENG.

B Stellar's returns explain 14.52% of the variability in CPIENG.

C changes in CPIENG explain 2.11% of the variability in Stellar's returns.

16 For Batten's regression model, 0.0710 is the standard deviation of:

A the dependent variable.

B the residuals from the regression.

C the predicted dependent variable from the regression.

17 For the analysis run by Batten, which of the following is an *incorrect* conclusion from the regression output?

A The estimated intercept from Batten's regression is statistically different from zero at the 0.05 level of significance.

B In the month after the CPIENG declines, Stellar's common stock is expected to exhibit a positive return.

C Viewed in combination, the slope and intercept coefficients from Batten's regression are not statistically different from zero at the 0.05 level of significance.

The following information relates to Questions 18–26

Anh Liu is an analyst researching whether a company's debt burden affects investors' decision to short the company's stock. She calculates the short interest ratio (the ratio of short interest to average daily share volume, expressed in days) for 50 companies as of the end of 2016 and compares this ratio with the companies' debt ratio (the ratio of total liabilities to total assets, expressed in decimal form).

Liu provides a number of statistics in Exhibit 1. She also estimates a simple regression to investigate the effect of the debt ratio on a company's short interest ratio. The results of this simple regression, including the analysis of variance (ANOVA), are shown in Exhibit 2.

In addition to estimating a regression equation, Liu graphs the 50 observations using a scatter plot, with the short interest ratio on the vertical axis and the debt ratio on the horizontal axis.

Exhibit 1 Summary Statistics

Statistic	Debt Ratio X_i	Short Interest Ratio Y_i
Sum	19.8550	192.3000
Sum of squared deviations from the mean	$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 2.2225.$	$\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = 412.2042.$
Sum of cross-products of deviations from the mean	$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = -9.2430.$	

Exhibit 2 Regression of the Short Interest Ratio on the Debt Ratio

ANOVA	Degrees of Freedom (df)	Sum of Squares	Mean Square
Regression	1	38.4404	38.4404
Residual	48	373.7638	7.7867
Total	49	412.2042	

Regression Statistics	
R^2	0.0933
Standard error of estimate	2.7905
Observations	50

	Coefficients	Standard Error	*t*-Statistic
Intercept	5.4975	0.8416	6.5322
Debt ratio (%)	–4.1589	1.8718	–2.2219

Critical t-values for a 0.05 level of significance:

One-sided, left side: −1.677

One-sided, right side: +1.677

Two-sided: ±2.011

Liu is considering three interpretations of these results for her report on the relationship between debt ratios and short interest ratios:

Interpretation 1	Companies' higher debt ratios cause lower short interest ratios.
Interpretation 2	Companies' higher short interest ratios cause higher debt ratios.
Interpretation 3	Companies with higher debt ratios tend to have lower short interest ratios.

She is especially interested in using her estimation results to predict the short interest ratio for MQD Corporation, which has a debt ratio of 0.40.

18 Based on Exhibits 1 and 2, if Liu were to graph the 50 observations, the scatter plot summarizing this relation would be *best* described as:

A horizontal.

B upward sloping.

C downward sloping.

19 Based on Exhibit 1, the sample covariance is *closest to*:

A −9.2430.

B −0.1886.

C 8.4123.

20 Based on Exhibits 1 and 2, the correlation between the debt ratio and the short interest ratio is *closest to*:

A −0.3054.

B 0.0933.

C 0.3054.

21 Which of the interpretations *best* describes Liu's findings?

A Interpretation 1

B Interpretation 2

C Interpretation 3

22 The dependent variable in Liu's regression analysis is the:

A intercept.

B debt ratio.

C short interest ratio.

23 Based on Exhibit 2, the degrees of freedom for the t-test of the slope coefficient in this regression are:

A 48.

B 49.

C 50.

24 Which of the following should Liu conclude from the results shown in Exhibit 2?

A The average short interest ratio is 5.4975.

B The estimated slope coefficient is different from zero at the 0.05 level of significance.

C The debt ratio explains 30.54% of the variation in the short interest ratio.

25 Based on Exhibit 2, the short interest ratio expected for MQD Corporation is *closest* to:

A 3.8339.

B 5.4975.

C 6.2462.

26 Based on Liu's regression results in Exhibit 2, the *F*-statistic for testing whether the slope coefficient is equal to zero is *closest* to:

A −2.2219.

B 3.5036.

C 4.9367.

The following information relates to Questions 27–31

Elena Vasileva recently joined EnergyInvest as a junior portfolio analyst. Vasileva's supervisor asks her to evaluate a potential investment opportunity in Amtex, a multinational oil and gas corporation based in the United States. Vasileva's supervisor suggests using regression analysis to examine the relation between Amtex shares and returns on crude oil.

Vasileva notes the following assumptions of regression analysis:

Assumption 1 The error term is uncorrelated across observations.

Assumption 2 The variance of the error term is the same for all observations.

Assumption 3 The dependent variable is normally distributed.

Vasileva runs a regression of Amtex share returns on crude oil returns using the monthly data she collected. Selected data used in the regression are presented in Exhibit 1, and selected regression output is presented in Exhibit 2. She uses a 1% level of significance in all her tests.

Exhibit 1 Selected Data for Crude Oil Returns and Amtex Share Returns

	Oil Return (X_i)	Amtex Return (Y_i)	Cross-Product $(X_i - \bar{X})(Y_i - \bar{Y})$	Predicted Amtex Return $\hat{Y}_i$	Regression Residual $Y_i - \hat{Y}_i$	Squared Residual $(Y_i - \hat{Y}_i)^2$
Month 1	−0.032000	0.033145	−0.000388	0.002011	−0.031134	0.000969
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Month 36	0.028636	0.062334	0.002663	0.016282	−0.046053	0.002121
Sum			0.085598			0.071475
Average	−0.018056	0.005293				

Exhibit 2 Selected Regression Output, Dependent Variable: Amtex Share Return

	Coefficient	Standard Error
Intercept	0.0095	0.0078
Oil return	0.2354	0.0760

Critical t-values for a 1% level of significance:

One-sided, left side: −2.441

One-sided, right side: +2.441

Two-sided: ±2.728

Vasileva expects the crude oil return next month, Month 37, to be −0.01. She computes the standard error of the forecast to be 0.0469.

27 Which of Vasileva's assumptions regarding regression analysis is *incorrect*?

A Assumption 1

B Assumption 2

C Assumption 3

28 Based on Exhibit 1, the standard error of the estimate is *closest* to:

A 0.04456.

B 0.04585.

C 0.05018.

29 Based on Exhibit 2, Vasileva should reject the null hypothesis that:

A the slope is less than or equal to 0.15.

B the intercept is less than or equal to zero.

C crude oil returns do not explain Amtex share returns.

30 Based on Exhibit 2 and Vasileva's prediction of the crude oil return for Month 37, the estimate of Amtex share return for Month 37 is *closest* to:

A −0.0024.

B 0.0071.

C 0.0119.

31 Using information from Exhibit 2, the 99% prediction interval for Amtex share return for Month 37 is *best* described as:

A $\hat{Y}_f \pm 0.0053$.

B $\hat{Y}_f \pm 0.0469$.

C $\hat{Y}_f \pm 0.1279$.

The following information relates to Questions 32–34

Doug Abitbol is a portfolio manager for Polyi Investments, a hedge fund that trades in the United States. Abitbol manages the hedge fund with the help of Robert Olabudo, a junior portfolio manager.

Abitbol looks at economists' inflation forecasts and would like to examine the relationship between the US Consumer Price Index (US CPI) consensus forecast and the actual US CPI using regression analysis. Olabudo estimates regression coefficients to test whether the consensus forecast is unbiased. If the consensus forecasts are unbiased, the intercept should be 0.0 and the slope will be equal to 1.0. Regression results are presented in Exhibit 1. Additionally, Olabudo calculates the 95% prediction interval of the actual CPI using a US CPI consensus forecast of 2.8.

Exhibit 1 Regression Output: Estimating US CPI

Regression Statistics			
R^2	0.9859		
Standard error of estimate	0.0009		
Observations	60		
	Coefficients	**Standard Error**	***t*-Statistic**
Intercept	0.0001	0.0002	0.5000
US CPI consensus forecast	0.9830	0.0155	63.4194

Notes:

1 The absolute value of the critical value for the t-statistic is 2.002 at the 5% level of significance.

2 The standard deviation of the US CPI consensus forecast is s_x = 0.7539.

3 The mean of the US CPI consensus forecast is $\overline{X}$ = 1.3350.

Finally, Abitbol and Olabudo discuss the forecast and forecast interval:

Observation 1 For a given confidence level, the forecast interval is the same no matter the US CPI consensus forecast.

Observation 2 A larger standard error of the estimate will result in a wider confidence interval.

32 Based on Exhibit 1, Olabudo should:

A conclude that the inflation predictions are unbiased.

B reject the null hypothesis that the slope coefficient equals one.

C reject the null hypothesis that the intercept coefficient equals zero.

33 Based on Exhibit 1, Olabudo should calculate a prediction interval for the actual US CPI *closest* to:

A 2.7506 to 2.7544.

B 2.7521 to 2.7529.

C 2.7981 to 2.8019.

34 Which of Olabudo's observations of forecasting is correct?

A Only Observation 1

B Only Observation 2

C Both Observation 1 and Observations 2

The following information relates to Questions 35–38

Espey Jones is examining the relation between the net profit margin (NPM) of companies, in percent, and their fixed asset turnover (FATO). He collected a sample of 35 companies for the most recent fiscal year and fit several different functional forms, settling on the following model:

$$\ln \text{NPM}_i = b_0 + b_1 \text{FATO}_i.$$

The results of this estimation are provided in Exhibit 1.

Exhibit 1 Results of Regressing NPM on FATO

Source	df	Sum of Squares	Mean Square	F	p-Value
Regression	1	102.9152	102.9152	1,486.7079	0.0000
Residual	32	2.2152	0.0692		
Total	33	105.1303			

	Coefficients	Standard Error	t- Statistic	p-Value
Intercept	0.5987	0.0561	10.6749	0.0000
FATO	0.2951	0.0077	38.5579	0.0000

35 The coefficient of determination is *closest* to:

A 0.0211.

B 0.9789.

C 0.9894.

36 The standard error of the estimate is *closest* to:

A 0.2631.

B 1.7849.

C 38.5579.

37 At a 0.01 level of significance, Jones should conclude that:

A the mean net profit margin is 0.5987%.

B the variation of the fixed asset turnover explains the variation of the natural log of the net profit margin.

C a change in the fixed asset turnover from 3 to 4 times is likely to result in a change in the net profit margin of 0.5987%.

38 The predicted net profit margin for a company with a fixed asset turnover of 2 times is *closest* to:

A 1.1889%.

B 1.8043%.

C 3.2835%

SOLUTIONS

1 **A** The coefficient of determination is 0.4279:

$$\frac{\text{Explained variation}}{\text{Total variation}} = \frac{60.16}{140.58} = 0.4279.$$

B $$F = \frac{60.16/1}{(140.58 - 60.16)/(60 - 2)} = \frac{60.16}{1.3866} = 43.3882.$$

C Begin with the sum of squares error of 140.58 − 60.16 = 80.42. Then calculate the mean square error of 80.42 ÷ (60 − 2) = 1.38655. The standard error of the estimate is the square root of the mean square error: $s_e = \sqrt{1.38655} = 1.1775$.

D The sample variance of the dependent variable uses the total variation of the dependent variable and divides it by the number of observations less one:

$$\sum_{i=1}^{n} \frac{\left(Y_i - \overline{Y}\right)^2}{n - 1} = \frac{\text{Total variation}}{n - 1} = \frac{140.58}{60 - 1} = 2.3827.$$

The sample standard deviation of the dependent variable is the square root of the variance, or $\sqrt{2.3827} = 1.544$.

2 C is correct. Homoskedasticity is the situation in which the variance of the residuals is constant across the observations.

3 A is correct. SHIFT is an indicator or dummy variable because it takes on only the values 0 and 1.

4 C is correct. In a simple regression with a single indicator variable, the intercept is the mean of the dependent variable when the indicator variable takes on a value of zero, which is before the shift in policy in this case.

5 C is correct. Whereas the intercept is the average of the dependent variable when the indicator variable is zero (that is, before the shift in policy), the slope is the difference in the mean of the dependent variable from before to after the change in policy.

6 A is correct. The null hypothesis of no difference in the annual growth rate is rejected at the 0.05 level: The calculated test statistic of −8.16188 is outside the bounds of ±2.048.

7

A The sample variance of the dependent variable is the sum of squares total divided by its degrees of freedom (n − 1 = 5 − 1 = 4, as given). Thus, the sample variance of the dependent variable is 95.2 ÷ 4 = 23.8.

B The coefficient of determination = 88.0 ÷ 95.2 = 0.92437.

C The F-statistic tests whether all the slope coefficients in a linear regression are equal to zero.

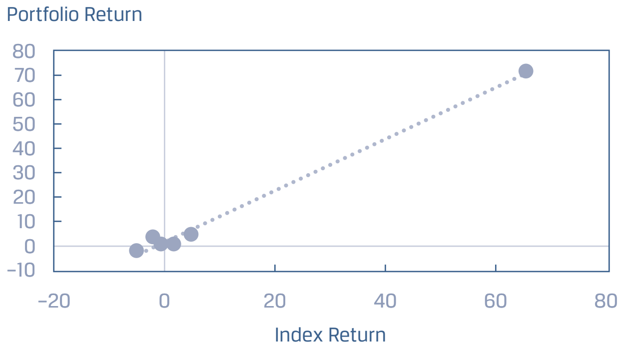
D The calculated value of the F-statistic is 36.667, as shown in the table. The corresponding p-value is less than 0.05, so you reject the null hypothesis of a slope equal to zero.

E The standard error of the estimate is the square root of the mean square error: $s_e = \sqrt{2.4} = 1.54919$.

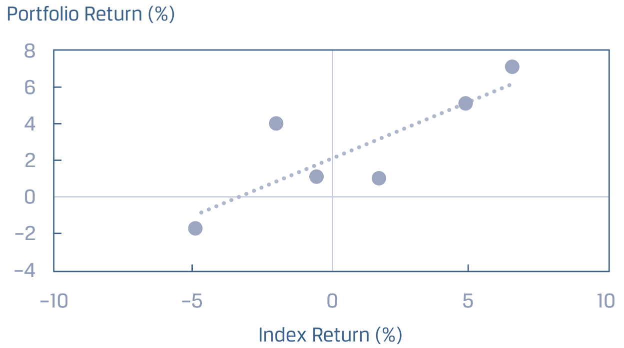
8 The Month 2 data point is an outlier, lying far away from the other data values. Because this outlier was caused by a data entry error, correcting the outlier improves the validity and reliability of the regression. In this case, revised R^2 is lower (from 0.9921 to 0.6784). The outliers created the illusion of a better fit from the higher R^2; the outliers altered the estimate of the slope. The standard error of the estimate is lower when the data error is corrected (from 2.8619 to 2.0624), as a result of the lower mean square error. However, at a 0.05 level of significance, both models fit well. The difference in the fit is illustrated in Exhibit 1.

Exhibit 1 The Fit of the Model with and without Data Errors

A. Before the Data Errors Are Corrected



B. After the Data Errors Are Corrected



9 A is correct. The coefficient of determination is the same as R^2, which is 0.7436 in the table.

10 C is correct. Because the slope is positive, the correlation between X and Y is simply the square root of the coefficient of determination: $\sqrt{0.7436} = 0.8623$.

11 C is correct. To make a prediction using the regression model, multiply the slope coefficient by the forecast of the independent variable and add the result to the intercept. Expected value of CFO to sales = 0.077 + (0.826 × 5) = 4.207.

12 C is correct. The p-value is the smallest level of significance at which the null hypotheses concerning the slope coefficient can be rejected. In this case, the p-value is less than 0.05, and thus the regression of the ratio of cash flow from operations to sales on the ratio of net income to sales is significant at the 5% level.

13 A is correct. The data are observations over time.

14 C is correct. From the regression equation, Expected return = 0.0138 + (–0.6486 × –0.01) = 0.0138 + 0.006486 = 0.0203, or 2.03%.

15 C is correct. R^2 is the coefficient of determination. In this case, it shows that 2.11% of the variability in Stellar's returns is explained by changes in CPIENG.

16 B is correct. The standard error of the estimate is the standard deviation of the regression residuals.

17 C is the correct response because it is a false statement. The slope and intercept are both statistically different from zero at the 0.05 level of significance.

18 C is correct. The slope coefficient (shown in Exhibit 2) is negative. We could also determine this by looking at the cross-product (Exhibit 1), which is negative.

19 B is correct. The sample covariance is calculated as

$$\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n-1} = -9.2430 \div 49 = -0.1886$$

20 A is correct. In simple regression, the R^2 is the square of the pairwise correlation. Because the slope coefficient is negative, the correlation is the negative of the square root of 0.0933, or –0.3054.

21 C is correct. Conclusions cannot be drawn regarding causation; they can be drawn only about association; therefore, Interpretations 1 and 2 are incorrect.

22 C is correct. Liu explains the variation of the short interest ratio using the variation of the debt ratio.

23 A is correct. The degrees of freedom are the number of observations minus the number of parameters estimated, which equals 2 in this case (the intercept and the slope coefficient). The number of degrees of freedom is 50 – 2 = 48.

24 B is correct. The t-statistic is –2.2219, which is outside the bounds created by the critical t-values of ±2.011 for a two-tailed test with a 5% significance level. The value of 2.011 is the critical t-value for the 5% level of significance (2.5% in one tail) for 48 degrees of freedom. A is incorrect because the mean of the short interest ratio is 192.3 ÷ 50 = 3.846. C is incorrect because the debt ratio explains 9.33% of the variation of the short interest ratio.

25 A is correct. The predicted value of the short interest ratio = 5.4975 + (–4.1589 × 0.40) = 5.4975 – 1.6636 = 3.8339.

26 C is correct because $F = \frac{\text{Mean square regression}}{\text{Mean square error}} = \frac{38.4404}{7.7867} = 4.9367$.

27 C is correct. The assumptions of the linear regression model are that (1) the relationship between the dependent variable and the independent variable is linear in the parameters b_0 and b_1, (2) the residuals are independent of one another, (3) the variance of the error term is the same for all observations, and (4) the error term is normally distributed. Assumption 3 is incorrect because the dependent variable need not be normally distributed.

28 B is correct. The standard error of the estimate for a linear regression model with one independent variable is calculated as the square root of the mean square error:

$$s_e = \sqrt{\frac{0.071475}{34}} = 0.04585.$$

29 C is correct. Crude oil returns explain the Amtex share returns if the slope coefficient is statistically different from zero. The slope coefficient is 0.2354, and the calculated t-statistic is

$$t = \frac{0.2354 - 0.0000}{0.0760} = 3.0974,$$

which is outside the bounds of the critical values of ±2.728.

Therefore, Vasileva should reject the null hypothesis that crude oil returns do not explain Amtex share returns, because the slope coefficient is statistically different from zero.

A is incorrect because the calculated t-statistic for testing the slope against 0.15 is $t = \frac{0.2354 - 0.1500}{0.0760} = 1.1237$, which is less than the critical value of +2.441.

B is incorrect because the calculated t-statistic is $t = \frac{0.0095 - 0.0000}{0.0078} = 1.2179$, which is less than the critical value of +2.441.

30 B is correct. The predicted value of the dependent variable, Amtex share return, given the value of the independent variable, crude oil return, −0.01, is calculated as $\hat{Y} = \hat{b}_0 + \hat{b}_1 X_i = 0.0095 + \left[0.2354 \times \left(-0.01\right)\right] = 0.0071$.

31 C is correct. The predicted share return is 0.0095 + [0.2354 × (−0.01)] = 0.0071. The lower limit for the prediction interval is 0.0071 − (2.728 × 0.0469) = −0.1208, and the upper limit for the prediction interval is 0.0071 + (2.728 × 0.0469) = 0.1350.

A is incorrect because the bounds of the interval should be based on the standard error of the forecast and the critical t-value, not on the mean of the dependent variable.

B is incorrect because bounds of the interval are based on the product of the standard error of the forecast *and* the critical t-value, not simply the standard error of the forecast.

32 A is correct. We fail to reject the null hypothesis of a slope equal to one, and we fail to reject the null hypothesis of an intercept equal to zero. The test of the slope equal to 1.0 is $t = \frac{0.9830 - 1.000}{0.0155} = -1.09677$. The test of the intercept equal to 0.0 is $t = \frac{0.0001 - 0.0000}{.00002} = 0.5000$. Therefore, we conclude that the forecasts are unbiased.

33 A is correct. The forecast interval for inflation is calculated in three steps:

Step 1. Make the prediction given the US CPI forecast of 2.8:

$$\begin{aligned}\widehat{Y} &= b_0 + b_1 X \\ &= 0.0001 + \left(0.9830 \times 2.8\right) \\ &= 2.7525.\end{aligned}$$

Step 2. Compute the variance of the prediction error:

$$s_f^2 = s_e^2\left\{1 + (1/n) + \left[\left(X_f - \bar{X}\right)^2\right]\Big/\left[(n-1) \times s_x^2\right]\right\}.$$

$$s_f^2 = 0.0009^2\left\{1 + (1/60) + \left[(2.8 - 1.3350)^2\right]\Big/\left[(60-1) \times 0.7539^2\right]\right\}.$$

$$s_f^2 = 0.00000088.$$

$$s_f = 0.0009.$$

Step 3. Compute the prediction interval:

$$\hat{Y} \pm t_c \times s_f$$

2.7525 ± (2.0 × 0.0009)

Lower bound: 2.7525 − (2.0 × 0.0009) = 2.7506.

Upper bound: 2.7525 + (2.0 × 0.0009) = 2.7544.

So, given the US CPI forecast of 2.8, the 95% prediction interval is 2.7506 to 2.7544.

34 B is correct. The confidence level influences the width of the forecast interval through the critical *t*-value that is used to calculate the distance from the forecasted value: The larger the confidence level, the wider the interval. Therefore, Observation 1 is not correct.

Observation 2 is correct. The greater the standard error of the estimate, the greater the standard error of the forecast.

35 B is correct. The coefficient of determination is 102.9152 ÷ 105.1303 = 0.9789.

36 A is correct. The standard error is the square root of the mean square error, or $\sqrt{0.0692} = 0.2631$.

37 B is correct. The *p*-value corresponding to the slope is less than 0.01, so we reject the null hypothesis of a zero slope, concluding that the fixed asset turnover explains the natural log of the net profit margin.

38 C is correct. The predicted natural log of the net profit margin is 0.5987 + (2 × 0.2951) = 1.1889. The predicted net profit margin is $e^{1.1889} = 3.2835\%$.

READING

2

Multiple Regression

by Richard A. DeFusco, PhD, CFA, Dennis W. McLeavey, DBA, CFA, Jerald E. Pinto, PhD, CFA, and David E. Runkle, PhD, CFA

Richard A. DeFusco, PhD, CFA, is at the University of Nebraska-Lincoln (USA). Dennis W. McLeavey, DBA, CFA, is at the University of Rhode Island (USA). Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). David E. Runkle, PhD, CFA, is at Jacobs Levy Equity Management (USA).

Contributions from Sanjiv Sabherwal, PhD, at the University of Texas (USA)

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** formulate a multiple regression equation to describe the relation between a dependent variable and several independent variables, and determine the statistical significance of each independent variable;
☐	**b.** interpret estimated regression coefficients and their *p*-values;
☐	**c.** formulate a null and an alternative hypothesis about the population value of a regression coefficient, calculate the value of the test statistic, and determine whether to reject the null hypothesis at a given level of significance;
☐	**d.** interpret the results of hypothesis tests of regression coefficients;
☐	**e.** calculate and interpret a predicted value for the dependent variable, given an estimated regression model and assumed values for the independent variables;
☐	**f.** explain the assumptions of a multiple regression model;
☐	**g.** calculate and interpret the *F*-statistic, and describe how it is used in regression analysis;
☐	**h.** contrast and interpret the R2 and adjusted R2 in multiple regression;
☐	**i.** evaluate how well a regression model explains the dependent variable by analyzing the output of the regression equation and an ANOVA table;
☐	**j.** formulate and interpret a multiple regression, including qualitative independent variables;

(continued)

© 2021 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**k.** explain the types of heteroskedasticity and how heteroskedasticity and serial correlation affect statistical inference;
☐	**l.** describe multicollinearity, and explain its causes and effects in regression analysis;
☐	**m.** describe how model misspecification affects the results of a regression analysis, and describe how to avoid common forms of misspecification;
☐	**n.** interpret an estimated logistic regression;
☐	**o.** evaluate and interpret a multiple regression model and its results.

1 MULTIPLE LINEAR REGRESSION ASSUMPTIONS, TESTING COEFFICIENTS, AND PREDICTION

a formulate a multiple regression equation to describe the relation between a dependent variable and several independent variables, and determine the statistical significance of each independent variable;

b interpret estimated regression coefficients and their *p*-values;

c formulate a null and an alternative hypothesis about the population value of a regression coefficient, calculate the value of the test statistic, and determine whether to reject the null hypothesis at a given level of significance;

d interpret the results of hypothesis tests of regression coefficients;

e calculate and interpret a predicted value for the dependent variable, given an estimated regression model and assumed values for the independent variables;

f explain the assumptions of a multiple regression model;

i evaluate how well a regression model explains the dependent variable by analyzing the output of the regression equation and an ANOVA table;

o evaluate and interpret a multiple regression model and its results.

As financial analysts, we often need to use more sophisticated statistical methods than correlation analysis or regression involving a single independent variable. For example, a mutual fund analyst might want to know whether returns to a technology mutual fund behaved more like the returns to a growth stock index or like the returns to a value stock index. An investor might be interested in the factors that determine whether analysts cover a stock. Or analysts researching individual companies may want to understand what factors (such as macroeconomic variables) drive the demand for the company's products or services. We can answer these questions using linear regression with more than one independent variable—multiple linear regression.

We first introduce and illustrate the basic concepts and models of multiple regression analysis. These models rest on assumptions that are sometimes violated in practice. We then discuss three commonly occurring violations of regression assumptions. We address practical concerns, such as how to diagnose an assumption violation and what remedial steps to take when a model assumption has been violated. The subsequent section outlines some guidelines for building good regression models and discusses ways that analysts sometimes go wrong in this endeavor. We then discuss a class of models whose dependent variable is qualitative in nature. Specifically, we discuss logistic regression that plays an important role in machine learning for Big Data analysis.

1.1 Multiple Linear Regression

As investment analysts, we often hypothesize that more than one variable explains the behavior of a variable in which we are interested. The variable we seek to explain is called the dependent variable. The variables that we believe explain the dependent variable are called the independent variables. They may also be termed explanatory variables, predictor variables, or simply regressors. A tool that permits us to examine the relationship (if any) between the two types of variables is multiple linear regression. **Multiple linear regression** allows us to determine the effect of more than one independent variable on a particular dependent variable.

A **multiple linear regression model** has the general form

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_kX_{ki} + \varepsilon_i,\ i = 1, 2, \ldots n, \tag{1}$$

where

Y_i = the ith observation of the dependent variable Y

X_{ji} = the ith observation of the independent variable X_j, $j = 1, 2, \ldots, k$

b_0 = the intercept of the equation

$b_1, \ldots, b_k$ = the slope coefficients for each of the independent variables

ε_i = the error term

n = the number of observations

A slope coefficient, b_j, measures how much the dependent variable, Y, changes when the independent variable, X_j, changes by one unit, holding all other independent variables constant. For example, if $b_1 = 1$ and all the other independent variables remain constant, then we predict that if X_1 increases by one unit, Y will also increase by one unit. If $b_1 = -1$ and all the other independent variables are held constant, then we predict that if X_1 increases by one unit, Y will decrease by one unit. Multiple linear regression estimates $b_0, \ldots, b_k$. In this reading, we will refer to both the intercept, b_0, and the slope coefficients, $b_1, \ldots, b_k$, as **regression coefficients**. As we proceed with our discussion, keep in mind that a regression equation has k slope coefficients and $k + 1$ regression coefficients.

Although Equation 1 may seem to apply only to cross-sectional data because the notation for the observations is the same ($i = 1, \ldots, n$), all these results apply to time-series data as well. For example, if we analyze data from many time periods for one company, we would typically use the notation $Y_t, X_{1t}, X_{2t}, \ldots, X_{kt}$, in which the first subscript denotes the variable and the second denotes the tth time period.

In practice, we use software to estimate a multiple regression model. Exhibit 1 presents an application of multiple regression analysis in investment practice. In the course of discussing a hypothesis test, Exhibit 1 presents typical regression output and its interpretation.

Exhibit 1 Explaining the Bid–Ask Spread

As the manager of the trading desk at an investment management firm, you have noticed that the average bid–ask spreads of different NASDAQ-listed stocks can vary widely. When the ratio of a stock's bid–ask spread to its price is higher than for another stock, your firm's costs of trading in that stock tend to be higher. You have formulated the hypothesis that NASDAQ stocks' percentage bid–ask

(continued)

Exhibit 1 (Continued)

spreads are related to the number of market makers and the company's stock market capitalization. You have decided to investigate your hypothesis using multiple regression analysis.

You specify a regression model in which the dependent variable measures the percentage bid–ask spread, and the independent variables measure the number of market makers and the company's stock market capitalization. The regression is estimated using data from 31 December 2013 for 2,587 NASDAQ-listed stocks. Based on earlier published research exploring bid–ask spreads, you express the dependent and independent variables as natural logarithms, a so-called **log-log regression model**. A log-log regression model may be appropriate when one believes that proportional changes in the dependent variable bear a constant relationship to proportional changes in the independent variable(s), as we illustrate next. You formulate the multiple regression:

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \varepsilon_i, \tag{2}$$

where

Y_i = the natural logarithm of (Bid–ask spread/Stock price) for stock i

X_{1i} = the natural logarithm of the number of NASDAQ market makers for stock i

X_{2i} = the natural logarithm of the market capitalization (measured in millions of US$) of company i

In a log-log regression, such as Equation 2, the slope coefficients are interpreted as elasticities assumed to be constant. For example, a value of $b_2 = -0.75$ would mean that for a 1% increase in the market capitalization, we expect Bid–ask spread/Stock price to decrease by 0.75%, holding all other independent variables constant [note that $\Delta (\ln X) \approx \Delta X/X$, where Δ represents "change in" and $\Delta X/X$ is a proportional change in X].

Reasoning that greater competition tends to lower costs, you suspect that the greater the number of market makers, the smaller the percentage bid–ask spread. Therefore, you formulate a first null hypothesis (H_0) and alternative hypothesis (H_a):

H_0: $b_1 \geq 0$

H_a: $b_1 < 0$

The null hypothesis is the hypothesis that the "suspected" condition is not true. If the evidence supports rejecting the null hypothesis and accepting the alternative hypothesis, you have statistically confirmed your suspicion. An alternative valid formulation is a two-sided test, H_0: $b_1 = 0$ versus H_a: $b_1 \neq 0$, which reflects the beliefs of the researcher less strongly.

You also believe that the stocks of companies with higher market capitalization may have more-liquid markets, tending to lower percentage bid–ask spreads. Therefore, you formulate a second null hypothesis and alternative hypothesis:

H_0: $b_2 \geq 0$

H_a: $b_2 < 0$

For both tests, we use a t-test, rather than a z-test, because we do not know the population variance of b_1 and b_2. Suppose that you choose a 0.01 significance level for both tests.

Exhibit 1 (Continued)

Results from Regressing ln(Bid–Ask Spread/Price) on ln(Number of Market Makers) and ln(Market Capitalization)

	Coefficient	Standard Error	*t*-Statistic
Intercept	1.5949	0.2275	7.0105
ln(Number of NASDAQ market makers)	−1.5186	0.0808	−18.7946
ln(Company's market capitalization)	−0.3790	0.0151	−25.0993

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	2	3,728.1334	1,864.0667	2,216.75	0.00
Residual	2,584	2,172.8870	0.8409		
Total	2,586	5,901.0204			
Residual standard error			0.9170		
Multiple R^2			0.6318		
Observations			2,587		

Note: "df" = degrees of freedom.
Source: Center for Research in Security Prices, University of Chicago.

The table shows the results of estimating this linear regression. If the regression result is not significant, we may follow the useful principle of not proceeding to interpret the individual regression coefficients. Thus, the analyst might look first at the **analysis of variance (ANOVA)** section, which addresses the regression's overall significance.

- The ANOVA section reports quantities related to the overall explanatory power and significance of the regression. SS stands for sum of squares, and MSS stands for mean sum of squares (SS divided by df). The *F*-test reports the overall significance of the regression. For example, an entry of 0.01 for the significance of *F* means that the regression is significant at the 0.01 level. In our illustration the regression is even more significant because the significance of *F* is 0 at two decimal places.

Having ascertained that the overall regression is highly significant, an analyst might turn to the first listed column in the first section of the regression output.

- The Coefficient column gives the estimates of the intercept, b_0, and the slope coefficients, b_1 and b_2. The estimated intercept is positive, but both estimated slope coefficients are negative. Are these estimated regression coefficients significantly different from zero? The Standard Error column gives the standard error (the standard deviation) of the estimated regression coefficients. The test statistic for hypotheses concerning the population value of a regression coefficient has the form (Estimated regression coefficient − Hypothesized population value of the regression coefficient)/(Standard error of the regression coefficient). This is a *t*-test. Under the null hypothesis, the hypothesized population value of the regression coefficient is 0. Thus (Estimated regression coefficient)/(Standard error of the regression coefficient) is the *t*-statistic given in the third column. For example, the *t*-statistic for the intercept is 1.5949/0.2275 = 7.0105. To evaluate the significance of the *t*-statistic, we need to determine a quantity called degrees of freedom (df). When calculating the degrees of freedom lost in the regression, we add 1 to the number of independent variables to account for the intercept term. The calculation is: Degrees of freedom = Number of observations − (Number of independent variables + 1) = $n - (k + 1)$.

- The final section of the regression results table presents two measures of how well the estimated regression fits or explains the data. The first is the standard deviation of the regression residual, the residual standard error. This standard deviation is called the standard error of estimate (SEE). The second measure quantifies the degree of linear association between the dependent variable and all the independent variables jointly. This measure is known as multiple R^2 or simply R^2 (the square of the correlation between predicted and actual values of the dependent variable). Multiple R^2 is also known as the multiple coefficient of determination, or simply the coefficient of determination. A value of 0 for R^2 indicates no linear association; a value of 1 indicates perfect linear association. The final item in Exhibit 1 is the number of observations in the sample (2,587).

Having reviewed the meaning of typical regression output, we can return to complete the hypothesis tests. The estimated regression supports the hypothesis that the greater the number of market makers, the smaller the percentage bid–ask spread: We reject H_0: $b_1 \geq 0$ in favor of H_a: $b_1 < 0$. The results also support the belief that the stocks of companies with higher market capitalization have lower percentage bid–ask spreads: We reject H_0: $b_2 \geq 0$ in favor of H_a: $b_2 < 0$.

To see that the null hypothesis is rejected for both tests, we can use t-test tables. For both tests, df = 2,587 − 3 = 2,584. The tables do not give critical values for degrees of freedom that large. The critical value for a one-tailed test with df = 200 at the 0.01 significance level is 2.345; for a larger number of degrees of freedom, the critical value would be even smaller in magnitude. Therefore, in our one-sided tests, we reject the null hypothesis in favor of the alternative hypothesis if

$$t = \frac{\hat{b}_j - b_j}{s_{\hat{b}_j}} = \frac{\hat{b}_j - 0}{s_{\hat{b}_j}} < -2.345$$

where

$\hat{b}_j$ = the regression estimate of b_j, $j = 1, 2$

b_j = the hypothesized value[1] of the coefficient (0)

$s_{\hat{b}_j}$ = the estimated standard error of $\hat{b}_j$

The t-values of −18.7946 and −25.0993 for the estimates of b_1 and b_2, respectively, are both less than −2.345.

Before proceeding further, we should address the interpretation of a prediction stated in natural logarithm terms. We can convert a natural logarithm to the original units by taking the antilogarithm. To illustrate this conversion, suppose that a particular stock has 20 NASDAQ market makers and a market capitalization of $100 million. The natural logarithm of the number of NASDAQ market makers is equal to ln 20 = 2.9957, and the natural logarithm of the company's market cap (in millions) is equal to ln 100 = 4.6052. With these values, the regression model predicts that the natural log of the ratio of the bid–ask spread to the stock price will be 1.5949 + (−1.5186 × 2.9957) + (−0.3790 × 4.6052) = −4.6997. We take the antilogarithm of −4.6997 by raising e to that power: $e^{-4.6997}$ = 0.0091. The predicted bid–ask spread will be 0.91% of the stock price. The

1 To economize on notation in stating test statistics, in this context we use b_j to represent the hypothesized value of the parameter (elsewhere we use it to represent the unknown population parameter).

operation illustrated (taking the antilogarithm) recovers the value of a variable in the original units as $e^{\ln X} = X$. Later we state the assumptions of the multiple regression model; before using an estimated regression to make predictions in actual practice, we should assure ourselves that those assumptions are satisfied.

In Exhibit 1, we presented output common to most regression software programs. Many software programs also report *p*-values for the regression coefficients (the entry 0.00 for the significance of *F* was a *p*-value for the *F*-test). For each regression coefficient, the *p*-value would be the smallest level of significance at which we can reject a null hypothesis that the population value of the coefficient is 0, in a two-sided test. The lower the *p*-value, the stronger the evidence against that null hypothesis. A *p*-value quickly allows us to determine if an independent variable is significant at a conventional significance level, such as 0.05, or at any other standard we believe is appropriate.

Having estimated Equation 1, we can write

$$\begin{aligned}\hat{Y}_i &= \hat{b}_0 + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} \\ &= 1.5949 - 1.5186 X_{1i} - 0.3790 X_{2i}\end{aligned}$$

where $\hat{Y}_i$ stands for the predicted value of Y_i, and $\hat{b}_0$, $\hat{b}_1$, and $\hat{b}_2$ stand for the estimated values of b_0, b_1, and b_2, respectively. How should we interpret the estimated slope coefficients −1.5186 and −0.3790?

Interpreting the slope coefficients in a multiple linear regression model is different than doing so in the one-independent-variable regressions explored in earlier coverage of the topic of simple regression. Suppose we have a one-independent-variable regression that we estimate as $\hat{Y}_i = 0.50 + 0.75 X_{1i}$. The interpretation of the slope estimate 0.75 is that for every 1-unit increase in X_1, we expect Y to increase by 0.75 units. If we were to add, however, a second independent variable to the equation, we would generally find that the estimated coefficient on X_1 is *not* 0.75 unless the second independent variable were uncorrelated with X_1. In other words, the slope coefficient of a dependent variable may depend upon other independent variables.

The slope coefficients in a multiple regression are known as **partial regression coefficients** or **partial slope coefficients** and need to be interpreted with care (the terminology comes from the fact that they correspond to the partial derivatives of Y with respect to the independent variables).

Suppose the coefficient on X_1 in a regression with the second independent variable was 0.60. Can we say that for every 1-unit increase in X_1, we expect Y to increase by 0.60 units? Not without qualification. For every 1-unit increase in X_1, we still expect Y to increase by 0.75 units when X_2 is not held constant. We would interpret 0.60 as the expected increase in Y for a 1-unit increase X_1 *holding the second independent variable constant.*

To explain what the shorthand reference "holding the second independent constant" refers to, if we were to regress X_1 on X_2, the residuals from that regression would represent the part of X_1 that is uncorrelated with X_2. We could then regress Y on those residuals in a one-independent-variable regression. We would find that the slope coefficient on the residuals would be 0.60; by construction, 0.60 would represent the expected effect on Y of a 1-unit increase in X_1 after removing the part of X_1 that is correlated with X_2. Consistent with this explanation, we can view 0.60 as the expected net effect on Y of a 1-unit increase in X_1, after accounting for any effects of the other independent variables on the expected value of Y. To reiterate, a partial regression coefficient measures the expected change in the dependent variable for a 1-unit increase in an independent variable, holding all the other independent variables constant.

To apply this process to the regression in Exhibit 1, we see that the estimated coefficient on the natural logarithm of market capitalization is −0.3790. Therefore, the model predicts that an increase of 1 in the natural logarithm of the company's market capitalization is associated with a −0.3790 change in the natural logarithm of the ratio of the bid–ask spread to the stock price, holding the natural logarithm of the number of market makers constant. We need to be careful not to expect that the natural logarithm of the ratio of the bid–ask spread to the stock price would differ by −0.3790 if we compared two stocks for which the natural logarithm of the company's market capitalization differed by 1, because in all likelihood the number of market makers for the two stocks would differ as well, which would affect the dependent variable. The value −0.3790 is the expected net effect of difference in log market capitalizations, net of the effect of the log number of market makers on the expected value of the dependent variable.

1.1.1 *Assumptions of the Multiple Linear Regression Model*

Before we can conduct correct statistical inference on a multiple linear regression model (a model with more than one independent variable estimated using ordinary least squares, OLS, an estimation method based on the criterion of minimizing the sum of the squared residuals of a regression), we need to know the assumptions underlying that model. Suppose we have n observations on the dependent variable, Y, and the independent variables, $X_1, X_2, \ldots, X_k$, and we want to estimate the equation $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_kX_{ki} + \varepsilon_i$.

In order to make a valid inference from a multiple linear regression model, we need to make the following six assumptions, which as a group define the classical normal multiple linear regression model:

1. The relationship between the dependent variable, Y, and the independent variables, $X_1, X_2, \ldots, X_k$, is linear as described in Equation 1.
2. The independent variables ($X_1, X_2, \ldots, X_k$) are not random, which means that they are fixed and known; no exact linear relation exists between two or more of the independent variables or combinations of independent variables.
3. The expected value of the error term, conditioned on the independent variables, is 0: $E(\varepsilon \mid X_1, X_2, \ldots, X_k) = 0$.
4. The variance of the error term is the same for all observations:[2] $E\left(\varepsilon_i^2\right) = \sigma_\varepsilon^2$.
5. The error term is uncorrelated across observations: $E(\varepsilon_i\varepsilon_j) = 0, j \neq i$.
6. The error term is normally distributed.

Note that these assumptions are almost exactly the same as those for the single-variable linear regression model. Assumption 2 is modified such that no exact linear relation exists between two or more independent variables or combinations of independent variables. If this part of Assumption 2 is violated, then we cannot compute linear regression estimates. Also, even if no exact linear relationship exists between two or more independent variables, or combinations of independent variables, linear regression may encounter problems if two or more of the independent variables or combinations thereof are highly correlated. Such a high correlation is known as multicollinearity, which we will discuss later. We will also discuss the consequences of conducting regression analysis premised on Assumptions 4 and 5 being met when, in fact, they are violated.

2 $\mathrm{Var}(\varepsilon) = E(\varepsilon^2)$ and $\mathrm{Cov}(\varepsilon_i\varepsilon_j) = E(\varepsilon_i\varepsilon_j)$ because $E(\varepsilon) = 0$.

Exhibit 2 Factors Explaining the Valuations of Multinational Corporations

Kyaw, Manley, and Shetty (2011) examined which factors affect the valuation of a multinational corporation (MNC). Specifically, they wanted to know whether political risk, transparency, and geographic diversification affected the valuations of MNCs. They used data for 450 US MNCs from 1998 to 2003. The valuations of these corporations were measured using Tobin's *q*, a commonly used measure of corporate valuation that is calculated as the ratio of the sum of the market value of a corporation's equity and the book value of long-term debt to the sum of the book values of equity and long-term debt. The authors regressed Tobin's *q* of MNCs on variables representing political risk, transparency, and geographic diversification. The authors also included some additional variables that may affect company valuation, including size, leverage, and beta. They used the equation

$$\text{Tobin's } q_{i,t} = b_0 + b_1(\text{Size}_{i,t}) + b_2(\text{Leverage}_{i,t}) + b_3(\text{Beta}_{i,t}) + b_4(\text{Political risk}_{i,t}) + b_5(\text{Transparency}_{i,t}) + b_6(\text{Geographic diversification}_{i,t}) + \varepsilon_{i,t},$$

where

Tobin's $q_{i,t}$ = the Tobin's *q* for MNC *i* in year *t*, with Tobin's *q* computed as (Market value of equity + Book value of long-term debt)/(Book value of equity + Book value of long-term debt)

$\text{Size}_{i,t}$ = the natural log of the total sales of MNC *i* in the year *t* in millions of US$

$\text{Leverage}_{i,t}$ = the ratio of total debt to total assets of MNC *i* in year *t*

$\text{Beta}_{i,t}$ = the beta of the stock of MNC *i* in year *t*

$\text{Political risk}_{i,t}$ = the at-risk-proportion of international operations of MNC *i* in year *t*, calculated as [1 – (number of safe countries/total number of foreign countries in which the firm has operations)], using national risk coding from *Euromoney*

$\text{Transparency}_{i,t}$ = the "transparency %" (representing the level of disclosure) of MNC *i* in year *t*, using survey data from *S&P Transparency & Disclosure*

$\text{Geographic diversification}_{i,t}$ = foreign sales of MNC *i* in year *t* expressed as a percentage of its total sales in that year

The following table shows the results of their analysis.

Results from Regressing Tobin's *q* on Factors Affecting the Value of Multinational Corporations

	Coefficient	Standard Error	*t*-Statistic
Intercept	19.829	4.798	4.133
Size	–0.712	0.228	–3.123
Leverage	–3.897	0.987	–3.948
Beta	–1.032	0.261	–3.954
Political risk	–2.079	0.763	–2.725

(continued)

Exhibit 2 (Continued)

	Coefficient	Standard Error	*t*-Statistic
Transparency	−0.129	0.050	−2.580
Geographic diversification	0.021	0.010	2.100

Notes: This study combines time series observations with cross-sectional observations; such data are commonly referred to as panel data. In such a setting, the standard errors need to be corrected for bias by using a clustered standard error approach as in Petersen (2009). The standard errors reported in this exhibit are clustered standard errors.
Size is the natural log of total sales. A log transformation (either natural log or log base 10) is commonly used for independent variables that can take a wide range of values; company size and fund size are two such variables. One reason to use the log transformation is to improve the statistical properties of the residuals. If the authors had not taken the log of sales and instead used sales as the independent variable, the regression model probably would not have explained Tobin's q as well.
Source: Kyaw, Manley, and Shetty (2011).

Suppose that we use the regression results to test the null hypothesis that the size of a multinational corporation has no effect on its value. Our null hypothesis is that the coefficient on the size variable equals 0 (H_0: $b_1 = 0$), and our alternative hypothesis is that the coefficient does not equal 0 (H_a: $b_1 \neq 0$). The t-statistic for testing that hypothesis is

$$t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{-0.712 - 0}{0.228} = -3.12$$

With 450 observations and 7 coefficients, the t-statistic has 450 − 7 = 443 degrees of freedom. At the 0.05 significance level, the critical value for t is about 1.97. The absolute value of computed t-statistic on the size coefficient is 3.12, which suggests strongly that we can reject the null hypothesis that size is unrelated to MNC value. In fact, the critical value for t is about 2.6 at the 0.01 significance level.

Because $\text{Size}_{i,t}$ is the natural (base e or 2.72) log of sales, an increase of 1 in $\text{Size}_{i,t}$ is the same as a 2.72-fold increase in sales. Thus, the estimated coefficient of approximately −0.7 for $\text{Size}_{i,t}$ implies that every 2.72-fold increase in sales of the MNC (an increase of 1 in $\text{Size}_{i,t}$) is associated with an expected decrease of 0.7 in Tobin's $q_{i,t}$ of the MNC, *holding constant the other five independent variables in the regression.*

Now suppose we want to test the null hypothesis that geographic diversification is not related to Tobin's q. We want to test whether the coefficient on geographic diversification equals 0 (H_0: $b_6 = 0$) against the alternative hypothesis that the coefficient on geographic diversification does not equal 0 (H_a: $b_6 \neq 0$). The t-statistic to test this hypothesis is

$$t = \frac{\hat{b}_6 - b_6}{s_{\hat{b}_6}} = \frac{0.021 - 0}{0.010} = 2.10$$

The critical value of the t-test is 1.97 at the 0.05 significance level. Therefore, at the 0.05 significance level, we can reject the null hypothesis that geographic diversification has no effect on MNC valuation. We can interpret the coefficient on geographic diversification of 0.021 as implying that an increase of 1 in the percentage of MNC's sales that are foreign sales is associated with an expected 0.021 increase in Tobin's q for the MNC, holding all other independent variables constant.

Exhibit 3 Explaining Returns to the Fidelity Select Technology Portfolio

Suppose you are considering an investment in the Fidelity Select Technology Portfolio (FSPTX), a US mutual fund specializing in technology stocks. You want to know whether the fund behaves more like a large-cap growth fund or a large-cap value fund. You decide to estimate the regression

$$Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + \varepsilon_t,$$

where

Y_t = the monthly return to the FSPTX
X_{1t} = the monthly return to the S&P 500 Growth Index
X_{2t} = the monthly return to the S&P 500 Value Index

The S&P 500 Growth and S&P 500 Value indexes represent predominantly large-cap growth and value stocks, respectively.

The regression results show the results of this linear regression using monthly data from August 2014 through August 2019. The estimated intercept in the regression is 0.0011. Thus, if both the return to the S&P 500 Growth Index and the return to the S&P 500 Value Index equal 0 in a specific month, the regression model predicts that the return to the FSPTX will be 0.11%. The coefficient on the large-cap growth index is 1.5850, and the coefficient on the large-cap value index return is −0.3902. Therefore, if in a given month the return to the S&P 500 Growth Index was 1% and the return to the S&P 500 Value Index was −2%, the model predicts that the return to the FSPTX would be 0.0011 + 1.5850(0.01) − 0.3902(−0.02) = 2.48%.

Results from Regressing the FSPTX Returns on the S&P 500 Growth and S&P 500 Value Indexes

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0011	0.0025	0.4405
S&P 500 Growth Index	1.5850	0.1334	11.88
S&P 500 Value Index	−0.3902	0.1332	−2.93

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	2	0.1198	0.0599	178.01	3.07E-25
Residual	57	0.0192	0.0003		
Total	59	0.1389			
Residual standard error			0.0183		
Multiple R^2			0.862		
Observations			60		

Source: finance.yahoo.com.

We may want to know whether the coefficient on the returns to the S&P 500 Value Index is statistically significant. Our null hypothesis states that the coefficient equals 0 (H_0: $b_2 = 0$); our alternative hypothesis states that the coefficient does not equal 0 (H_a: $b_2 \neq 0$).

Our test of the null hypothesis uses a t-test constructed as follows:

$$t = \frac{\hat{b}_2 - b_2}{s_{\hat{b}_2}} = \frac{-0.3902 - 0}{0.1332} = -2.93,$$

where

$\hat{b}_2$ = the regression estimate of b_2

b_2 = the hypothesized value of the coefficient (0)

$s_{\hat{b}_2}$ = the estimated standard error of $\hat{b}_2$ This regression has 60 observations and three coefficients (two independent variables and the intercept); therefore, the t-test has 60 − 3 = 57 degrees of freedom. At the 0.05 significance level, the critical value for the test statistic is about 2.00. The absolute value of the test statistic is 2.93. Because the test statistic's absolute value is more than the critical value (2.93 > 2.00), we reject the null hypothesis that $b_2 = 0$. (Note that the t-tests reported in the regression results table, as well as the other regression tables, are tests of the null hypothesis that the population value of a regression coefficient equals 0.)

Similar analysis shows that at the 0.05 significance level, we cannot reject the null hypothesis that the intercept equals 0 (H_0: $b_0 = 0$) in favor of the alternative hypothesis that the intercept does not equal 0 (H_a: $b_0 \neq 0$). The results also show that the t-statistic for testing that hypothesis is 0.4405, a result smaller in absolute value than the critical value of 2.00. However, at the 0.05 significance level we *can* reject the null hypothesis that the coefficient on the S&P 500 Growth Index equals 0 (H_0: $b_1 = 0$) in favor of the alternative hypothesis that the coefficient does not equal 0 (H_a: $b_1 \neq 0$). The t-statistic for testing that hypothesis is 11.88, a result far above the critical value of 2.00. Thus, multiple regression analysis suggests that returns to the FSPTX are very closely associated with the returns to the S&P 500 Growth Index, but they are negatively related to S&P 500 Value Index. This regression is related to return-based style analysis, one of the most frequent applications of regression analysis in the investment profession. For more information, see Sharpe (1988), who pioneered this field, and Buetow, Johnson, and Runkle (2000).

1.1.2 *Predicting the Dependent Variable in a Multiple Regression Model*

Financial analysts often want to predict the value of the dependent variable in a multiple regression based on assumed values of the independent variables. We have previously discussed how to make such a prediction in the case of only one independent variable. The process for making that prediction with multiple linear regression is very similar.

To predict the value of a dependent variable using a multiple linear regression model, we follow these three steps:

1 Obtain estimates $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$, …, $\hat{b}_k$ of the regression parameters b_0, b_1, b_2, …, b_k.

2 Determine the assumed values of the independent variables, $\hat{X}_{1i}$, $\hat{X}_{2i}$, …, $\hat{X}_{ki}$.

3 Compute the predicted value of the dependent variable, $\hat{Y}_i$, using the equation

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 \hat{X}_{1i} + \hat{b}_2 \hat{X}_{2i} + \ldots + \hat{b}_k \hat{X}_{ki} \qquad (3)$$

Two practical points concerning using an estimated regression to predict the dependent variable are in order. First, we should be confident that the assumptions of the regression model are met. Second, we should be cautious about predictions based on values of the independent variables that are outside the range of the data on which the model was estimated; such predictions are often unreliable.

EXAMPLE 1

Predicting a Multinational Corporation's Tobin's q

In Exhibit 2, we explained the Tobin's q for US multinational corporations (MNC) based on the natural log of sales, leverage, beta, political risk, transparency, and geographic diversification. To review the regression equation:

$$\text{Tobin's } q_{i,t} = b_0 + b_1(\text{Size}_{i,t}) + b_2(\text{Leverage}_{i,t}) + b_3(\text{Beta}_{i,t}) + b_4(\text{Political risk}_{i,t}) + b_5(\text{Transparency}_{i,t}) + b_6(\text{Geographic diversification}_{i,t}) + \varepsilon_i.$$

Now we can use the results of the regression (excerpted here) to predict the Tobin's q for a US MNC.

Regression results

	Coefficient
Intercept	19.829
Size	−0.712
Leverage	−3.897
Beta	−1.032
Political risk	−2.079
Transparency	−0.129
Geographic diversification	0.021

Suppose that a particular MNC has the following data for a given year:

- Total sales of \$7,600 million. The natural log of total sales in millions of US\$ equals ln(7,600) = 8.94.
- Leverage (Total debt/Total assets) of 0.45.
- Beta of 1.30.
- Political risk of 0.47, implying that the ratio of the number of safe countries to the total number of foreign countries in which the MNC has operations is 0.53.
- Transparency score of 65, indicating 65% "yes" answers to survey questions related to the corporation's transparency.
- Geographic diversification of 30, indicating that 30% of the corporation's sales are in foreign countries.

What is the predicted Tobin's q for the above MNC?

Solution:

The predicted Tobin's q for the MNC, based on the regression, is:

$$19.829 + (-0.712 \times 8.94) + (-3.897 \times 0.45) + (-1.032 \times 1.30) + (-2.079 \times 0.47) + (-0.129 \times 65) + (0.021 \times 30) = 1.64.$$

When predicting the dependent variable using a linear regression model, we encounter two types of uncertainty: uncertainty in the regression model itself, as reflected in the standard error of estimate, and uncertainty about the estimates of the regression model's parameters. In earlier coverage of the regression topic, we presented procedures for constructing a prediction interval for linear regression with one independent variable. For multiple regression, however, computing a prediction interval to properly incorporate both types of uncertainty requires matrix algebra, which is outside the scope of our discussion (See Greene 2018 for more information).

2 TESTING THE WHOLE MULTIPLE LINEAR REGRESSION MODEL AND ADJUSTED R-SQUARE

g calculate and interpret the F-statistic, and describe how it is used in regression analysis;

h contrast and interpret the R2 and adjusted R2 in multiple regression;

i evaluate how well a regression model explains the dependent variable by analyzing the output of the regression equation and an ANOVA table;

o evaluate and interpret a multiple regression model and its results.

Earlier, we illustrated how to conduct hypothesis tests on regression coefficients individually. What if we now want to test the significance of the regression as a whole? As a group, do the independent variables help explain the dependent variable? To address this question, we test the null hypothesis that all the slope coefficients in a regression are simultaneously equal to 0. In this section, we further discuss ANOVA with regard to a regression's explanatory power and the inputs for an F-test of the above null hypothesis.

If none of the independent variables in a regression model helps explain the dependent variable, the slope coefficients should all equal 0. In a multiple regression, however, we cannot test the null hypothesis that *all* slope coefficients equal 0 based on t-tests that *each individual* slope coefficient equals 0, because the individual tests do not account for the effects of interactions among the independent variables. For example, a classic symptom of multicollinearity is that we can reject the hypothesis that all the slope coefficients equal 0 even though none of the t-statistics for the individual estimated slope coefficients is significant. Conversely, we can construct unusual examples in which the estimated slope coefficients are significantly different from 0 although jointly they are not.

To test the null hypothesis that all the slope coefficients in the multiple regression model are jointly equal to 0 (H_0: $b_1 = b_2 = ... = b_k = 0$) against the alternative hypothesis that at least one slope coefficient is not equal to 0, we must use an F-test. The F-test is viewed as a test of the regression's overall significance.

To correctly calculate the test statistic for the null hypothesis, we need four inputs:

- Total number of observations, n.

- Total number of regression coefficients to be estimated, $k + 1$, where k is the number of slope coefficients.
- Sum of squared errors or residuals, $\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$, abbreviated SSE, also known as the residual sum of squares, the unexplained variation. In a table of regression output, this is the number under the "SS" column in the row "Residual."
- Regression sum of squares, $\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$, abbreviated RSS. This amount is the variation in Y from its mean that the regression equation explains (explained variation). In a table of regression output, this is the number under the "SS" column in the row "Regression."

The F-test for determining whether the slope coefficients equal 0 is based on an F-statistic calculated using the four values listed above. The F-statistic measures how well the regression equation explains the variation in the dependent variable; it is the ratio of the mean regression sum of squares to the mean squared error.

We compute the mean regression sum of squares by dividing the regression sum of squares by the number of slope coefficients estimated, k. We compute the mean squared error by dividing the sum of squared errors by the number of observations, n, minus $(k + 1)$. The two divisors in these computations are the degrees of freedom for calculating an F-statistic. For n observations and k slope coefficients, the F-test for the null hypothesis that the slope coefficients are all equal to 0 is denoted $F_{k,n-(k+1)}$. The subscript indicates that the test should have k degrees of freedom in the numerator (numerator degrees of freedom) and $n - (k + 1)$ degrees of freedom in the denominator (denominator degrees of freedom).

The formula for the F-statistic is

$$F = \frac{\text{RSS}/k}{\text{SSE}/\left[n - (k+1)\right]} = \frac{\text{Mean regression sum of squares}}{\text{Mean squared error}} = \frac{\text{MSR}}{\text{MSE}} \tag{4}$$

where MSR is the mean regression sum of squares and MSE is the mean squared error. In our regression output tables, MSR and MSE are the first and second quantities under the MSS (mean sum of squares) column in the ANOVA section of the output. If the regression model does a good job of explaining variation in the dependent variable, then the ratio MSR/MSE will be large.

What does this F-test tell us when the independent variables in a regression model explain none of the variation in the dependent variable? In this case, each predicted value in the regression model, $\hat{Y}_i$, has the average value of the dependent variable, $\overline{Y}$, and the regression sum of squares, $\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$ is 0. Therefore, the F-statistic for testing the null hypothesis (that all the slope coefficients are equal to 0) has a value of 0 when the independent variables do not explain the dependent variable at all.

To specify the details of making the statistical decision when we have calculated F, we reject the null hypothesis at the α significance level if the calculated value of F is greater than the upper α critical value of the F distribution with the specified numerator and denominator degrees of freedom. Note that we use a one-tailed F-test (because MSR necessarily increases relative to MSE as the explanatory power of the regression increases.)

We can illustrate the test using Exhibit 1, in which we investigated whether the natural log of the number of NASDAQ market makers and the natural log of the stock's market capitalization explained the natural log of the bid–ask spread divided by price. Assume that we set the significance level for this test to $\alpha = 0.05$ (i.e., a 5% probability that we will mistakenly reject the null hypothesis if it is true). Excerpt from Exhibit 1 presents the results of variance computations for this regression.

Excerpt from Exhibit 1:

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	2	3,728.1334	1,864.0667	2,216.7505	0.00
Residual	2,584	2,172.8870	0.8409		
Total	2,586	5,901.0204			

This model has two slope coefficients (k = 2), so two degrees of freedom are in the numerator of this F-test. With 2,587 observations in the sample, the number of degrees of freedom in the denominator of the F-test is $n - (k + 1) = 2{,}587 - 3 = 2{,}584$. The sum of the squared errors is 2,172.8870. The regression sum of squares is 3,728.1334. Therefore, the F-test for the null hypothesis that the two slope coefficients in this model equal 0 is

$$\frac{3{,}728.1334/2}{2{,}172.8870/2{,}584} = 2{,}216.7505$$

This test statistic is distributed as an $F_{2,2{,}584}$ random variable under the null hypothesis that the slope coefficients are equal to 0. In Exhibit 1 for the 0.05 significance level, we look at the second column, which shows F-distributions with two degrees of freedom in the numerator. Near the bottom of the column, we find that the critical value of the F-test needed to reject the null hypothesis is between 3.00 and 3.07. (We see a range of values because the denominator has more than 120 degrees of freedom but less than an infinite number of degrees of freedom.) The actual value of the F-test statistic at 2,216.75 is much greater, so we reject the null hypothesis that coefficients of both independent variables equal 0. In fact, regression results in Exhibit 1 under "Significance F," reports a p-value of 0. This p-value means that the smallest level of significance at which the null hypothesis can be rejected is practically 0. The large value for this F-statistic implies a very small probability of incorrectly rejecting the null hypothesis (a mistake known as a Type I error).

2.1 Adjusted R^2

In our coverage of simple regression, we presented the coefficient of determination, R^2, as a measure of the goodness of fit of an estimated regression to the data. In a multiple linear regression, however, R^2 is less appropriate as a measure of whether a regression model fits the data well (goodness of fit). Recall that R^2 is defined as

$$\frac{\text{Total variation} - \text{Unexplained variation}}{\text{Total variation}}$$

The numerator equals the regression sum of squares, RSS. Thus, R^2 states RSS as a fraction of the total sum of squares, $\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$. If we add regression variables to the model, the amount of unexplained variation will decrease; RSS will increase if the new independent variable explains any of the unexplained variation in the model. Such a reduction occurs when the new independent variable is even slightly correlated with the dependent variable and is not a linear combination of other independent variables in the regression (note that we say that variable y is a linear combination of variables x and z, or even more variables, if $y = ax + bz$ for some constants a and b). Consequently, we can increase R^2 simply by including many additional independent variables that explain even a slight amount of the previously unexplained variation, even if the amount they explain is not statistically significant.

Some financial analysts use an alternative measure of goodness of fit called **adjusted R^2**, or $\bar{R}^2$. This measure of fit does not automatically increase when another variable is added to a regression; it is adjusted for degrees of freedom. Adjusted R^2 is typically part of the multiple regression output produced by statistical software packages.

The relation between R^2 and $\bar{R}^2$ is

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\left(1 - R^2\right)$$

where n is the number of observations and k is the number of independent variables (the number of slope coefficients). Note that if $k \geq 1$, then R^2 is strictly greater than adjusted R^2. When a new independent variable is added, $\bar{R}^2$ can decrease if adding that variable results in only a small increase in R^2. In fact, $\bar{R}^2$ can be negative, although R^2 is always nonnegative. When $\bar{R}^2$ is negative, we can effectively consider its value to be 0. If we use $\bar{R}^2$ to compare regression models, it is important that the dependent variable be defined the same way in both models and that the sample sizes used to estimate the models are the same. For example, it makes a difference for the value of $\bar{R}^2$ if the dependent variable is GDP (gross domestic product) or ln(GDP), even if the independent variables are identical. Furthermore, we should be aware that a high $\bar{R}^2$ does not necessarily indicate that the regression is well specified in the sense of including the correct set of variables. One reason for caution is that a high $\bar{R}^2$ may reflect peculiarities of the dataset used to estimate the regression. To evaluate a regression model, we need to take many other factors into account, as we discuss later in the section on model specification.

3 DUMMY VARIABLES IN A MULTIPLE LINEAR REGRESSION

- **i.** evaluate how well a regression model explains the dependent variable by analyzing the output of the regression equation and an ANOVA table;
- **j.** formulate and interpret a multiple regression, including qualitative independent variables;
- **o.** evaluate and interpret a multiple regression model and its results.

Financial analysts often need to use qualitative variables as independent variables in a regression. One such type of variable that we will focus on is called a **dummy variable**. It takes on a value of 1 if a particular condition is true and 0 if that condition is false. We will see that one purpose of using dummy variables is to distinguish between "groups" or "categories" of data.

3.1 Defining a Dummy Variable

A dummy variable may arise in several ways in datasets:

- **i.** It may reflect an inherent property of the data (e.g., belonging to an industry or a region). For example, a company belongs to health care industry (dummy variable = 1) or it does not (dummy variable = 0). The data on such variables are collected directly along with the rest of the independent variables for each observation.

ii. It may be an identified characteristic of the data. We may introduce such a binary variable by a condition that is either true or false. For example, the date may be before 2008 (prior to the onset of financial crisis, dummy variable = 0) or after 2008 (after the onset of the financial crisis, dummy variable = 1).

iii. Alternatively, it may be constructed from some characteristic of the data. The dummy variable would reflect a condition that is either true or false. Examples would include satisfying a condition, such as particular company size (dummy = 1 if revenues exceed €1bn, otherwise it equals 0).

We need to exercise care when choosing the number of dummy variables in a regression. If we want to distinguish among n categories, we need $n - 1$ dummy variables. So, if we use dummy variables to denote companies belonging to one of 11 industries, we would use 10 dummies. We still apply the analysis to 11 categories, but the one to which we do not assign a dummy will be referred to as the "base" or "control" group. If, for example, we wish to analyze a dataset with three mutual fund types, we need two dummies. The reason for the need to use $n - 1$ dummy variables is that we must not violate assumption 2 that no exact linear relationship must exist between two or more of the independent variables. If we were to make the mistake of including dummy variables for all n categories rather than $n - 1$, the regression procedure would fail due to the complete violation of assumption 2.

3.2 Visualizing and Interpreting Dummy Variables

One of the most common types of dummy variables are so-called intercept dummies. Consider a regression model for the dependent variable Y that involves one continuous independent variable (X) and one dummy variable (D).

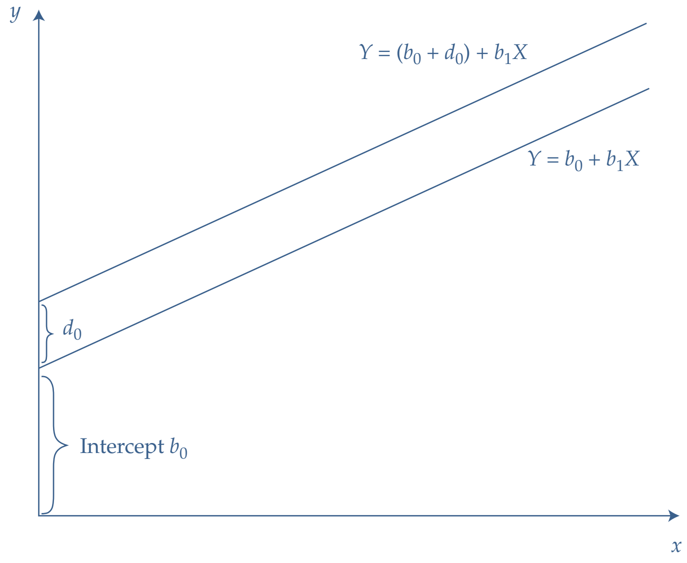
$$Y = b_0 + d_0D + b_1X + \varepsilon. \tag{5}$$

This single regression model can be shown to estimate two lines of best fit corresponding to the value of the dummy variable:

- If $D = 0$, then the equation becomes $Y = b_0 + b_1X + \varepsilon$.
- If $D = 1$, then the equation becomes $Y = (b_0 + d_0) + b_1X + \varepsilon$.

Exhibit 4 shows a graphical illustration. This scenario can be interpreted as an intercept shift shown by the d_0 distance. The shift can be positive or negative (it is positive in the illustration). The line where the dummy takes the value of zero ($D = 0$) relates to the base category; the other line where the dummy variable takes the value of 1 ($D = 1$) relates to the category to which we apply the dummy variable.

Exhibit 4 Intercept Dummy



A different scenario would reflect dummies that allow for slope differences. We will refer to those as slope dummies. They can be explained using a simple model with one continuous variable (x) and one dummy variable (D).

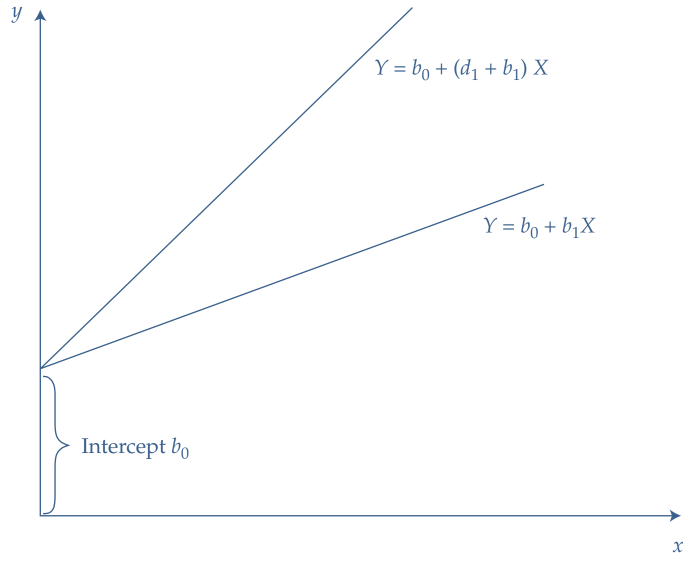
$$Y = b_0 + b_1X + d_1(D \times X) + \varepsilon. \quad (6)$$

The presence of such slope dummy can be interpreted as a change in the slope between the categories captured by the dummy variable:

If $D = 0$, then $Y = b_0 + b_1X + \varepsilon$.
If $D = 1$, then $Y = b_0 + (b_1 + d_1)\, X + \varepsilon$.

As before, the case of $D = 0$ is the base or control group. The dummy variable allows for slopes to differ between the two categories. For the base category, the relationship between x and y is shown by the less steep line $Y = b_0 + b_1X$. For the other category, the relationship between Y and X is shown by the steeper sloping line $Y = b_0 + (b_1 + d_1)X$. This difference between slopes may be positive or negative depending on the scenario.

Exhibit 5 Slope Dummy Variables



It is also possible to use dummies in both slope and intercept. To do so we combine the two previous models. We let the dummy variable "interact" with the continuous independent variable *x*.

$$Y = b_0 + d_0D + b_1X + d_1(D \times X) + \varepsilon. \tag{7}$$

If $D = 0$, then $Y = b_0 + b_1X + \varepsilon$.
If $D = 1$, then $Y = (b_0 + d_0) + (b_1 + d_1)X + \varepsilon$.

This allows for both a change in intercept and a change in slope across the two groups. In this more complex treatment, the difference between the two categories now depends on both an intercept effect (d_0) and a slope effect (d_1X) that vary with the size of the independent variable.

Exhibit 6 Slope and Intercept Dummies

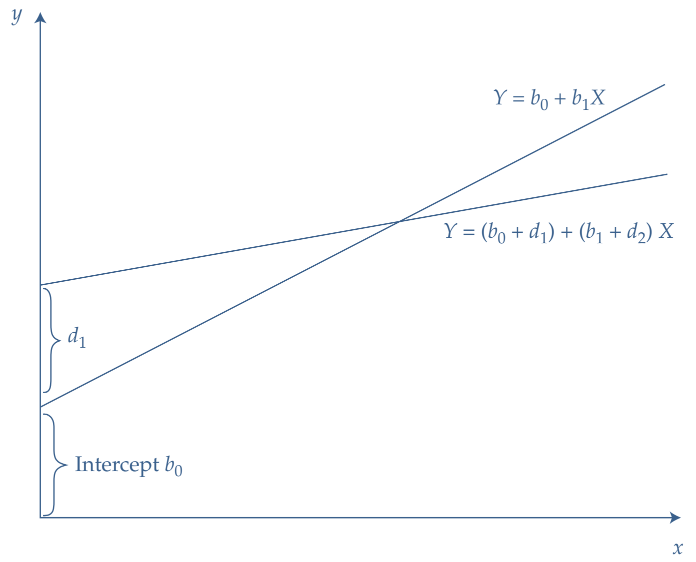


Exhibit 6 (Continued)

Note: The graph shows a scenario where $d_1 > 0$ and $d_2 < 0$.

These scenarios are based on only two data categories. We may, however, have more categories with more dummies and more independent variables. The graphs would simply show more lines of best fit, one relating to each category.

3.3 Testing for Statistical Significance

One of the purposes of using dummy variables is to distinguish between "groups" or "categories" of data. To test whether a regression function is different for one group versus another is straightforward with dummy variables with the help of *t*-tests. Individual *t*-tests on the dummy variable coefficients indicate if that difference is significantly different from zero. Exhibit 7 illustrates the use of dummy variables in a regression using a cross-section of mutual fund data.

Exhibit 7 Analysis of Mutual Funds in Different Categories

William James is a fund analyst at an investment consultancy firm. He has been tasked with analyzing how mutual fund characteristics affect fund returns measured as the average returns over the past 5 years. He uses a large database of US mutual funds that include a number of style classes. The dependent variable is the average annual return over the last 5 years. The independent variables that the analyst chose to focus on are fund expense ratio, the natural logarithm of fund size and fund age, and two dummy variables to denote fund style (style being Value, Blend, or Growth).

As there are three possible style categories, he uses $n - 1 = 2$ dummy variables. The dummy variable BLEND has a value of 1 if the observation (the mutual fund) is "Blend" and a value of 0 if it is not. The GROWTH dummy has a value of 1 if the fund is labelled as "Growth"; otherwise, it equals zero. The base or "control" category, for which we do not use a specific dummy, is the "Value" category. In this regression, for simplicity we are only allowing for an effect on the intercept of the regression, not the slopes of the independent variables.

He estimates the following cross-sectional regression model:

$$\begin{aligned}\text{Fund returns} = {} & b_0 + d_1\text{BLEND} + d_2\text{GROWTH} \\ & + b_1\text{Expense ratio} + b_2\text{Portfolio cash} \\ & + b_3\text{Fund age} + b_4\text{Log of assets} + e.\end{aligned}$$

The regression output, shown next, suggests that while the R^2 is relatively low at 0.12, the slope coefficients are statistically significant. The results suggest that fund returns are negatively impacted by the level of expenses and cash holdings (coefficients of −0.58 and −0.03, respectively), which we would intuitively expect. The results also indicate that older funds perform better, with a positive age coefficient of 0.074.

The estimated coefficients for the dummy variables show the estimated difference between the returns on different types of funds. At 0.66 and 2.50, the coefficients of the dummy variables suggest that Blend funds deliver average returns that exceed those in the Value category by 0.66% per year, while Growth funds deliver 2.50% over and above the base, "Value," category. The intercept coefficient, also statistically significant, suggests that average annual negative return of 2.91% is unexplained by the independent variables in the model.

(continued)

Exhibit 7 (Continued)

Mutual Funds Regression Output

Regression Statistics	
R^2	0.1230
Adjusted R^2	0.1228
Standard Error	4.224
Observations	23,025

ANOVA	df	SS	MS	*F*	Significance *F*
Regression	6	57,636.46	9,606	538	0
Residual	23,018	410,816.9	17.85		
Total	23,024	468,453.3			

	Coefficients	Standard Error	t-Statistic	*P*-value
Intercept	−2.909	0.299	−9.7376	2.30177E-22
Annual expense ratio	−0.586	0.0495	−11.824	3.623E-32
Portfolio cash	−0.032	0.0029	−11.168	6.93514E-29
Fund age	0.074	0.0033	22.605	6.3821E-112
Log of assets	0.267	0.0141	18.924	2.92142E-79
Blend dummy	0.661	0.0678	9.749	2.0673E-22
Growth dummy	2.498	0.0748	33.394	8.2581E-239

We can also use the *F*-test to analyze the null hypothesis that jointly the independent variables (including the dummies) all equal 0. Regression results shows the value of the *F*-statistic at 538, which we compare to the critical value. Appendix D (the *F*-distribution table) shows the critical values for this *F*-test. If we choose a significance level of 0.01 and look in column 6 (because the numerator has 6 degrees of freedom), we see that the critical value is 2.96 when the denominator has 120 degrees of freedom. The denominator actually has 23,024 degrees of freedom, so the critical value of the *F*-statistic is smaller than 2.96 (for df = 120) but larger than 2.8 (for an infinite number of degrees of freedom). The value of the *F*-test statistic is 538, so we can reject the null hypothesis that the coefficients jointly are equal to 0.

James decides to extend his study of mutual funds by introducing slope dummies. The initial results indicated a relationship between returns and fund age, although the magnitude was small at 0.07% for each year of age. He wonders whether this relationship between performance and age differs between different fund types. For example, does the age factor affect Growth or Blend funds more than it affects Value funds? In other words, is the improvement in performance with fund age different for the different types of funds?

To explore this hypothesis—that the impact of Fund age is different across types of funds—he introduces two additional independent variables, one that is a multiple of "Fund age x Blend dummy" and the second "Fund age x Growth dummy." He estimates the following model:

Fund Returns (avg over 5 years) = b_0 + b_1 Expense ratio + b_2 Portfolio cash + b_3 Fund age + b_4 Log of assets + d_1 (Blend dummy) + d_2 (Growth dummy) + "Fund age x Blend dummy" + "Fund age x Growth dummy" + ε.

When the Blend dummy is equal to 1, the interaction term takes on the value of "Fund age." For observations when the Growth dummy is equal to 1, the second interaction term takes on the value of "Fund age." The regression results are as follows:

Regression Statistics	
R^2	0.123
Adjusted R^2	0.123
Standard Error	4.224
Observations	23,025

ANOVA	df	SS	MS	*F*	Significance *F*
Regression	8	57,760.46	7,220	404.6	0
Residual	23,016	410,692.9	17.84		
Total	23,024	468,453.3			

	Coefficients	Standard Error	*t*-Statistic	*P*-value
Intercept	−2.81	0.306	−9.183	4.54531E-20
Annual_expense_ratio	−0.587	0.0496	−11.839	3.0289E-32
Portfolio_cash	−0.032	0.0029	−11.211	4.28797E-29
Fund age	0.065	0.0059	11.012	3.91371E-28
Log of assets	0.267	0.0141	18.906	4.05994E-79
Blend dummy	0.603	0.1088	5.546	2.95478E-08
Growth dummy	2.262	0.1204	18.779	4.27618E-78
Age x Blend	0.0049	0.0077	0.627	0.530817435
Age x Growth	0.0201	0.0081	2.478	0.01323

The regression results feature the original intercept and slope coefficient variables plus the new slope coefficients of the interaction dummies. The values and statistical significance of the intercept and slope coefficient show little change. But the revised model provides more information about the Fund age variable. For our default base or control group—Value funds—we observe the Fund age slope coefficient of 0.065, suggesting that those funds see extra return with the passage of time (i.e., fund age).

In this model, we also have the interaction dummies, of which "Fund age x Growth" has a statistically significant coefficient. For Growth funds, the extra annual return with each additional year of age is the sum of the "Age" and "Fund age x Growth" coefficients (i.e., 0.065 % plus 0.02 %). So, the overall "slope" coefficient for the performance of Growth (with respect to Age) is the sum of the two coefficients. One can interpret the overall output as suggesting that

Growth funds deliver returns that exceed those of Value funds by 2.26% (the Growth Intercept) plus 0.085% for each year of age. Another way to interpret this result is to imagine a two-dimensional space similar to the one in Exhibit 6. The coefficient of the "Fund age x Growth" variable would give the extra slope implied by growth over and above the slope coefficient of the "Fund Age" variable.

4 VIOLATIONS OF REGRESSION ASSUMPTIONS: HETEROSKEDASTICITY

k explain the types of heteroskedasticity and how heteroskedasticity and serial correlation affect statistical inference;

Earlier we presented the assumptions of the multiple linear regression model. Inference based on an estimated regression model rests on those assumptions being satisfied. In applying regression analysis to financial data, analysts need to be able to diagnose violations of regression assumptions, understand the consequences of violations, and know the remedial steps to take. In the following sections, we discuss three regression violations: **heteroskedasticity**, serial correlation, and multicollinearity.

4.1 Heteroskedasticity

So far, we have made an important assumption that the variance of error in a regression is constant across observations. In statistical terms, we assumed that the errors were homoskedastic. Errors in financial data, however, are often **heteroskedastic**: The variance of the errors differs across observations. In this section, we discuss how heteroskedasticity affects statistical analysis, how to test for heteroskedasticity, and how to correct for it.

We can see the difference between homoskedastic and heteroskedastic errors by comparing two graphs. Exhibit 8 shows the values of the dependent and independent variables and a fitted regression line for a model with homoskedastic errors. There is no systematic relationship between the value of the independent variable and the regression residuals (the vertical distance between a plotted point and the fitted regression line). Exhibit 9 shows the values of the dependent and independent variables and a fitted regression line for a model with heteroskedastic errors. Here, a systematic relationship is visually apparent: On average, the regression residuals grow much larger as the size of the independent variable increases.

4.1.1 *The Consequences of Heteroskedasticity*

What are the consequences when the assumption of constant error variance is violated? Although heteroskedasticity does not affect the consistency (Greene 2018) of the regression parameter estimators, it can lead to mistakes in inference. Informally, an estimator of a regression parameter is consistent if the probability that estimates of a regression parameter differ from the true value of the parameter decreases as the number of observations used in the regression increases. When errors are heteroskedastic, the F-test for the overall significance of the regression is unreliable. This unreliability occurs because the mean squared error is a biased estimator of the true population variance given heteroskedasticity. Furthermore, t-tests for the significance of individual regression coefficients are unreliable because heteroskedasticity introduces bias into estimators of the standard error of regression coefficients. If a regression shows significant heteroskedasticity, the standard errors and test statistics computed by regression programs will be incorrect unless they are adjusted for heteroskedasticity.

Exhibit 8 Regression with Homoskedasticity

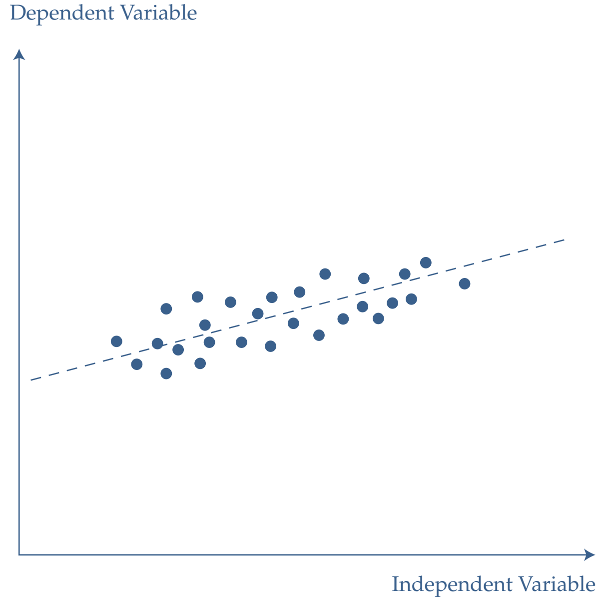
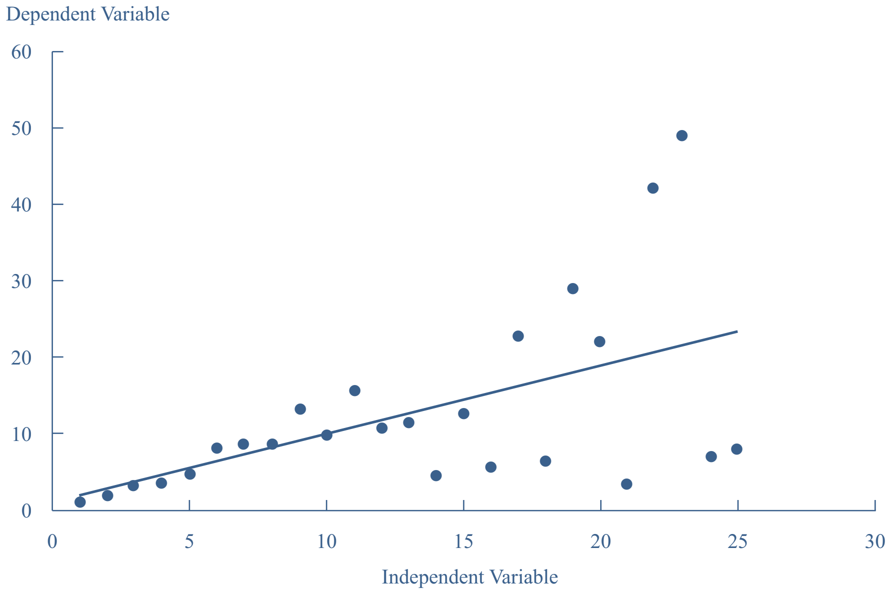


Exhibit 9 Regression with Heteroskedasticity



In regressions with financial data, the most likely results of heteroskedasticity are that the estimated standard errors will be underestimated and the t-statistics inflated. When we ignore heteroskedasticity, we tend to find significant relationships where none actually exist. Sometimes, however, failure to adjust for heteroskedasticity results in standard errors that are too large (and t-statistics that are too small). The consequences in practice may be serious if we are using regression analysis in the development of investment strategies. As Exhibit 10 shows, the issue impinges even on our understanding of financial models.

Exhibit 10 Heteroskedasticity and Tests of an Asset Pricing Model

MacKinlay and Richardson (1991) examined how heteroskedasticity affects tests of the capital asset pricing model (CAPM). These authors argued that if the CAPM is correct, they should find no significant differences between the risk-adjusted returns for holding small stocks versus large stocks. To implement their test, MacKinlay and Richardson grouped all stocks on the New York Stock Exchange and the American Stock Exchange (now called NYSE MKT) by market-value decile with annual reassignment. They then tested for systematic differences in risk-adjusted returns across market-capitalization-based stock portfolios. They estimated the following regression:

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \varepsilon_{i,t},$$

where

$r_{i,t}$ = excess return (return above the risk-free rate) to portfolio i in period t

$r_{m,t}$ = excess return to the market as a whole in period t

The CAPM formulation hypothesizes that excess returns on a portfolio are explained by excess returns on the market as a whole. That hypothesis implies that $\alpha_i = 0$ for every portfolio i; on average, no excess return accrues to any portfolio after taking into account its systematic (market) risk.

Using data from January 1926 to December 1988 and a market index based on equal-weighted returns, MacKinlay and Richardson failed to reject the CAPM at the 0.05 level when they assumed that the errors in the regression model are normally distributed and homoskedastic. They found, however, that they could reject the CAPM when they corrected their test statistics to account for heteroskedasticity. They rejected the hypothesis that there are no size-based, risk-adjusted excess returns in historical data.

We have stated that effects of heteroskedasticity on statistical inference can be severe. To be more precise about this concept, we should distinguish between two broad kinds of heteroskedasticity: unconditional and conditional.

Unconditional heteroskedasticity occurs when heteroskedasticity of the error variance is not correlated with the independent variables in the multiple regression. Although this form of heteroskedasticity violates Assumption 4 of the linear regression model, it creates no major problems for statistical inference.

The type of heteroskedasticity that causes the most problems for statistical inference is **conditional heteroskedasticity**—heteroskedasticity in the error variance that is correlated with (conditional on) the values of the independent variables in the regression. Fortunately, many statistical software packages easily test and correct for conditional heteroskedasticity.

4.1.2 *Testing for Heteroskedasticity*

Because of conditional heteroskedasticity's consequences on inference, the analyst must be able to diagnose its presence. The Breusch–Pagan test is widely used in finance research because of its generality.

Breusch and Pagan (1979) suggested the following test for conditional heteroskedasticity: Regress the squared residuals from the estimated regression equation on the independent variables in the regression. If no conditional heteroskedasticity exists, the independent variables will not explain much of the variation in the squared residuals. If conditional heteroskedasticity is present in the original regression, however, the independent variables will explain a significant portion of the variation in the squared residuals. The independent variables can explain the variation because each observation's squared residual will be correlated with the independent variables if the independent variables affect the variance of the errors.

Breusch and Pagan showed that under the null hypothesis of no conditional heteroskedasticity, nR^2 (from the regression of the squared residuals on the independent variables from the original regression) will be a χ^2 random variable with the number of degrees of freedom equal to the number of independent variables in the regression (for more on the Breusch–Pagan test, see Greene 2018). Therefore, the null hypothesis states that the regression's squared error term is uncorrelated with the independent variables. The alternative hypothesis states that the squared error term is correlated with the independent variables. Exhibit 11 illustrates the Breusch–Pagan test for conditional heteroskedasticity.

Exhibit 11 Testing for Conditional Heteroskedasticity in the Relation between Interest Rates and Expected Inflation

Suppose an analyst wants to know how closely nominal interest rates are related to expected inflation to determine how to allocate assets in a fixed-income portfolio. The analyst wants to test the Fisher effect, the hypothesis suggested by Irving Fisher that nominal interest rates increase by 1 percentage point for every 1 percentage point increase in expected inflation. The Fisher effect assumes the following relation between nominal interest rates, real interest rates, and expected inflation:

$$i = r + \pi^e,$$

where

i = the nominal rate

r = the real interest rate (assumed constant)

π^e = the expected rate of inflation

To test the Fisher effect using time-series data, we could specify the following regression model for the nominal interest rate:

$$i_t = b_0 + b_1\pi_t^e + \varepsilon_t \quad \textbf{(8)}$$

Noting that the Fisher effect predicts that the coefficient on the inflation variable is 1, we can state the null and alternative hypotheses as

H_0: $b_1 = 1$ and

H_a: $b_1 \neq 1$.

We might also specify a 0.05 significance level for the test. Before we estimate Equation 8 we must decide how to measure expected inflation (π_t^e) and the nominal interest rate (i_t).

The Survey of Professional Forecasters (SPF) has compiled data on the quarterly inflation expectations of professional forecasters using annualized median SPF prediction of current-quarter growth in the GDP deflator. We use those data as our measure of expected inflation. We use three-month US Treasury bill

(continued)

Exhibit 11 (Continued)

returns as our measure of the (risk-free) nominal interest rate. We use quarterly data from the fourth quarter of 1968 to the fourth quarter of 2013 to estimate Equation 8 The regression results are shown next.

To make the statistical decision on whether the data support the Fisher effect, we calculate the following *t*-statistic, which we then compare to its critical value:

$$t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{1.1744 - 1}{0.0761} = 2.29$$

With a 0.05 significance level and 181 − 2 = 179 degrees of freedom, the critical *t*-value is about 1.97. If we have conducted a valid test, we can reject at the 0.05 significance level the hypothesis that the true coefficient in this regression is 1 and that the Fisher effect holds. The *t*-test assumes that the errors are homoskedastic. Before we accept the validity of the *t*-test, therefore, we should test whether the errors are conditionally heteroskedastic. If those errors prove to be conditionally heteroskedastic, then the test is invalid.

Results from Regressing T-Bill Returns on Predicted Inflation

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0116	0.0033	3.5152
Inflation prediction	1.1744	0.0761	15.4323
Residual standard error	0.0233		
Multiple R^2	0.5708		
Observations	181		
Durbin–Watson statistic	0.2980		

Note: The Durbin–Watson statistic will be explained in the section on serial correlation.
Source: Federal Reserve Bank of Philadelphia, US Department of Commerce.

We can perform the **Breusch–Pagan test** for conditional heteroskedasticity on the squared residuals from the Fisher effect regression. The test regresses the squared residuals on the predicted inflation rate. The R^2 in the squared residuals regression (not shown here) is 0.0666. The test statistic from this regression, nR^2, is 181 × 0.0666 = 12.0546. Under the null hypothesis of no conditional heteroskedasticity, this test statistic is a χ^2 random variable with one degree of freedom (because there is only one independent variable).

We should be concerned about heteroskedasticity only for large values of the test statistic. Therefore, we should use a one-tailed test to determine whether we can reject the null hypothesis. The critical value of the test statistic for a variable from a χ^2 distribution with one degree of freedom at the 0.05 significance level is 3.84. The test statistic from the Breusch–Pagan test is 12.0546, so we can reject the hypothesis of no conditional heteroskedasticity at the 0.05 level. In fact, we can even reject the hypothesis of no conditional heteroskedasticity at the 0.01 significance level, because the critical value of the test statistic in the case is 6.63. As a result, we conclude that the error term in the Fisher effect regression is conditionally heteroskedastic. The standard errors computed in the original regression are not correct, because they do not account for heteroskedasticity. Therefore, we cannot accept the *t*-test as valid.

In Exhibit 11, we concluded that a t-test that we might use to test the Fisher effect was not valid. Does that mean that we cannot use a regression model to investigate the Fisher effect? Fortunately, no. A methodology is available to adjust regression coefficients' standard error to correct for heteroskedasticity. Using an adjusted standard error for $\hat{b}_1$, we can reconduct the t-test. As we shall see in the next section, using this valid t-test we will not reject the null hypothesis in Exhibit 11. That is, our statistical conclusion will change after we correct for heteroskedasticity.

4.1.3 *Correcting for Heteroskedasticity*

Financial analysts need to know how to correct for heteroskedasticity, because such a correction may reverse the conclusions about a particular hypothesis test—and thus affect a particular investment decision. In Exhibit 10, for instance, MacKinlay and Richardson reversed their investment conclusions after correcting their model's significance tests for heteroskedasticity.

We can use two different methods to correct for the effects of conditional heteroskedasticity in linear regression models. The first method, computing **robust standard errors**, corrects the standard errors of the linear regression model's estimated coefficients to account for the conditional heteroskedasticity. The second method, **generalized least squares**, modifies the original equation to eliminate the heteroskedasticity. The new, modified regression equation is then estimated under the assumption that heteroskedasticity is no longer a problem. The technical details behind these two methods of correcting for conditional heteroskedasticity are outside the scope of this discussion. Many statistical software packages can easily compute robust standard errors, however, and we recommend using them. Note that robust standard errors are also known as **heteroskedasticity-consistent standard errors** or **White-corrected standard errors**.

Returning to the subject of Exhibit 11 concerning the Fisher effect, recall that we concluded that the error variance was heteroskedastic. If we correct the regression coefficients' standard errors for conditional heteroskedasticity, we get the results shown in Exhibit 12. In comparing the standard errors with those in Exhibit 11, we see that the standard error for the intercept changes very little but the standard error for the coefficient on predicted inflation (the slope coefficient) increases by about 22% (from 0.0761 to 0.0931). Note also that the regression coefficients are the same in both tables because the results correct only the standard errors in Exhibit 11.

Exhibit 12 Results from Regressing T-Bill Returns on Predicted Inflation (Standard Errors Corrected for Conditional Heteroskedasticity)

	Coefficients	Standard Error	t-Statistic
Intercept	0.0116	0.0034	3.4118
Inflation prediction	1.1744	0.0931	12.6144
Residual standard error	0.0233		
Multiple R^2	0.5708		
Observations	181		

Source: Federal Reserve Bank of Philadelphia, US Department of Commerce.

We can now conduct a valid t-test of the null hypothesis that the slope coefficient has a true value of 1 by using the robust standard error for $\hat{b}_1$. We find that $t = (1.1744 - 1)/0.0931 = 1.8733$. This number is smaller than the critical value of 1.97 needed to reject the null hypothesis that the slope equals 1 (remember, this is a two-tailed test).

So, we can no longer reject the null hypothesis that the slope equals 1 because of the greater uncertainty (standard error) around the coefficient estimate. Thus, in this example, correcting for the statistically significant conditional heteroskedasticity had an effect on the result of the hypothesis test about the slope of the predicted inflation coefficient. Exhibit 10 concerning tests of the CAPM is a similar case. In other cases, however, our statistical decision might not change based on using robust standard errors in the *t*-test.

5 VIOLATIONS OF REGRESSION ASSUMPTIONS: SERIAL CORRELATION

k explain the types of heteroskedasticity and how heteroskedasticity and serial correlation affect statistical inference;

A more common—and potentially more serious—problem than violation of the homoskedasticity assumption is the violation of the assumption that regression errors are uncorrelated across observations. Trying to explain a particular financial relation over a number of periods is risky because errors in financial regression models are often correlated through time.

When regression errors are correlated across observations, we say that they are **serially correlated** (or autocorrelated). Serial correlation most typically arises in time-series regressions. In this section, we discuss three aspects of serial correlation: its effect on statistical inference, tests for it, and methods to correct for it.

5.1 The Consequences of Serial Correlation

As with heteroskedasticity, the principal problem caused by serial correlation in a linear regression is an incorrect estimate of the regression coefficient standard errors computed by statistical software packages. As long as none of the independent variables is a lagged value of the dependent variable (a value of the dependent variable from a previous period), then the estimated parameters themselves will be consistent and need not be adjusted for the effects of serial correlation. If, however, one of the independent variables is a lagged value of the dependent variable—for example, if the T-bill return from the previous month was an independent variable in the Fisher effect regression—then serial correlation in the error term will cause all the parameter estimates from linear regression to be inconsistent and they will not be valid estimates of the true parameters (we will address this later).

In none of the regressions examined so far is an independent variable a lagged value of the dependent variable. Thus, in these regressions any effect of serial correlation appears in the regression coefficient standard errors. We will examine here the positive serial correlation case because that case is so common. **Positive serial correlation** is serial correlation in which a positive error for one observation increases the chance of a positive error for another observation. Positive serial correlation also means that a negative error for one observation increases the chance of a negative error for another observation. In contrast, with **negative serial correlation**, a positive error for one observation increases the chance of a negative error for another observation, and a negative error for one observation increases the chance of a positive error for another. In examining positive serial correlation, we make the common assumption that serial correlation takes the form of **first-order serial correlation**, or serial correlation between adjacent observations. In a time-series context, that assumption means the sign of the error term tends to persist from one period to the next.

Although positive serial correlation does not affect the consistency of the estimated regression coefficients, it does affect our ability to conduct valid statistical tests. First, the *F*-statistic to test for overall significance of the regression may be inflated because the mean squared error (MSE) will tend to underestimate the population error variance. Second, positive serial correlation typically causes the ordinary least squares (OLS) standard errors for the regression coefficients to underestimate the true standard errors. Consequently, if positive serial correlation is present in the regression, standard linear regression analysis will typically lead us to compute artificially small standard errors for the regression coefficient. These small standard errors will cause the estimated *t*-statistics to be inflated, suggesting significance where perhaps there is none. The inflated *t*-statistics may, in turn, lead us to incorrectly reject null hypotheses about population values of the parameters of the regression model more often than we would if the standard errors were correctly estimated. This Type I error could lead to improper investment recommendations.

5.2 Testing for Serial Correlation

We can choose from a variety of tests for serial correlation in a regression model (see Greene 2018), but the most common is based on a statistic developed by Durbin and Watson (1951); in fact, many statistical software packages compute the Durbin–Watson statistic automatically. The equation for the Durbin–Watson test statistic is

$$\mathrm{DW} = \frac{\sum_{t=2}^{T}\left(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}\right)^2}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2} \tag{9}$$

where $\hat{\varepsilon}_t$ is the regression residual for period *t*. We can rewrite this equation as

$$\frac{\frac{1}{T-1}\sum_{t=2}^{T}\left(\hat{\varepsilon}_t^2 - 2\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} + \hat{\varepsilon}_{t-1}^2\right)}{\frac{1}{T-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2} \approx \frac{Var\left(\hat{\varepsilon}_t\right) - 2\,\mathrm{Cov}\left(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}\right) + \mathrm{Var}\left(\hat{\varepsilon}_{t-1}\right)}{\mathrm{Var}\left(\hat{\varepsilon}_t\right)}$$

If the variance of the error is constant through time, then we expect $\mathrm{Var}\left(\hat{\varepsilon}_t\right) = \hat{\sigma}_\varepsilon^2$ for all *t*, where we use $\hat{\sigma}_\varepsilon^2$ to represent the estimate of the constant error variance. If the errors are also not serially correlated, then we expect $\mathrm{Cov}\left(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}\right) = 0$. In that case, the Durbin–Watson statistic is approximately equal to

$$\frac{\hat{\sigma}_\varepsilon^2 - 0 + \hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2} = 2$$

This equation tells us that if the errors are homoskedastic and not serially correlated, then the Durbin–Watson statistic will be close to 2. Therefore, we can test the null hypothesis that the errors are not serially correlated by testing whether the Durbin–Watson statistic differs significantly from 2.If the sample is very large, the Durbin–Watson statistic will be approximately equal to $2(1 - r)$, where *r* is the sample correlation between the regression residuals from one period and those from the previous period. This approximation is useful because it shows the value of the

Durbin–Watson statistic for differing levels of serial correlation. The Durbin–Watson statistic can take on values ranging from 0 (in the case of serial correlation of +1) to 4 (in the case of serial correlation of −1):

- If the regression has no serial correlation, then the regression residuals will be uncorrelated through time and the value of the Durbin–Watson statistic will be equal to 2(1 − 0) = 2.
- If the regression residuals are positively serially correlated, then the Durbin–Watson statistic will be less than 2. For example, if the serial correlation of the errors is 1, then the value of the Durbin–Watson statistic will be 0.
- If the regression residuals are negatively serially correlated, then the Durbin–Watson statistic will be greater than 2. For example, if the serial correlation of the errors is −1, then the value of the Durbin–Watson statistic will be 4.

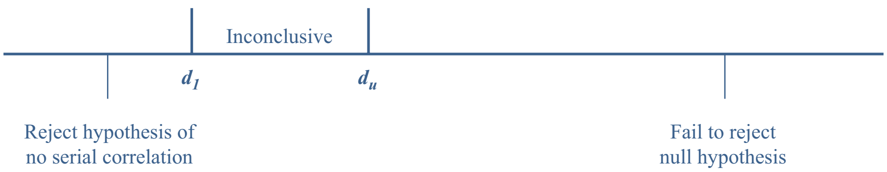
Returning to Exhibit 11, which explored the Fisher effect, the Durbin–Watson statistic for the OLS regression is 0.2980. This result means that the regression residuals are positively serially correlated:

$$\begin{aligned} DW &= 0.2980 \\ &\approx 2(1 - r) \\ r &\approx 1 - \mathrm{DW}/2 \\ &= 1 - 0.2980/2 \\ &= 0.8510 \end{aligned}$$

This outcome raises the concern that OLS standard errors may be incorrect because of positive serial correlation. Does the observed Durbin–Watson statistic (0.2980) provide enough evidence to warrant rejecting the null hypothesis of no positive serial correlation?

We should reject the null hypothesis of no serial correlation if the Durbin–Watson statistic is below a critical value, d^*. Unfortunately, Durbin and Watson also showed that for a given sample we cannot know the true critical value, d^*. Instead, we can determine only that d^* lies either between two values, d_u (an upper value) and d_l (a lower value), or outside those values. Exhibit 13 depicts the upper and lower values of d^* as they relate to the results of the Durbin–Watson statistic.

Exhibit 13 Value of the Durbin–Watson Statistic



From Exhibit 13 we learn the following:

- When the Durbin–Watson (DW) statistic is less than d_l, we reject the null hypothesis of no positive serial correlation.
- When the DW statistic falls between d_l and d_u, the test results are inconclusive.
- When the DW statistic is greater than d_u, we fail to reject the null hypothesis of no positive serial correlation (sometimes serial correlation in a regression model is negative rather than positive). For a null hypothesis of no serial correlation, the null hypothesis is rejected if DW < d_l, indicating significant positive serial correlation) or if DW > 4 − d_l, indicating significant negative serial correlation).

Returning to Exhibit 11, the Fisher effect regression has one independent variable and 181 observations. The Durbin–Watson statistic is 0.2980. We can reject the null hypothesis of no correlation in favor of the alternative hypothesis of positive serial correlation at the 0.05 level because the Durbin–Watson statistic is far below d_l for $k = 1$ and $n = 100$ (1.65). The level of d_l would be even higher for a sample of 181 observations. This finding of significant positive serial correlation suggests that the OLS standard errors in this regression probably significantly underestimate the true standard errors.

5.3 Correcting for Serial Correlation

We have two alternative remedial steps when a regression has significant serial correlation. First, we can adjust the coefficient standard errors for the linear regression parameter estimates to account for the serial correlation. Second, we can modify the regression equation itself to eliminate the serial correlation. We recommend using the first method for dealing with serial correlation; the second method may result in inconsistent parameter estimates unless implemented with extreme care.

Two of the most prevalent methods for adjusting standard errors were developed by Hansen (1982) and Newey and West (1987). These methods are standard features in many statistical software packages and the correction is known by various names, including serial-correlation consistent standard errors, serial correlation and heteroskedasticity adjusted standard errors, and robust standard errors. An additional advantage of these methods is that they simultaneously correct for conditional heteroskedasticity.

Exhibit 14 shows the results of correcting the standard errors from Exhibit 11 for serial correlation and heteroskedasticity using the Newey–West method. Note that the coefficients for both the intercept and the slope are the same as in the original regression. The robust standard errors are now much larger, however—more than twice the OLS standard errors in Exhibit 11. Because of the severe serial correlation in the regression error, OLS greatly underestimates the uncertainty about the estimated parameters in the regression.

Note also that the serial correlation has not been eliminated, but the standard error has been corrected to account for the serial correlation.

Exhibit 14 Results from Regressing T-Bill Returns on Predicted Inflation (Standard Errors Corrected for Conditional Heteroskedasticity and Serial Correlation)

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0116	0.0067	1.7313
Inflation prediction	1.1744	0.1751	6.7070
Residual standard error	0.0233		
Multiple R^2	0.5708		
Observations	181		

Source: Federal Reserve Bank of Philadelphia, US Department of Commerce.

Now suppose we want to test our original null hypothesis (the Fisher effect) that the coefficient on the predicted inflation term equals 1 (H_0: $b_1 = 1$) against the alternative that the coefficient on the inflation term is not equal to 1 (H_a: $b_1 \neq 1$). With the corrected standard errors, the value of the test statistic for this null hypothesis is

$$\frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{1.1744 - 1}{0.1751} = 0.996$$

The critical values for both the 0.05 and 0.01 significance level are much larger than 0.996 (the t-test statistic), so we cannot reject the null hypothesis. This conclusion is the same as that reached in Exhibit 10, where the correction was only for heteroskedasticity.

This shows that for some hypotheses, serial correlation and conditional heteroskedasticity could have a big effect on whether we accept or reject those hypotheses. In addition, serial correlation can also affect forecast accuracy.

6 VIOLATIONS OF REGRESSION ASSUMPTIONS: MULTICOLLINEARITY

l. describe multicollinearity, and explain its causes and effects in regression analysis;

The second assumption of the multiple linear regression model is that no exact linear relationship exists between two or more of the independent variables. When one of the independent variables is an exact linear combination of other independent variables, it becomes mechanically impossible to estimate the regression. Suppose we tried to explain a company's credit ratings with a regression that included net sales, cost of goods sold, and gross profit as independent variables. Because Gross profit = Net sales – Cost of goods sold, by definition there is an exact linear relationship between these variables. This type of blunder is relatively obvious (and easy to avoid). The problem just described, known as perfect collinearity, is much less of a practical concern than multicollinearity. **Multicollinearity** occurs when two or more independent variables (or combinations of independent variables) are highly (but not perfectly) correlated with each other. With multicollinearity we can estimate the regression, but the interpretation of the regression output becomes problematic. Multicollinearity is a serious practical concern because approximate linear relationships among financial variables are common.

6.1 The Consequences of Multicollinearity

Although the presence of multicollinearity does not affect the consistency of the OLS estimates of the regression coefficients, the estimates become extremely imprecise and unreliable. Furthermore, it becomes practically impossible to distinguish the individual impacts of the independent variables on the dependent variable. These consequences are reflected in inflated OLS standard errors for the regression coefficients. With inflated standard errors, t-tests on the coefficients have little power (ability to reject the null hypothesis).

6.2 Detecting Multicollinearity

In contrast to the cases of heteroskedasticity and serial correlation, we shall not provide a formal statistical test for multicollinearity. In practice, multicollinearity is often a matter of degree rather than of absence or presence.

The analyst should be aware that using the magnitude of pairwise correlations among the independent variables to assess multicollinearity, as has occasionally been suggested, is generally inadequate. Although very high pairwise correlations among independent variables can indicate multicollinearity, it is not necessary for such pairwise correlations to be high for there to be a problem of multicollinearity. Stated another way, high pairwise correlations among the independent variables are not a necessary condition for multicollinearity, and low pairwise correlations do not mean that multicollinearity is not a problem. Even if pairs of independent variables have low correlation, there may be linear combinations of the independent variables that are very highly correlated, creating a multicollinearity problem. The only case in which correlation between independent variables may be a reasonable indicator of multicollinearity occurs in a regression with exactly two independent variables.

The classic symptom of multicollinearity is a high R^2 (and significant F-statistic), even though the t-statistics on the estimated slope coefficients are not significant. The insignificant t-statistics reflect inflated standard errors. Although the coefficients might be estimated with great imprecision, as reflected in low t-statistics, the independent variables *as a group* may do a good job of explaining the dependent variable. A high R^2 would reflect this effectiveness. Exhibit 15 illustrates this diagnostic.

Exhibit 15 Multicollinearity in Explaining Returns to the Fidelity Select Technology Portfolio

In Exhibit 3 we regressed returns to the Fidelity Select Technology Portfolio (FSPTX) on returns to the S&P 500 Growth Index and the S&P 500 Value Index using data from August 2014 through August 2019. The regression results are reproduced next. The t-statistic of 11.88 on the growth index return is greater than 2, indicating that the coefficient on the growth index differs significantly from 0 at standard significance levels. The t-statistic on the value index return is −2.93 and is therefore also statistically significant. This result suggests that the returns to the FSPTX are linked to the returns to the growth index and negatively associated with the returns to the value index. Note that the coefficient on the growth index, however, is 1.585. This result implies that returns on the FSPTX are more volatile than are returns on the growth index.

Results from Regressing the FSPTX Returns on the S&P 500 Growth and Value Indexes

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0011	0.0025	0.4406
S&P 500 Growth Index	1.5850	0.1334	11.8843
S&P 500 Value Index	−0.3902	0.1332	−2.93

ANOVA	df	SS	MSS	F	Significance F
Regression	2	0.1198	0.0599	178	0.000
Residual	57	0.0192	0.0003		
Total	59	0.1390			
Residual standard error			0.0183		

(continued)

Exhibit 15 (Continued)

ANOVA	df	SS	MSS	F	Significance F
Multiple R^2			0.862		
Observations			60		

Source: Finance.yahoo.com.

Note also that this regression explains a significant amount of the variation in the returns to the FSPTX. Specifically, the R^2 from this regression is 0.8627. Thus, approximately 86% of the variation in the returns to the FSPTX is explained by returns to the S&P 500 Growth and S&P 500 Value Indexes.

Now suppose we run another linear regression that adds returns to the S&P 500 itself to the returns to the S&P 500 Growth and S&P 500 Value Indexes. The S&P 500 includes the component stocks of these two style indexes, so we are introducing a severe multicollinearity problem.

The regression results are shown next. Note that the R^2 in this regression has changed almost imperceptibly from the R^2 in the previous regression (increasing from 0.8620 to 0.8624), but now the standard errors of the coefficients of the independent variables are much larger. Adding the return to the S&P 500 to the previous regression does not explain any more of the variance in the returns to the FSPTX than the previous regression did, but now none of the coefficients is statistically significant. This is the classic case of multicollinearity.

Results from Regressing the FSPTX Returns on Returns to the S&P 500 Growth and S&P 500 Value Indexes and the S&P 500 Index

	Coefficient	Standard Error	t-Statistic
Intercept	0.0008	0.0025	0.4047
S&P 500 Growth Index	−0.1873	4.1890	−0.0447
S&P 500 Value Index	−1.8717	3.7387	−0.5274
S&P 500 Index	3.3522	7.9194	−0.4233

ANOVA	df	SS	MSS	F	Significance F
Regression	3	0.1198	0.0399	117.02	4.26E-24
Residual	56	0.0191	0.0003		
Total	59	0.1389			
Residual standard error			0.0185		
Multiple R^2			0.8624		
Observations			60		

Source: finance.yahoo.com.

Multicollinearity may be a problem even when we do not observe the classic symptom of insignificant t-statistics but a highly significant F-test. Advanced textbooks provide further tools to help diagnose multicollinearity (Greene 2018).

6.3 Correcting for Multicollinearity

The most direct solution to multicollinearity is excluding one or more of the regression variables. In the previous example, we can see that the S&P 500 total returns should not be included if both the S&P 500 Growth and S&P 500 Value Indexes are included because the returns to the entire S&P 500 Index are a weighted average of the return to growth stocks and value stocks. In many cases, however, no easy solution is available to the problem of multicollinearity, and you will need to experiment with including or excluding different independent variables to determine the source of multicollinearity.

6.4 Heteroskedasticity, Serial Correlation, Multicollinearity: Summarizing the Issues

We have discussed some of the problems that heteroskedasticity, serial correlation, and multicollinearity may cause in interpreting regression results. These violations of regression assumptions, we have noted, all lead to problems in making valid inferences. The analyst should check that model assumptions are fulfilled before interpreting statistical tests.

Exhibit 16 gives a summary of these problems, the effect they have on the linear regression results (an analyst can see these effects using regression software), and the solutions to these problems.

Exhibit 16 Problems in Linear Regression and Their Solutions

Problem	Effect	Solution
Heteroskedasticity	Incorrect standard errors	Use robust standard errors (corrected for conditional heteroskedasticity)
Serial correlation	Incorrect standard errors (additional problems if a lagged value of the dependent variable is used as an independent variable)	Use robust standard errors (corrected for serial correlation)
Multicollinearity	High R^2 and low t-statistics	Remove one or more independent variables; often no solution based in theory

MODEL SPECIFICATION ERRORS 7

m describe how model misspecification affects the results of a regression analysis, and describe how to avoid common forms of misspecification;

i evaluate how well a regression model explains the dependent variable by analyzing the output of the regression equation and an ANOVA table;

o evaluate and interpret a multiple regression model and its results.

Until now, we have assumed that whatever regression model we estimate is correctly specified. **Model specification** refers to the set of variables included in the regression and the regression equation's functional form. In the following, we first give some broad guidelines for correctly specifying a regression. Then, we turn to three types of model misspecification: misspecified functional form, regressors that are correlated with the error term, and additional time-series misspecification. Each of these types of misspecification invalidates statistical inference using OLS; most of these misspecifications will cause the estimated regression coefficients to be inconsistent.

7.1 Principles of Model Specification

In discussing the principles of model specification, we need to acknowledge that there are competing philosophies about how to approach model specification. Furthermore, our purpose for using regression analysis may affect the specification we choose. The following principles have fairly broad application, however.

- *The model should be grounded in cogent economic reasoning.* We should be able to supply the economic reasoning behind the choice of variables, and the reasoning should make sense. When this condition is fulfilled, we increase the chance that the model will have predictive value with new data. This approach contrasts to the variable-selection process known as **data mining**. With data mining, the investigator essentially develops a model that maximally exploits the characteristics of a specific dataset. "Data mining" is used in the different sense of discovering patterns in large datasets.
- *The functional form chosen for the variables in the regression should be appropriate given the nature of the variables.* As one illustration, consider studying mutual fund **market timing** based on fund and market returns alone. One might reason that for a successful timer, a plot of mutual fund returns against market returns would show curvature because a successful timer would tend to increase (decrease) beta when market returns were high (low). The model specification should reflect the expected nonlinear relationship. In other cases, we may transform the data such that a regression assumption is better satisfied.
- *The model should be parsimonious.* In this context, "parsimonious" means accomplishing a lot with a little. We should expect each variable included in a regression to play an essential role.
- *The model should be examined for violations of regression assumptions before being accepted.* We have already discussed detecting the presence of heteroskedasticity, serial correlation, and multicollinearity. As a result of such diagnostics, we may conclude that we need to revise the set of included variables and/or their functional form.
- *The model should be tested and be found useful out of sample before being accepted.* The term "out of sample" refers to observations outside the dataset on which the model was estimated. A plausible model may not perform well out of sample because economic relationships have changed since the sample period. That possibility is itself useful to know. A second explanation, however, may be that relationships have not changed but that the model explains only a specific dataset.

Having given some broad guidance on model specification, we turn to a discussion of specific model specification errors. Understanding these errors will help an analyst develop better models and be a more informed consumer of investment research.

7.2 Misspecified Functional Form

Whenever we estimate a regression, we must assume that the regression has the correct functional form. This assumption can fail in several ways:

- Omitted variable(s). One or more important variables could be omitted from regression.
- Inappropriate variable scaling. One or more of the regression variables may need to be transformed (for example, by taking the natural logarithm of the variable) before estimating the regression.
- Inappropriate data pooling. The regression model pools data from different samples that should not be pooled.

First, consider the effects of omitting an important independent variable from a regression (omitted variable bias). If the true regression model was

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \varepsilon_i, \quad (10)$$

but we estimate the model

$$Y_i = a_0 + a_1X_{1i} + \varepsilon_i,$$

then our regression model would be misspecified (note the different notation when X_{2i} is omitted, because the intercept term and slope coefficient on X_{1i} will generally not be the same as when X_{2i} is included). What is wrong with the model?

If the omitted variable (X_2) is correlated with the remaining variable (X_1), then the error term in the model will be correlated with (X_1) and the estimated values of the regression coefficients a_0 and a_1 would be biased and inconsistent. In addition, the estimates of the standard errors of those coefficients will also be inconsistent. So, we can use neither the coefficient estimates nor the estimated standard errors to make statistical tests.

Exhibit 17 Omitted Variable Bias and the Bid–Ask Spread

In this example, we extend our examination of the bid–ask spread to show the effect of omitting an important variable from a regression. In Example 1, we showed that the natural logarithm of the ratio [(Bid–ask spread)/Price] was significantly related to both the natural logarithm of the number of market makers and the natural logarithm of the market capitalization of the company. We repeat the regression results from Exhibit 1 next.

Results from Regressing ln(Bid–Ask Spread/Price) on ln(Number of Market Makers) and ln(Market Capitalization) (repeated)

	Coefficients	Standard Error	*t*-Statistic
Intercept	1.5949	0.2275	7.0105
ln(Number of NASDAQ market makers)	−1.5186	0.0808	−18.7946
ln(Company's market capitalization)	−0.3790	0.0151	−25.0993

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	2	3,728.1334	1,864.0667	2,216.7505	0.00
Residual	2,584	2,172.8870	0.8409		

(continued)

Exhibit 17 (Continued)

ANOVA	df	SS	MSS	*F*	Significance *F*
Total	2,586	5,901.0204			
Residual standard error			0.9170		
Multiple R^2			0.6318		
Observations			2,587		

Source: Center for Research in Security Prices, University of Chicago.

If we did not include the natural log of market capitalization as an independent variable in the regression and regressed the natural logarithm of the ratio [(Bid–ask spread)/Price] only on the natural logarithm of the number of market makers for the stock, the results would be as shown next.

Results from Regressing ln(Bid–Ask Spread/Price) on ln(Number of Market Makers)

	Coefficients	Standard Error	*t*-Statistic
Intercept	5.0707	0.2009	25.2399
ln(Number of NASDAQ market makers)	−3.1027	0.0561	−55.3066

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	1	3,200.3918	3,200.3918	3,063.3655	0.00
Residual	2,585	2,700.6287	1.0447		
Total	2,586	5,901.0204			
Residual standard error			1.0221		
Multiple R^2			0.5423		
Observations			2,587		

Source: Center for Research in Security Prices, University of Chicago.

Note that the coefficient on ln(Number of NASDAQ market makers) changed from −1.5186 in the original (correctly specified) regression to −3.1027 in the misspecified regression. Also, the intercept changed from 1.5949 in the correctly specified regression to 5.0707 in the misspecified regression. These results illustrate that omitting an independent variable that should be in the regression can cause the remaining regression coefficients to be inconsistent.

A second common cause of misspecification in regression models is the use of the wrong form of the data in a regression when a transformed version of the data is appropriate. For example, sometimes analysts fail to account for curvature or nonlinearity in the relationship between the dependent variable and one or more of the independent variables, instead specifying a linear relation among variables. When we are specifying a regression model, we should consider whether economic theory suggests a nonlinear relation. We can often confirm the nonlinearity by plotting the data, as we will illustrate in Example 2. If the relationship between the variables becomes linear when one or more of the variables is represented as a proportional change in the variable, we may be able to correct the misspecification by taking the natural logarithm of the variable(s) we want to represent as a proportional change. Other times, analysts use unscaled data in regressions when scaled data (such as dividing net income or cash flow by sales) are more appropriate. In Exhibit 1, we scaled the bid–ask spread by stock price because what a given bid–ask spread means in terms of transactions costs for a given size investment depends on the price of the stock. If we had not scaled the bid–ask spread, the regression would have been misspecified.

EXAMPLE 2

Nonlinearity and the Bid–Ask Spread

In Exhibit 1, we showed that the natural logarithm of the ratio [(Bid–ask spread)/Price] was significantly related to both the natural logarithm of the number of market makers and the natural logarithm of the company's market capitalization. But why did we take the natural logarithm of each of the variables in the regression? We began a discussion of this question in Exhibit 1, which we continue now.

What does theory suggest about the nature of the relationship between the ratio (Bid–ask spread)/Price, or the percentage bid–ask spread, and its determinants (the independent variables)? Stoll (1978) builds a theoretical model of the determinants of percentage bid–ask spread in a dealer market. In his model, the determinants enter multiplicatively in a particular fashion. In terms of the independent variables introduced in Exhibit 1, the functional form assumed is

$$\left[\left(\text{Bid–ask spread}\right)/\text{Price}\right]_i = c\left(\text{Number of market makers}\right)_i^{b_1} \times \left(\text{Market capitalization}\right)_i^{b_2}$$

where c is a constant. The relationship of the percentage bid–ask spread with the number of market makers and market capitalization is not linear in the original variables (the form of the model is analogous to the Cobb–Douglas production function in economics). If we take the natural log of both sides of this model, however, we have a log-log regression that is linear in the transformed variables:

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \varepsilon_i,$$

where

Y_i = the natural logarithm of the ratio (Bid–ask spread)/Price for stock i

b_0 = a constant that equals $\ln(c)$

X_{1i} = the natural logarithm of the number of market makers for stock i

X_{2i} = the natural logarithm of the market capitalization of company i

ε_i = the error term (note: added to the model)

As mentioned in Exhibit 1, a slope coefficient in the log-log model is interpreted as an elasticity—precisely, the partial elasticity of the dependent variable with respect to the independent variable ("partial" means holding the other independent variables constant).

We can plot the data to assess whether the variables are linearly related after the logarithmic transformation. For example, in Exhibit 18 we show a scatterplot of the natural logarithm of the number of market makers for a stock (on the X axis) and the natural logarithm of (Bid–ask spread)/Price (on the Y axis) as well as a regression line showing the linear relation between the two transformed variables. The relation between the two transformed variables is clearly linear.

Exhibit 18 Linear Regression When Two Variables Have a Linear Relation

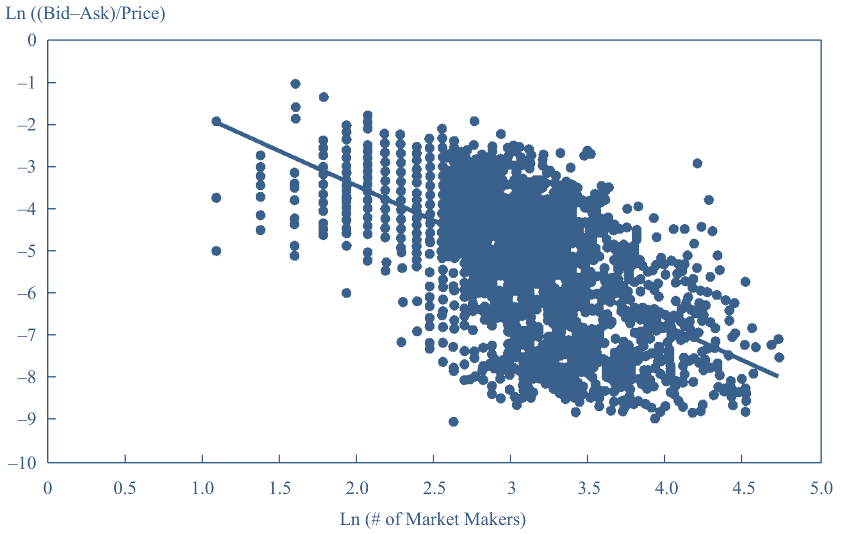
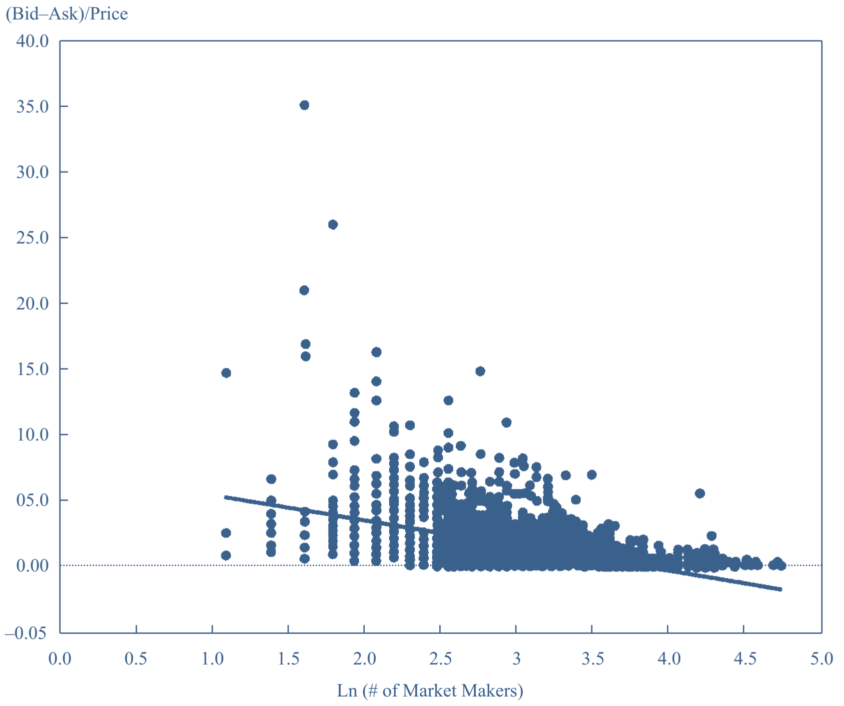


Exhibit 19 Linear Regression When Two Variables Have a Nonlinear Relation



If we do not take log of the ratio (Bid–ask spread)/Price, the plot is not linear. Exhibit 19 shows a plot of the natural logarithm of the number of market makers for a stock (on the X axis) and the ratio (Bid–ask spread)/Price expressed as a percentage (on the Y axis) as well as a regression line that attempts to show a linear relation between the two variables. We see that the relation between the two variables is very nonlinear. Note that the relation between (Bid–ask spread)/Price and ln(Market cap) is also nonlinear, while the relation between ln(Bid–ask spread)/Price and ln(Market cap) is linear; we omit these scatterplots to save space. Consequently, we should not estimate a regression with (Bid–ask spread)/Price as the dependent variable. Consideration of the need to ensure that predicted bid–ask spreads are positive would also lead us to not use (Bid–ask spread)/Price as the dependent variable. If we use the non-transformed ratio (Bid–ask spread)/Price as the dependent variable, the estimated model could predict negative values of the bid–ask spread. This result would be nonsensical; in reality, no bid–ask spread is negative (it is hard to motivate traders to simultaneously buy high and sell low), so a model that predicts negative bid–ask spreads is certainly misspecified. In our data sample, the bid–ask spread for each of the 2,587 companies is positive. We illustrate the problem of negative values of the predicted bid–ask spreads now.

Exhibit 20 shows the results of a regression with (Bid–ask spread)/ Price as the dependent variable and the natural logarithm of the number of market makers and the natural logarithm of the company's market capitalization as the independent variables.

Exhibit 20 Results from Regressing Bid–Ask Spread/Price on ln(Number of Market Makers) and ln(Market Cap)

	Coefficients	Standard Error	*t*-Statistic
Intercept	0.0674	0.0035	19.2571
ln(Number of NASDAQ market makers)	−0.0142	0.0012	−11.8333
ln(Company's market cap)	−0.0016	0.0002	−8.0000

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	2	0.1539	0.0770	392.3338	0.00
Residual	2,584	0.5068	0.0002		
Total	2,586	0.6607			
Residual standard error			0.0140		
Multiple R^2			0.2329		
Observations			2,587		

Source: Center for Research in Security Prices, University of Chicago.

1 Suppose that for a particular NASDAQ-listed stock, the number of market makers is 50 and the market capitalization is $6 billion. What is the predicted ratio of bid–ask spread to price for this stock based on the model just shown?

Solution to 1:

The natural log of the number of market makers equals ln 50 = 3.9120, and the natural log of the stock's market capitalization (in millions) is ln 6,000 = 8.6995. In this case, the predicted ratio of bid–ask spread to price is 0.0674 + (−0.0142 × 3.9120) + (−0.0016 × 8.6995) = −0.0021. Therefore, the model predicts that the ratio of bid–ask spread to stock price is −0.0021 or −0.21% of the stock price.

2 Does the predicted bid–ask spread for this stock make sense? If not, how could this problem be avoided?

Solution to 2:

The predicted bid–ask spread is negative, which does not make economic sense. This problem could be avoided by using log of (Bid–ask spread)/Price as the dependent variable. Whether the natural log of the percentage bid–ask spread, Y, is positive or negative, the percentage bid–ask spread found as e^Y is positive because a positive number raised to any power is positive. The constant e is positive ($e \approx 2.7183$).

Often, analysts must decide whether to scale variables before they compare data across companies. For example, in financial statement analysis, analysts often compare companies using **common size statements**. In a common size income statement, all the line items in a company's income statement are divided by the company's revenues. Common size statements make comparability across companies much easier. An analyst can use common size statements to quickly compare trends in gross margins (or other income statement variables) for a group of companies.

Issues of comparability also appear for analysts who want to use regression analysis to compare the performance of a group of companies. Exhibit 21 illustrates this issue.

Exhibit 21 Scaling and the Relation between Cash Flow from Operations and Free Cash Flow

Suppose we go back to the year 2001 and want to explain free cash flow to the firm as a function of cash flow from operations in 2001 for 11 family clothing stores in the United States with market capitalizations of more than \$100 million as of the end of 2001.

To investigate this issue, the analyst might use free cash flow as the dependent variable and cash flow from operations as the independent variable in single-independent-variable linear regression. Next, we show the results of that regression. Note that the t-statistic for the slope coefficient for cash flow from operations is quite high (6.5288), the significance level for the F-statistic for the regression is very low (0.0001), and the R^2 is quite high. We might be tempted to believe that this regression is a success and that for a family clothing store, if cash flow from operations increased by \$1.00, we could confidently predict that free cash flow to the firm would increase by \$0.3579.

Results from Regressing the Free Cash Flow on Cash Flow from Operations for Family Clothing Stores

	Coefficients	Standard Error	t-Statistic
Intercept	0.7295	27.7302	0.0263
Cash flow from operations	0.3579	0.0548	6.5288

Exhibit 21 (Continued)

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	1	245,093.7836	245,093.7836	42.6247	0.0001
Residual	9	51,750.3139	5,750.0349		
Total	10	296,844.0975			
Residual standard error			75.8290		
Multiple R^2			0.8257		
Observations			11		

Source: Compustat.

But is this specification correct? The regression does not account for size differences among the companies in the sample.

We can account for size differences by using common size cash flow results across companies. We scale the variables by dividing cash flow from operations and free cash flow to the firm by the company's sales before using regression analysis. We will use (Free cash flow to the firm/Sales) as the dependent variable and (Cash flow from operations/Sales) as the independent variable. The results are shown next. Note that the t-statistic for the slope coefficient on (Cash flow from operations/Sales) is 1.6262, so it is not significant at the 0.05 level. Note also that the significance level of the F-statistic is 0.1383, so we cannot reject at the 0.05 level the hypothesis that the regression does not explain variation in (Free cash flow/Sales) among family clothing stores. Finally, note that the R^2 in this regression is much lower than that of the previous regression.

Results from Regressing the Free Cash Flow/Sales on Cash Flow from Operations/Sales for Family Clothing Stores

	Coefficient	Standard Error	*t*-Statistic
Intercept	−0.0121	0.0221	−0.5497
Cash flow from operations/Sales	0.4749	0.2920	1.6262

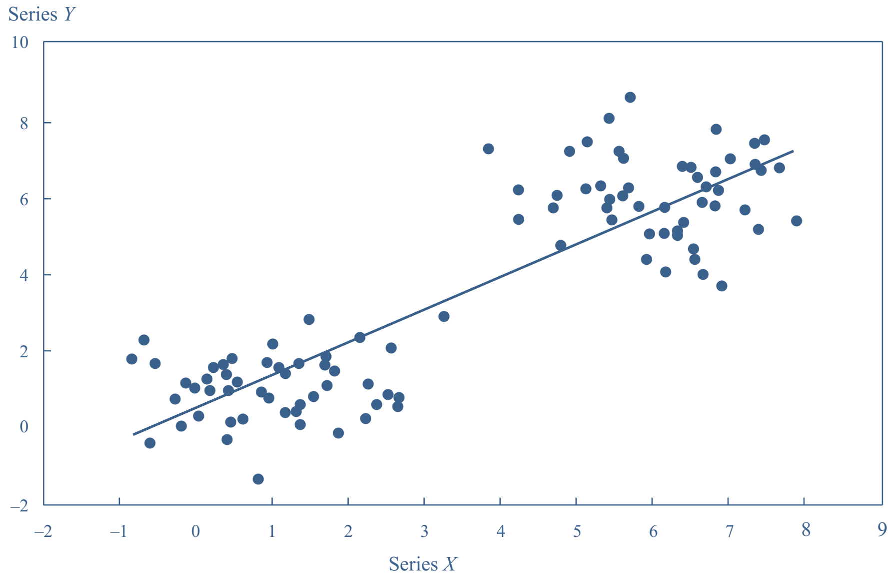
ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	1	0.0030	0.0030	2.6447	0.1383
Residual	9	0.0102	0.0011		
Total	10	0.0131			
Residual standard error		0.0336			
Multiple R^2		0.2271			
Observations		11			

Source: Compustat.

Which regression makes more sense? Usually, the scaled regression makes more sense. We want to know what happens to free cash flow (as a fraction of sales) if a change occurs in cash flow from operations (as a fraction of sales). Without scaling, the results of the regression can be based solely on scale differences across companies rather than on the companies' underlying economics.

A third common form of misspecification in regression models is pooling data from different samples that should not be pooled. This type of misspecification can best be illustrated graphically. Exhibit 22 shows two clusters of data on variables *X* and *Y* with a fitted regression line. The data could represent the relationship between two financial variables at two different time periods, for example.

Exhibit 22 Plot of Two Series with Changing Means



In each cluster of data on *X* and *Y*, the correlation between the two variables is virtually 0. Because the means of both *X* and *Y* are different for the two clusters of data in the combined sample, *X* and *Y* are highly correlated. The correlation is spurious (misleading), however, because it reflects differences in the relationship between *X* and *Y* during two different time periods.

7.3 Time-Series Misspecification (Independent Variables Correlated with Errors)

In the previous section, we discussed the misspecification that arises when a relevant independent variable is omitted from a regression. In this section, we discuss problems that arise from the kinds of variables included in the regression, particularly in a time-series context. In models that use time-series data to explain the relations among different variables, it is particularly easy to violate Regression Assumption 3: that the error term has mean 0, conditioned on the independent variables. If this assumption is violated, the estimated regression coefficients will be biased and inconsistent.

Three common problems that create this type of time-series misspecification are:

- including lagged dependent variables as independent variables in regressions with serially correlated errors;
- including a function of a dependent variable as an independent variable, sometimes as a result of the incorrect dating of variables; and
- independent variables that are measured with error.

The next examples demonstrate these problems.

Suppose that an analyst includes the first lagged value of the dependent variable in a multiple regression that, as a result, has significant serial correlation in the errors. For example, the analyst might use the regression equation

$$Y_t = b_0 + b_1X_{1t} + b_2Y_{t-1} + \varepsilon_t. \qquad (11)$$

Because we assume that the error term is serially correlated, by definition the error term is correlated with the dependent variable. Consequently, the lagged dependent variable, Y_{t-1}, will be correlated with the error term, violating the assumption that the independent variables are uncorrelated with the error term. As a result, the estimates of the regression coefficients will be biased and inconsistent.

Exhibit 23 Fisher Effect with a Lagged Dependent Variable

In our discussion of serial correlation, we concluded from a test using the Durbin–Watson test that the error term in the Fisher effect equation (Equation 8) showed positive (first-order) serial correlation, using three-month T-bill returns as the dependent variable and inflation expectations of professional forecasters as the independent variable. Observations on the dependent and independent variables were quarterly. We now modify that regression by including the previous quarter's three-month T-bill returns as an additional independent variable.

Results from Regressing T-Bill Returns on Predicted Inflation and Lagged T-Bill Returns

	Coefficient	Standard Error	*t*-Statistic
Intercept	−0.0005	0.0014	−0.3571
Inflation prediction	0.1843	0.0455	4.0505
Lagged T-bill return	0.8796	0.0295	29.8169
Residual standard error	0.0095		
Multiple R^2	0.9285		
Observations	181		

Source: Federal Reserve Bank of Philadelphia, US Department of Commerce.

At first glance, these regression results look very interesting: The coefficient on the lagged T-bill return appears to be highly significant. But on closer consideration, we must ignore these regression results because the regression is fundamentally misspecified. As long as the error term is serially correlated, including lagged T-bill returns as an independent variable in the regression will cause all the coefficient estimates to be biased and inconsistent. Therefore, this regression is not usable for either testing a hypothesis or for forecasting.

A second common time-series misspecification in investment analysis is to forecast the past. What does that mean? If we forecast the future (say we predict at time t the value of variable Y in period $t + 1$), we must base our predictions on information we knew at time t. We could use a regression to make that forecast using the equation

$$Y_{t+1} = b_0 + b_1X_{1t} + \varepsilon_{t+1}. \tag{12}$$

In this equation, we predict the value of Y in time $t + 1$ using the value of X in time t. The error term, ε_{t+1}, is unknown at time t and thus should be uncorrelated with X_{1t}.

Unfortunately, analysts sometimes use regressions that try to forecast the value of a dependent variable at time $t + 1$ based on independent variable(s) that are functions of the value of the dependent variable at time $t + 1$. In such a model, the independent variable(s) would be correlated with the error term, so the equation would be misspecified. As an example, an analyst may try to explain the cross-sectional returns for a group of companies during a particular year using the market-to-book ratio and the market capitalization for those companies at the end of the year. ("Market-to-book ratio" is the ratio of price per share divided by book value per share.) If the analyst believes that such a regression predicts whether companies with high market-to-book ratios or high market capitalizations will have high returns, the analyst is mistaken. This is because for any given period, the higher the return during the period, the higher the market capitalization and the market-to-book period will be at the end of the period. In this case, if all the cross-sectional data come from period $t + 1$, a high value of the dependent variable (returns) actually causes a high value of the independent variables (market capitalization and the market-to-book ratio) rather than the other way around. In this type of misspecification, the regression model effectively includes the dependent variable on both the right-and left-hand sides of the regression equation.

The third common time-series misspecification arises when an independent variable is measured with error. Suppose a financial theory tells us that a particular variable X_t, such as expected inflation, should be included in the regression model. But we cannot directly observe X_t; instead, we can observe actual inflation, $Z_t = X_t + u_t$, where we assume u_t is an error term that is uncorrelated with X_t. Even in this best of circumstances, using Z_t in the regression instead of X_t will cause the regression coefficient estimates to be biased and inconsistent. To see why, assume we want to estimate the regression

$$Y_t = b_0 + b_1X_t + \varepsilon_t,$$

but we substitute Z_t for X_t. Then we would estimate

$$Y_t = b_0 + b_1Z_t + (-b_1u_t + \varepsilon_t).$$

But $Z_t = X_t + u_t$, Z_t is correlated with the error term $(-b_1u_t + \varepsilon_t)$. Therefore, our estimated model violates the assumption that the error term is uncorrelated with the independent variable. Consequently, the estimated regression coefficients will be biased and inconsistent.

Exhibit 24 The Fisher Effect with Measurement Error

Recall from Exhibit 11 on the Fisher effect that based on our initial analysis in which we did not correct for heteroskedasticity and serial correlation, we rejected the hypothesis that three-month T-bill returns moved one-for-one with expected inflation.

Exhibit 24 (Continued)

Results from Regressing T-Bill Returns on Predicted Inflation (repeated)

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0116	0.0033	3.5152
Inflation prediction	1.1744	0.0761	15.4323
Residual standard error	0.0223		
Multiple R^2	0.5708		
Observations	181		
Durbin–Watson statistic	0.2980		

Source: Federal Reserve Bank of Philadelphia, US Department of Commerce.

What if we used actual inflation instead of expected inflation as the independent variable? Note first that

$$\pi = \pi^e + \nu,$$

where

π = actual rate of inflation
π^e = expected rate of inflation
ν = the difference between actual and expected inflation

Because actual inflation measures expected inflation with error, the estimators of the regression coefficients using T-bill yields as the dependent variable and actual inflation as the independent variable will not be consistent. (Note that a consistent estimator is one for which the probability of estimates close to the value of the population parameter increases as sample size increases.)

The following regression output shows the results of using actual inflation as the independent variable. The estimates in this exhibit are quite different from those presented in the previous exhibit. Note that the slope coefficient on actual inflation is much lower than the slope coefficient on predicted inflation in the previous regression. This result is an illustration of a general proposition: In a single-independent-variable regression, if we select a version of that independent variable that is measured with error, the estimated slope coefficient on that variable will be biased toward 0. Note that this proposition does not generalize to regressions with more than one independent variable. Of course, we ignore serially-correlated errors in this example, but because the regression coefficients are inconsistent (due to measurement error), testing or correcting for serial correlation is not worthwhile.

Results from Regressing T-Bill Returns on Actual Inflation

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0227	0.0034	6.6765
Actual inflation	0.8946	0.0761	11.7556
Residual standard error	0.0267		

(continued)

(Continued)

	Coefficient	Standard Error	*t*-Statistic
Multiple R^2	0.4356		
Observations	181		

Source: Federal Reserve Bank of Philadelphia, US Department of Commerce.

7.4 Other Types of Time-Series Misspecification

By far the most frequent source of misspecification in linear regressions that use time series from two or more different variables is nonstationarity. Very roughly, **nonstationarity** means that a variable's properties, such as mean and variance, are not constant through time. We will postpone our discussion about stationarity to the later coverage on time-series analysis, but we can list some examples in which we need to use stationarity tests before we use regression statistical inference.

- Relations among time series with trends (for example, the relation between consumption and GDP).
- Relations among time series that may be **random walks** (time series for which the best predictor of next period's value is this period's value). Exchange rates are often random walks.

The time-series examples in our discussion were carefully chosen such that nonstationarity was unlikely to be an issue for any of them. But nonstationarity can be a very severe problem for analyzing the relations among two or more time series in practice. Analysts must understand these issues before they apply linear regression to analyzing the relations among time series. Otherwise, they may rely on invalid statistical inference.

8 MULTIPLE LINEAR REGRESSION WITH QUALITATIVE DEPENDENT VARIABLES

n interpret an estimated logistic regression;

i evaluate how well a regression model explains the dependent variable by analyzing the output of the regression equation and an ANOVA table;

o evaluate and interpret a multiple regression model and its results.

In this section, we explain what qualitative dependent variables are, how the regression models that feature such variables work, and how the regression results can be interpreted.

8.1 Models with Qualitative Dependent Variables

Qualitative dependent variables (also called **categorical dependent variables**) are outcome variables that describe data that fit into categories. For example, to predict whether or not a company will go bankrupt, we need to use a qualitative dependent variable (bankrupt or not) as the dependent variable and use data on the company's financial performance (e.g., return on equity, debt-to-equity ratio, or debt rating) as independent variables. The qualitative dependent variable in this example here is a binary variable. This is one of many potential scenarios in which financial analysts need to be able to explain the outcomes of a qualitative dependent variable that describes data that belong to two categories. It is also possible to carry out analysis where the dependent variable can fall into more than two categories. For example, Moody's Bank Financial Strength Rating is a qualitative variable that indicates the Moody's rating or category—A, B, C, D, or E—of a bank.

In contrast to a linear regression, the dependent variable here is not continuous in nature but is discrete and in the simple scenario has two categories. Unfortunately, for estimating such a model, linear regression is not the best statistical method to use. If we use the qualitative dependent variable Y = {bankrupt (= 1) or not bankrupt (= 0)} as the dependent variable in a regression with financial variables as the independent variables, then we are estimating a linear probability model:

$$l_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + \varepsilon_i. \quad (13)$$

Unfortunately, the predicted value of the dependent variable could be much greater than 1 or much lower than 0 depending on the estimated coefficients b_i and the value of observed X_is. Of course, these results would be invalid. The probability of bankruptcy (or of anything, for that matter) cannot be greater than 1.0 or less than 0. Another issue with the use of linear regression is that it assumes the relationship between the probability of bankruptcy and each financial variable to be linear throughout the range of the financial variable. However, we may not expect that. For example, we may expect that the probability of bankruptcy and debt-to-equity ratio are not linearly related for very low or high levels of debt-to-equity ratio.

To address these issues associated with linear regression, we should apply a nonlinear transformation to the probability of bankruptcy and relate the transformed probabilities linearly to the independent variables. There are many possible transformations, which are closely related except when the probability is quite low or high.

The most commonly used transformation is the logistic transformation. Denote by "p" the probability that a company goes bankrupt, or more generally, a condition is fulfilled or an event happens. The logistic transformation is $\ln\left(\frac{p}{1-p}\right)$. The ratio $\frac{p}{1-p}$ is a ratio of probabilities—the probability that the event of interest happens (p) divided by the probability that it does not happen ($1 - p$). This ratio is called the odds of an event happening. For example, if the probability of a company going bankrupt is 0.75, then $\frac{p}{1-p}$ is $0.75/(1 - 0.75) = 0.75/0.25 = 3$. So, the odds of bankruptcy is 3, which indicates that the probability of bankruptcy is three times as large as the probability of the company not going bankrupt. The natural logarithm of the odds of an event happening is called log odds or logit.

The logistic transformation tends to linearize the relationship between the dependent and independent variables. Instead of a linear regression to estimate the probability of bankruptcy, we should use **logistic regression (logit model)** or **discriminant analysis** for this kind of estimation.

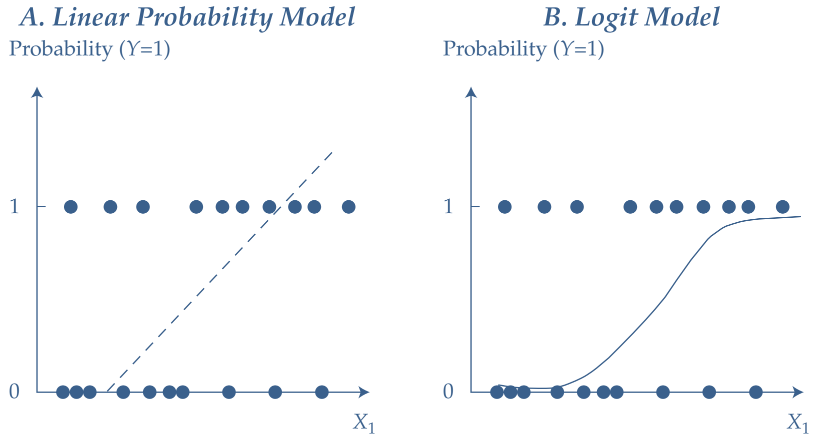
Logistic regression models are used to estimate the probability of a discrete outcome given the values of the independent variables used to explain that outcome. Logistic regression is widely used in machine learning where the objective is classification. Logistic regression involves using the logistic transformation of the event probability as the dependent variable:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \varepsilon. \quad (14)$$

The event probability can be derived from Equation 7 as:

$$p = \frac{1}{1 + exp\left[-\left(b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3\right)\right]} \quad (15)$$

Exhibit 25 Linear Probability Models versus Logit Models



This nonlinear function takes on the sigmoidal shape shown in Exhibit 25. The shape is approximately linear except when probability estimates are close to zero or one. One can see in the exhibit that the nonlinear transformation constrains probability estimates to be between 0 and 1. We can also see that mathematically. As (b_0 + b_1X_1 + b_2X_2 + b_3X_3) approaches positive infinity, p approaches 1; and as (b_0 + b_1X_1 + b_2X_2 + b_3X_3) approaches negative infinity, p approaches 0. Logistic regression assumes a logistic distribution for the error term; this distribution is similar in shape to the normal distribution but has heavier tails.

For a binary p, $\ln\left(\frac{p}{1-p}\right)$ is undefined for both $p = 0$ and $p = 1$. In such a case, logistic regression coefficients are estimated by maximum likelihood method rather than by least squares. The maximum likelihood method estimates logistic regression coefficients that make it most likely that the choices in the sample would have occurred by maximizing the likelihood function for the data. We need to assume the probability distribution of p to construct the likelihood function. Because p is binary, the Bernoulli distribution is chosen as its probability distribution. Maximum likelihood method is an iterative method in which the goal is to maximize log likelihood. Each iteration results in a higher log likelihood, and the iterating process stops when the difference in the log likelihood of two successive iterations is quite small.

Because the logit model has the logistic transformation of event probability as the dependent variable, the interpretation of regression coefficients is not as simple or intuitive as an ordinary linear regression. In a linear regression, the slope coefficient of an independent variable is the change in the dependent variable per unit change in the independent variable, holding all other independent variables constant. In the logit model, the slope coefficient is the change in the "log odds" that the event happens per unit change in the independent variable, holding all other independent variables constant. The exponent of the slope coefficient is the "odds ratio," which is the ratio of odds that the event happens with a unit increase in the independent variable to the odds that the event happens without the increase in the independent variable. The test of the hypothesis that a logit regression coefficient is significantly different from zero is similar to the test in an ordinary linear regression.

We can evaluate the overall performance of a logit regression by examining the likelihood ratio chi-square test statistic. Most statistical analysis packages report this statistic along with the *p*-value or the probability of obtaining this statistic if there is no collective effect of the independent variables on the probability of the event. The *p*-value helps us evaluate the overall statistical significance of the model.

There is no equivalent measure in logistic regression of the R^2 statistic of an ordinary linear regression since logistic regression cannot be fitted using a least square approach. However, researchers have proposed different measures for logistic regression to capture the explained variation. These measures are called pseudo-R^2 and must be interpreted with caution. The pseudo-R^2 in logistic regression may be used to compare different specifications of the same model but is not appropriate for comparing models based on different datasets.

Qualitative dependent variable models can be useful not only for portfolio management but also for business management. For example, we might want to predict whether a client is likely to continue investing in a company or to withdraw assets from the company. We might also want to explain how particular demographic characteristics might affect the probability that a potential investor will sign on as a new client or evaluate the effectiveness of a particular direct-mail advertising campaign based on the demographic characteristics of the target audience. These issues can be analyzed with a logit model. Logistic regression also plays an important role in Big Data analysis—for example, in binary classification problems in machine learning and in neural networks, a topic explained at a later stage.

Exhibit 26 Explaining Financing Choice

Grundy and Verwijmeren (2019) investigate what investment characteristics determine the financing choice of a company. We can employ a logit model to address the question. The sample consists of 680 investments financed with debt or equity by US firms between 1995 and 2017 for which the information on investments' characteristics could be obtained. Because the dependent variable in the regression analysis is a binary variable, a logit model is used.

(continued)

Exhibit 26 (Continued)

The variables in the logit model are as follows:

Dependent variable:

EQUITY = a binary variable that takes on a value of 1 if equity is used to finance an investment and 0 if debt is used

Independent variables:

TANGIBLE & NON-UNIQUE = a binary variable that takes on a value of 1 if the investment is in a tangible asset and the asset is redeployable

R&D = a binary variable that takes on a value of 1 if the investment has R&D-like characteristics

LN(INVESTMENT LIFE) = natural log of the expected life span of the investment in years

TIME UNTIL PAYOFFS = an ordered categorical variable to capture the time until positive payoffs from the investment begin

INVESTMENT LIFE UNCERTAINTY = a binary variable that takes on a value of 1 if the investment has a relatively uncertain lifespan

VOLATILITY = a binary variable that takes on a value of 1 if the investment is relatively more risky

NEED FOR MONITORING = an ordered categorical variable based on an assessment of the need for monitoring of the investment. Takes one of three values: low, medium, or high.

The authors of the study are examining whether the choice of financing type, either equity or debt, is related to characteristics of the investment being financed. One hypothesis is that equity is more likely to be used when the investment being undertaken has more uncertainty associated with its payoffs. Correspondingly, debt financing is more likely to be used for investments in tangible or investments with a greater need for monitoring. Neither of these hypotheses provides a clear prediction regarding the relationship between investment life and the financing method used. The following table shows an excerpt from the results of the logit estimation: Model 1 of Table II of Grundy and Verwijmeren (2019).

Explaining Financing Choice Using a Logit Model

DEPENDENT VARIABLE EQUITY INDEPENDENT VARIABLES	Coefficient	Standard Error	*z*-Statistic
TANGIBLE & NON-UNIQUE	−1.18	0.29	−4.07
R&D	0.90	0.46	1.96

Exhibit 26 (Continued)

DEPENDENT VARIABLE EQUITY INDEPENDENT VARIABLES	Coefficient	Standard Error	*z*-Statistic
LN(INVESTMENT LIFE)	−0.39	0.26	−1.50
TIME UNTIL PAYOFFS	1.49	0.31	4.81
INVESTMENT LIFE UNCERTAINTY	0.13	0.39	0.33
VOLATILITY	1.29	0.35	3.69
NEED FOR MONITORING	−0.98	0.31	−3.16
Pseudo-R^2		0.28	

Notes: The research paper does not include the *z*-statistics. However, we can compute them as the ratio of Coefficient and Standard Error)

As the results in the table indicate, the absolute value of *z*-statistics for TANGIBLE & NON-UNIQUE, R&D, TIME UNTIL PAYOFFS, VOLATILITY, and NEED FOR MONITORING is equal to or higher than the critical value at the 0.05 level for the *z*-statistic (1.96). For each of these variables at the 0.05 level of significance, we can reject the null hypothesis that the coefficient equals 0 in favor of the alternative hypothesis that the coefficient is not equal to 0. The statistically significant coefficients suggest that investments with R&D-like characteristics, more time until positive payoffs begin, and more volatility are likely to be equity-financed; investments in tangible and non-unique assets and investments with greater need for monitoring are more likely to be debt-financed. Thus, both of the original hypotheses are confirmed with respect to the factors that determine the choice of financing by a firm.

Neither of the two remaining independent variables is statistically significant at the 0.05 level in this logit analysis. The absolute values of *z*-statistics on these two variables are 1.50 or less, so neither one reaches the critical value of 1.96 needed to reject the null hypothesis (that the associated coefficient is significantly different from 0). This result shows that once we take into account the factors included in the analysis, the other factors—life of the investment and uncertainty of investment life—have no power to explain the financing choice.

The estimated regression coefficient for an independent variable is the change in "log odds" that the investment is financed by equity per unit change in that independent variable, holding all other independent variables constant. Consider an investment that does not have R&D-like characteristics. So, R&D takes a value of 0. Suppose that for this investment, after we input the values of R&D and all the other independent variables in the estimated logit model, we get −0.6577. So, the log odds for this investment, that is not R&D-like, being financed by equity equal −0.6577.

The estimated regression coefficient of 0.90 for R&D implies that if this investment had R&D-like characteristics while other characteristics were held constant, the log odds for this investment being financed by equity would increase to −0.6577 + 0.90 = 0.2423. Therefore, the odds of this investment with R&D-like characteristics being financed by equity would be exp(0.2423) = 1.2742. In other words, the probability of equity financing is about 1.27 times as large as the probability of debt financing: $\frac{p}{1-p}$ = 1.2742, where p is the probability of the investment being financed by equity. Solving this equation for p results in 0.5603 or 56.03%. We could have also computed this using Equation 8:

$$p = \frac{1}{1 + \exp[-(0.2423)]} = 0.5603.$$

The exponent of estimated regression coefficient of 0.90 for R&D = exp(0.90) = 2.4596 (or 1.2742 / 0,5180). This "odds ratio" is the ratio of odds that the investment is equity financed if it has R&D-like characteristics to the odds that the investment is equity financed if it does not have R&D-like characteristics.

EXAMPLE 3

Explaining CEO Awards

Use the following information to answer Questions 1–8.

CEOs receive substantial attention in the media. Various publications, such as *Bloomberg Businessweek* and *Forbes,* confer prestigious business awards to a small number of CEOs. Studies, such as those by Malmendier and Tate (2009), find that after receiving an award, the performance of the CEO, as measured by the firm's stock return and return on assets, declines. They also find that award-winning CEOs spend more time on activities outside their companies and underperform relative to non-winning CEOs. Kim Dalton is a financial analyst interested in determining which CEOs are likely to win an award. Her sample consists of observations of company characteristics for each month in which an award is given to CEOs of companies in the S&P 1500 index for a 10-year period in the 2000s. Dalton employs a logistic regression for her analysis.

The dependent variable in the logistic regression is the logistic transformation of AWARD, a binary variable that takes on a value of 1 for a CEO winning an award in the award month and 0 for non-winning CEOs. The independent variables include BOOK-TO-MARKET (the ratio of the company's book equity and market capitalization); LNSIZE (the natural log of the market value of the company's equity); RETURN-1TO3, RETURN-4TO6, RETURN-7TO12 (total return during months 1–3, 4–6, and 7–12 prior to the award month, respectively); LNTENURE (the natural log of the CEO's tenure with a firm in number of years); and FEMALE (a dummy variable that takes on a value of 1 if the CEO is a female).

In this attempt to explain CEO award winning, Dalton is examining whether CEOs of companies with a low book-to-market ratio, larger companies (as captured by their market values), and companies with higher returns in recent months are more likely to win an award. Dalton is also examining if female CEOs and older CEOs are more likely to receive an award. The following table shows the results of the logit estimation.

Explaining CEO Award Winning Using a Logit Model

	Coefficient	Standard Error	*z*-Statistic	*p*-Value
Intercept	−2.5169	2.2675	−1.11	0.267
BOOK-TO-MARKET	−0.0618	0.0243	−2.54	0.011
LNSIZE	1.3515	0.5201	2.60	0.009
RETURN-1TO3	0.3684	0.5731	0.64	0.520
RETURN-4TO6	0.1734	0.5939	0.29	0.770
RETURN-7TO12	0.9345	0.2250	4.15	0.000
LNTENURE	1.2367	0.5345	2.31	0.021
FEMALE	0.8100	0.3632	2.23	0.026
Likelihood ratio chi-square		323.16		
Prob > chi-square		0.000		
Pseudo R^2		0.226		

1 Which of the following is the reason for Kim Dalton choosing a logit regression for her analysis?
   - **A** AWARD is a binary variable.
   - **B** FEMALE is a binary variable.
   - **C** Two binary variables are in the model.

2 CEOs of which of the following companies are most likely to win an award?
   - **A** Large companies with a high book-to-market ratio that have achieved high stock returns in months 7–12 before the award month.
   - **B** Large companies with a low book-to-market ratio that have achieved high stock returns in months 4–6 before the award month.
   - **C** Large companies with a low book-to-market ratio that have achieved high stock returns in months 7–12 before the award month.

3 Which of the following types of CEOs are most likely to win an award?
   - **A** Females with a long tenure with the company.
   - **B** Females with a short tenure with the company.
   - **C** Males with a long tenure with the company.

4 Consider a company for which the log odds of its CEO winning an award based on the estimated regression model work out to −2.3085. The CEO of the company is a male. What would be the log odds of the CEO winning an award if the CEO was a female, while all the other variables are held constant?
   - **A** −4.0154
   - **B** −3.1185
   - **C** −1.4985

5 What are the odds of the male CEO mentioned in the previous question winning an award?
   - **A** 0.0807
   - **B** 0.0994

C 0.2235

6 Assuming the odds of the male CEO winning an award are 0.0994, what is the probability of the male CEO winning an award?

A 9.04%

B 9.94%

C 18.27%

7 What is the ratio of odds that a female CEO wins an award to the odds that a male CEO wins an award?

A 0.0807

B 0.4449

C 2.2479

8 In estimating the logit regression model, Dalton has used returns expressed in fractions. For example, a return of 10% is entered as 0.10. Therefore, one unit is 1 or 100%. Consider the company with a male CEO discussed earlier. For this company, the total return during months 7 to 12 prior to the award month was 11%. We know that the log odds of its CEO winning an award based on the estimated regression model work out to −2.3085. What would be the log odds of its CEO winning an award if the total return during months 7 to 12 prior to the award month was 12%?

A −1.3740

B −2.2991

C −2.3178

Solution to 1

A is correct. AWARD being a binary dependent variable requires that we use a nonlinear estimation model, such as logit.

B is incorrect because FEMALE is an independent variable. Having a binary independent variable does not make ordinary linear regression inappropriate for estimating the model.

C is incorrect. The total number of binary variables in the model is not relevant to the choice of the estimation procedure. What matters in the context of this question is whether the dependent variable is binary or not.

Solution to 2

C is correct. LNSIZE and RETURN-7TO12 have a significantly positive relationship, while BOOK-TO-MARKET has a significantly negative relationship with log odds of a CEO winning an award.

A is incorrect because the book-to-market ratio has a significantly negative relationship with log odds of a CEO winning an award.

B is incorrect because RETURN-4TO6 has a z-statistic of only 0.29 and is not statistically significant at the 10% level of significance.

Solution to 3

A is correct. The binary variable FEMALE is positive and statistically significant, indicating that female CEOs are more likely to win an award than male CEOs. LNTENURE is also positive and statistically significant, indicating that CEOs with a longer tenure with the company are more likely to win an award. Therefore, female CEOs with longer tenure are most likely to win an award.

Solution to 4

C is correct. The binary variable FEMALE has a slope coefficient of 0.8100. Therefore, the log odds for a female CEO instead of a male CEO, while other variables are held constant, will be −2.3085 + 0.8100 = −1.4985.

Solution to 5

B is correct. The log odds of the CEO winning an award are −2.3085. This means that the odds of the CEO winning an award are exp(−2.3085) = 0.0994.

Solution to 6

A is correct. Given that the odds (of the CEO winning an award) are 0.0994, we know that $p/(1 - p) = 0.0994$, where p is the probability of the CEO winning an award. Solving this equation for p results in 0.0904 or 9.04%. We could have also computed this using Equation 15: $p = 1/(1 + \exp[-(2.3085)]) = 0.0904$.

Solution to 7

C is correct. The binary variable FEMALE has a slope coefficient of 0.8100. Therefore, the odds ratio for a female CEO winning an award to a male CEO winning an award is exp(0.8100) = 2.2479. In other words, the odds of a female CEO winning an award are about 2.25 times the odds of a male CEO winning an award.

Solution to 8

B is correct. The variable RETURN-7TO12 has a slope coefficient of 0.9345. Therefore, for every 1 unit or 100% increase in this variable, log odds increase by 0.9345. In the previous question, the variable increases by 0.01 unit or 1%. Accordingly, log odds would increase by 0.01 x 0.9345 = 0.009345. So, the log odds would be the −2.3085 + 0.009345 = −2.2991.

SUMMARY

We have presented the multiple linear regression model and discussed violations of regression assumptions, model specification and misspecification, and models with qualitative variables.

- The general form of a multiple linear regression model is $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_kX_{ki} + \varepsilon_i$.
- We conduct hypothesis tests concerning the population values of regression coefficients using t-tests of the form

$$t = \frac{\hat{b}_j - b_j}{s_{\hat{b}_j}}$$

- The lower the p-value reported for a test, the more significant the result.
- The assumptions of classical normal multiple linear regression model are as follows:
  1. A linear relation exists between the dependent variable and the independent variables.
  2. The independent variables are not random. Also, no exact linear relation exists between two or more of the independent variables.

3 The expected value of the error term, conditioned on the independent variables, is 0.

4 The variance of the error term is the same for all observations.

5 The error term is uncorrelated across observations.

6 The error term is normally distributed.

- To make a prediction using a multiple linear regression model, we take the following three steps:

  1 Obtain estimates of the regression coefficients.

  2 Determine the assumed values of the independent variables.

  3 Compute the predicted value of the dependent variable.

- When predicting the dependent variable using a linear regression model, we encounter two types of uncertainty: uncertainty in the regression model itself, as reflected in the standard error of estimate, and uncertainty about the estimates of the regression coefficients.

- The F-test is reported in an ANOVA table. The F-statistic is used to test whether at least one of the slope coefficients on the independent variables is significantly different from 0.

$$F = \frac{\text{RSS}/k}{\text{SSE}/\left[n - \left(k + 1\right)\right]} = \frac{\text{Mean regression sum of squares}}{\text{Mean squared error}}$$

  Under the null hypothesis that all the slope coefficients are jointly equal to 0, this test statistic has a distribution of $F_{k,n-(k+1)}$, where the regression has n observations and k independent variables. The F-test measures the overall significance of the regression.

- R^2 is nondecreasing in the number of independent variables, so it is less reliable as a measure of goodness of fit in a regression with more than one independent variable than in a one-independent-variable regression.

  Analysts often choose to use adjusted R^2 because it does not necessarily increase when one adds an independent variable.

- Dummy variables in a regression model can help analysts determine whether a particular qualitative independent variable explains the model's dependent variable. A dummy variable takes on the value of 0 or 1. If we need to distinguish among n categories, the regression should include n – 1 dummy variables.

- When using intercept dummies, the intercept of the regression measures the average value of the dependent variable of the omitted category, and the coefficient on each dummy variable measures the average incremental effect of that dummy variable on the dependent variable.

- When using slope dummies, the coefficient on each dummy measures the average incremental effect on the slope coefficient of the independent variable.

- If a regression shows significant conditional heteroskedasticity, the standard errors and test statistics computed by regression programs will be incorrect unless they are adjusted for heteroskedasticity.

- One simple test for conditional heteroskedasticity is the Breusch–Pagan test. Breusch and Pagan showed that, under the null hypothesis of no conditional heteroskedasticity, nR^2 (from the regression of the squared residuals on the independent variables from the original regression) will be a χ^2 random variable with the number of degrees of freedom equal to the number of independent variables in the regression.

- The principal effect of serial correlation in a linear regression is that the standard errors and test statistics computed by regression programs will be incorrect unless adjusted for serial correlation. Positive serial correlation typically inflates the *t*-statistics of estimated regression coefficients as well as the *F*-statistic for the overall significance of the regression.
- The most commonly used test for serial correlation is based on the Durbin–Watson statistic. If the Durbin–Watson statistic differs sufficiently from 2, then the regression errors have significant serial correlation.
- Multicollinearity occurs when two or more independent variables (or combinations of independent variables) are highly (but not perfectly) correlated with each other. With multicollinearity, the regression coefficients may not be individually statistically significant even when the overall regression is significant, as judged by the *F*-statistic.
- Model specification refers to the set of variables included in the regression and the regression equation's functional form. The following principles can guide model specification:
  - The model should be grounded in cogent economic reasoning.
  - The functional form chosen for the variables in the regression should be appropriate given the nature of the variables.
  - The model should be parsimonious.
  - The model should be examined for violations of regression assumptions before being accepted.
  - The model should be tested and found useful out of sample before being accepted.
- If a regression is misspecified, then statistical inference using OLS is invalid and the estimated regression coefficients may be inconsistent.
- Assuming that a model has the correct functional form when in fact it does not is one example of misspecification. This assumption may be violated in several ways:
  - One or more important variables could be omitted from the regression.
  - One or more of the regression variables may need to be transformed before estimating the regression.
  - The regression model pools data from different samples that should not be pooled.
- Another type of misspecification occurs when independent variables are correlated with the error term. This is a violation of Regression Assumption 3, that the error term has a mean of 0, and causes the estimated regression coefficients to be biased and inconsistent. Three common problems that create this type of time-series misspecification are:
  - including lagged dependent variables as independent variables in regressions with serially correlated errors;
  - including a function of the dependent variable as an independent variable, sometimes as a result of the incorrect dating of variables; and
  - independent variables that are measured with error.
- Logit models estimate the probability of a discrete outcome (the value of a qualitative dependent variable, such as whether a company enters bankruptcy) given the values of the independent variables used to explain that outcome. The logit model, which is based on the logistic distribution, estimates the probability that $Y = 1$ (a condition is fulfilled) given the values of the independent variables.

REFERENCES

Altman, Edward I. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance*, vol. 23: 589–609.

Altman, Edward I., R. Halderman, and P. Narayanan. 1977. "Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations." *Journal of Banking & Finance*, vol. 1: 29–54.

Bhabra, Harjeet S., and Jiayin Huang. 2013. "An Empirical Investigation of Mergers and Acquisitions by Chinese Listed Companies, 1997–2007." *Journal of Multinational Financial Management*, vol. 23: 186–207.

Breusch, T., and A. Pagan. 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica*, vol. 47: 1287–1294.

Buetow, Gerald W., Jr, Robert R. Johnson, and David E. Runkle. 2000. "The Inconsistency of Return-Based Style Analysis." *Journal of Portfolio Management*, vol. 26, no. 3: 61–77.

Durbin, J., and G.S. Watson. 1951. "Testing for Serial Correlation in Least Squares Regression, II." *Biometrika*, vol. 38: 159–178.

Greene, William H. 2018. *Economic Analysis*, 8th edition. New York: Pearson Education.

Grundy, BruceD., and Patrick Verwijmeren. . 2019. "The External Financing of Investment" (3 April). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2986127.

Gujarati, Damodar N., Dawn C. Porter, and Sangeetha Gunasekar. 2011. *Basic Econometrics*, 5th edition. New York: McGraw-Hill Irwin.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, vol. 50, no. 4: 1029–1054.

Keane, Michael P., and David E. Runkle. 1998. "Are Financial Analysts' Forecasts of Corporate Profits Rational?" *Journal of Political Economy*, vol. 106, no. 4: 768–805.

Kyaw, NyoNyo A., John Manley, and Anand Shetty. 2011. "Factors in Multinational Valuations: Transparency, Political Risk, and Diversification." *Journal of Multinational Financial Management*, vol. 21: 55–67.

MacKinlay, A. Craig, and Matthew P. Richardson. 1991. "Using Generalized Methods of Moments to Test Mean–Variance Efficiency." *Journal of Finance*, vol. 46, no. 2: 511–527.

Malmendier, Ulrike, and Geoffrey Tate. 2009. "Superstar CEOs." *Quarterly Journal of Economics* 124 (4): 1593–638.

Mankiw, N. Gregory. 2015. *Macroeconomics*, 9th edition. New York: Worth Publishers.

Mitchell, T. 1997. *Machine Learning*. McGraw-Hill.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, vol. 55, no. 3: 703–708.

Petersen, Mitchell A. 2009. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Financial Studies*, vol. 22, no. 1: 435–480.

Samuel, A.July 1959. "Some Studies in Machine Learning Using the Game of Checkers." *IBM Journal of Research and Development* 3 (3): 210–29.

Sharpe, William F. 1988. "Determining a Fund's Effective Asset Mix." *Investment Management Review*, November/December:59–69.

Siegel, Jeremy J. 2014. *Stocks for the Long Run*, 5th edition. New York: McGraw-Hill.

Stoll, Hans R. 1978. "The Pricing of Security Dealer Services: An Empirical Study of Nasdaq Stocks." *Journal of Finance*, vol. 33, no. 4: 1153–1172.

Theobald, O. 2017. *Machine Learning for Absolute Beginners*. Independently published.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B. Methodological* 58 (1): 267–88.

Wang, W., Y. Huang, Y. Wang, and L. Wang. 2014. "Generalized Autoencoder: A Neural Network Framework for Dimensionality Reduction." *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*: 490–497.

PRACTICE PROBLEMS

1 With many US companies operating globally, the effect of the US dollar's strength on a US company's returns has become an important investment issue. You would like to determine whether changes in the US dollar's value and overall US equity market returns affect an asset's returns. You decide to use the S&P 500 Index to represent the US equity market.

A Write a multiple regression equation to test whether changes in the value of the dollar and equity market returns affect an asset's returns. Use the notations below.

R_{it} = return on the asset in period t
R_{Mt} = return on the S&P 500 in period t
ΔX_t = change in period t in the log of a trade-weighted index of the foreign exchange value of US dollar against the currencies of a broad group of major US trading partners.

B You estimate the regression for Archer Daniels Midland Company (NYSE: ADM). You regress its monthly returns for the period January 1990 to December 2002 on S&P 500 Index returns and changes in the log of the trade-weighted exchange value of the US dollar. The table below shows the coefficient estimates and their standard errors.

Coefficient Estimates from Regressing ADM's Returns: Monthly Data, January 1990–December 2002

	Coefficient	Standard Error
Intercept	0.0045	0.0062
R_{Mt}	0.5373	0.1332
ΔX_t	−0.5768	0.5121
n = 156		

Source: FactSet, Federal Reserve Bank of Philadelphia.

Determine whether S&P 500 returns affect ADM's returns. Then determine whether changes in the value of the US dollar affect ADM's returns. Use a 0.05 significance level to make your decisions.

C Based on the estimated coefficient on R_{Mt}, is it correct to say that "for a 1 percentage point increase in the return on the S&P 500 in period t, we expect a 0.5373 percentage point increase in the return on ADM"?

2 One of the most important questions in financial economics is what factors determine the cross-sectional variation in an asset's returns. Some have argued that book-to-market ratio and size (market value of equity) play an important role.

A Write a multiple regression equation to test whether book-to-market ratio and size explain the cross-section of asset returns. Use the notations below.

$(B/M)_i$ = book-to-market ratio for asset i

© 2017 CFA Institute. All rights reserved.

R_i = return on asset i in a particular month

Size_i = natural log of the market value of equity for asset i

B The table below shows the results of the linear regression for a cross-section of 66 companies. The size and book-to-market data for each company are for December 2001. The return data for each company are for January 2002.

Results from Regressing Returns on the Book-to-Market Ratio and Size

	Coefficient	Standard Error
Intercept	0.0825	0.1644
$(\text{B/M})_i$	−0.0541	0.0588
Size_i	−0.0164	0.0350
n = 66		

Source: FactSet.

Determine whether the book-to-market ratio and size are each useful for explaining the cross-section of asset returns. Use a 0.05 significance level to make your decision.

3 There is substantial cross-sectional variation in the number of financial analysts who follow a company. Suppose you hypothesize that a company's size (market cap) and financial risk (debt-to-equity ratios) influence the number of financial analysts who follow a company. You formulate the following regression model:

$$(\text{Analyst following})_i = b_0 + b_1\text{Size}_i + b_2(\text{D/E})_i + \varepsilon_i$$

where

$(\text{Analyst following})_i$ = the natural log of (1 + n), where n_i is the number of analysts following company i

Size_i = the natural log of the market capitalization of company i in millions of dollars

$(\text{D/E})_i$ = the debt-to-equity ratio for company i

In the definition of Analyst following, 1 is added to the number of analysts following a company because some companies are not followed by any analysts, and the natural log of 0 is indeterminate. The following table gives the coefficient estimates of the above regression model for a randomly selected sample of 500 companies. The data are for the year 2002.

Coefficient Estimates from Regressing Analyst Following on Size and Debt-to-Equity Ratio

	Coefficient	Standard Error	*t*-Statistic
Intercept	−0.2845	0.1080	−2.6343
Size_i	0.3199	0.0152	21.0461

(Continued)

	Coefficient	Standard Error	*t*-Statistic
$(D/E)_i$	−0.1895	0.0620	−3.0565
n = 500			

Source: First Call/Thomson Financial, Compustat.

A Consider two companies, both of which have a debt-to-equity ratio of 0.75. The first company has a market capitalization of $100 million, and the second company has a market capitalization of $1 billion. Based on the above estimates, how many more analysts will follow the second company than the first company?

B Suppose the p-value reported for the estimated coefficient on $(D/E)_i$ is 0.00236. State the interpretation of 0.00236.

4 In early 2001, US equity marketplaces started trading all listed shares in minimal increments (ticks) of $0.01 (decimalization). After decimalization, bid–ask spreads of stocks traded on the NASDAQ tended to decline. In response, spreads of NASDAQ stocks cross-listed on the Toronto Stock Exchange (TSE) tended to decline as well. Researchers Oppenheimer and Sabherwal (2003) hypothesized that the percentage decline in TSE spreads of cross-listed stocks was related to company size, the predecimalization ratio of spreads on NASDAQ to those on the TSE, and the percentage decline in NASDAQ spreads. The following table gives the regression coefficient estimates from estimating that relationship for a sample of 74 companies. Company size is measured by the natural logarithm of the book value of company's assets in thousands of Canadian dollars.

Coefficient Estimates from Regressing Percentage Decline in TSE Spreads on Company Size, Predecimalization Ratio of NASDAQ to TSE Spreads, and Percentage Decline in NASDAQ Spreads

	Coefficient	*t*-Statistic
Intercept	−0.45	−1.86
Size_i	0.05	2.56
$(\text{Ratio of spreads})_i$	−0.06	−3.77
$(\text{Decline in NASDAQ spreads})_i$	0.29	2.42
n = 74		

Source: Oppenheimer and Sabherwal (2003).

The average company in the sample has a book value of assets of C$900 million and a predecimalization ratio of spreads equal to 1.3. Based on the above model, what is the predicted decline in spread on the TSE for a company with these average characteristics, given a 1 percentage point decline in NASDAQ spreads?

5 The "neglected-company effect" claims that companies that are followed by fewer analysts will earn higher returns on average than companies that are followed by many analysts. To test the neglected-company effect, you have collected data on 66 companies and the number of analysts providing earnings estimates for each company. You decide to also include size as an independent variable, measuring size as the log of the market value of the company's equity, to try to distinguish any small-company effect from a neglected-company effect. The small-company effect asserts that small-company stocks may earn average higher risk-adjusted returns than large-company stocks.

The table below shows the results from estimating the model $R_i = b_0 + b_1\text{Size}_i + b_2(\text{Number of analysts})_i + \varepsilon_i$ for a cross-section of 66 companies. The size and number of analysts for each company are for December 2001. The return data are for January 2002.

Results from Regressing Returns on Size and Number of Analysts

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0388	0.1556	0.2495
Size_i	–0.0153	0.0348	–0.4388
$(\text{Number of analysts})_i$	0.0014	0.0015	0.8995

ANOVA	df	SS	MSS
Regression	2	0.0094	0.0047
Residual	63	0.6739	0.0107
Total	65	0.6833	
Residual standard error	0.1034		
R-squared	0.0138		
Observations	66		

Source: First Call/Thomson Financial, FactSet.

A What test would you conduct to see whether the two independent variables are *jointly* statistically related to returns (H_0: $b_1 = b_2 = 0$)?

B What information do you need to conduct the appropriate test?

C Determine whether the two variables jointly are statistically related to returns at the 0.05 significance level.

D Explain the meaning of adjusted R^2 and state whether adjusted R^2 for the regression would be smaller than, equal to, or larger than 0.0138.

6 Some developing nations are hesitant to open their equity markets to foreign investment because they fear that rapid inflows and outflows of foreign funds will increase volatility. In July 1993, India implemented substantial equity market reforms, one of which allowed foreign institutional investors into the Indian equity markets. You want to test whether the volatility of returns of stocks traded on the Bombay Stock Exchange (BSE) increased after July 1993, when foreign institutional investors were first allowed to invest in India. You have collected monthly return data for the BSE from February 1990 to December 1997. Your dependent variable is a measure of return volatility of stocks traded on the BSE; your independent variable is a dummy variable that is coded 1 if foreign investment was allowed during the month and 0 otherwise.

You believe that market return volatility actually *decreases* with the opening up of equity markets. The table below shows the results from your regression.

Results from Dummy Regression for Foreign Investment in India with a Volatility Measure as the Dependent Variable

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0133	0.0020	6.5351
Dummy	−0.0075	0.0027	−2.7604
n = 95			

Source: FactSet.

A State null and alternative hypotheses for the slope coefficient of the dummy variable that are consistent with testing your stated belief about the effect of opening the equity markets on stock return volatility.

B Determine whether you can reject the null hypothesis at the 0.05 significance level (in a one-sided test of significance).

C According to the estimated regression equation, what is the level of return volatility before and after the market-opening event?

7 Both researchers and the popular press have discussed the question as to which of the two leading US political parties, Republicans or Democrats, is better for the stock market.

A Write a regression equation to test whether overall market returns, as measured by the annual returns on the S&P 500 Index, tend to be higher when the Republicans or the Democrats control the White House. Use the notations below.

R_{Mt} = return on the S&P 500 in period t

Party_t = the political party controlling the White House (1 for a Republican president; 0 for a Democratic president) in period t

B The table below shows the results of the linear regression from Part A using annual data for the S&P 500 and a dummy variable for the party that controlled the White House. The data are from 1926 to 2002.

Results from Regressing S&P 500 Returns on a Dummy Variable for the Party That Controlled the White House, 1926-2002

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.1494	0.0323	4.6270
Party_t	−0.0570	0.0466	−1.2242

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	1	0.0625	0.0625	1.4987	0.2247
Residual	75	3.1287	0.0417		
Total	76	3.1912			
Residual standard error		0.2042			

(continued)

(Continued)

ANOVA	df	SS	MSS	*F*	Significance *F*
R-squared		0.0196			
Observations		77			

Source: FactSet.

Based on the coefficient and standard error estimates, verify to two decimal places the *t*-statistic for the coefficient on the dummy variable reported in the table.

C Determine at the 0.05 significance level whether overall US equity market returns tend to differ depending on the political party controlling the White House.

8 Problem 3 addressed the cross-sectional variation in the number of financial analysts who follow a company. In that problem, company size and debt-to-equity ratios were the independent variables. You receive a suggestion that membership in the S&P 500 Index should be added to the model as a third independent variable; the hypothesis is that there is greater demand for analyst coverage for stocks included in the S&P 500 because of the widespread use of the S&P 500 as a benchmark.

A Write a multiple regression equation to test whether analyst following is systematically higher for companies included in the S&P 500 Index. Also include company size and debt-to-equity ratio in this equation. Use the notations below.

$(\text{Analyst following})_i$ = natural log of (1 + Number of analysts following company i)

Size_i = natural log of the market capitalization of company i in millions of dollars

$(\text{D/E})_i$ = debt-to-equity ratio for company i

$\text{S\&P}_i$ = inclusion of company i in the S&P 500 Index (1 if included, 0 if not included)

In the above specification for analyst following, 1 is added to the number of analysts following a company because some companies are not followed by any analyst, and the natural log of 0 is indeterminate.

B State the appropriate null hypothesis and alternative hypothesis in a two-sided test of significance of the dummy variable.

C The following table gives estimates of the coefficients of the above regression model for a randomly selected sample of 500 companies. The data are for the year 2002. Determine whether you can reject the null hypothesis at the 0.05 significance level (in a two-sided test of significance).

Coefficient Estimates from Regressing Analyst Following on Size, Debt-to-Equity Ratio, and S&P 500 Membership, 2002

	Coefficient	Standard Error	*t*-Statistic
Intercept	−0.0075	0.1218	−0.0616
Size_i	0.2648	0.0191	13.8639

(Continued)

	Coefficient	Standard Error	t-Statistic
$(D/E)_i$	−0.1829	0.0608	−3.0082
$S\&P_i$	0.4218	0.0919	4.5898
n = 500			

Source: First Call/Thomson Financial, Compustat.

D Consider a company with a debt-to-equity ratio of 2/3 and a market capitalization of $10 billion. According to the estimated regression equation, how many analysts would follow this company if it were not included in the S&P 500 Index, and how many would follow if it were included in the index?

E In Problem 3, using the sample, we estimated the coefficient on the size variable as 0.3199, versus 0.2648 in the above regression. Discuss whether there is an inconsistency in these results.

9 You believe there is a relationship between book-to-market ratios and subsequent returns. The output from a cross-sectional regression and a graph of the actual and predicted relationship between the book-to-market ratio and return are shown below.

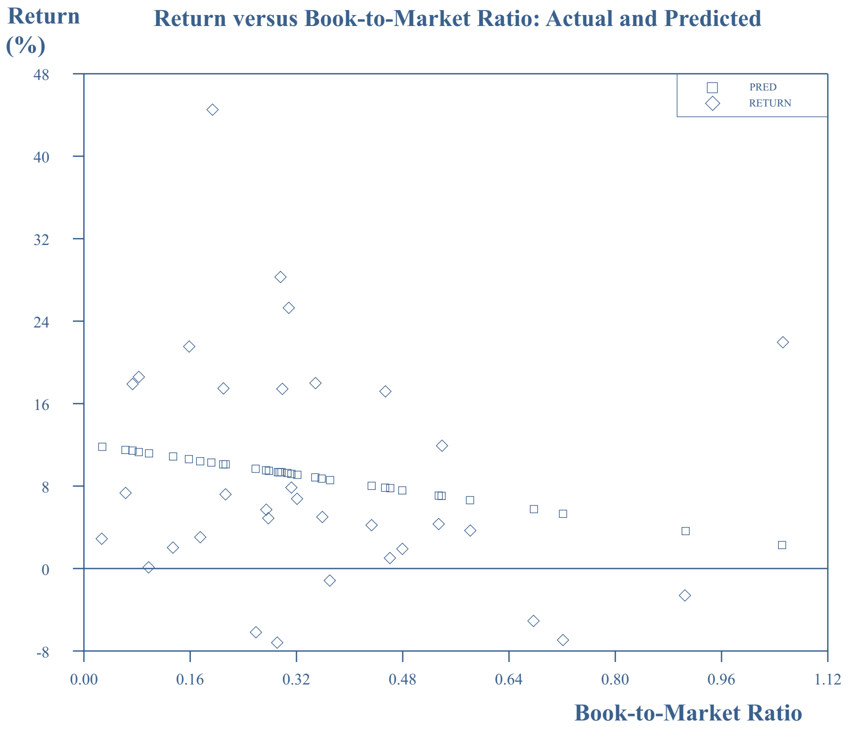
Results from Regressing Returns on the Book-to-Market Ratio

	Coefficient	Standard Error	t-Statistic
Intercept	12.0130	3.5464	3.3874
$\left(\frac{\text{Book value}}{\text{Market value}}\right)_i$	−9.2209	8.4454	−1.0918

ANOVA	df	SS	MSS	F	Significance F
Regression	1	154.9866	154.9866	1.1921	0.2831
Residual	32	4162.1895	130.0684		
Total	33	4317.1761			
Residual standard error		11.4048			
R-squared		0.0359			
Observations		34			

(continued)

(Continued)



A You are concerned with model specification problems and regression assumption violations. Focusing on assumption violations, discuss symptoms of conditional heteroskedasticity based on the graph of the actual and predicted relationship.

B Describe in detail how you could formally test for conditional heteroskedasticity in this regression.

C Describe a recommended method for correcting for conditional heteroskedasticity.

10 You are examining the effects of the January 2001 NYSE implementation of the trading of shares in minimal increments (ticks) of $0.01 (decimalization). In particular, you are analyzing a sample of 52 Canadian companies cross-listed on both the NYSE and the Toronto Stock Exchange (TSE). You find that the bid–ask spreads of these shares decline on both exchanges after the NYSE decimalization. You run a linear regression analyzing the decline in spreads on the TSE, and find that the decline on the TSE is related to company size, pre-decimalization ratio of NYSE to TSE spreads, and decline in the NYSE spreads. The relationships are statistically significant. You want to be sure, however, that the results are not influenced by conditional heteroskedasticity. Therefore, you regress the squared residuals of the regression model on the three independent variables. The R^2 for this regression is 14.1 percent. Perform a statistical test to determine if conditional heteroskedasticity is present.

11 You are analyzing if institutional investors such as mutual funds and pension funds prefer to hold shares of companies with less volatile returns. You have the percentage of shares held by institutional investors at the end of 1998 for a random sample of 750 companies. For these companies, you compute the standard deviation of daily returns during that year. Then you regress the institutional holdings on the standard deviation of returns. You find that the regression is significant at the 0.01 level and the F-statistic is 12.98. The R^2 for this regression is 1.7 percent. As expected, the regression coefficient of the standard deviation of returns is negative. Its t-statistic is −3.60, which is also significant at the 0.01 level. Before concluding that institutions prefer to hold shares of less volatile stocks, however, you want to be sure that the regression results are not influenced by conditional heteroskedasticity. Therefore, you regress the squared residuals of the regression model on the standard deviation of returns. The R^2 for this regression is 0.6 percent.

A Perform a statistical test to determine if conditional heteroskedasticity is present at the 0.05 significance level.

B In view of your answer to Part A, what remedial action, if any, is appropriate?

12 In estimating a regression based on monthly observations from January 1987 to December 2002 inclusive, you find that the coefficient on the independent variable is positive and significant at the 0.05 level. You are concerned, however, that the t-statistic on the independent variable may be inflated because of serial correlation between the error terms. Therefore, you examine the Durbin–Watson statistic, which is 1.8953 for this regression.

A Based on the value of the Durbin–Watson statistic, what can you say about the serial correlation between the regression residuals? Are they positively correlated, negatively correlated, or not correlated at all?

B Compute the sample correlation between the regression residuals from one period and those from the previous period.

C Perform a statistical test to determine if serial correlation is present. Assume that the critical values for 192 observations when there is a single independent variable are about 0.09 above the critical values for 100 observations.

13 The book-to-market ratio and the size of a company's equity are two factors that have been asserted to be useful in explaining the cross-sectional variation in subsequent returns. Based on this assertion, you want to estimate the following regression model:

$$R_i = b_0 + b_1\left(\frac{\text{Book}}{\text{Market}}\right)_i + b_2\text{Size}_i + \varepsilon_i$$

where

R_i = Return of company i's shares (in the following period)

$\left(\frac{\text{Book}}{\text{Market}}\right)_i$ = company i's book-to-market ratio

Size_i = Market value of company i's equity

A colleague suggests that this regression specification may be erroneous, because he believes that the book-to-market ratio may be strongly related to (correlated with) company size.

A To what problem is your colleague referring, and what are its consequences for regression analysis?

B With respect to multicollinearity, critique the choice of variables in the regression model above.

Regression of Return on Book-to-Market and Size

	Coefficient	Standard Error	*t*-Statistic
Intercept	14.1062	4.220	3.3427
$\left(\frac{\text{Book}}{\text{Market}}\right)_i$	−12.1413	9.0406	−1.3430
Size_i	−0.00005502	0.00005977	−0.92047
R-squared	0.06156		
Observations	34		

Correlation Matrix

	Book-to-Market Ratio	Size
Book-to-Market Ratio	1.0000	
Size	−0.3509	1.0000

C State the classic symptom of multicollinearity and comment on that basis whether multicollinearity appears to be present, given the additional fact that the F-test for the above regression is not significant.

14 You are analyzing the variables that explain the returns on the stock of the Boeing Company. Because overall market returns are likely to explain a part of the returns on Boeing, you decide to include the returns on a value-weighted index of all the companies listed on the NYSE, AMEX, and NASDAQ as an independent variable. Further, because Boeing is a large company, you also decide to include the returns on the S&P 500 Index, which is a value-weighted index of the larger market-capitalization companies. Finally, you decide to include the changes in the US dollar's value. To conduct your test, you have collected the following data for the period 1990–2002.

R_t = monthly return on the stock of Boeing in month t

R_{ALLt} = monthly return on a value-weighted index of all the companies listed on the NYSE, AMEX, and NASDAQ in month t

R_{SPt} = monthly return on the S&P 500 Index in month t

ΔX_t = change in month t in the log of a trade-weighted index of the foreign exchange value of the US dollar against the currencies of a broad group of major US trading partners

The following table shows the output from regressing the monthly return on Boeing stock on the three independent variables.

Regression of Boeing Returns on Three Explanatory Variables: Monthly Data, January 1990–December 2002

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0026	0.0066	0.3939
R_{ALLt}	−0.1337	0.6219	−0.2150

(Continued)

	Coefficient	Standard Error	*t*-Statistic
R_{SPt}	0.8875	0.6357	1.3961
ΔX_t	0.2005	0.5399	0.3714

ANOVA	df	SS	MSS
Regression	3	0.1720	0.0573
Residual	152	0.8947	0.0059
Total	155	1.0667	
Residual standard error	0.0767		
R-squared	0.1610		
Observations	156		

Source: FactSet, Federal Reserve Bank of Philadelphia.

From the *t*-statistics, we see that none of the explanatory variables is statistically significant at the 5 percent level or better. You wish to test, however, if the three variables *jointly* are statistically related to the returns on Boeing.

A Your null hypothesis is that all three population slope coefficients equal 0—that the three variables *jointly* are statistically not related to the returns on Boeing. Conduct the appropriate test of that hypothesis.

B Examining the regression results, state the regression assumption that may be violated in this example. Explain your answer.

C State a possible way to remedy the violation of the regression assumption identified in Part B.

15 You are analyzing the cross-sectional variation in the number of financial analysts that follow a company (also the subject of Problems 3 and 8). You believe that there is less analyst following for companies with a greater debt-to-equity ratio and greater analyst following for companies included in the S&P 500 Index. Consistent with these beliefs, you estimate the following regression model.

$$(\text{Analysts following})_i = b_0 + b_1(\text{D/E})_i + b_2(\text{S\&P})_i + \varepsilon_i$$

where

$(\text{Analysts following})_i$ = natural log of (1 + Number of analysts following company i)

$(\text{D/E})_i$ = debt-to-equity ratio for company i

$\text{S\&P}_i$ = inclusion of company i in the S&P 500 Index (1 if included; 0 if not included)

In the preceding specification, 1 is added to the number of analysts following a company because some companies are not followed by any analysts, and the natural log of 0 is indeterminate. The following table gives the coefficient estimates of the above regression model for a randomly selected sample of 500 companies. The data are for the year 2002.

Coefficient Estimates from Regressing Analyst Following on Debt-to-Equity Ratio and S&P 500 Membership, 2002

	Coefficient	Standard Error	*t*-Statistic
Intercept	1.5367	0.0582	26.4038
$(D/E)_i$	−0.1043	0.0712	−1.4649
$S\&P_i$	1.2222	0.0841	14.5327
n = 500			

Source: First Call/Thomson Financial, Compustat.

You discuss your results with a colleague. She suggests that this regression specification may be erroneous, because analyst following is likely to be also related to the size of the company.

A What is this problem called, and what are its consequences for regression analysis?

B To investigate the issue raised by your colleague, you decide to collect data on company size also. You then estimate the model after including an additional variable, Size *i*, which is the natural log of the market capitalization of company *i* in millions of dollars. The following table gives the new coefficient estimates.

Coefficient Estimates from Regressing Analyst Following on Size, Debt-to-Equity Ratio, and S&P 500 Membership, 2002

	Coefficient	Standard Error	*t*-Statistic
Intercept	−0.0075	0.1218	−0.0616
$Size_i$	0.2648	0.0191	13.8639
$(D/E)_i$	−0.1829	0.0608	−3.0082
$S\&P_i$	0.4218	0.0919	4.5898
n = 500			

Source: First Call/Thomson Financial, Compustat.

What do you conclude about the existence of the problem mentioned by your colleague in the original regression model you had estimated?

16 You have noticed that hundreds of non-US companies are listed not only on a stock exchange in their home market but also on one of the exchanges in the United States. You have also noticed that hundreds of non-US companies are listed only in their home market and not in the United States. You are trying to predict whether or not a non-US company will choose to list on a US exchange. One of the factors that you think will affect whether or not a company lists in the United States is its size relative to the size of other companies in its home market.

A What kind of a dependent variable do you need to use in the model?

B What kind of a model should be used?

The following information relates to Questions 17–22

Gary Hansen is a securities analyst for a mutual fund specializing in small-capitalization growth stocks. The fund regularly invests in initial public offerings (IPOs). If the fund subscribes to an offer, it is allocated shares at the offer price. Hansen notes that IPOs frequently are underpriced, and the price rises when open market trading begins. The initial return for an IPO is calculated as the change in price on the first day of trading divided by the offer price. Hansen is developing a regression model to predict the initial return for IPOs. Based on past research, he selects the following independent variables to predict IPO initial returns:

Underwriter rank	=	1–10, where 10 is highest rank
Pre-offer price adjustment[a]	=	(Offer price − Initial filing price)/Initial filing price
Offer size ($ millions)	=	Shares sold × Offer price
Fraction retained[a]	=	Fraction of total company shares retained by insiders

[a]Expressed as a decimal

Hansen collects a sample of 1,725 recent IPOs for his regression model. Regression results appear in Exhibit 1, and ANOVA results appear in Exhibit 2.

Exhibit 1 Hansen's Regression Results Dependent Variable: IPO Initial Return (Expressed in Decimal Form, i.e., 1% = 0.01)

Variable	Coefficient (b_j)	Standard Error	*t*-Statistic
Intercept	0.0477	0.0019	25.11
Underwriter rank	0.0150	0.0049	3.06
Pre-offer price adjustment	0.4350	0.0202	21.53
Offer size	−0.0009	0.0011	−0.82
Fraction retained	0.0500	0.0260	1.92

Exhibit 2 Selected ANOVA Results for Hansen's Regression

	Degrees of Freedom (df)	Sum of Squares (SS)
Regression	4	51.433
Residual	1,720	91.436
Total	1,724	142.869
	Multiple *R*-squared = 0.36	

Hansen wants to use the regression results to predict the initial return for an upcoming IPO. The upcoming IPO has the following characteristics:

- underwriter rank = 6;
- pre-offer price adjustment = 0.04;

- offer size = $40 million;
- fraction retained = 0.70.

Because he notes that the pre-offer price adjustment appears to have an important effect on initial return, Hansen wants to construct a 95 percent confidence interval for the coefficient on this variable. He also believes that for each 1 percent increase in pre-offer price adjustment, the initial return will increase by less than 0.5 percent, holding other variables constant. Hansen wishes to test this hypothesis at the 0.05 level of significance.

Before applying his model, Hansen asks a colleague, Phil Chang, to review its specification and results. After examining the model, Chang concludes that the model suffers from two problems: 1) conditional heteroskedasticity, and 2) omitted variable bias. Chang makes the following statements:

Statement 1 "Conditional heteroskedasticity will result in consistent coefficient estimates, but both the *t*-statistics and *F*-statistic will be biased, resulting in false inferences."

Statement 2 "If an omitted variable is correlated with variables already included in the model, coefficient estimates will be biased and inconsistent and standard errors will also be inconsistent."

Selected values for the *t*-distribution and *F*-distribution appear in Exhibits 3 and 4, respectively.

Exhibit 3 Selected Values for the *t*-Distribution (df = ∞)

Area in Right Tail	*t*-Value
0.050	1.645
0.025	1.960
0.010	2.326
0.005	2.576

Exhibit 4 Selected Values for the *F*-Distribution (α = 0.01) (df1/df2: Numerator/Denominator Degrees of Freedom)

		df1	
		4	∞
df2	4	16.00	13.50
	∞	3.32	1.00

17 Based on Hansen's regression, the predicted initial return for the upcoming IPO is *closest* to:

A 0.0943.

B 0.1064.

C 0.1541.

18 The 95 percent confidence interval for the regression coefficient for the pre-offer price adjustment is *closest* to:

A 0.156 to 0.714.

B 0.395 to 0.475.

C 0.402 to 0.468.

19 The *most* appropriate null hypothesis and the *most* appropriate conclusion regarding Hansen's belief about the magnitude of the initial return relative to that of the pre-offer price adjustment (reflected by the coefficient b_j) are:

	Null Hypothesis	Conclusion about b_j (0.05 Level of Significance)
A	H_0: $b_j = 0.5$	Reject H_0
B	H_0: $b_j \geq 0.5$	Fail to reject H_0
C	H_0: $b_j \geq 0.5$	Reject H_0

20 The *most* appropriate interpretation of the multiple *R*-squared for Hansen's model is that:

A unexplained variation in the dependent variable is 36 percent of total variation.

B correlation between predicted and actual values of the dependent variable is 0.36.

C correlation between predicted and actual values of the dependent variable is 0.60.

21 Is Chang's Statement 1 correct?

A Yes.

B No, because the model's *F*-statistic will not be biased.

C No, because the model's *t*-statistics will not be biased.

22 Is Chang's Statement 2 correct?

A Yes.

B No, because the model's coefficient estimates will be unbiased.

C No, because the model's coefficient estimates will be consistent.

The following information relates to Questions 23–28

Adele Chiesa is a money manager for the Bianco Fund. She is interested in recent findings showing that certain business condition variables predict excess US stock market returns (one-month market return minus one-month T-bill return). She is also familiar with evidence showing how US stock market returns differ by the political party affiliation of the US President. Chiesa estimates a multiple regression model to predict monthly excess stock market returns accounting for business conditions and the political party affiliation of the US President:

$$\text{Excess stock market return}_t = a_0 + a_1\text{Default spread}_{t-1} + a_2\text{Term spread}_{t-1} + a_3\text{Pres party dummy}_{t-1} + e_t$$

Default spread is equal to the yield on Baa bonds minus the yield on Aaa bonds. Term spread is equal to the yield on a 10-year constant-maturity US Treasury index minus the yield on a 1-year constant-maturity US Treasury index. Pres party dummy is equal to 1 if the US President is a member of the Democratic Party and 0 if a member of the Republican Party.

Chiesa collects 432 months of data (all data are in percent form, i.e., 0.01 = 1 percent). The regression is estimated with 431 observations because the independent variables are lagged one month. The regression output is in Exhibit 1. Exhibits 2 through 5 contain critical values for selected test statistics.

Exhibit 1 Multiple Regression Output (the Dependent Variable Is the One-Month Market Return in Excess of the One-Month T-Bill Return)

	Coefficient	*t*-Statistic	*p*-Value
Intercept	−4.60	−4.36	<0.01
Default spread$_{t-1}$	3.04	4.52	<0.01
Term spread$_{t-1}$	0.84	3.41	<0.01
Pres party dummy$_{t-1}$	3.17	4.97	<0.01
Number of observations		431	
Test statistic from Breusch–Pagan (BP) test		7.35	
R^2		0.053	
Adjusted R^2		0.046	
Durbin–Watson (DW)		1.65	
Sum of squared errors (SSE)		19,048	
Regression sum of squares (SSR)		1,071	

An intern working for Chiesa has a number of questions about the results in Exhibit 1:

Question 1 How do you test to determine whether the overall regression model is significant?

Question 2 Does the estimated model conform to standard regression assumptions? For instance, is the error term serially correlated, or is there conditional heteroskedasticity?

Question 3 How do you interpret the coefficient for the Pres party dummy variable?

Question 4 Default spread appears to be quite important. Is there some way to assess the precision of its estimated coefficient? What is the economic interpretation of this variable?

After responding to her intern's questions, Chiesa concludes with the following statement: "Predictions from Exhibit 1 are subject to parameter estimate uncertainty, but not regression model uncertainty."

Exhibit 2 Critical Values for the Durbin–Watson Statistic (α = 0.05)

	K = 3	
N	d_l	d_u
420	1.825	1.854
430	1.827	1.855
440	1.829	1.857

Exhibit 3 Table of the Student's *t*-Distribution (One-Tailed Probabilities for df = ∞)

P	*t*
0.10	1.282
0.05	1.645
0.025	1.960
0.01	2.326
0.005	2.576

Exhibit 4 Values of χ^2

	Probability in Right Tail			
df	**0.975**	**0.95**	**0.05**	**0.025**
1	0.0001	0.0039	3.841	5.024
2	0.0506	0.1026	5.991	7.378
3	0.2158	0.3518	7.815	9.348
4	0.4840	0.7110	9.488	11.14

Exhibit 5 Table of the *F*-Distribution (Critical Values for Right-Hand Tail Area Equal to 0.05) Numerator: df1 and Denominator: df2

	df1				
df2	**1**	**2**	**3**	**4**	**427**
1	161	200	216	225	254
2	18.51	19.00	19.16	19.25	19.49
3	10.13	9.55	9.28	9.12	8.53

(continued)

Exhibit 5 (Continued)

	df1				
df2	**1**	**2**	**3**	**4**	**427**
4	7.71	6.94	6.59	6.39	5.64
427	3.86	3.02	2.63	2.39	1.17

23 Regarding the intern's Question 1, is the regression model as a whole significant at the 0.05 level?

A No, because the calculated F-statistic is less than the critical value for F.

B Yes, because the calculated F-statistic is greater than the critical value for F.

C Yes, because the calculated χ^2 statistic is greater than the critical value for χ^2.

24 Which of the following is Chiesa's *best* response to Question 2 regarding serial correlation in the error term? At a 0.05 level of significance, the test for serial correlation indicates that there is:

A no serial correlation in the error term.

B positive serial correlation in the error term.

C negative serial correlation in the error term.

25 Regarding Question 3, the Pres party dummy variable in the model indicates that the mean monthly value for the excess stock market return is:

A 1.43 percent larger during Democratic presidencies than Republican presidencies.

B 3.17 percent larger during Democratic presidencies than Republican presidencies.

C 3.17 percent larger during Republican presidencies than Democratic presidencies.

26 In response to Question 4, the 95 percent confidence interval for the regression coefficient for the default spread is *closest* to:

A 0.13 to 5.95.

B 1.72 to 4.36.

C 1.93 to 4.15.

27 With respect to the default spread, the estimated model indicates that when business conditions are:

A strong, expected excess returns will be higher.

B weak, expected excess returns will be lower.

C weak, expected excess returns will be higher.

28 Is Chiesa's concluding statement correct regarding parameter estimate uncertainty and regression model uncertainty?

A Yes.

B No, predictions are not subject to parameter estimate uncertainty.

C No, predictions are subject to regression model uncertainty and parameter estimate uncertainty.

The following information relates to Questions 29–36

Doris Honoré is a securities analyst with a large wealth management firm. She and her colleague Bill Smith are addressing three research topics: how investment fund characteristics affect fund total returns, whether a fund rating system helps predict fund returns, and whether stock and bond market returns explain the returns of a portfolio of utility shares run by the firm.

To explore the first topic, Honoré decides to study US mutual funds using a sample of 555 large-cap US equity funds. The sample includes funds in style classes of value, growth, and blend (i.e., combining value and growth characteristics). The dependent variable is the average annualized rate of return (in percent) over the past five years. The independent variables are fund expense ratio, portfolio turnover, the natural logarithm of fund size, fund age, and three dummy variables. The multiple manager dummy variable has a value of 1 if the fund has multiple managers (and a value of 0 if it has a single manager). The fund style is indicated by a growth dummy (value of 1 for growth funds and 0 otherwise) and a blend dummy (value of 1 for blend funds and 0 otherwise). If the growth and blend dummies are both zero, the fund is a value fund. The regression output is given in Exhibit 1.

Exhibit 1 Multiple Regression Output for Large-Cap Mutual Fund Sample

	Coefficient	Standard Error	*t*-Statistic
Intercept	10.9375	1.3578	8.0551
Expense ratio (%)	−1.4839	0.2282	−6.5039
Portfolio turnover (%)	0.0017	0.0016	1.0777
ln (fund size in $)	0.1467	0.0612	2.3976
Manager tenure (years)	−0.0098	0.0102	−0.9580
Multiple manager dummy	0.0628	0.1533	0.4100
Fund age (years)	−0.0123	0.0047	−2.6279
Growth dummy	2.4368	0.1886	12.9185
Blend dummy	0.5757	0.1881	3.0611

ANOVA	df	SS	MSS
Regression	8	714.169	89.2712
Residual	546	1583.113	2.8995
Total	554	2297.282	
Multiple R	0.5576		
R^2	0.3109		
Adjusted R^2	0.3008		
Standard error (%)	1.7028		
Observations	555		

Based on the results shown in Exhibit 1, Honoré wants to test the hypothesis that all of the regression coefficients are equal to zero. For the 555 fund sample, she also wants to compare the performance of growth funds with the value funds.

Honoré is concerned about the possible presence of multicollinearity in the regression. She states that adding a new independent variable that is highly correlated with one or more independent variables already in the regression model, has three potential consequences:

1 The R^2 is expected to decline.
2 The regression coefficient estimates can become imprecise and unreliable.
3 The standard errors for some or all of the regression coefficients will become inflated.

Another concern for the regression model (in Exhibit 1) is conditional heteroskedasticity. Honoré is concerned that the presence of heteroskedasticity can cause both the *F*-test for the overall significance of the regression and the *t*-tests for significance of individual regression coefficients to be unreliable. She runs a regression of the squared residuals from the model in Exhibit 1 on the eight independent variables, and finds the R^2 is 0.0669.

As a second research project, Honoré wants to test whether including Morningstar's rating system, which assigns a one- through five-star rating to a fund, as an independent variable will improve the predictive power of the regression model. To do this, she needs to examine whether values of the independent variables in a given period predict fund return in the next period. Smith suggests three different methods of adding the Morningstar ratings to the model:

- Method 1: Add an independent variable that has a value equal to the number of stars in the rating of each fund.
- Method 2: Add five dummy variables, one for each rating.
- Method 3: Add dummy variables for four of the five ratings.

As a third research project, Honoré wants to establish whether bond market returns (proxied by returns of long-term US Treasuries) and stock market returns (proxied by returns of the S&P 500 Index) explain the returns of a portfolio of utility stocks being recommended to clients. Exhibit 2 presents the results of a regression of 10 years of monthly percentage total returns for the utility portfolio on monthly total returns for US Treasuries and the S&P 500.

Exhibit 2 Regression Analysis of Utility Portfolio Returns

	Coefficient	Standard Error	*t*-Statistic	*p*-Value
Intercept	–0.0851	0.2829	–0.3008	0.7641
US Treasury	0.4194	0.0848	4.9474	<0.0001
S&P 500	0.6198	0.0666	9.3126	<0.0001

ANOVA	df	SS	MSS	*F*	Significance *F*
Regression	2	827.48	413.74	46.28	<0.0001
Residual	117	1045.93	8.94		
Total	119	1873.41			
Multiple R	0.6646				
R^2	0.4417				
Adjusted R^2	0.4322				

Exhibit 2 (Continued)

ANOVA	df	SS	MSS	*F*	Significance *F*
Standard error (%)	2.99				
Observations	120				

For the time-series model in Exhibit 2, Honoré says that positive serial correlation would not require that the estimated coefficients be adjusted, but that the standard errors of the regression coefficients would be underestimated. This issue would cause the *t*-statistics of the regression coefficients to be inflated. Honoré tests the null hypothesis that the there is no serial correlation in the regression residuals and finds that the Durbin–Watson statistic is equal to 1.81. The critical values at the 0.05 significance level for the Durbin–Watson statistic are $d_l = 1.63$ and $d_u = 1.72$.

Smith asks whether Honoré should have estimated the models in Exhibit 1 and Exhibit 2 using a probit or logit model instead of using a traditional regression analysis.

29 Considering Exhibit 1, the *F*-statistic is closest to:

A 3.22.

B 8.06.

C 30.79.

30 Based on Exhibit 1, the difference between the predicted annualized returns of a growth fund and an otherwise similar value fund is *closest* to:

A 1.86%.

B 2.44%.

C 3.01%.

31 Honoré describes three potential consequences of multicollinearity. Are all three consequences correct?

A Yes

B No, 1 is incorrect

C No, 2 is incorrect

32 Which of the three methods suggested by Smith would *best* capture the ability of the Morningstar rating system to predict mutual fund performance?

A Method 1

B Method 2

C Method 3

33 Honoré is concerned about the consequences of heteroskedasticity. Is she correct regarding the effect of heteroskedasticity on the reliability of the *F*-test and *t*-tests?

A Yes

B No, she is incorrect with regard to the *F*-test

C No, she is incorrect with regard to the *t*-tests

34 Is Honoré's description of the effects of positive serial correlation (in Exhibit 2) correct regarding the estimated coefficients and the standard errors?

A Yes

B No, she is incorrect about only the estimated coefficients

C No, she is incorrect about only the standard errors of the regression coefficients

35 Based on her estimated Durbin–Watson statistic, Honoré should:

A fail to reject the null hypothesis.

B reject the null hypothesis because there is significant positive serial correlation.

C reject the null hypothesis because there is significant negative serial correlation.

36 Should Honoré have estimated the models in Exhibit 1 and Exhibit 2 using probit or logit models instead of traditional regression analysis?

A Both should be estimated with probit or logit models.

B Neither should be estimated with probit or logit models.

C Only the analysis in Exhibit 1 should be done with probit or logit models.

The following information relates to Questions 37–45

Brad Varden, a junior analyst at an actively managed mutual fund, is responsible for research on a subset of the 500 large-cap equities the fund follows. Recently, the fund has been paying close attention to management turnover and to publicly available environmental, social, and governance (ESG) ratings. Varden is given the task of investigating whether any significant relationship exists between a company's profitability and either of these two characteristics. Colleen Quinni, a senior analyst at the fund, suggests that as an initial step in his investigation, Varden should perform a multiple regression analysis on the variables and report back to her.

Varden knows that Quinni is an expert at quantitative research, and she once told Varden that after you get an idea, you should formulate a hypothesis, test the hypothesis, and analyze the results. Varden expects to find that ESG rating is negatively related to ROE and CEO tenure is positively related to ROE. He considers a relationship meaningful when it is statistically significant at the 0.05 level. To begin, Varden collects values for ROE, CEO tenure, and ESG rating for a sample of 40 companies from the large-cap security universe. He performs a multiple regression with ROE (in percent) as the dependent variable and ESG rating and CEO tenure (in years) as the independent variables: $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \varepsilon_i$.

Exhibit 1 shows the regression results.

Exhibit 1 Regression Statistics

$$\hat{Y}_i = 9.442 + 0.069X_{1i} + 0.681X_{2i}$$

	Coefficient	Standard Error	*t*-Statistic	*p*-Value
Intercept	9.442	3.343	2.824	0.008
b_1 (ESG variable)	0.069	0.058	1.201	0.238
b_2 (Tenure variable)	0.681	0.295	2.308	0.027

Exhibit 1 (Continued)

ANOVA	df	SS	MSS	F	Significance F
Regression	2	240.410	120.205	4.161	0.023
Residual	37	1069.000	28.892		
Total	39	1309.410			
Multiple R	0.428				
R^2	0.183				
Adjusted R^2	0.139				
Standard error (%)	5.375				
Observations	40				

DF Associates is one of the companies Varden follows. He wants to predict its ROE using his regression model. DF Associates' corporate ESG rating is 55, and the company's CEO has been in that position for 10.5 years.

Varden also wants to check on the relationship between these variables and the dividend growth rate (divgr), so he completes the correlation matrix shown in Exhibit 2.

Exhibit 2 Correlation Matrix

	ROE	ESG	Tenure	Divgr
ROE	1.0			
ESG	0.446	1.0		
Tenure	0.369	0.091	1.0	
Divgr	0.117	0.046	0.028	1.0

Investigating further, Varden determines that dividend growth is not a linear combination of CEO tenure and ESG rating. He is unclear about how additional independent variables would affect the significance of the regression, so he asks Quinni, "Given this correlation matrix, will both R^2 and adjusted R^2 automatically increase if I add dividend growth as a third independent variable?"

The discussion continues, and Quinni asks two questions.

1 What does your F-statistic of 4.161 tell you about the regression?

2 In interpreting the overall significance of your regression model, which statistic do you believe is most relevant: R^2, adjusted R^2, or the F-statistic?

Varden answers both questions correctly and says he wants to check two more ideas. He believes the following:

1 ROE is less correlated with the dividend growth rate in firms whose CEO has been in office more than 15 years, and

2 CEO tenure is a normally distributed random variable.

Later, Varden includes the dividend growth rate as a third independent variable and runs the regression on the fund's entire group of 500 large-cap equities. He finds that the adjusted R^2 is much higher than the results in Exhibit 1. He reports this to Quinni and says, “Adding the dividend growth rate gives a model with a higher adjusted R^2. The three-variable model is clearly better.” Quinni cautions, “I don’t think you can conclude that yet.”

37 Based on Exhibit 1 and given Varden’s expectations, which is the *best* null hypothesis and conclusion regarding CEO tenure?

A $b_2 \leq 0$; reject the null hypothesis

B $b_2 = 0$; cannot reject the null hypothesis

C $b_2 \geq 0$; reject the null hypothesis

38 At a significance level of 1%, which of the following is the *best* interpretation of the regression coefficients with regard to explaining ROE?

A ESG is significant, but tenure is not.

B Tenure is significant, but ESG is not.

C Neither ESG nor tenure is significant.

39 Based on Exhibit 1, which independent variables in Varden’s model are significant at the 0.05 level?

A ESG only

B Tenure only

C Neither ESG nor tenure

40 Based on Exhibit 1, the predicted ROE for DF Associates is *closest* to:

A 10.957%.

B 16.593%.

C 20.388%.

41 Based on Exhibit 2, Quinni’s *best* answer to Varden’s question about the effect of adding a third independent variable is:

A no for R^2 and no for adjusted R^2.

B yes for R^2 and no for adjusted R^2.

C yes for R^2 and yes for adjusted R^2.

42 Based on Exhibit 1, Varden’s *best* answer to Quinni’s question about the F-statistic is:

A both independent variables are significant at the 0.05 level.

B neither independent variable is significant at the 0.05 level.

C at least one independent variable is significant at the 0.05 level.

43 Varden’s *best* answer to Quinni’s question about overall significance is:

A R^2.

B adjusted R^2.

C the F-statistic.

44 If Varden’s beliefs about ROE and CEO tenure are true, which of the following would violate the assumptions of multiple regression analysis?

A The assumption about CEO tenure distribution only

B The assumption about the ROE/dividend growth correlation only

C The assumptions about both the ROE/dividend growth correlation and CEO tenure distribution

45 The *best* rationale for Quinni’s caution about the three-variable model is that the:

A dependent variable is defined differently.

B sample sizes are different in the two models.

C dividend growth rate is positively correlated with the other independent variables.

SOLUTIONS

1 **A** $R_{it} = b_0 + b_1R_{Mt} + b_2\Delta X_t + \varepsilon_{it}$

B We can test whether the coefficient on the S&P 500 Index returns is statistically significant. Our null hypothesis is that the coefficient is equal to 0 (H_0: $b_1 = 0$); our alternative hypothesis is that the coefficient is not equal to 0 (H_a: $b_1 \neq 0$). We construct the t-test of the null hypothesis as follows:

$$\frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{0.5373 - 0}{0.1332} = 4.0338$$

where

$\hat{b}_1$ = regression estimate of b_1

b_1 = the hypothesized value of the coefficient (here, 0)

$s_{\hat{b}_1}$ = the estimated standard error of $\hat{b}_1$

Because this regression has 156 observations and three regression coefficients, the t-test has 156 − 3 = 153 degrees of freedom. At the 0.05 significance level, the critical value for the test statistic is between 1.98 and 1.97. The absolute value of the test statistic is 4.0338; therefore, we can reject the null hypothesis that $b_1 = 0$.

Similarly, we can test whether the coefficient on the change in the value of the US dollar is statistically significant in this regression. Our null hypothesis is that the coefficient is equal to 0 (H_0: $b_2 = 0$); our alternative hypothesis is that the coefficient is not equal to 0 (H_a: $b_2 \neq 0$). We construct the t-test as follows:

$$\frac{\hat{b}_2 - b_2}{s_{\hat{b}_2}} = \frac{-0.5768 - 0}{0.5121} = -1.1263$$

As before, the t-test has 153 degrees of freedom, and the critical value for the test statistic is between 1.98 and 1.97 at the 0.05 significance level. The absolute value of the test statistic is 1.1263; therefore, we cannot reject the null hypothesis that $b_2 = 0$.

Based on the above t-tests, we conclude that S&P 500 Index returns do affect ADM's returns but that changes in the value of the US dollar do not affect ADM's returns.

C The statement is not correct. To make it correct, we need to add the qualification "holding ΔX constant" to the end of the quoted statement.

2 **A** $R_i = b_0 + b_1(\text{B/M})_i + b_2\text{Size}_i + \varepsilon_i$

B We can test whether the coefficients on the book-to-market ratio and size are individually statistically significant using t-tests. For the book-to-market ratio, our null hypothesis is that the coefficient is equal to 0 (H_0: $b_1 = 0$); our alternative hypothesis is that the coefficient is not equal to 0 (H_a: $b_1 \neq 0$). We can test the null hypothesis using a t-test constructed as follows:

$$\frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{-0.0541 - 0}{0.0588} = -0.9201$$

where

$\hat{b}_1$ = regression estimate of b_1

b_1 = the hypothesized value of the coefficient (here, 0)

$s_{\hat{b}_1}$ = the estimated standard error of $\hat{b}_1$

This regression has 66 observations and three coefficients, so the t-test has 66 − 3 = 63 degrees of freedom. At the 0.05 significance level, the critical value for the test statistic is about 2.0. The absolute value of the test statistic is 0.9201; therefore, we cannot reject the null hypothesis that $b_1 = 0$. We can conclude that the book-to-market ratio is not useful in explaining the cross-sectional variation in returns for this sample.

We perform the same analysis to determine whether size (as measured by the log of the market value of equity) can help explain the cross-sectional variation in asset returns. Our null hypothesis is that the coefficient is equal to 0 (H_0: $b_2 = 0$); our alternative hypothesis is that the coefficient is not equal to 0 (H_a: $b_2 \neq 0$). We can test the null hypothesis using a t-test constructed as follows:

$$\frac{\hat{b}_2 - b_2}{s_{\hat{b}_2}} = \frac{-0.0164 - 0}{0.0350} = -0.4686$$

where

$\hat{b}_2$ = regression estimate of b_2

b_2 = the hypothesized value of the coefficient (here, 0)

$s_{\hat{b}_2}$ = the estimated standard error of $\hat{b}_2$

Again, because this regression has 66 observations and three coefficients, the t-test has 66 − 3 = 63 degrees of freedom. At the 0.05 significance level, the critical value for the test statistic is about 2.0. The absolute value of the test statistic is 0.4686; therefore, we cannot reject the null hypothesis that $b_2 = 0$. We can conclude that asset size is not useful in explaining the cross-sectional variation of asset returns in this sample.

3 A The estimated regression is (Analyst following)$_i$ = −0.2845 + 0.3199Size$_i$ − 0.1895(D/E)$_i$ + ε_i. Therefore, the prediction for the first company is

$$\begin{aligned}(\text{Analyst following})_i &= -0.2845 + 0.3199(\ln 100) - 0.1895(0.75)\\ &= -0.2845 + 1.4732 - 0.1421 = 1.0466\end{aligned}$$

Recalling that (Analyst following)$_i$ is the natural log of $(1 + n_i)$, where n_i is the number of analysts following company i; it follows that $1 + n_1 = e^{1.0466} = 2.848$, approximately. Therefore, $n_1 = 2.848 - 1 = 1.848$, or about two analysts. Similarly, the prediction for the second company is as follows:

$$\begin{aligned}(\text{Analyst following})_i &= -0.2845 + 0.3199(\ln 1{,}000) - 0.1895(0.75)\\ &= -0.2845 + 2.2098 - 0.1421\\ &= 1.7832\end{aligned}$$

Thus, $1 + n_2 = e^{1.7832} = 5.949$, approximately. Therefore, $n_2 = 5.949 - 1 = 4.949$, or about five analysts.

The model predicts that 5 − 2 = 3 more analysts will follow the second company than the first company.

B We would interpret the p-value of 0.00236 as the smallest level of significance at which we can reject a null hypothesis that the population value of the coefficient is 0, in a two-sided test. Clearly, in this regression the debt-to-equity ratio is a highly significant variable.

4 The estimated model is

Percentage decline in TSE spread of company i = −0.45 + 0.05Size$_i$ − 0.06(Ratio of spreads)$_i$ + 0.29(Decline in NASDAQ spreads)$_i$

Therefore, the prediction is

$$\begin{aligned}\text{Percentage decline in TSE spread} &= -0.45 + 0.05(\ln 900{,}000) - \\ &\quad 0.06(1.3) + 0.29(1) \\ &= -0.45 + 0.69 - 0.08 + 0.29 \\ &= 0.45\end{aligned}$$

The model predicts that for a company with average sample characteristics, the spread on the TSE declines by 0.45 percent for a 1 percent decline in NASDAQ spreads.

5 **A** To test the null hypothesis that all the slope coefficients in the regression model are equal to 0 (H_0: $b_1 = b_2 = 0$) against the alternative hypothesis that at least one slope coefficient is not equal to 0, we must use an F-test.

B To conduct the F-test, we need four inputs, all of which are found in the ANOVA section of the table in the statement of the problem:

i. total number of observations, n

ii. total number of regression coefficients to be estimated, $k + 1$

iii. sum of squared errors or residuals, $\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$ abbreviated SSE, and

iv. regression sum of squares, $\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$ abbreviated RSS

C The F-test formula is

$$F = \frac{\text{RSS}/k}{\text{SSE}/\left[n - (k+1)\right]} = \frac{0.0094/2}{0.6739/\left[66 - (2+1)\right]} = 0.4394$$

The F-statistic has degrees of freedom $F\{k, [n - (k + 1)]\} = F(2, 63)$. From the F-test table, for the 0.05 significance level, the critical value for $F(2, 63)$ is about 3.15, so we cannot reject the hypothesis that the slope coefficients are both 0. The two independent variables are jointly statistically unrelated to returns.

D Adjusted R^2 is a measure of goodness of fit that takes into account the number of independent variables in the regression, in contrast to R^2. We can assert that adjusted R^2 is smaller than $R^2 = 0.0138$ without the need to perform any calculations. (However, adjusted R^2 can be shown to equal −0.0175 using an expression in the text on the relationship between adjusted R^2 and R^2.)

6 **A** You believe that opening markets actually reduces return volatility; if that belief is correct, then the slope coefficient would be negative, $b_1 < 0$. The null hypothesis is that the belief is not true: H_0: $b_1 \geq 0$. The alternative hypothesis is that the belief is true: H_a: $b_1 < 0$.

B The critical value for the t-statistic with 95 − 2 = 93 degrees of freedom at the 0.05 significance level in a one-sided test is about 1.66. For the one-sided test stated in Part A, we reject the null hypothesis if the t-statistic on the slope coefficient is less than −1.66. As the t-statistic of $-2.7604 < -1.66$, we reject the null. Because the dummy variable takes on a value of 1 when foreign investment is allowed, we can conclude that the volatility was lower with foreign investment.

C According to the estimated regression, average return volatility was 0.0133 (the estimated value of the intercept) before July 1993 and 0.0058 (= 0.0133 − 0.0075) after July 1993.

7 A The appropriate regression model is $R_{Mt} = b_0 + b_1 \text{ Party}_t + \varepsilon_t$.

B The t-statistic reported in the table for the dummy variable tests whether the coefficient on Party_t is significantly different from 0. It is computed as follows:

$$\frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{-0.0570 - 0}{0.0466} = -1.22$$

where

$\hat{b}_1$ = regression estimate of b_1

b_1 = the hypothesized value of the coefficient (here, 0)

$s_{\hat{b}_1}$ = the estimated standard error of $\hat{b}_1$

To two decimal places, this value is the same as the t-statistic reported in the table for the dummy variable, as expected. The problem specified two decimal places because the reported regression output reflects rounding; for this reason, we often cannot exactly reproduce reported t-statistics.

C Because the regression has 77 observations and two coefficients, the t-test has 77 − 2 = 75 degrees of freedom. At the 0.05 significance level, the critical value for the two-tailed test statistic is about 1.99. The absolute value of the test statistic is 1.2242; therefore, we do not reject the null hypothesis that $b_1 = 0$. We can conclude that the political party in the White House does not, on average, affect the annual returns of the overall market as measured by the S&P 500.

8 A The regression model is as follows:

$$(\text{Analyst following})_i = b_0 + b_1\text{Size}_i + b_2(\text{D/E})_i + b_3\text{S\&P}_i + \varepsilon_i$$

where $(\text{Analyst following})_i$ is the natural log of (1 + number of analysts following company i); Size_i is the natural log of the market capitalization of company i in millions of dollars; $(\text{D/E})_i$ is the debt-to-equity ratio for company i, and $\text{S\&P}_i$ is a dummy variable with a value of 1 if the company i belongs to the S&P 500 Index and 0 otherwise.

B The appropriate null and alternative hypotheses are H_0: $b_3 = 0$ and H_a: $b_3 \neq 0$, respectively.

C The t-statistic to test the null hypothesis can be computed as follows:

$$\frac{\hat{b}_3 - b_3}{s_{\hat{b}_3}} = \frac{0.4218 - 0}{0.0919} = 4.5898$$

This value is, of course, the same as the value reported in the table. The regression has 500 observations and 4 regression coefficients, so the t-test has 500 − 4 = 496 degrees of freedom. At the 0.05 significance level, the critical value for the test statistic is between 1.96 and 1.97. Because the value of the test statistic is 4.5898 we can reject the null hypothesis that $b_3 = 0$. Thus a company's membership in the S&P 500 appears to significantly influence the number of analysts who cover that company.

D The estimated model is

$$\left(\text{Analyst following}\right)_i = -0.0075 + 0.2648\text{Size}_i - 0.1829\left(\text{D/E}\right)_i + 0.4218\text{S\&P}_i + \varepsilon_i$$

Therefore the prediction for number of analysts following the indicated company that is not part of the S&P 500 Index is

$$\begin{aligned}(\text{Analyst following})_i &= -0.0075 + 0.2648(\ln 10{,}000) - 0.1829(2/3) + 0.4218(0)\\ &= -0.0075 + 2.4389 - 0.1219 + 0\\ &= 2.3095\end{aligned}$$

Recalling that $(\text{Analyst following})_i$ is the natural log of $(1 + n_i)$, where n_i is the number of analysts following company i; it ensues (coding the company under consideration as 1) that $1 + n_1 = e^{2.3095} = 10.069$, approximately. Therefore, the prediction is that $n_1 = 10.069 - 1 = 9.069$, or about nine analysts.

Similarly, the prediction for the company that is included in the S&P 500 Index is

$$\begin{aligned}(\text{Analyst following})_i &= -0.0075 + 0.2648(\ln 10{,}000) - 0.1829(2/3) + 0.4218(1)\\ &= -0.0075 + 2.4389 - 0.1219 + 0.4218\\ &= 2.7313\end{aligned}$$

Coding the company that does belong to the S&P 500 as 2, $1 + n_2 = e^{2.7313} = 15.353$. Therefore, the prediction is that $n_2 = 15.353 - 1 = 14.353$, or about 14 analysts.

E There is no inconsistency in the coefficient on the size variable differing between the two regressions. The regression coefficient on an independent variable in a multiple regression model measures the expected net effect on the expected value of the dependent variable for a one-unit increase in that independent variable, after accounting for any effects of the other independent variables on the expected value of the dependent variable. The earlier regression had one fewer independent variable; after the effect of S&P 500 membership on the expected value of the dependent variable is taken into account, it is to be expected that the effect of the size variable on the dependent variable will change. What the regressions appear to indicate is that the net effect of the size variable on the expected analyst following diminishes when S&P 500 membership is taken into account.

9 A In a well-specified regression, the differences between the actual and predicted relationship should be random; the errors should not depend on the value of the independent variable. In this regression, the errors seem larger for smaller values of the book-to-market ratio. This finding indicates that we may have conditional heteroskedasticity in the errors, and consequently, the standard errors may be incorrect. We cannot proceed with hypothesis testing until we test for and, if necessary, correct for heteroskedasticity.

B A test for heteroskedasticity is to regress the squared residuals from the estimated regression equation on the independent variables in the regression. As seen in Section 4.1.2, Breusch and Pagan showed that, under the null hypothesis of no conditional heteroskedasticity, $n \times R^2$ (from the regression of the squared residuals on the independent variables from the original regression) will be a χ^2 random variable, with the number of degrees of freedom equal to the number of independent variables in the regression.

C One method to correct for heteroskedasticity is to use robust standard errors. This method uses the parameter estimates from the linear regression model but corrects the standard errors of the estimated parameters to account for the heteroskedasticity. Many statistical software packages can easily compute robust standard errors.

10 The test statistic is nR^2, where n is the number of observations and R^2 is the R^2 of the regression of squared residuals. So, the test statistic is 52 × 0.141 = 7.332. Under the null hypothesis of no conditional heteroskedasticity, this test statistic is a χ^2 random variable. There are three degrees of freedom, the number of independent variables in the regression. Appendix C, at the end of this volume, shows that for a one-tailed test, the test statistic critical value for a variable from a χ^2 distribution with 3 degrees of freedom at the 0.05 significance level is 7.815. The test statistic from the Breusch–Pagan test is 7.332. So, we cannot reject the hypothesis of no conditional heteroskedasticity at the 0.05 level. Therefore, we do not need to correct for conditional heteroskedasticity.

11 A The test statistic is nR^2, where n is the number of observations and R^2 is the R^2 of the regression of squared residuals. So, the test statistic is 750 × 0.006 = 4.5. Under the null hypothesis of no conditional heteroskedasticity, this test statistic is a χ^2 random variable. Because the regression has only one independent variable, the number of degrees of freedom is equal to 1. Appendix C, at the end of this volume, shows that for a one-tailed test, the test statistic critical value for a variable from a χ^2 distribution with one degree of freedom at the 0.05 significance level is 3.841. The test statistic is 4.5. So, we can reject the hypothesis of no conditional heteroskedasticity at the 0.05 level. Therefore, we need to correct for conditional heteroskedasticity.

B Two different methods can be used to correct for the effects of conditional heteroskedasticity in linear regression models. The first method involves computing robust standard errors. This method corrects the standard errors of the linear regression model's estimated parameters to account for the conditional heteroskedasticity. The second method is generalized least squares. This method modifies the original equation in an attempt to eliminate the heteroskedasticity. The new, modified regression equation is then estimated under the assumption that heteroskedasticity is no longer a problem.

Many statistical software packages can easily compute robust standard errors (the first method), and we recommend using them.

12 A Because the value of the Durbin–Watson statistic is less than 2, we can say that the regression residuals are positively correlated. Because this statistic is fairly close to 2, however, we cannot say without a statistical test if the serial correlation is statistically significant.

B From January 1987 through December 2002, there are 16 years, or 16 × 12 = 192 monthly returns. Thus the sample analyzed is quite large. Therefore, the Durbin–Watson statistic is approximately equal to $2(1 - r)$, where r is the sample correlation between the regression residuals from one period and those from the previous period.

$$\text{DW} = 1.8953 \approx 2(1 - r)$$

So, $r \approx 1 - \text{DW}/2 = 1 - 1.8953/2 = 0.0524$. Consistent with our answer to Part A, the correlation coefficient is positive.

C Appendix E indicates that the critical values d_l and d_u for 100 observations when there is one independent variable are 1.65 and 1.69, respectively. Based on the information given in the problem, the critical values d_l and d_u for about 200 observations when there is one independent variable are about 1.74 and 1.78, respectively. Because the DW statistic of 1.8953 for our regression is above d_u, we fail to reject the null hypothesis of no positive serial correlation. Therefore, we conclude that there is no evidence of positive serial correlation for the error term.

13 A This problem is known as multicollinearity. When some linear combinations of the independent variables in a regression model are highly correlated, the standard errors of the independent coefficient estimates become quite large, even though the regression equation may fit rather well.

B The choice of independent variables presents multicollinearity concerns because market value of equity appears in both variables.

C The classic symptom of multicollinearity is a high R^2 (and significant F-statistic) even though the t-statistics on the estimated slope coefficients are insignificant. Here a significant F-statistic does not accompany the insignificant t-statistics, so the classic symptom is not present.

14 A To test the null hypothesis that all of the regression coefficients except for the intercept in the multiple regression model are equal to 0 (H_0: $b_1 = b_2 = b_3 = 0$) against the alternative hypothesis that at least one slope coefficient is not equal to 0, we must use an F-test.

$$F = \frac{\text{RSS}/k}{\text{SSE}/\left[n - (k + 1)\right]} = \frac{0.1720/3}{0.8947/\left[156 - (3 + 1)\right]} = 9.7403$$

The F-statistic has degrees of freedom $F\{k, [n - (k + 1)]\} = F(3, 152)$. From the F-test table, the critical value for $F(3, 120) = 2.68$ and $F(3, 152)$ will be less than $F(3, 120)$, so we can reject at the 0.05 significance level the null hypothesis that the slope coefficients are all 0. Changes in the three independent variables are jointly statistically related to returns.

B None of the t-statistics are significant, but the F-statistic is significant. This suggests the possibility of multicollinearity in the independent variables.

C The apparent multicollinearity is very likely related to the inclusion of *both* the returns on the S&P 500 Index *and* the returns on a value-weighted index of all the companies listed on the NYSE, AMEX, and NASDAQ as independent variables. The value-weighting of the latter index, giving relatively high weights to larger companies such as those included in the S&P 500, may make one return series an approximate linear function of the other. By dropping one or the other of these two variables, we might expect to eliminate the multicollinearity.

15 A Your colleague is indicating that you have omitted an important variable from the regression. This problem is called the omitted variable bias. If the omitted variable is correlated with an included variable, the estimated values of the regression coefficients would be biased and inconsistent. Moreover, the estimates of standard errors of those coefficients would also be inconsistent. So, we cannot use either the coefficient estimates or the estimates of their standard errors to perform statistical tests.

B A comparison of the new estimates with the original estimates clearly indicates that the original model suffered from the omitted variable bias due to the exclusion of company size from that model. As the *t*-statistics of the new model indicate, company size is statistically significant. Further, for the debt-to-equity ratio, the absolute value of the estimated coefficient substantially increases from 0.1043 to 0.1829, while its standard error declines. Consequently, it becomes significant in the new model, in contrast to the original model, in which it is not significant at the 5 percent level. The value of the estimated coefficient of the S&P 500 dummy substantially declines from 1.2222 to 0.4218. These changes imply that size should be included in the model.

16 A You need to use a qualitative dependent variable. You could give a value of 1 to this dummy variable for a listing in the United States and a value of 0 for not listing in the United States.

B Because you are using a qualitative dependent variable, linear regression is not the right technique to estimate the model. One possibility is to use either a probit or a logit model. Both models are identical, except that the logit model is based on logistic distribution while the probit model is based on normal distribution. Another possibility is to use discriminant analysis.

17 C is correct. The predicted initial return (IR) is:

$$\begin{aligned} IR &= 0.0477 + (0.0150 \times 6) + (0.435 \times 0.04) - (0.0009 \times 40) + (0.05 \times 0.70) \\ &= 0.1541 \end{aligned}$$

18 B is correct. The 95% confidence interval is 0.435 ± (0.0202 × 1.96) = (0.395, 0.475).

19 C is correct. To test Hansen's belief about the direction and magnitude of the initial return, the test should be a one-tailed test. The alternative hypothesis is H_1: $b_j < 0.5$, and the null hypothesis is H_0: $b_j \geq 0.5$. The correct test statistic is: $t = (0.435 - 0.50)/0.0202 = -3.22$, and the critical value of the *t*-statistic for a one-tailed test at the 0.05 level is −1.645. The test statistic is significant, and the null hypothesis can be rejected at the 0.05 level of significance.

20 C is correct. The multiple *R*-squared for the regression is 0.36; thus, the model explains 36 percent of the variation in the dependent variable. The correlation between the predicted and actual values of the dependent variable is the square root of the *R*-squared or $\sqrt{0.36} = 0.60$.

21 A is correct. Chang is correct because the presence of conditional heteroskedasticity results in consistent parameter estimates, but biased (up or down) standard errors, *t*-statistics, and *F*-statistics.

22 A is correct. Chang is correct because a correlated omitted variable will result in biased and inconsistent parameter estimates and inconsistent standard errors.

23 B is correct.

The *F*-test is used to determine if the regression model as a whole is significant.

F = Mean square regression (MSR) ÷ Mean squared error (MSE)

MSE = SSE/[n − (k + 1)] = 19,048 ÷ 427 = 44.60

MSR = SSR/*k* = 1071 ÷ 3 = 357

F = 357 ÷ 44.60 = 8.004

The critical value for degrees of freedom of 3 and 427 with α = 0.05 (one-tail) is F = 2.63 from Exhibit 5. The calculated F is greater than the critical value, and Chiesa should reject the null hypothesis that all regression coefficients are equal to zero.

24 B is correct. The Durbin–Watson test used to test for serial correlation in the error term, and its value reported in Exhibit 1 is 1.65. For no serial correlation, DW is approximately equal to 2. If DW < d_l, *the error terms are positively serially correlated.* Because the DW = 1.65 is less than d_l = 1.827 for n = 431 (see Exhibit 2), Chiesa should reject the null hypothesis of no serial correlation and conclude that there is evidence of positive serial correlation among the error terms.

25 B is correct. The coefficient for the Pres party dummy variable (3.17) represents the increment in the mean value of the dependent variable related to the Democratic Party holding the presidency. In this case, the excess stock market return is 3.17 percent greater in Democratic presidencies than in Republican presidencies.

26 B is correct. The confidence interval is computed as $a_1 \pm s(a_1) \times t(95\%,\infty)$. From Exhibit 1, a_1 = 3.04 and $t(a_1)$ = 4.52, resulting in a standard error of $a_1 = s(a_1)$ = 3.04/4.52 = 0.673. The critical value for t from Exhibit 3 is 1.96 for p = 0.025. The confidence interval for a_1 is 3.04 ± 0.673 × 1.96 = 3.04 ± 1.31908 or from 1.72092 to 4.35908.

27 C is correct. The default spread is typically larger when business conditions are poor, i.e., a greater probability of default by the borrower. The positive sign for default spread (see Exhibit 1) indicates that expected returns are positively related to default spreads, meaning that excess returns are greater when business conditions are poor.

28 C is correct. Predictions in a multiple regression model are subject to both parameter estimate uncertainty and regression model uncertainty.

29 C is correct. The F-statistic is

$$F = \frac{\text{RSS}/k}{\text{SSE}/\left[n - (k+1)\right]} = \frac{714.169/8}{1583.113/546} = \frac{89.2712}{2.8995} = 30.79$$

Because F = 30.79 exceeds the critical F of 1.96, the null hypothesis that the regression coefficients are all 0 is rejected at the 0.05 significance level.

30 B is correct. The estimated coefficients for the dummy variables show the estimated difference between the returns on different types of funds. The growth dummy takes the value of 1 for growth funds and 0 for the value fund. Exhibit 1 shows a growth dummy coefficient of 2.4368. The estimated difference between the return of growth funds and value funds is thus 2.4368.

31 B is correct. The R^2 is expected to increase, not decline, with a new independent variable. The other two potential consequences Honoré describes are correct.

32 C is correct. Using dummy variables to distinguish among n categories would best capture the ability of the Morningstar rating system to predict mutual fund performance. We need n − 1 dummy variables to distinguish among n categories. In this case, there are five possible ratings and we need four dummy variables. Adding an independent variable that has a value equal to the number of stars in the rating of each fund is not appropriate because if the coefficient for this variable is positive, this method assumes that the extra return for a two-star fund is twice that of a one-star fund, the extra return for a three-star fund is three times that of a one-star fund, and so forth, which is not a reasonable assumption.

33 A is correct. Heteroskedasticity causes the *F*-test for the overall significance of the regression to be unreliable. It also causes the *t*-tests for the significance of individual regression coefficients to be unreliable because heteroskedasticity introduces bias into estimators of the standard error of regression coefficients.

34 A is correct. The model in Exhibit 2 does not have a lagged dependent variable. Positive serial correlation will, for such a model, not affect the consistency of the estimated coefficients. Thus, the coefficients will not need to be corrected for serial correlation. Positive serial correlation will, however, cause the standard errors of the regression coefficients to be understated; thus, the corresponding *t*-statistics will be inflated.

35 A is correct. The critical Durbin–Watson (D–W) values are $d_l = 1.63$ and $d_u = 1.72$. Because the estimated D–W value of 1.81 is greater than $d_u = 1.73$ (and less than 2), she fails to reject the null hypothesis of no serial correlation.

36 B is correct. Probit and logit models are used for models with qualitative dependent variables, such as models in which the dependent variable can have one of two discreet outcomes (i.e., 0 or 1). The analysis in the two exhibits are explaining security returns, which are continuous (not 0 or 1) variables.

37 A is correct. Varden expects to find that CEO tenure is positively related to the firm's ROE. If he is correct, the regression coefficient for tenure, b_2, will be greater than zero ($b_2 > 0$) and statistically significant. The null hypothesis supposes that the "suspected" condition is not true, so the null hypothesis should state the variable is less than or equal to zero. The *t*-statistic for tenure is 2.308, significant at the 0.027 level, meeting Varden's 0.05 significance requirement. Varden should reject the null hypothesis.

38 C is correct. The *t*-statistic for tenure is 2.308, indicating significance at the 0.027 level but not the 0.01 level. The *t*-statistic for ESG is 1.201, with a *p*-value of 0.238, which means we fail to reject the null hypothesis for ESG at the 0.01 significance level.

39 B is correct. The *t*-statistic for tenure is 2.308, which is significant at the 0.027 level. The *t*-statistic for ESG is 1.201, with a *p*-value of 0.238. This result is not significant at the 0.05 level.

40 C is correct. The regression equation is as follows:

$$\hat{Y}_i = 9.442 + 0.069X_{1i} + 0.681X_{2i}$$

$$\begin{aligned}\text{ROE} &= 9.442 + 0.069(\text{ESG}) + 0.681(\text{Tenure})\\ &= 9.442 + 0.069(55) + 0.681(10.5)\\ &= 9.442 + 3.795 + 7.151\\ &= 20.388.\end{aligned}$$

41 B is correct. When you add an additional independent variable to the regression model, the amount of unexplained variance will decrease, provided the new variable explains any of the previously unexplained variation. This result occurs as long as the new variable is even slightly correlated with the dependent variable. Exhibit 2 indicates the dividend growth rate is correlated with the dependent variable, ROE. Therefore, R^2 will increase.

Adjusted R^2, however, may not increase and may even decrease if the relationship is weak. This result occurs because in the formula for adjusted R^2, the new variable increases k (the number of independent variables) in the denominator, and the increase in R^2 may be insufficient to increase the value of the formula.

$$\text{adjusted } R^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\left(1 - R^2\right)$$

42 C is correct. Exhibit 1 indicates that the F-statistic of 4.161 is significant at the 0.05 level. A significant F-statistic means at least one of the independent variables is significant.

43 C is correct. In a multiple linear regression (as compared with simple regression), R^2 is less appropriate as a measure of whether a regression model fits the data well. A high adjusted R^2 does not necessarily indicate that the regression is well specified in the sense of including the correct set of variables. The F-test is an appropriate test of a regression's overall significance in either simple or multiple regressions.

44 C is correct. Multiple linear regression assumes that the relationship between the dependent variable and each of the independent variables is linear. Varden believes that this is not true for dividend growth because he believes the relationship may be different in firms with a long-standing CEO. Multiple linear regression also assumes that the independent variables are not random. Varden states that he believes CEO tenure is a random variable.

45 B is correct. If we use adjusted R^2 to compare regression models, it is important that the dependent variable be defined the same way in both models and that the sample sizes used to estimate the models are the same. Varden's first model was based on 40 observations, whereas the second model was based on 500.

READING

3

Time-Series Analysis

by Richard A. DeFusco, PhD, CFA, Dennis W. McLeavey, DBA, CFA, Jerald E. Pinto, PhD, CFA, and David E. Runkle, PhD, CFA

Richard A. DeFusco, PhD, CFA, is at the University of Nebraska-Lincoln (USA). Dennis W. McLeavey, DBA, CFA, is at the University of Rhode Island (USA). Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). David E. Runkle, PhD, CFA, is at Jacobs Levy Equity Management (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** calculate and evaluate the predicted trend value for a time series, modeled as either a linear trend or a log-linear trend, given the estimated trend coefficients;
☐	**b.** describe factors that determine whether a linear or a log-linear trend should be used with a particular time series and evaluate limitations of trend models;
☐	**c.** explain the requirement for a time series to be covariance stationary and describe the significance of a series that is not stationary;
☐	**d.** describe the structure of an autoregressive (AR) model of order *p* and calculate one- and two-period-ahead forecasts given the estimated coefficients;
☐	**e.** explain how autocorrelations of the residuals can be used to test whether the autoregressive model fits the time series;
☐	**f.** explain mean reversion and calculate a mean-reverting level;
☐	**g.** contrast in-sample and out-of-sample forecasts and compare the forecasting accuracy of different time-series models based on the root mean squared error criterion;
☐	**h.** explain the instability of coefficients of time-series models;
☐	**i.** describe characteristics of random walk processes and contrast them to covariance stationary processes;
☐	**j.** describe implications of unit roots for time-series analysis, explain when unit roots are likely to occur and how to test for them, and demonstrate how a time series with a unit root can be transformed so it can be analyzed with an AR model;

(continued)

© 2015 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**k.** describe the steps of the unit root test for nonstationarity and explain the relation of the test to autoregressive time-series models;
☐	**l.** explain how to test and correct for seasonality in a time-series model and calculate and interpret a forecasted value using an AR model with a seasonal lag;
☐	**m.** explain autoregressive conditional heteroskedasticity (ARCH) and describe how ARCH models can be applied to predict the variance of a time series;
☐	**n.** explain how time-series variables should be analyzed for nonstationarity and/or cointegration before use in a linear regression; and
☐	**o.** determine an appropriate time-series model to analyze a given investment problem and justify that choice.

1 INTRODUCTION TO TIME-SERIES ANALYSIS AND CHALLENGES OF WORKING WITH TIME SERIES

As financial analysts, we often use time-series data to make investment decisions. A **time series** is a set of observations on a variable's outcomes in different time periods: the quarterly sales for a particular company during the past five years, for example, or the daily returns on a traded security. In this reading, we explore the two chief uses of time-series models: to explain the past and to predict the future of a time series. We also discuss how to estimate time-series models, and we examine how a model describing a particular time series can change over time. The following two examples illustrate the kinds of questions we might want to ask about time series.

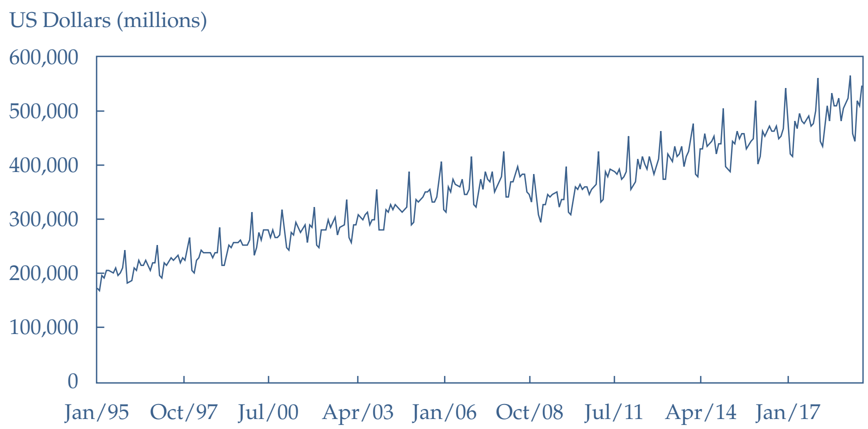
Suppose it is the beginning of 2020 and we are managing a US-based investment portfolio that includes Swiss stocks. Because the value of this portfolio would decrease if the Swiss franc depreciates with respect to the dollar, and vice versa, holding all else constant, we are considering whether to hedge the portfolio's exposure to changes in the value of the franc. To help us in making this decision, we decide to model the time series of the franc/dollar exchange rate. Exhibit 1 shows monthly data on the franc/dollar exchange rate. The data are monthly averages of daily exchange rates. Has the exchange rate been more stable since 1987 than it was in previous years? Has the exchange rate shown a long-term trend? How can we best use past exchange rates to predict future exchange rates?

Exhibit 1 Swiss Franc/US Dollar Exchange Rate, Monthly Average of Daily Data

Source: Board of Governors of the Federal Reserve System.

As another example, suppose it is the beginning of 2020. We cover retail stores for a sell-side firm and want to predict retail sales for the coming year. Exhibit 2 shows monthly data on US retail sales. The data are not seasonally adjusted, hence the spikes around the holiday season at the turn of each year. Because the reported sales in the stores' financial statements are not seasonally adjusted, we model seasonally unadjusted retail sales. How can we model the trend in retail sales? How can we adjust for the extreme seasonality reflected in the peaks and troughs occurring at regular intervals? How can we best use past retail sales to predict future retail sales?

Exhibit 2 Monthly US Retail Sales



Source: US Department of Commerce, Census Bureau.

Some fundamental questions arise in time-series analysis: How do we model trends? How do we predict the future value of a time series based on its past values? How do we model seasonality? How do we choose among time-series models? And how do we model changes in the variance of time series over time? We address each of these issues in this reading.

We first describe typical challenges in applying the linear regression model to time-series data. We present linear and log-linear trend models, which describe, respectively, the value and the natural log of the value of a time series as a linear function of time. We then present autoregressive time-series models—which explain the current value of a time series in terms of one or more lagged values of the series. Such models are among the most commonly used in investments, and the section addresses many related concepts and issues. We then turn our attention to random walks. Because such time series are not covariance stationary, they cannot be modeled using autoregressive models unless they can be transformed into stationary series. We therefore explore appropriate transformations and tests of stationarity. The subsequent sections address moving-average time-series models and discuss the problem of seasonality in time series and how to address it. We also cover autoregressive moving-average models, a more complex alternative to autoregressive models. The last two topics are modeling changing variance of the error term in a time series and the consequences of regression of one time series on another when one or both time series may not be covariance stationary.

1.1 Challenges of Working with Time Series

Throughout the reading, our objective will be to apply linear regression to a given time series. Unfortunately, in working with time series, we often find that the assumptions of the linear regression model are not satisfied. To apply time-series analysis, we need to assure ourselves that the linear regression model assumptions are met. When those assumptions are not satisfied, in many cases we can transform the time series or specify the regression model differently, so that the assumptions of the linear regression model are met.

We can illustrate assumption difficulties in the context of a common time-series model, an autoregressive model. Informally, an autoregressive model is one in which the independent variable is a lagged (that is, past) value of the dependent variable, such as the model $x_t = b_0 + b_1x_{t-1} + \varepsilon_t$ (we could also write the equation as $y_t = b_0 + b_1y_{t-1} + \varepsilon_t$). Specific problems that we often encounter in dealing with time series include the following:

- The residual errors are correlated instead of being uncorrelated. In the calculated regression, the difference between x_t and $b_0 + b_1x_{t-1}$ is called the residual error (ε_t). The linear regression assumes that this error term is not correlated across observations. The violation of that assumption is frequently more critical in terms of its consequences in the case of time-series models involving past values of the time series as independent variables than for other models (such as cross-sectional models) in which the dependent and independent variables are distinct. As we discussed in the reading on multiple regression, in a regression in which the dependent and independent variables are distinct, serial correlation of the errors in this model does not affect the consistency of our estimates of intercept or slope coefficients. By contrast, in an autoregressive time-series regression, such as $x_t = b_0 + b_1 x_{t-1} + \varepsilon_t$, serial correlation in the error term causes estimates of the intercept (b_0) and slope coefficient (b_1) to be inconsistent.

- The mean or variance of the time series changes over time. Regression results are invalid if we estimate an autoregressive model for a time series with mean or variance that changes over time.

Before we try to use time series for forecasting, we may need to transform the time-series model so that it is well specified for linear regression. With this objective in mind, you will observe that time-series analysis is relatively straightforward and logical.

LINEAR TREND MODELS

2

a calculate and evaluate the predicted trend value for a time series, modeled as either a linear trend or a log-linear trend, given the estimated trend coefficients;

Estimating a trend in a time series and using that trend to predict future values of the time series is the simplest method of forecasting. For example, we saw in Exhibit 2 that monthly US retail sales show a long-term pattern of upward movement—that is, a **trend**. In this section, we examine two types of trends—linear trends and log-linear trends—and discuss how to choose between them.

2.1 Linear Trend Models

The simplest type of trend is a **linear trend**, one in which the dependent variable changes at a constant rate with time. If a time series, y_t, has a linear trend, then we can model the series using the following regression equation:

$$y_t = b_0 + b_1 t + \varepsilon_t,\ t = 1, 2, \ldots, T, \qquad (1)$$

where

y_t = the value of the time series at time t (value of the dependent variable)
b_0 = the y-intercept term
b_1 = the slope coefficient
t = time, the independent or explanatory variable
ε_t = a random error term

In Equation 1, the trend line, $b_0 + b_1 t$, predicts the value of the time series at time t (where t takes on a value of 1 in the first period of the sample and increases by 1 in each subsequent period). Because the coefficient b_1 is the slope of the trend line, we refer to b_1 as the trend coefficient. We can estimate the two coefficients, b_0 and b_1, using ordinary least squares, denoting the estimated coefficients as $\hat{b}_0$ and $\hat{b}_1$. Recall that ordinary least squares is an estimation method based on the criterion of minimizing the sum of a regression's squared residuals.

Now we demonstrate how to use these estimates to predict the value of the time series in a particular period. Recall that t takes on a value of 1 in Period 1. Therefore, the predicted or fitted value of y_t in Period 1 is $\hat{y}_1 = \hat{b}_0 + \hat{b}_1(1)$. Similarly, in a subsequent period—say, the sixth period—the fitted value is $\hat{y}_6 = \hat{b}_0 + \hat{b}_1(6)$. Now suppose that we want to predict the value of the time series for a period outside the sample—say, period $T + 1$. The predicted value of y_t for period $T + 1$ is $\hat{y}_{T+1} = \hat{b}_0 + \hat{b}_1(T+1)$. For example, if $\hat{b}_0$ is 5.1 and $\hat{b}_1$ is 2, then at $t = 5$ the predicted value of y_5 is 15.1 and at $t = 6$ the predicted value of y_6 is 17.1. Note that each consecutive observation in this time series increases by $\hat{b}_1 = 2$, irrespective of the level of the series in the previous period.

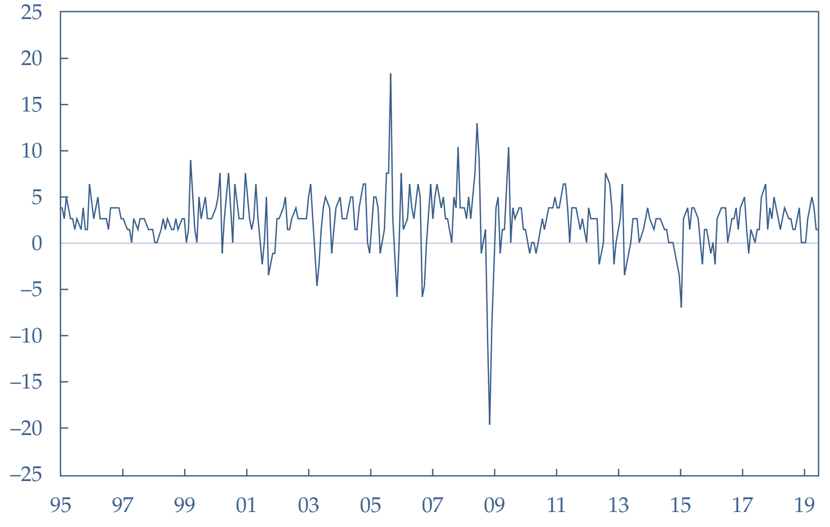
EXAMPLE 1

The Trend in the US Consumer Price Index

It is January 2020. As a fixed-income analyst in the trust department of a bank, Lisette Miller is concerned about the future level of inflation and how it might affect portfolio value. Therefore, she wants to predict future inflation rates. For this purpose, she first needs to estimate the linear trend in inflation. To do so, she uses the monthly US Consumer Price Index (CPI) inflation data, expressed as an annual percentage rate, (1% is represented as 1.0) shown in Exhibit 3. The data include 228 months from January 1995 through June 2019, and the model to be estimated is $y_t = b_0 + b_1 t + \varepsilon_t$, $t = 1, 2, \ldots, 294$. The table in Exhibit 4 shows the results of estimating this equation. With 294 observations and two parameters, this model has 292 degrees of freedom. At the 0.05 significance level, the critical value for a t-statistic is 1.97. The intercept $\left(\hat{b}_0 = 2.7845\right)$ is statistically significant because the value of the t-statistic for the coefficient is well above the critical value. The trend coefficient is negative $\left(\hat{b}_1 = -0.0037\right)$, suggesting a slightly declining trend in inflation during the sample time period. However, the trend is not statistically significant because the absolute value of the t-statistic for the coefficient is below the critical value. The estimated regression equation can be written as

$$y_t = 2.7845 - 0.0037t.$$

Exhibit 3 Monthly CPI Inflation, Not Seasonally Adjusted



Source: Bureau of Labor Statistics.

Exhibit 4 Estimating a Linear Trend in Inflation: Monthly Observations, January 1995–June 2019

Regression Statistics

R^2	0.0099
Standard error	3.1912
Observations	294
Durbin–Watson	1.2145

	Coefficient	Standard Error	*t*-Statistic
Intercept	2.7845	0.3732	7.4611
t (Trend)	−0.0037	0.0022	−1.68

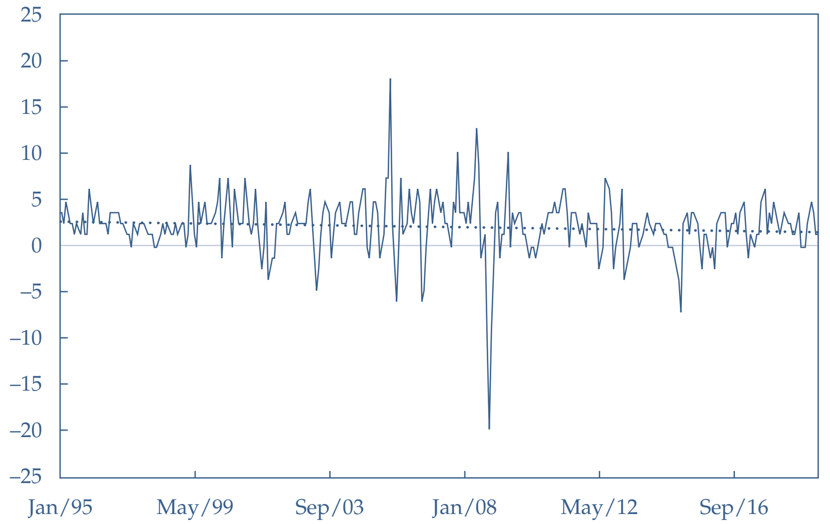
Source: US Bureau of Labor Statistics.

Because the trend line slope is estimated to be −0.0037, Miller concludes that the linear trend model's best estimate is that the annualized rate of inflation declined at a rate of about 37 bps per month during the sample time period. The decline is not statistically significantly different from zero.

In January 1995, the first month of the sample, the predicted value of inflation is $\hat{y}_1 = 2.7845 - 0.0037(1) = 2.7808\%$. In June 2019, the 294th, or last, month of the sample, the predicted value of inflation is $\hat{y}_{228} = 2.7845 - 0.0037(294) = 1.697\%$. Note, though, that these predicted values are for in-sample periods. A comparison of these values with the actual values indicates how well Miller's model fits the data; however, a main purpose of the estimated model is to predict the level of inflation for out-of-sample periods. For example, for June 2020 (12 months after the end of the sample), $t = 294 + 12 = 306$, and the predicted level of inflation is $\hat{y}_{306} = 2.7845 - 0.0037(306) = 1.6523\%$.

Exhibit 5 shows the inflation data along with the fitted trend. Consistent with the negative but small and statistically insignificant trend coefficient, the fitted trend line is slightly downward sloping. Note that inflation does not appear to be above or below the trend line for a long period of time. No persistent differences exist between the trend and actual inflation. The residuals (actual minus trend values) appear to be unpredictable and uncorrelated in time. Therefore, using a linear trend line to model inflation rates from 1995 through 2019 does not appear to violate the assumptions of the linear regression model. Note also that the R^2 in this model is quite low, indicating great uncertainty in the inflation forecasts from this model. In fact, the estimated model explains only 0.99% of the variation in monthly inflation. Although linear trend models have their uses, they are often inappropriate for economic data. Most economic time series reflect trends with changing slopes and/or intercepts over time. The linear trend model identifies the slope and intercept that provides the best linear fit for all past data. The model's deviation from the actual data can be greatest near the end of a data series, which can compromise forecasting accuracy. Later in this reading, we will examine whether we can build a better model of inflation than a model that uses only a trend line.

Exhibit 5 Monthly CPI Inflation with Trend



Source: US Bureau of Labor Statistics.

3 LOG-LINEAR TREND MODELS

a calculate and evaluate the predicted trend value for a time series, modeled as either a linear trend or a log-linear trend, given the estimated trend coefficients;

b describe factors that determine whether a linear or a log-linear trend should be used with a particular time series and evaluate limitations of trend models;

Sometimes a linear trend does not correctly model the growth of a time series. In those cases, we often find that fitting a linear trend to a time series leads to persistent rather than uncorrelated errors. If the residuals from a linear trend model are persistent, then we need to employ an alternative model satisfying the conditions of linear regression. For financial time series, an important alternative to a linear trend is a log-linear trend. Log-linear trends work well in fitting time series that have exponential growth.

Exponential growth means constant growth at a particular rate. For example, annual growth at a constant rate of 5% is exponential growth. How does exponential growth work? Suppose we describe a time series by the following equation:

$$y_t = e^{b_0 + b_1 t}, \; t = 1, 2, \ldots, T. \tag{2}$$

Exponential growth is growth at a constant rate $\left(e^{b_1} - 1\right)$ with continuous compounding. For instance, consider values of the time series in two consecutive periods. In Period 1, the time series has the value $y_1 = e^{b_0 + b_1(1)}$, and in Period 2, it has the value $y_2 = e^{b_0 + b_1(2)}$. The resulting ratio of the values of the time series in the first two periods is $y_2/y_1 = \left(e^{b_0 + b_1(2)}\right) / \left(e^{b_0 + b_1(1)}\right) = e^{b_1(1)}$. Generally, in any period t, the time series has the value $y_t = e^{b_0 + b_1(t)}$. In period $t + 1$, the time series has the value $y_{t+1} = e^{b_0 + b_1(t+1)}$.

The ratio of the values in the periods $(t + 1)$ and t is $y_{t+1}/y_t = e^{b_0 + b_1(t+1)} \big/ e^{b_0 + b_1(t)} = e^{b_1(1)}$. Thus, the proportional rate of growth in the time series over two consecutive periods is always the same: $(y_{t+1} - y_t)/y_t = y_{t+1}/y_t - 1 = e^{b_1} - 1$. For example, if we use annual periods and $e^{b_1} = 1.04$ for a particular series, then that series grows by $1.04 - 1 = 0.04$, or 4% a year. Therefore, exponential growth is growth at a constant rate. Continuous compounding is a mathematical convenience that allows us to restate the equation in a form that is easy to estimate.

If we take the natural log of both sides of Equation 2, the result is the following equation:

$$\ln y_t = b_0 + b_1 t,\ t = 1, 2, \ldots, T.$$

Therefore, if a time series grows at an exponential rate, we can model the natural log of that series using a linear trend (an exponential growth rate is a compound growth rate with continuous compounding). Of course, no time series grows exactly at a constant rate. Consequently, if we want to use a **log-linear model**, we must estimate the following equation:

$$\ln y_t = b_0 + b_1 t + \varepsilon_t,\ t = 1, 2, \ldots, T. \tag{3}$$

Note that this equation is linear in the coefficients b_0 and b_1. In contrast to a linear trend model, in which the predicted trend value of y_t is $\hat{b}_0 + \hat{b}_1 t$, the predicted trend value of y_t in a log-linear trend model is $e^{\hat{b}_0 + \hat{b}_1 t}$ because $e^{\ln y_t} = y_t$.

Examining Equation 3, we see that a log-linear model predicts that $\ln y_t$ will increase by b_1 from one time period to the next. The model predicts a constant growth rate in y_t of $e^{b_1} - 1$. For example, if $b_1 = 0.05$, then the predicted growth rate of y_t in each period is $e^{0.05} - 1 = 0.051271$, or 5.13%. In contrast, the linear trend model (Equation 1) predicts that y_t grows by a constant amount from one period to the next.

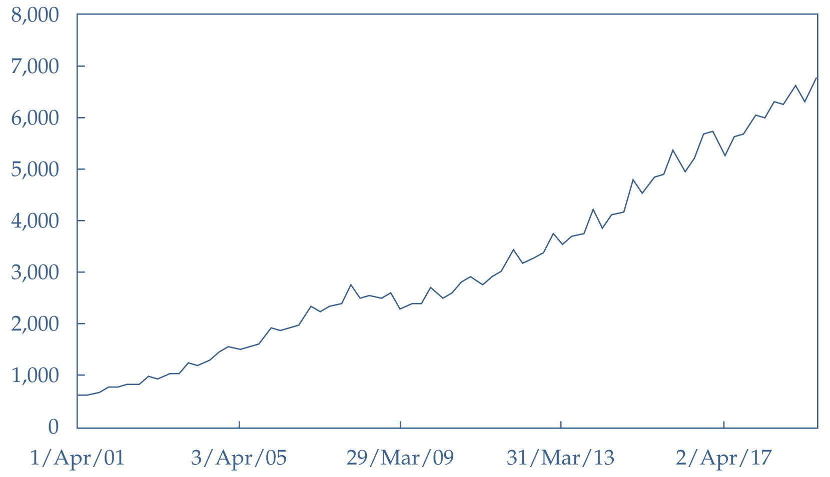
Example 2 illustrates the problem of nonrandom residuals in a linear trend model, and Example 3 shows a log-linear regression fit to the same data.

EXAMPLE 2

A Linear Trend Regression for Quarterly Sales at Starbucks

In September 2019, technology analyst Ray Benedict wants to use Equation 1 to fit the data on quarterly sales for Starbucks Corporation shown in Exhibit 6. Starbucks' fiscal year ends in June. Benedict uses 74 observations on Starbucks' sales from the second quarter of fiscal year 2001 (starting in April 2001) to the third quarter of fiscal year 2019 (ending in June 2019) to estimate the linear trend regression model $y_t = b_0 + b_1 t + \varepsilon_t$, $t = 1, 2, \ldots, 74$. Exhibit 7 shows the results of estimating this equation.

Exhibit 6 Starbucks Quarterly Sales by Fiscal Year



Source: Bloomberg.

Exhibit 7 Estimating a Linear Trend in Starbucks Sales

Regression Statistics

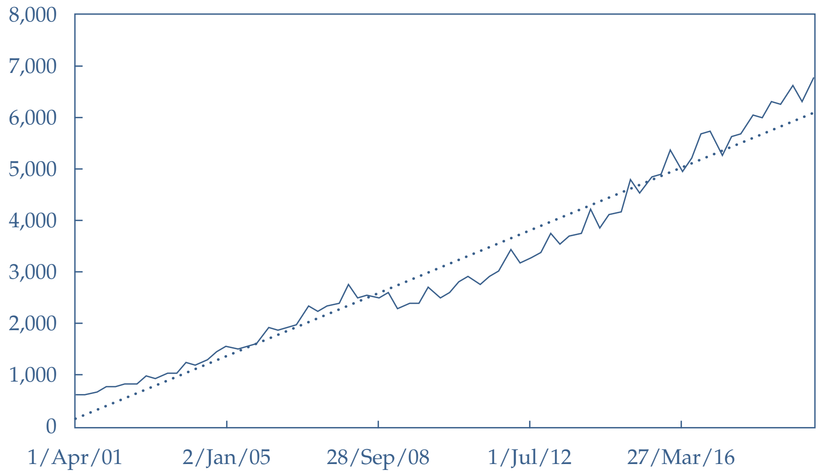
R^2	0.9603
Standard error	353.36
Observations	74
Durbin–Watson	0.40

	Coefficient	Standard Error	*t*-Statistic
Intercept	137.4213	82.99	1.6559
t (Trend)	80.2060	1.9231	41.7066

Source: Bloomberg.

At first glance, the results shown in Exhibit 7 seem quite reasonable: The trend coefficient is highly statistically significant. When Benedict plots the data on Starbucks' sales and the trend line, however, he sees a different picture. As Exhibit 8 shows, before 2008 the trend line is persistently below sales. Subsequently, until 2015, the trend line is persistently above sales and then varies somewhat thereafter.

Exhibit 8 Starbucks Quarterly Sales with Trend

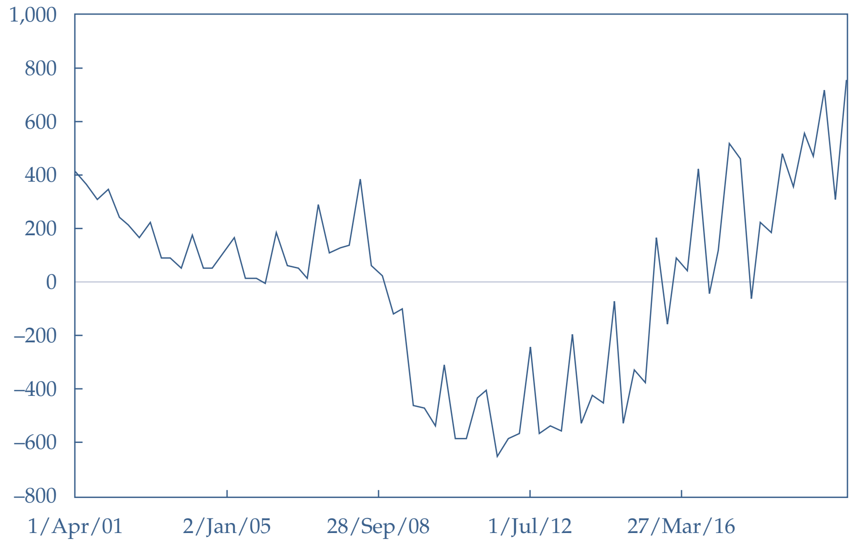


Source: Bloomberg.

Recall a key assumption underlying the regression model: that the regression errors are not correlated across observations. If a trend is persistently above or below the value of the time series, however, the residuals (the difference between the time series and the trend) are serially correlated. Exhibit 9 shows the residuals (the difference between sales and the trend) from estimating a linear trend model with the raw sales data. The figure shows that the residuals are persistent: They are consistently negative from 2008 to 2015 and consistently positive from 2001 to 2008 and from 2017 to 2019.

Because of this persistent serial correlation in the errors of the trend model, using a linear trend to fit sales at Starbucks would be inappropriate, even though the R^2 of the equation is high (0.96). The assumption of uncorrelated residual errors has been violated. Because the dependent and independent variables are not distinct, as in cross-sectional regressions, this assumption violation is serious and causes us to search for a better model.

Exhibit 9 Residual from Predicting Starbucks Sales with a Trend



Source: Bloomberg.

EXAMPLE 3

A Log-Linear Regression for Quarterly Sales at Starbucks

Having rejected a linear trend model in Example 2, technology analyst Benedict now tries a different model for the quarterly sales for Starbucks Corporation from the second quarter of 2001 to the third quarter of 2019. The curvature in the data plot shown in Exhibit 6 provides a hint that an exponential curve may fit the data. Consequently, he estimates the following linear equation:

$$\ln y_t = b_0 + b_1 t + \varepsilon_t,\ t = 1, 2, \ldots, 74.$$

This equation seems to fit the sales data well. As Exhibit 10 shows, the R^2 for this equation is 0.95. An R^2 of 0.95 means that 95% of the variation in the natural log of Starbucks' sales is explained solely by a linear trend.

Exhibit 10 Estimating a Linear Trend in Lognormal Starbucks Sales

Regression Statistics	
R^2	0.9771
Standard error	0.1393
Observations	74
Durbin–Watson	0.26

Exhibit 10 (Continued)

	Coefficient	Standard Error	*t*-Statistic
Intercept	6.7617	0.0327	206.80
t (Trend)	0.0295	0.0008	36.875

Source: Compustat.

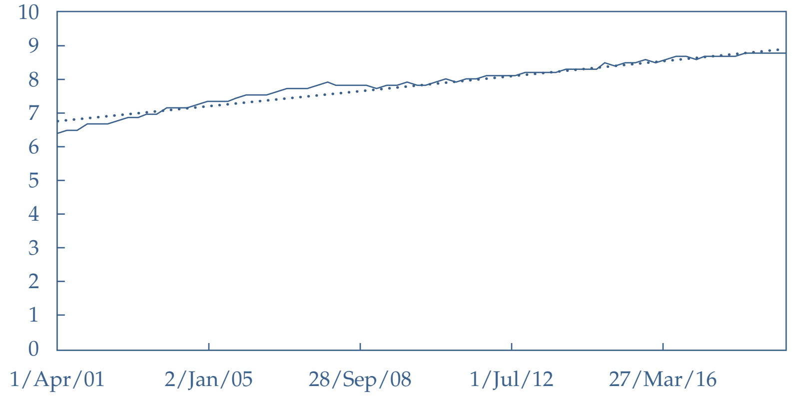
Although both Equations 1 and 3 have a high R^2, Exhibit 11 shows how well a linear trend fits the natural log of Starbucks' sales (Equation 3). The natural logs of the sales data lie very close to the linear trend during the sample period, and log sales are not substantially above or below the trend for long periods of time. Thus, a log-linear trend model seems better suited for modeling Starbucks' sales than a linear trend model is.

1 Benedict wants to use the results of estimating Equation 3 to predict Starbucks' sales in the future. What is the predicted value of Starbucks' sales for the fourth quarter of 2019?

Solution to 1:

The estimated value $\hat{b}_0$ is 6.7617, and the estimated value $\hat{b}_1$ is 0.0295. Therefore, for fourth quarter of 2019 (t = 75), the estimated model predicts that $\ln \hat{y}_{75}$ = 6.7617 + 0.0295(75) = 8.9742 and that sales will be $\hat{y} = e^{\ln \hat{y}_{75}} = e^{8.9742}$ = \$7,896.7 million. Note that a $\hat{b}_1$ of 0.0295 implies that the exponential growth rate per quarter in Starbucks' sales will be 2.99475% ($e^{0.0464} - 1 = 0.0299475$).

Exhibit 11 Natural Log of Starbucks Quarterly Sales



Source: Compustat.

2 How much different is the previous forecast from the prediction of the linear trend model?

Solution to 2:

Exhibit 7 showed that for the linear trend model, the estimated value of $\hat{b}_0$ is 137.4213 and the estimated value of $\hat{b}_1$ is 80.2060. Thus, if we predict Starbucks' sales for the fourth quarter of 2019 (t = 75) using the linear trend model, the forecast is $\hat{y}_{75}$ = 137.4213 + 80.2060(75) = \$6,152.87 million. This forecast is far below the prediction made by the log-linear regression model. Later we will examine whether we can build a better model of Starbucks' quarterly sales than a model that uses only a log-linear trend.

4 TREND MODELS AND TESTING FOR CORRELATED ERRORS

b describe factors that determine whether a linear or a log-linear trend should be used with a particular time series and evaluate limitations of trend models;

Both the linear trend model and the log-linear trend model are single-variable regression models. If they are to be correctly specified, the regression model assumptions must be satisfied. In particular, the regression error for one period must be uncorrelated with the regression error for all other periods. In Example 2 in the previous section, we could infer an obvious violation of that assumption from a visual inspection of a plot of residuals (Exhibit 9). The log-linear trend model of Example 3 appeared to fit the data much better, but we still need to confirm that the uncorrelated errors assumption is satisfied. To address that question formally, we must carry out a Durbin–Watson test on the residuals.

Logical Ordering of Time-Series Observations

In contrast to cross-sectional observations, time-series observations have a logical ordering. They must be processed in chronological order of the time periods involved. For example, we should not make a prediction of the inflation rate using a CPI series in which the order of the observations had been scrambled, because time patterns such as growth in the independent variables can negatively affect the statistical properties of the estimated regression coefficients.

In the reading on regression analysis, we showed how to test whether regression errors are serially correlated using the Durbin–Watson statistic. For example, if the trend models shown in Examples 1 and 3 really capture the time-series behavior of inflation and the log of Starbucks' sales, then the Durbin–Watson statistic for both of those models should not differ significantly from 2.0. Otherwise, the errors in the model are either positively or negatively serially correlated, and that correlation can be used to build a better forecasting model for those time series.

In Example 1, estimating a linear trend in the monthly CPI inflation yielded a Durbin–Watson statistic of 1.09. Is this result significantly different from 2.0? To find out, we need to test the null hypothesis of no positive serial correlation. For a sample with 228 observations and one independent variable, the critical value, d_l, for the Durbin–Watson test statistic at the 0.05 significance level is above 1.77. Because the value of the Durbin–Watson statistic (1.09) is below this critical value, we can reject the hypothesis of no positive serial correlation in the errors. (Remember that significantly small values of the Durbin–Watson statistic indicate positive serial correlation; significantly large values point to negative serial correlation; here the Durbin–Watson statistic of 1.09 indicates positive serial correlation.) We can conclude that a regression equation that uses a linear trend to model inflation has positive serial correlation in the errors. We will need a different kind of regression model because this one violates the least squares assumption of no serial correlation in the errors.

In Example 3, estimating a linear trend with the natural logarithm of sales for the Starbucks example yielded a Durbin–Watson statistic of 0.12. Suppose we wish to test the null hypothesis of no positive serial correlation. The critical value, d_l, is above 1.60 at the 0.05 significance level. The value of the Durbin–Watson statistic (0.12) is below this critical value, so we can reject the null hypothesis of no positive serial correlation in the errors. We can conclude that a regression equation that uses a trend to model the log of Starbucks' quarterly sales has positive serial correlation in the errors. So, for this series as well, we need to build a different kind of model.

Overall, we conclude that the trend models sometimes have the limitation that errors are serially correlated. Existence of serial correlation suggests that we can build better forecasting models for such time series than trend models.

AUTOREGRESSIVE (AR) TIME-SERIES MODELS AND COVARIANCE-STATIONARY SERIES

5

c explain the requirement for a time series to be covariance stationary and describe the significance of a series that is not stationary;

A key feature of the log-linear model's depiction of time series, and a key feature of time series in general, is that current-period values are related to previous-period values. For example, Starbucks' sales for the current period are related to its sales in the previous period. An **autoregressive model (AR)**, a time series regressed on its own past values, represents this relationship effectively. When we use this model, we can drop the normal notation of y as the dependent variable and x as the independent variable because we no longer have that distinction to make. Here we simply use x_t. For example, Equation 4 shows a first-order autoregression, AR(1), for the variable x_t:

$$x_t = b_0 + b_1 x_{t-1} + \varepsilon_t. \tag{4}$$

Thus, in an AR(1) model, we use only the most recent past value of x_t to predict the current value of x_t. In general, a pth-order autoregression, AR(p), for the variable x_t is shown by

$$x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + b_p x_{t-p} + \varepsilon_t. \tag{5}$$

In this equation, p past values of x_t are used to predict the current value of x_t. In the next section, we discuss a key assumption of time-series models that include lagged values of the dependent variable as independent variables.

5.1 Covariance-Stationary Series

Note that the independent variable (x_{t-1}) in Equation 4 is a random variable. This fact may seem like a mathematical subtlety, but it is not. If we use ordinary least squares to estimate Equation 4 when we have a randomly distributed independent variable that is a lagged value of the dependent variable, our statistical inference may be invalid. To make a valid statistical inference, we must make a key assumption in time-series analysis: We must assume that the time series we are modeling is **covariance stationary**.[1]

What does it mean for a time series to be covariance stationary? The basic idea is that a time series is covariance stationary if its properties, such as mean and variance, do not change over time. A covariance stationary series must satisfy three principal requirements. First, the expected value of the time series must be constant and finite in all periods: $E(y_t) = \mu$ and $|\mu| < \infty$, $t = 1, 2, \ldots, T$ (for this first requirement, we use the absolute value to rule out the case in which the mean is negative without limit—i.e., minus infinity). Second, the variance of the time series must be constant and finite in all periods. Third, the covariance of the time series with itself for a fixed number of periods in the past or future must be constant and finite in all periods. The second and third requirements can be summarized as follows:

$$\text{cov}(y_t, y_{t-s}) = \lambda_s;\ |\lambda_s| < \infty;\ t = 1, 2, \ldots, T;\ s = 0, \pm 1, \pm 2, \ldots, \pm T,$$

where λ signifies a constant. (Note that when s in this equation equals 0, this equation imposes the condition that the variance of the time series is finite, because the covariance of a random variable with itself is its variance: $\text{cov}(y_t, y_t) = \text{var}(y_t)$.) What happens if a time series is not covariance stationary but we model it using Equation 4? The estimation results will have no economic meaning. For a non-covariance-stationary time series, estimating the regression in Equation 4 will yield spurious results. In particular, the estimate of b_1 will be biased, and any hypothesis tests will be invalid.

How can we tell if a time series is covariance stationary? We can often answer this question by looking at a plot of the time series. If the plot shows roughly the same mean and variance over time without any significant seasonality, then we may want to assume that the time series is covariance stationary.

Some of the time series we looked at in the exhibits appear to be covariance stationary. For example, the inflation data shown in Exhibit 3 appear to have roughly the same mean and variance over the sample period. Many of the time series one encounters in business and investments, however, are not covariance stationary. For example, many time series appear to grow (or decline) steadily over time and thus have a mean that is nonconstant, which implies that they are nonstationary. As an example, the time series of quarterly sales in Exhibit 8 clearly shows the mean increasing as time passes. Thus, Starbucks' quarterly sales are not covariance stationary (in general, any time series accurately described with a linear or log-linear trend model is not covariance stationary, although a transformation of the original series might be covariance stationary). Macroeconomic time series such as those relating to income and consumption are often strongly trending as well. A time series with seasonality (regular patterns of movement with the year) also has a nonconstant mean, as do other types of time series that we discuss later (in particular, random walks are not covariance stationary).

Exhibit 2 showed that monthly retail sales (not seasonally adjusted) are also not covariance stationary. Sales in December are always much higher than sales in other months (these are the regular large peaks), and sales in January are always much lower (these are the regular large drops after the December peaks). On average, sales also increase over time, so the mean of sales is not constant.

Later we will show that we can often transform a nonstationary time series into a stationary time series. But whether a stationary time series is original or transformed, a warning is necessary: Stationarity in the past does not guarantee stationarity in the

1 "Weakly stationary" is a synonym for covariance stationary. Note that the terms "stationary" and "stationarity" are often used to mean "covariance stationary" or "covariance stationarity," respectively. You may also encounter the more restrictive concept of "strictly" stationary, which has little practical application. For details, see Diebold (2008).

future. There is always the possibility that a well-specified model will fail when the state of the world changes and yields a different underlying model that generates the time series.

DETECTING SERIALLY CORRELATED ERRORS IN AN AUTOREGRESSIVE MODEL

6

d describe the structure of an autoregressive (AR) model of order p and calculate one- and two-period-ahead forecasts given the estimated coefficients;

e explain how autocorrelations of the residuals can be used to test whether the autoregressive model fits the time series;

We can estimate an autoregressive model using ordinary least squares if the time series is covariance stationary and the errors are uncorrelated. Unfortunately, our previous test for serial correlation, the Durbin–Watson statistic, is invalid when the independent variables include past values of the dependent variable. Therefore, for most time-series models, we cannot use the Durbin–Watson statistic. Fortunately, we can use other tests to determine whether the errors in a time-series model are serially correlated. One such test reveals whether the autocorrelations of the error term are significantly different from 0. This test is a t-test involving a residual autocorrelation and the standard error of the residual autocorrelation. As background for the test, we next discuss autocorrelation in general before moving to residual autocorrelation.

The **autocorrelations** of a time series are the correlations of that series with its own past values. The order of the correlation is given by k, where k represents the number of periods lagged. When $k = 1$, the autocorrelation shows the correlation of the variable in one period with its occurrence in the previous period. For example, the ***k*th-order autocorrelation** (ρ_k) is

$$\rho_k = \frac{\operatorname{cov}(x_t, x_{t-k})}{\sigma_x^2} = \frac{E[(x_t - \mu)(x_{t-k} - \mu)]}{\sigma_x^2},$$

where E stands for the expected value. Note that we have the relationship $\operatorname{cov}(x_t, x_{t-k}) \le \sigma_x^2$, with equality holding when $k = 0$. This means that the absolute value of ρ_k is less than or equal to 1.

Of course, we can never directly observe the autocorrelations, ρ_k. Instead, we must estimate them. Thus, we replace the expected value of x_t, μ, with its estimated value, $\bar{x}$, to compute the estimated autocorrelations. The kth-order estimated autocorrelation of the time series x_t, which we denote $\hat{\rho}_k$, is

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}\left[(x_t - \bar{x})(x_{t-k} - \bar{x})\right]}{\sum_{t=1}^{T}(x_t - \bar{x})^2}.$$

Analogous to the definition of autocorrelations for a time series, we can define the autocorrelations of the error term for a time-series model as[2]

$$\begin{aligned}\rho_{\varepsilon,k} &= \frac{\operatorname{cov}\left(\varepsilon_t,\varepsilon_{t-k}\right)}{\sigma_\varepsilon^2}\\ &= \frac{E\left[\left(\varepsilon_t - 0\right)\left(\varepsilon_{t-k} - 0\right)\right]}{\sigma_\varepsilon^2}\\ &= \frac{E\left(\varepsilon_t\varepsilon_{t-k}\right)}{\sigma_\varepsilon^2}.\end{aligned}$$

We assume that the expected value of the error term in a time-series model is 0.[3]

We can determine whether we are using the correct time-series model by testing whether the autocorrelations of the error term (**error autocorrelations**) differ significantly from 0. If they do, the model is not specified correctly. We estimate the error autocorrelation using the sample autocorrelations of the residuals (**residual autocorrelations**) and their sample variance.

A test of the null hypothesis that an error autocorrelation at a specified lag equals 0 is based on the residual autocorrelation for that lag and the standard error of the residual correlation, which is equal to $1/\sqrt{T}$, where T is the number of observations in the time series (Diebold 2008). Thus, if we have 100 observations in a time series, the standard error for each of the estimated autocorrelations is 0.1. We can compute the t-test of the null hypothesis that the error correlation at a particular lag equals 0 by dividing the residual autocorrelation at that lag by its standard error $\left(1/\sqrt{T}\right)$.

How can we use information about the error autocorrelations to determine whether an autoregressive time-series model is correctly specified? We can use a simple three-step method. First, estimate a particular autoregressive model—say, an AR(1) model. Second, compute the autocorrelations of the residuals from the model.[4] Third, test to see whether the residual autocorrelations differ significantly from 0. If significance tests show that the residual autocorrelations differ significantly from 0, the model is not correctly specified; we may need to modify it in ways that we will discuss shortly.[5] We now present an example to demonstrate how this three-step method works.

EXAMPLE 4

Predicting Gross Margins for Intel Corporation

Analyst Melissa Jones decides to use a time-series model to predict Intel Corporation's gross margin [(Sales – Cost of goods sold)/Sales] using quarterly data from the first quarter of 1999 through the second quarter of 2019. She does not know the best model for gross margin but believes that the current-period value will be related to the previous-period value. She decides to start out with a first-order autoregressive model, AR(1): Gross margin$_t$ = b_0 + b_1(Gross margin$_{t-1}$)

2 Whenever we refer to autocorrelation without qualification, we mean autocorrelation of the time series itself rather than autocorrelation of the error term or residuals.

3 This assumption is similar to the one made in earlier coverage of regression analysis about the expected value of the error term.

4 We can compute these residual autocorrelations easily with most statistical software packages. In Microsoft Excel, for example, to compute the first-order residual autocorrelation, we compute the correlation of the residuals from Observations 1 through T – 1 with the residuals from Observations 2 through T.

5 Often, econometricians use additional tests for the significance of residual autocorrelations. For example, the Box–Pierce Q-statistic is frequently used to test the joint hypothesis that all autocorrelations of the residuals are equal to 0. For further discussion, see Diebold (2008).

+ ε_t. Her observations on the dependent variable are 1Q 2003 through 2Q 2019. Exhibit 12 shows the results of estimating this AR(1) model, along with the autocorrelations of the residuals from that model.

Exhibit 12 Autoregression: AR(1) Model Gross Margin of Intel Quarterly Observations, January 2003–June 2019

Regression Statistics	
R^2	0.5746
Standard error	0.03002
Observations	65
Durbin–Watson	1.743

	Coefficient	Standard Error	t-Statistic
Intercept	0.1513	0.0480	3.15
Gross margin$_{t-1}$	0.7462	0.0809	9.2236

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	t-Statistic
1	0.1308	0.1240	1.0545
2	−0.2086	0.1240	−1.6818
3	0.0382	0.1240	0.3080
4	0.0608	0.1240	0.4903

Source: Bloomberg.

The first thing to note about Exhibit 12 is that both the intercept ($\hat{b}_0$ = 0.1513) and the coefficient on the first lag ($\hat{b}_1$ = 0.7462) of the gross margin are highly significant in the regression equation. The first lag of a time series is the value of the time series in the previous period. The t-statistic for the intercept is about 3.2, whereas the t-statistic for the first lag of the gross margin is more than 9. With 65 observations and two parameters, this model has 63 degrees of freedom. At the 0.05 significance level, the critical value for a t-statistic is about 2.0. Therefore, Jones must reject the null hypotheses that the intercept is equal to 0 (b_0 = 0) and the coefficient on the first lag is equal to 0 (b_1 = 0) in favor of the alternative hypothesis that the coefficients, individually, are not equal to 0. But are these statistics valid? Although the Durbin–Watson statistic is presented in Exhibit 12, it cannot be used to test serial correlation when the independent variables include past values of the dependent variable. The correct approach is to test whether the residuals from this model are serially correlated.

At the bottom of Exhibit 12, the first four autocorrelations of the residual are displayed along with the standard error and the t-statistic for each of those autocorrelations.[6] The sample has 65 observations, so the standard error for each of the autocorrelations is $1/\sqrt{65}$ = 0.1240. Exhibit 12 shows that none of

6 For seasonally unadjusted data, analysts often compute the same number of autocorrelations as there are observations in a year (for example, four for quarterly data). The number of autocorrelations computed also often depends on sample size, as discussed in Diebold (2008).

the first four autocorrelations has a t-statistic larger than 1.6818 in absolute value. Therefore, Jones can conclude that none of these autocorrelations differs significantly from 0. Consequently, she can assume that the residuals are not serially correlated and that the model is correctly specified, and she can validly use ordinary least squares to estimate the parameters and the parameters' standard errors in the autoregressive model (for other tests for serial correlation of residuals, see Diebold 2008).

Now that Jones has concluded that this model is correctly specified, how can she use it to predict Intel's gross margin in the next period? The estimated equation is Gross margin$_t$ = 0.1513 + 0.7462(Gross margin$_{t-1}$) + ε_t. The expected value of the error term is 0 in any period. Therefore, this model predicts that gross margin in period t + 1 will be Gross margin$_{t+1}$ = 0.1513 + 0.7462(Gross margin$_t$). For example, if gross margin is 55% in this quarter (0.55), the model predicts that in the next quarter gross margin will increase to 0.1513 + 0.7462(0.55) = 0.5617, or 56.17%. However, if gross margin is currently 65% (0.65), the model predicts that in the next quarter, gross margin will fall to 0.1513 + 0.7462(0.65) = 0.6363, or 63.63%. As we show in the following section, the model predicts that gross margin will increase if it is below a certain level (59.61%) and decrease if it is above that level.

7 MEAN REVERSION AND MULTIPERIOD FORECASTS AND THE CHAIN RULE OF FORECASTING

f explain mean reversion and calculate a mean-reverting level;

d describe the structure of an autoregressive (AR) model of order p and calculate one- and two-period-ahead forecasts given the estimated coefficients;

We say that a time series shows **mean reversion** if it tends to fall when its level is above its mean and rise when its level is below its mean. Much like the temperature in a room controlled by a thermostat, a mean-reverting time series tends to return to its long-term mean. How can we determine the value that the time series tends toward? If a time series is currently at its mean-reverting level, then the model predicts that the value of the time series will be the same in the next period. At its mean-reverting level, we have the relationship $x_{t+1} = x_t$. For an AR(1) model ($x_{t+1} = b_0 + b_1x_t$), the equality $x_{t+1} = x_t$ implies the level $x_t = b_0 + b_1x_t$ or that the mean-reverting level, x_t, is given by

$$x_t = \frac{b_0}{1 - b_1}.$$

So the AR(1) model predicts that the time series will stay the same if its current value is $b_0/(1 - b_1)$, increase if its current value is below $b_0/(1 - b_1)$, and decrease if its current value is above $b_0/(1 - b_1)$.

In the case of gross margins for Intel, the mean-reverting level for the model shown in Exhibit 12 is 0.1513/(1 − 0.7462) = 0.5961. If the current gross margin is above 0.5961, the model predicts that the gross margin will fall in the next period. If the current gross margin is below 0.5961, the model predicts that the gross margin will rise in the next period. As we will discuss later, all covariance-stationary time series have a finite mean-reverting level.

7.1 Multiperiod Forecasts and the Chain Rule of Forecasting

Often, financial analysts want to make forecasts for more than one period. For example, we might want to use a quarterly sales model to predict sales for a company for each of the next four quarters. To use a time-series model to make forecasts for more than one period, we must examine how to make multiperiod forecasts using an AR(1) model. The one-period-ahead forecast of x_t from an AR(1) model is as follows:

$$\hat{x}_{t+1} = \hat{b}_0 + \hat{b}_1 x_t \tag{6}$$

If we want to forecast x_{t+2} using an AR(1) model, our forecast will be based on

$$\hat{x}_{t+2} = \hat{b}_0 + \hat{b}_1 x_{t+1} \tag{7}$$

Unfortunately, we do not know x_{t+1} in period t, so we cannot use Equation 7 directly to make a two-period-ahead forecast. We can, however, use our forecast of x_{t+1} and the AR(1) model to make a prediction of x_{t+2}. The **chain rule of forecasting** is a process in which the next period's value, predicted by the forecasting equation, is substituted into the equation to give a predicted value two periods ahead. Using the chain rule of forecasting, we can substitute the predicted value of x_{t+1} into Equation 7 to get $\hat{x}_{t+2} = \hat{b}_0 + \hat{b}_1 \hat{x}_{t+1}$. We already know $\hat{x}_{t+1}$ from our one-period-ahead forecast in Equation 6. Now we have a simple way of predicting x_{t+2}.

Multiperiod forecasts are more uncertain than single-period forecasts because each forecast period has uncertainty. For example, in forecasting x_{t+2}, we first have the uncertainty associated with forecasting x_{t+1} using x_t, and then we have the uncertainty associated with forecasting x_{t+2} using the forecast of x_{t+1}. In general, the more periods a forecast has, the more uncertain it is. Note that if a forecasting model is well specified, the prediction errors from the model will not be serially correlated. If the prediction errors for each period are not serially correlated, then the variance of a multiperiod forecast will be higher than the variance of a single-period forecast.

EXAMPLE 5

Multiperiod Prediction of Intel's Gross Margin

Suppose that at the beginning of 2020, we want to predict Intel's gross margin in two periods using the model shown in Exhibit 12. Assume that Intel's gross margin in the current period is 63%. The one-period-ahead forecast of Intel's gross margin from this model is 0.6214 = 0.1513 + 0.7462(0.63). By substituting the one-period-ahead forecast, 0.6214, back into the regression equation, we can derive the following two-period-ahead forecast: 0.6150 = 0.1513 + 0.7462(0.6214). Therefore, if the current gross margin for Intel is 63%, the model predicts that Intel's gross margin in two quarters will be 61.50%.

EXAMPLE 6

Modeling US CPI Inflation

Analyst Lisette Miller has been directed to build a time-series model for monthly US inflation. Inflation and expectations about inflation, of course, have a significant effect on bond returns. For a 24-year period beginning January 1995 and ending December 2018, she selects as data the annualized monthly percentage change in the CPI. Which model should Miller use?

The process of model selection parallels that of Example 4 relating to Intel's gross margins. The first model Miller estimates is an AR(1) model, using the previous month's inflation rate as the independent variable: $\text{Inflation}_t = b_0 + b_1(\text{Inflation}_{t-1}) + \varepsilon_t$, $t = 1, 2, \ldots, 359$. To estimate this model, she uses monthly CPI inflation data from January 1995 to December 2018 ($t = 1$ denotes February 1995). Exhibit 13 shows the results of estimating this model.

Exhibit 13 Monthly CPI Inflation at an Annual Rate: AR(1) Model—Monthly Observations, February 1995–December 2018

Regression Statistics	
R^2	0.1586
Standard error	2.9687
Observations	287
Durbin–Watson	1.8442

	Coefficient	Standard Error	*t*-Statistic
Intercept	1.3346	0.2134	6.2540
Inflation_{t-1}	0.3984	0.0544	7.3235

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.0777	0.0590	1.3175
2	−0.1653	0.0590	−2.8013
3	−0.1024	0.0590	−1.7362
4	−0.0845	0.0590	1.4324

Source: US Bureau of Labor Statistics.

As Exhibit 13 shows, both the intercept ($\hat{b}_0 = 1.3346$) and the coefficient on the first lagged value of inflation ($\hat{b}_1 = 0.3984$) are highly statistically significant, with large *t*-statistics. With 287 observations and two parameters, this model has 285 degrees of freedom. The critical value for a *t*-statistic at the 0.05 significance level is about 1.97. Therefore, Miller can reject the individual null hypotheses that the intercept is equal to 0 ($b_0 = 0$) and the coefficient on the first lag is equal to 0 ($b_1 = 0$) in favor of the alternative hypothesis that the coefficients, individually, are not equal to 0.

Are these statistics valid? Miller will know when she tests whether the residuals from this model are serially correlated. With 287 observations in this sample, the standard error for each of the estimated autocorrelations is $1/\sqrt{287} = 0.0590$. The critical value for the *t*-statistic is 1.97. Because the second estimated autocorrelation has *t*-statistic larger than 1.97 in absolute value, Miller concludes that the autocorrelations are significantly different from 0. This model is thus misspecified because the residuals are serially correlated.

If the residuals in an autoregressive model are serially correlated, Miller can eliminate the correlation by estimating an autoregressive model with more lags of the dependent variable as explanatory variables. Exhibit 14 shows the result of estimating a second time-series model, an AR(2) model using the same data as in the analysis shown in Exhibit 13. With 286 observations and three parameters, this model has 283 degrees of freedom. Because the degrees of freedom are almost the same as those for the estimates shown in Exhibit 13, the critical value of the t-statistic at the 0.05 significance level also is almost the same (1.97). If she estimates the equation with two lags—$\text{Inflation}_t = b_0 + b_1(\text{Inflation}_{t-1}) + b_2(\text{Inflation}_{t-2}) + \varepsilon_t$—Miller finds that all three of the coefficients in the regression model (an intercept and the coefficients on two lags of the dependent variable) differ significantly from 0. The bottom portion of Exhibit 14 shows that none of the first four autocorrelations of the residual has a t-statistic greater in absolute value than the critical value of 1.97. Therefore, Miller fails to reject the hypothesis that the individual autocorrelations of the residual equal 0. She concludes that this model is correctly specified because she finds no evidence of serial correlation in the residuals.

Exhibit 14 Monthly CPI Inflation at an Annual Rate: AR(2) Model—Monthly Observations, March 1995–December 2018

Regression Statistics

R^2	0.1907
Standard error	2.9208
Observations	286
Durbin–Watson	1.9934

	Coefficient	Standard Error	t-Statistic
Intercept	1.5996	0.2245	7.1252
Inflation_{t-1}	0.4759	0.0583	8.1636
Inflation_{t-2}	−0.1964	0.0583	−3.368

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	t-Statistic
1	0.0032	0.0591	0.0536
2	0.0042	0.0591	0.0707
3	−0.0338	0.0591	−0.5696
4	0.0155	0.0591	1.7692

Source: US Bureau of Labor Statistics.

1 The analyst selected an AR(2) model because the residuals from the AR(1) model were serially correlated. Suppose that in a given month, inflation had been 4% at an annual rate in the previous month and 3% in the month before that. What would be the difference in the analyst forecast of inflation for that month if she had used an AR(1) model instead of the AR(2) model?

Solution to 1:

The AR(1) model shown in Exhibit 13 predicted that inflation in the next month would be 1.3346 + 0.3984(4) = 2.93%, approximately, whereas the AR(2) model shown in Exhibit 14 predicts that inflation in the next month will be 1.5996 +

0.4759(4) – 0.1964(3) = 2.91% approximately. If the analyst had used the incorrect AR(1) model, she would have predicted inflation to be 2 bps higher (2.93% versus 2.91%) than when using the AR(2) model. Although in this case the difference in the predicted inflation is actually very small, this kind of scenario illustrates that using an incorrect forecast could adversely affect the quality of her company's investment choices.

8 COMPARING FORECAST MODEL PERFORMANCE

g contrast in-sample and out-of-sample forecasts and compare the forecasting accuracy of different time-series models based on the root mean squared error criterion;

One way to compare the forecast performance of two models is to compare the variance of the forecast errors that the two models make. The model with the smaller forecast error variance will be the more accurate model, and it will also have the smaller standard error of the time-series regression. (This standard error usually is reported directly in the output for the time-series regression.)

In comparing forecast accuracy among models, we must distinguish between in-sample forecast errors and out-of-sample forecast errors. **In-sample forecast errors** are the residuals from a fitted time-series model. For example, when we estimated a linear trend with raw inflation data from January 1995 to December 2018, the in-sample forecast errors were the residuals from January 1995 to December 2018. If we use this model to predict inflation outside this period, the differences between actual and predicted inflation are **out-of-sample forecast errors**.

EXAMPLE 7

In-Sample Forecast Comparisons of US CPI Inflation

In Example 6, the analyst compared an AR(1) forecasting model of monthly US inflation with an AR(2) model of monthly US inflation and decided that the AR(2) model was preferable. Exhibit 13 showed that the standard error from the AR(1) model of inflation is 2.9687, and Exhibit 14 showed that the standard error from the AR(2) model is 2.9208. Therefore, the AR(2) model had a lower in-sample forecast error variance than the AR(1) model had, which is consistent with our belief that the AR(2) model was preferable. Its standard error is 2.9208/2.9687 = 98.39% of the forecast error of the AR(1) model.

Often, we want to compare the forecasting accuracy of different models after the sample period for which they were estimated. We wish to compare the out-of-sample forecast accuracy of the models. Out-of-sample forecast accuracy is important because the future is always out of sample. Although professional forecasters distinguish between out-of-sample and in-sample forecasting performance, many articles that analysts read contain only in-sample forecast evaluations. Analysts should be aware that out-of-sample performance is critical for evaluating a forecasting model's real-world contribution.

Typically, we compare the out-of-sample forecasting performance of forecasting models by comparing their **root mean squared error (RMSE)**, which is the square root of the average squared error. The model with the smallest RMSE is judged the most accurate. The following example illustrates the computation and use of RMSE in comparing forecasting models.

EXAMPLE 8

Out-of-Sample Forecast Comparisons of US CPI Inflation

Suppose we want to compare the forecasting accuracy of the AR(1) and AR(2) models of US inflation estimated over 1995 to 2018, using data on US inflation from January 2019 to September 2019.

Exhibit 15 Out-of-Sample Forecast Error Comparisons: January 2019–September 2019 US CPI Inflation (Annualized)

Date	Infl(*t*)	Infl(*t*–1)	Infl(*t*–2)	AR(1) Error	Squared Error	AR(2) Error	Squared Error
2019							
January	0.0000	0.0000	0.0000	0.1335	0.0178	−1.6000	2.5599
February	2.4266	0.0000	0.0000	−2.2931	5.2585	0.8266	0.6833
March	4.9070	2.4266	0.0000	−3.8068	14.4916	2.1522	4.6320
April	3.6600	4.9070	2.4266	−1.5716	2.4699	0.2014	0.0406
May	1.2066	3.6600	4.9070	0.3850	0.1482	−1.1714	1.3722
June	1.2066	1.2066	3.6600	−0.5924	0.3510	−0.2488	0.0619
July	3.6600	1.2066	1.2066	−3.0458	9.2770	1.7228	2.9680
August	1.2066	3.6600	1.2066	0.3850	0.1482	−1.8982	3.6030
September	0.0000	1.2066	3.6600	0.6142	0.3772	−1.4554	2.1181
				Average	3.6155	Average	2.0043
				RMSE	1.9014	RMSE	1.4157

Note: Any apparent discrepancies between error and squared error results are due to rounding.
Source: US Bureau of Labor Statistics.

For each month from January 2019 to September 2019, the first column of numbers in Exhibit 15 shows the actual annualized inflation rate during the month. The second and third columns show the rate of inflation in the previous two months. The fourth column shows the out-of-sample errors (Actual – Forecast) from the AR(1) model shown in Exhibit 13. The fifth column shows the squared errors from the AR(1) model. The sixth column shows the out-of-sample errors from the AR(2) model shown in Exhibit 14. The final column shows the squared errors from the AR(2) model. The bottom of the table displays the average squared error and the RMSE. According to these measures, the AR(2) model was slightly more accurate than the AR(1) model in its out-of-sample forecasts of inflation from January 2019 to September 2019. The RMSE from the AR(2) model was only 1.4157/1.9014 = 74.46% as large as the RMSE from the AR(1) model. Therefore, the AR(2) model was more accurate both in sample and out of sample. Of course, this was a small sample to use in evaluating out-of-sample forecasting performance. Sometimes, an analyst may have conflicting information about whether to choose an AR(1) or an AR(2) model. We must also consider regression coefficient stability. We will continue the comparison between these two models in the following section.

9 INSTABILITY OF REGRESSION COEFFICIENTS

h explain the instability of coefficients of time-series models;

One of the important issues an analyst faces in modeling a time series is the sample period to use. The estimates of regression coefficients of the time-series model can change substantially across different sample periods used for estimating the model. Often, the regression coefficient estimates of a time-series model estimated using an earlier sample period can be quite different from those of a model estimated using a later sample period. Similarly, the estimates can be different between models estimated using relatively shorter and longer sample periods. Further, the choice of model for a particular time series can also depend on the sample period. For example, an AR(1) model may be appropriate for the sales of a company in one particular sample period, but an AR(2) model may be necessary for an earlier or later sample period (or for a longer or shorter sample period). Thus, the choice of a sample period is an important decision in modeling a financial time series.

Unfortunately, there is usually no clear-cut basis in economic or financial theory for determining whether to use data from a longer or shorter sample period to estimate a time-series model. We can get some guidance, however, if we remember that our models are valid only for covariance-stationary time series. For example, we should not combine data from a period when exchange rates were fixed with data from a period when exchange rates were floating. The exchange rates in these two periods would not likely have the same variance because exchange rates are usually much more volatile under a floating-rate regime than when rates are fixed. Similarly, many US analysts consider it inappropriate to model US inflation or interest-rate behavior since the 1960s as a part of one sample period, because the Federal Reserve had distinct policy regimes during this period. A simple way to determine appropriate samples for time-series estimation is to look at graphs of the data to see whether the time series looks stationary before estimation begins. If we know that a government policy changed on a specific date, we might also test whether the time-series relation was the same before and after that date.

In the following example, we illustrate how the choice of a longer versus a shorter period can affect the decision of whether to use, for example, a first- or second-order time-series model. We then show how the choice of the time-series model (and the associated regression coefficients) affects our forecast. Finally, we discuss which sample period, and accordingly which model and corresponding forecast, is appropriate for the time series analyzed in the example.

EXAMPLE 9

Instability in Time-Series Models of US Inflation

In Example 6, the analyst Lisette Miller concluded that US CPI inflation should be modeled as an AR(2) time series. A colleague examined her results and questioned estimating one time-series model for inflation in the United States since 1984, given that the Federal Reserve responded aggressively to the financial crisis that emerged in 2007. He argues that the inflation time series from 1995 to 2018 has two **regimes** or underlying models generating the time series: one running from 1995 through 2007 and another starting in 2008. Therefore, the colleague suggests that Miller estimate a new time-series model for US inflation starting in 2008. Because of his suggestion, Miller first estimates an AR(1) model for inflation using data for a sample period from 2008 to 2018. Exhibit 16 shows her AR(1) estimates.

Exhibit 16 Autoregression: AR(1) Model Monthly CPI Inflation at an Annual Rate, January 2008–December 2018

Regression Statistics

R^2	0.2536
Standard error	3.0742
Observations	132
Durbin–Watson	1.8164

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.8431	0.2969	2.8397
Inflation$_{t-1}$	0.5036	0.0758	6.6438

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.0999	0.087	1.1479
2	−0.1045	0.087	−1.2015
3	−0.1568	0.087	−1.8051
4	0.0500	0.087	0.5750

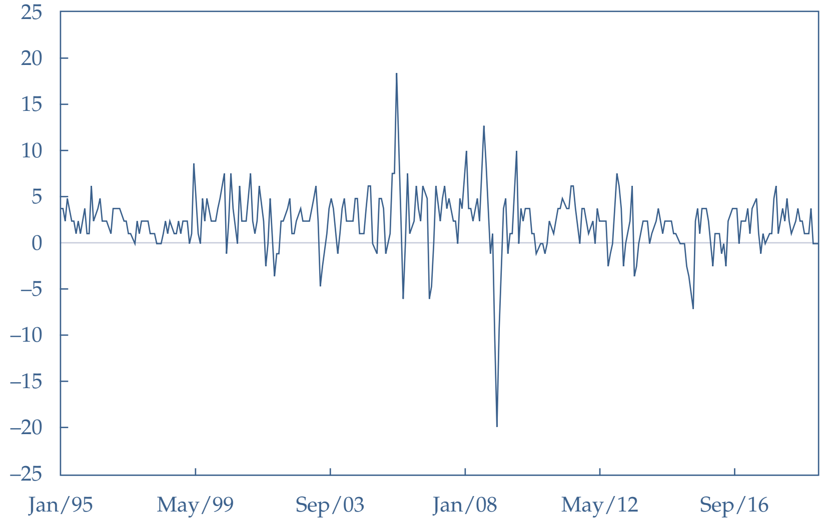
Source: US Bureau of Labor Statistics.

The bottom part of Exhibit 16 shows that the first four autocorrelations of the residuals from the AR(1) model are quite small. None of these autocorrelations has a *t*-statistic larger than 1.99, the critical value for significance. Consequently, Miller cannot reject the null hypothesis that the residuals are serially uncorrelated. The AR(1) model is correctly specified for the sample period from 2008 to 2018, so there is no need to estimate the AR(2) model. This conclusion is very different from that reached in Example 6 using data from 1995 to 2018. In that example, Miller initially rejected the AR(1) model because its residuals exhibited serial correlation. When she used a larger sample, an AR(2) model initially appeared to fit the data much better than did an AR(1) model.

How deeply does our choice of sample period affect our forecast of future inflation? Suppose that in a given month, inflation was 4% at an annual rate, and the month before that it was 3%. The AR(1) model shown in Exhibit 16 predicts that inflation in the next month will be $0.8431 + 0.5036(4) \approx 2.86\%$. Therefore, the forecast of the next month's inflation using the 2008 to 2018 sample is 2.86%. Remember from the analysis following Example 6 that the AR(2) model for the 1995 to 2018 sample predicts inflation of 2.91% in the next month. Thus, using the correctly specified model for the shorter sample produces an inflation forecast 0.05 pps below the forecast made from the correctly specified model for the longer sample period. Such a difference might substantially affect a particular investment decision.

Which model is correct? Exhibit 17 suggests an answer. Monthly US inflation was so much more volatile during the middle part of the study period than in the earlier or later years that inflation is probably not a covariance-stationary time series from 1995 to 2018. Therefore, we can reasonably believe that the data have more than one regime and Miller should estimate a separate model for inflation from 2009 to 2018, as shown previously. In fact, the standard deviation of annualized monthly inflation rates is just 2.86% for 1995–2007 but 3.54% for 2008–2018, largely because of volatility during the 2008 crisis. As the example shows, experience (such as knowledge of government policy changes) and judgment play a vital role in determining how to model a time series. Simply relying on autocorrelations of the residuals from a time-series model cannot tell us the correct sample period for our analysis.

Exhibit 17 Monthly CPI Inflation



Source: US Bureau of Labor Statistics.

10 RANDOM WALKS

i. describe characteristics of random walk processes and contrast them to covariance stationary processes;

f. explain mean reversion and calculate a mean-reverting level;

So far, we have examined those time series in which the time series has a tendency to revert to its mean level as the change in a variable from one period to the next follows a mean-reverting pattern. In contrast, there are many financial time series in which the changes follow a random pattern. We discuss these "random walks" in the following section.

10.1 Random Walks

A random walk is one of the most widely studied time-series models for financial data. A **random walk** is a time series in which the value of the series in one period is the value of the series in the previous period plus an unpredictable random error. A random walk can be described by the following equation:

$$x_t = x_{t-1} + \varepsilon_t,\ E(\varepsilon_t) = 0,\ E\left(\varepsilon_t^2\right) = \sigma^2,\ \mathrm{cov}(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) = 0 \text{ if } t \neq s. \quad \textbf{(8)}$$

Equation 8 means that the time series x_t is in every period equal to its value in the previous period plus an error term, ε_t, that has constant variance and is uncorrelated with the error term in previous periods. Note two important points. First, this equation is a special case of an AR(1) model with $b_0 = 0$ and $b_1 = 1$.[7] Second, the expected value of ε_t is zero. Therefore, the best forecast of x_t that can be made in period $t - 1$ is x_{t-1}. In fact, in this model, x_{t-1} is the best forecast of x in every period after $t - 1$.

Random walks are quite common in financial time series. For example, many studies have tested whether and found that currency exchange rates follow a random walk. Consistent with the second point made in the previous paragraph, some studies have found that sophisticated exchange rate forecasting models cannot outperform forecasts made using the random walk model and that the best forecast of the future exchange rate is the current exchange rate.

Unfortunately, we cannot use the regression methods we have discussed so far to estimate an AR(1) model on a time series that is actually a random walk. To see why this is so, we must determine why a random walk has no finite mean-reverting level or finite variance. Recall that if x_t is at its mean-reverting level, then $x_t = b_0 + b_1 x_t$, or $x_t = b_0/(1 - b_1)$. In a random walk, however, $b_0 = 0$ and $b_1 = 1$, so $b_0/(1 - b_1) = 0/0$. Therefore, a random walk has an undefined mean-reverting level.

What is the variance of a random walk? Suppose that in Period 1, the value of x_1 is 0. Then we know that $x_2 = 0 + \varepsilon_2$. Therefore, the variance of $x_2 = \mathrm{var}(\varepsilon_2) = \sigma^2$. Now $x_3 = x_2 + \varepsilon_3 = \varepsilon_2 + \varepsilon_3$. Because the error term in each period is assumed to be uncorrelated with the error terms in all other periods, the variance of $x_3 = \mathrm{var}(\varepsilon_2) + \mathrm{var}(\varepsilon_3) = 2\sigma^2$. By a similar argument, we can show that for any period t, the variance of $x_t = (t - 1)\sigma^2$. But this means that as t grows large, the variance of x_t grows without an upper bound: It approaches infinity. This lack of upper bound, in turn, means that a random walk is not a covariance-stationary time series, because a covariance-stationary time series must have a finite variance.

What is the practical implication of these issues? *We cannot use standard regression analysis on a time series that is a random walk.* We can, however, attempt to convert the data to a covariance-stationary time series if we suspect that the time series is a random walk. In statistical terms, we can difference it.

We difference a time series by creating a new time series—say, y_t—that in each period is equal to the difference between x_t and x_{t-1}. This transformation is called **first-differencing** because it subtracts the value of the time series in the first prior period from the current value of the time series. Sometimes the first difference of x_t is written as $\Delta x_t = x_t - x_{t-1}$. Note that the first difference of the random walk in Equation 8 yields

$$y_t = x_t - x_{t-1} = \varepsilon_t,\ E(\varepsilon_t) = 0,\ E\left(\varepsilon_t^2\right) = \sigma^2,\ \mathrm{cov}(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) = 0 \text{ for } t \neq s.$$

The expected value of ε_t is 0. Therefore, the best forecast of y_t that can be made in period $t - 1$ is 0. This implies that the best forecast is that there will be no change in the value of the current time series, x_{t-1}.

---

7 Equation 8 with a nonzero intercept added (as in Equation 9, given later) is sometimes referred to as a random walk with drift.

The first-differenced variable, y_t, is covariance stationary. How is this so? First, note that this model ($y_t = \varepsilon_t$) is an AR(1) model with $b_0 = 0$ and $b_1 = 0$. We can compute the mean-reverting level of the first-differenced model as $b_0/(1 - b_1) = 0/1 = 0$. Therefore, a first-differenced random walk has a mean-reverting level of 0. Note also that the variance of y_t in each period is $\text{var}(\varepsilon_t) = \sigma^2$. Because the variance and the mean of y_t are constant and finite in each period, y_t is a covariance-stationary time series and we can model it using linear regression. Of course, modeling the first-differenced series with an AR(1) model does not help us predict the future, because $b_0 = 0$ and $b_1 = 0$. We simply conclude that the original time series is, in fact, a random walk.

Had we tried to estimate an AR(1) model for a time series that was a random walk, our statistical conclusions would have been incorrect because AR models cannot be used to estimate random walks or any time series that is not covariance stationary. The following example illustrates this issue with exchange rates.

EXAMPLE 10

The Yen/US Dollar Exchange Rate

Financial analysts often assume that exchange rates are random walks. Consider an AR(1) model for the Japanese yen/US dollar exchange rate (JPY/USD). Exhibit 18 shows the results of estimating the model using month-end observations from October 1980 through August 2019.

Exhibit 18 Yen/US Dollar Exchange Rate: AR(1) Model Month-End Observations, October 1980–August 2019

Regression Statistics	
R^2	0.9897
Standard error	4.5999
Observations	467
Durbin–Watson	1.9391

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.8409	0.6503	1.2931
JPY/USD_{t-1}	0.9919	0.0047	211.0426

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.0302	0.0465	0.6495
2	0.0741	0.0465	1.5935
3	0.0427	0.0465	0.9183
4	−0.0034	0.0465	0.0731

Source: US Federal Reserve Board of Governors.

The results in Exhibit 18 suggest that the yen/US dollar exchange rate is a random walk because the estimated intercept does not appear to be significantly different from 0 and the estimated coefficient on the first lag of the exchange rate is very close to 1. Can we use the *t*-statistics in Exhibit 18 to test whether the exchange rate is a random walk? Unfortunately, no, because the standard errors in an AR model are invalid if the model is estimated using a data series that is a random walk (remember, a random walk is not covariance stationary). If the exchange rate is, in fact, a random walk, we might come to an incorrect conclusion based on faulty statistical tests and then invest incorrectly. We can use a test presented in the next section to test whether the time series is a random walk.

Suppose the exchange rate is a random walk, as we now suspect. If so, the first-differenced series, $y_t = x_t - x_{t-1}$, will be covariance stationary. We present the results from estimating $y_t = b_0 + b_1 y_{t-1} + \varepsilon_t$ in Exhibit 19. If the exchange rate is a random walk, then $b_0 = 0$, $b_1 = 0$, and the error term will not be serially correlated.

Exhibit 19 First-Differenced Yen/US Dollar Exchange Rate: AR(1) Model Month-End Observations, November 1980–August 2019

Regression Statistics

R^2	0.0008
Standard error	4.6177
Observations	466
Durbin–Watson	2.0075

	Coefficient	Standard Error	*t*-Statistic
Intercept	−0.2185	0.2142	−1.0200
$JPY/USD_{t-1} - JPY/USD_{t-2}$	0.0287	0.0464	0.6185

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	−0.0023	0.0463	−0.0501
2	0.0724	0.0463	1.5643
3	0.0387	0.0463	0.8361
4	−0.0062	0.0463	−0.1329

Source: US Federal Reserve Board of Governors.

In Exhibit 19, neither the intercept nor the coefficient on the first lag of the first-differenced exchange rate differs significantly from 0, and no residual autocorrelations differ significantly from 0. These findings are consistent with the yen/US dollar exchange rate being a random walk.

We have concluded that the differenced regression is the model to choose. Now we can see that we would have been seriously misled if we had based our model choice on an R^2 comparison. In Exhibit 18, the R^2 is 0.9897, whereas in Exhibit 19, the R^2 is 0.0008. How can this be, if we just concluded that the model in Exhibit 19 is the one that we should use? In Exhibit 18, the R^2 measures how well the exchange rate in one period predicts the exchange rate in the next period. If the exchange rate is a random walk, its current value will be an extremely good predictor of its value in the next period, and thus the R^2 will be extremely high. At the same time, if the exchange rate is a random walk, then changes in the exchange rate should be completely unpredictable. Exhibit 19 estimates whether changes in the exchange rate from one month to the next can be predicted by changes in the exchange rate over the previous month. If they cannot be predicted, the R^2 in Exhibit 19 should be very low. In fact, it is low (0.0008). This comparison provides a good example of the general rule that we cannot necessarily choose which model is correct solely by comparing the R^2 from the two models.

The exchange rate is a random walk, and changes in a random walk are by definition unpredictable. Therefore, we cannot profit from an investment strategy that predicts changes in the exchange rate.

To this point, we have discussed only simple random walks—that is, random walks without drift. In a random walk without drift, the best predictor of the time series in the next period is its current value. A random walk with drift, however, should increase or decrease by a constant amount in each period. The equation describing a random walk with drift is a special case of the AR(1) model:

$$\begin{aligned} x_t &= b_0 + b_1 x_{t-1} + \varepsilon_t, \\ b_1 &= 1,\ b_0 \neq 0, \text{ or} \\ x_t &= b_0 + x_{t-1} + \varepsilon_t,\ E(\varepsilon_t) = 0. \end{aligned} \tag{9}$$

A random walk with drift has $b_0 \neq 0$, compared to a simple random walk, which has $b_0 = 0$.

We have already seen that $b_1 = 1$ implies an undefined mean-reversion level and thus nonstationarity. Consequently, we cannot use an AR model to analyze a time series that is a random walk with drift until we transform the time series by taking first differences. If we first-difference Equation 9, the result is $y_t = x_t - x_{t-1}$, $y_t = b_0 + \varepsilon_t$, $b_0 \neq 0$.

11 THE UNIT ROOT TEST OF NONSTATIONARITY

j describe implications of unit roots for time-series analysis, explain when unit roots are likely to occur and how to test for them, and demonstrate how a time series with a unit root can be transformed so it can be analyzed with an AR model;

k describe the steps of the unit root test for nonstationarity and explain the relation of the test to autoregressive time-series models;

In this section, we discuss how to use random walk concepts to determine whether a time series is covariance stationary. This approach focuses on the slope coefficient in the random-walk-with-drift case of an AR(1) model in contrast with the traditional autocorrelation approach, which we discuss first.

The examination of the autocorrelations of a time series at various lags is a well-known prescription for inferring whether or not a time series is stationary. Typically, for a stationary time series, either autocorrelations at all lags are statistically indistinguishable from zero or the autocorrelations drop off rapidly to zero as the number of lags becomes large. Conversely, the autocorrelations of a nonstationary time series do not exhibit those characteristics. However, this approach is less definite than a currently more popular test for nonstationarity known as the Dickey–Fuller test for a unit root.

We can explain what is known as the unit root problem in the context of an AR(1) model. If a time series comes from an AR(1) model, then to be covariance stationary, the absolute value of the lag coefficient, b_1, must be less than 1.0. We could not rely on the statistical results of an AR(1) model if the absolute value of the lag coefficient were greater than or equal to 1.0 because the time series would not be covariance stationary. If the lag coefficient is equal to 1.0, the time series has a **unit root**: It is a random walk and is not covariance stationary (note that when b_1 is greater than 1 in absolute value, we say that there is an "explosive root"). By definition, all random walks, with or without a drift term, have unit roots.

How do we test for unit roots in a time series? If we believed that a time series, x_t, was a random walk with drift, it would be tempting to estimate the parameters of the AR(1) model $x_t = b_0 + b_1x_{t-1} + \varepsilon_t$ using linear regression and conduct a t-test of the hypothesis that $b_1 = 1$. Unfortunately, if $b_1 = 1$, then x_t is not covariance stationary and the t-value of the estimated coefficient, $\hat{b}_1$, does not actually follow the t-distribution; consequently, a t-test would be invalid.

Dickey and Fuller (1979) developed a regression-based unit root test based on a transformed version of the AR(1) model $x_t = b_0 + b_1x_{t-1} + \varepsilon_t$. Subtracting x_{t-1} from both sides of the AR(1) model produces

$$x_t - x_{t-1} = b_0 + (b_1 - 1)x_{t-1} + \varepsilon_t,$$

or

$$x_t - x_{t-1} = b_0 + g_1x_{t-1} + \varepsilon_t, E(\varepsilon_t) = 0, \quad \textbf{(10)}$$

where $g_1 = (b_1 - 1)$. If $b_1 = 1$, then $g_1 = 0$ and thus a test of $g_1 = 0$ is a test of $b_1 = 1$. If there is a unit root in the AR(1) model, then g_1 will be 0 in a regression where the dependent variable is the first difference of the time series and the independent variable is the first lag of the time series. The null hypothesis of the Dickey–Fuller test is H_0: $g_1 = 0$—that is, that the time series has a unit root and is nonstationary—and the alternative hypothesis is H_a: $g_1 < 0$, that the time series does not have a unit root and is stationary.

To conduct the test, one calculates a t-statistic in the conventional manner for $\hat{g}_1$ but instead of using conventional critical values for a t-test, one uses a revised set of values computed by Dickey and Fuller; the revised critical values are larger in absolute value than the conventional critical values. A number of software packages incorporate Dickey–Fuller tests.

EXAMPLE 11 (HISTORICAL EXAMPLE)

AstraZeneca's Quarterly Sales (1)

In January 2012, equity analyst Aron Berglin is building a time-series model for the quarterly sales of AstraZeneca, a British/Swedish biopharmaceutical company headquartered in London. He is using AstraZeneca's quarterly sales in US dollars for January 2000 to December 2011 and any lagged sales data that he may need prior to 2000 to build this model. He finds that a log-linear trend model seems better suited for modeling AstraZeneca's sales than does a linear trend model. However, the Durbin–Watson statistic from the log-linear regression is just 0.7064, which causes him to reject the hypothesis that the errors in the regression are serially uncorrelated. He concludes that he cannot model the log of AstraZeneca's quarterly sales using only a time trend line. He decides to model the log of AstraZeneca's quarterly sales using an AR(1) model. He uses $\ln \text{Sales}_t = b_0 + b_1(\ln \text{Sales}_{t-1}) + \varepsilon_t$.

Before he estimates this regression, the analyst should use the Dickey–Fuller test to determine whether there is a unit root in the log of AstraZeneca's quarterly sales. If he uses the sample of quarterly data on AstraZeneca's sales from the first quarter of 2000 through the fourth quarter of 2011, takes the natural log of each observation, and computes the Dickey–Fuller *t*-test statistic, the value of that statistic might cause him to fail to reject the null hypothesis that there is a unit root in the log of AstraZeneca's quarterly sales.

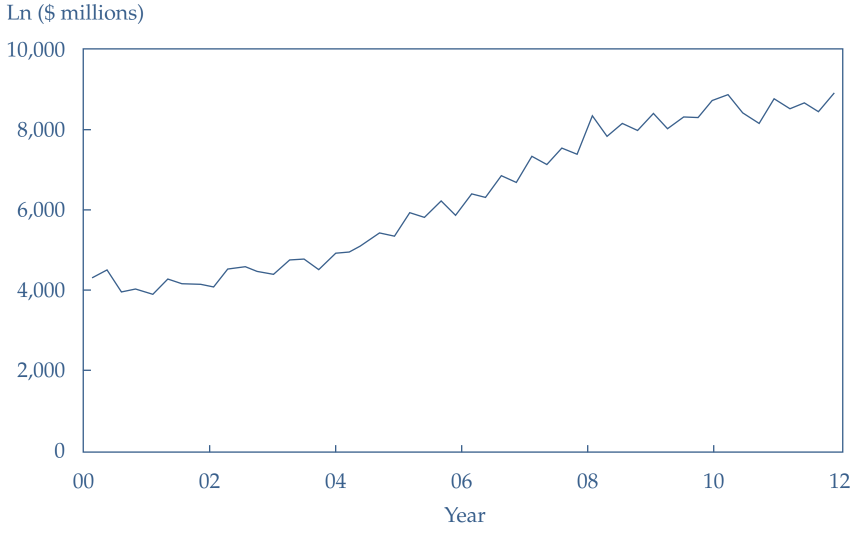
If a time series appears to have a unit root, how should we model it? One method that is often successful is to model the first-differenced series as an autoregressive time series. The following example demonstrates this method.

EXAMPLE 12

AstraZeneca's Quarterly Sales (2)

The plot of the log of AstraZeneca's quarterly sales is shown in Exhibit 20. By looking at the plot, Berglin is convinced that the log of quarterly sales is not covariance stationary (that it has a unit root).

Exhibit 20 Log of AstraZeneca's Quarterly Sales

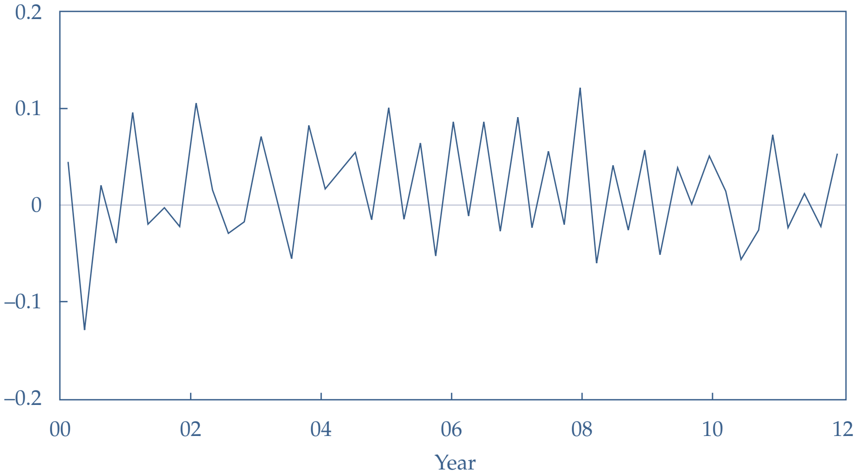


Source: Compustat.

So he creates a new series, y_t, that is the first difference of the log of AstraZeneca's quarterly sales. Exhibit 21 shows that series.

Berglin compares Exhibit 21 to Exhibit 20 and notices that first-differencing the log of AstraZeneca's quarterly sales eliminates the strong upward trend that was present in the log of AstraZeneca's sales. Because the first-differenced series has no strong trend, Berglin is better off assuming that the differenced series is covariance stationary rather than assuming that AstraZeneca's sales or the log of AstraZeneca's sales is a covariance-stationary time series.

Exhibit 21 Log Difference, AstraZeneca's Quarterly Sales



Source: Compustat.

Now suppose Berglin decides to model the new series using an AR(1) model. Berglin uses $\ln(\text{Sales}_t) - \ln(\text{Sales}_{t-1}) = b_0 + b_1[\ln(\text{Sales}_{t-1}) - \ln(\text{Sales}_{t-2})] + \varepsilon_t$. Exhibit 22 shows the results of that regression.

Exhibit 22 Log Differenced Sales: AR(1) Model of AstraZeneca Quarterly Observations, January 2000–December 2011

Regression Statistics

R^2	0.3005
Standard error	0.0475
Observations	48
Durbin–Watson	1.6874

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0222	0.0071	3.1268
$\ln \text{Sales}_{t-1} - \ln \text{Sales}_{t-2}$	–0.5493	0.1236	–4.4442

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.2809	0.1443	1.9466
2	–0.0466	0.1443	–0.3229
3	0.0081	0.1443	0.0561
4	0.2647	0.1443	1.8344

Source: Compustat.

The lower part of Exhibit 22 suggests that the first four autocorrelations of residuals in this model are not statistically significant. With 48 observations and two parameters, this model has 46 degrees of freedom. The critical value for a t-statistic in this model is above 2.0 at the 0.05 significance level. None of the t-statistics for these autocorrelations has an absolute value larger than 2.0. Therefore, we fail to reject the null hypotheses that each of these autocorrelations is equal to 0 and conclude instead that no significant autocorrelation is present in the residuals.

This result suggests that the model is well specified and that we could use the estimates. Both the intercept ($\hat{b}_0$ = 0.0222) and the coefficient ($\hat{b}_1$ = −0.5493) on the first lag of the new first-differenced series are statistically significant.

1 Explain how to interpret the estimated coefficients in the model.

Solution to 1:

The value of the intercept (0.0222) implies that if sales have not changed in the current quarter (y_t = ln Sales$_t$ − ln Sales$_{t-1}$ = 0), sales will grow by 2.22% next quarter.[8] If sales have changed during this quarter, however, the model predicts that sales will grow by 2.22% minus 0.5493 times the sales growth in this quarter.

2 AstraZeneca's sales in the third and fourth quarters of 2011 were $8,405 million and $8,872 million, respectively. If we use the previous model soon after the end of the fourth quarter of 2011, what will be the predicted value of AstraZeneca's sales for the first quarter of 2012?

Solution to 2:

Let us say that t is the fourth quarter of 2011, so t − 1 is the third quarter of 2011 and t + 1 is the first quarter of 2012. Then we would have to compute $\hat{y}_{t+1}$ = 0.0222 − 0.5493y_t. To compute $\hat{y}_{t+1}$, we need to know y_t = ln Sales$_t$ − ln Sales$_{t-1}$. In the third quarter of 2011, AstraZeneca's sales were $8,405 million, so ln Sales$_{t-1}$ = ln 8,405 = 9.0366. In the fourth quarter of 2011, AstraZeneca's sales were $8,872 million, so ln Sales$_t$ = ln 8,872 = 9.0907. Thus y_t = 9.0907 − 9.0366 = 0.0541. Therefore, $\hat{y}_{t+1}$ = 0.0222 − 0.5493(0.0541) = −0.0075. If $\hat{y}_{t+1}$ = −0.0075, then −0.0075 = ln Sales$_{t+1}$ − ln Sales$_t$ = ln(Sales$_{t+1}$/Sales$_t$). If we exponentiate both sides of this equation, the result is

$$
\begin{aligned}
e^{-0.0075} &= \left(\frac{\text{Sales}_{t+1}}{\text{Sales}_t}\right).\\
\text{Sales}_{t+1} &= \text{Sales}_t e^{-0.0075}\\
&= \$8{,}872 \text{ million} \times 0.9925\\
&= \$8{,}805 \text{ million}.
\end{aligned}
$$

Thus, based on fourth quarter sales for 2011, this model would have predicted that AstraZeneca's sales in the first quarter of 2012 would be $8,805 million. This sales forecast might have affected our decision to buy AstraZeneca's stock at the time.

8 Note that 2.22 percent is the exponential growth rate, not [(Current quarter sales/Previous quarter sales) − 1]. The difference between these two methods of computing growth is usually small.

MOVING-AVERAGE TIME-SERIES MODELS

12

k describe the steps of the unit root test for nonstationarity and explain the relation of the test to autoregressive time-series models;

So far, many of the forecasting models we have used have been autoregressive models. Because most financial time series have the qualities of an autoregressive process, autoregressive time-series models are probably the most frequently used time-series models in financial forecasting. Some financial time series, however, seem to more closely follow another kind of time-series model, called a moving-average model. For example, as we will show, returns on the S&P BSE 100 Index can be better modeled as a moving-average process than as an autoregressive process.

In this section, we present the fundamentals of moving-average models so that you can ask the right questions when considering their use. We first discuss how to smooth past values with a moving average and then how to forecast a time series using a moving-average model. Even though both methods include the words "moving average" in the name, they are very different.

12.1 Smoothing Past Values with an *n*-Period Moving Average

Suppose you are analyzing the long-term trend in the past sales of a company. In order to focus on the trend, you may find it useful to remove short-term fluctuations or noise by smoothing out the time series of sales. One technique to smooth out period-to-period fluctuations in the value of a time series is an ***n*-period moving average**. An n-period moving average of the current and past $n - 1$ values of a time series, x_t, is calculated as

$$\frac{x_t + x_{t-1} + \cdots + x_{t-(n-1)}}{n}. \tag{11}$$

The following example demonstrates how to compute a moving average of AstraZeneca's quarterly sales.

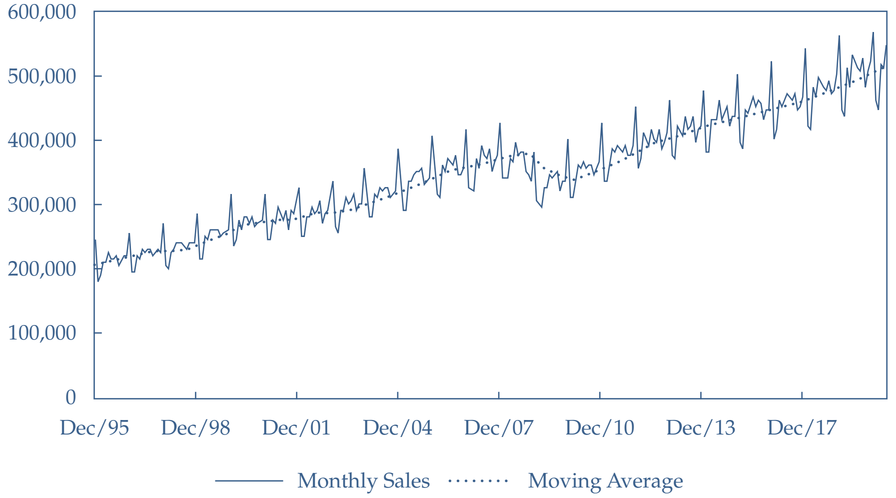
EXAMPLE 13

AstraZeneca's Quarterly Sales (3)

Suppose we want to compute the four-quarter moving average of AstraZeneca's sales as of the beginning of the first quarter of 2012. AstraZeneca's sales in the previous four quarters were as follows: 1Q 2011, \$8,490 million; 2Q 2011, \$8,601 million; 3Q 2011, \$8,405 million; and 4Q 2011, \$8,872 million. The four-quarter moving average of sales as of the beginning of the first quarter of 2012 is thus (8,490 + 8,601 + 8,405 + 8,872)/4 = \$8,592 million.

We often plot the moving average of a series with large fluctuations to help discern any patterns in the data. Exhibit 23 shows monthly retail sales for the United States from December 1995 to June 2019, along with a 12-month moving average of the data (data from January 1995 are used to compute the 12-month moving average).

Exhibit 23 Monthly US Real Retail Sales and 12-Month Moving Average of Retail Sales

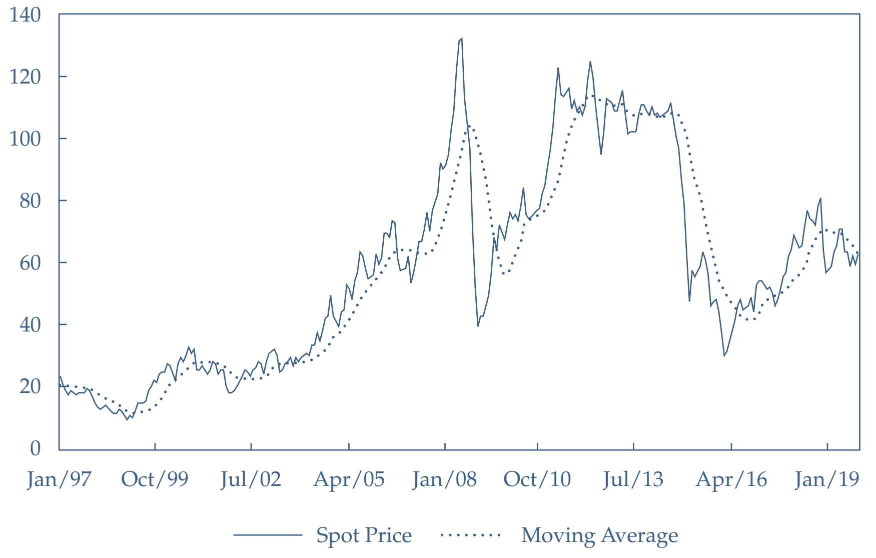


Source: Bloomberg.

As Exhibit 23 shows, each year has a very strong peak in retail sales (December) followed by a sharp drop in sales (January). Because of the extreme seasonality in the data, a 12-month moving average can help us focus on the long-term movements in retail sales instead of seasonal fluctuations. Note that the moving average does not have the sharp seasonal fluctuations of the original retail sales data. Rather, the moving average of retail sales grows steadily—for example, from 1995 through the second half of 2008—and then declines for about a year and grows steadily thereafter. We can see that trend more easily by looking at a 12-month moving average than by looking at the time series itself.

Exhibit 24 shows monthly Europe Brent Crude Oil spot prices along with a 12-month moving average of oil prices. Although these data do not have the same sharp regular seasonality displayed in the retail sales data in Exhibit 23, the moving average smooths out the monthly fluctuations in oil prices to show the longer-term movements.

Exhibit 24 Monthly Europe Brent Crude Oil Price and 12-Month Moving Average of Prices



Source: US Energy Information Administration.

Exhibit 24 also shows one weakness with a moving average: It always lags large movements in the actual data. For example, when oil prices rose quickly in late 2007 and the first half of 2008, the moving average rose only gradually. When oil prices fell sharply toward the end of 2008, the moving average also lagged. Consequently, a simple moving average of the recent past, though often useful in smoothing out a time series, may not be the best predictor of the future. A main reason for this is that a simple moving average gives equal weight to all the periods in the moving average. In order to forecast the future values of a time series, it is often better to use a more sophisticated moving-average time-series model. We discuss such models below.

12.2 Moving-Average Time-Series Models for Forecasting

Suppose that a time series, x_t, is consistent with the following model:

$$x_t = \varepsilon_t + \theta\varepsilon_{t-1},\ E(\varepsilon_t) = 0,\ E\left(\varepsilon_t^2\right) = \sigma^2,$$
$$\text{cov}(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) = 0 \text{ for } t \neq s. \tag{12}$$

This equation is called a moving-average model of order 1, or simply an MA(1) model. Theta (θ) is the parameter of the MA(1) model.[9]

Equation 12 is a moving-average model because in each period, x_t is a moving average of ε_t and ε_{t-1}, two uncorrelated random variables that each have an expected value of zero. Unlike the simple moving-average model of Equation 11, this moving-average model places different weights on the two terms in the moving average (1 on ε_t, and θ on ε_{t-1}).

9 Note that a moving-average time-series model is very different from a simple moving average, as discussed in Section 12. The simple moving average is based on observed values of a time series. In a moving-average time-series model, we never directly observe ε_t or any other ε_{t-j}, but we can infer how a particular moving-average model will imply a particular pattern of serial correlation for a time series, as we will discuss.

We can see if a time series fits an MA(1) model by looking at its autocorrelations to determine whether x_t is correlated only with its preceding and following values. First, we examine the variance of x_t in Equation 12 and its first two autocorrelations. Because the expected value of x_t is 0 in all periods and ε_t is uncorrelated with its own past values, the first autocorrelation is not equal to 0, but the second and higher autocorrelations are equal to 0. Further analysis shows that all autocorrelations except for the first will be equal to 0 in an MA(1) model. Thus for an MA(1) process, any value x_t is correlated with x_{t-1} and x_{t+1} but with no other time-series values; we could say that an MA(1) model has a memory of one period.

Of course, an MA(1) model is not the most complex moving-average model. A qth-order moving-average model, denoted MA(q) and with varying weights on lagged terms, can be written as

$$x_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q},\; E(\varepsilon_t) = 0,\; E\left(\varepsilon_t^2\right) = \sigma^2,$$
$$\text{cov}(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t\varepsilon_s) = 0 \text{ for } t \neq s. \tag{13}$$

How can we tell whether an MA(q) model fits a time series? We examine the autocorrelations. For an MA(q) model, the first q autocorrelations will be significantly different from 0, and all autocorrelations beyond that will be equal to 0; an MA(q) model has a memory of q periods. This result is critical for choosing the right value of q for an MA model. We discussed this result previously for the specific case of q = 1 that all autocorrelations except for the first will be equal to 0 in an MA(1) model.

How can we distinguish an autoregressive time series from a moving-average time series? Once again, we do so by examining the autocorrelations of the time series itself. The autocorrelations of most autoregressive time series start large and decline gradually, whereas the autocorrelations of an MA(q) time series suddenly drop to 0 after the first q autocorrelations. We are unlikely to know in advance whether a time series is autoregressive or moving average. Therefore, the autocorrelations give us our best clue about how to model the time series. Most time series, however, are best modeled with an autoregressive model.

EXAMPLE 14 (HISTORICAL EXAMPLE)

A Time-Series Model for Monthly Returns on the S&P BSE 100 Index

The S&P BSE 100 Index is designed to reflect the performance of India's top 100 large-cap companies listed on the BSE Ltd. (formerly Bombay Stock Exchange). Are monthly returns on the S&P BSE 100 Index autocorrelated? If so, we may be able to devise an investment strategy to exploit the autocorrelation. What is an appropriate time-series model for S&P BSE 100 monthly returns?

Exhibit 25 shows the first six autocorrelations of returns to the S&P BSE 100 using monthly data from January 2000 through December 2013. Note that all of the autocorrelations are quite small. Do they reach significance? With 168 observations, the critical value for a t-statistic in this model is about 1.98 at the 0.05 significance level. None of the autocorrelations has a t-statistic larger in absolute value than the critical value of 1.98. Consequently, we fail to reject the null hypothesis that those autocorrelations, individually, do not differ significantly from 0.

Exhibit 25 Annualized Monthly Returns to the S&P BSE 100, January 2000–December 2013

Autocorrelations

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.1103	0.0772	1.4288
2	−0.0045	0.0772	−0.0583
3	0.0327	0.0772	0.4236
4	0.0370	0.0772	0.4793
5	−0.0218	0.0772	−0.2824
6	0.0191	0.0772	0.2474
Observations	168		

Source: BSE Ltd.

If returns on the S&P BSE 100 were an MA(q) time series, then the first q autocorrelations would differ significantly from 0. None of the autocorrelations is statistically significant, however, so returns to the S&P BSE 100 appear to come from an MA(0) time series. An MA(0) time series in which we allow the mean to be nonzero takes the following form:[10]

$$x_t = \mu + \varepsilon_t,\ E(\varepsilon_t) = 0,\ E\left(\varepsilon_t^2\right) = \sigma^2, \\ \text{cov}(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) = 0 \text{ for } t \neq s, \tag{14}$$

which means that the time series is not predictable. This result should not be surprising, because most research suggests that short-term returns to stock indexes are difficult to predict.

We can see from this example how examining the autocorrelations allowed us to choose between the AR and MA models. If returns to the S&P BSE 100 had come from an AR(1) time series, the first autocorrelation would have differed significantly from 0 and the autocorrelations would have declined gradually. Not even the first autocorrelation is significantly different from 0, however. Therefore, we can be sure that returns to the S&P BSE 100 do not come from an AR(1) model—or from any higher-order AR model, for that matter. This finding is consistent with our conclusion that the S&P BSE 100 series is MA(0).

SEASONALITY IN TIME-SERIES MODELS 13

l. explain how to test and correct for seasonality in a time-series model and calculate and interpret a forecasted value using an AR model with a seasonal lag;

10 On the basis of investment theory and evidence, we expect that the mean monthly return on the S&P BSE 100 is positive ($\mu > 0$). We can also generalize Equation 13 for an MA(q) time series by adding a constant term, μ. Including a constant term in a moving-average model does not change the expressions for the variance and autocovariances of the time series. A number of early studies of weak-form market efficiency used Equation 14 as the model for stock returns. See Garbade (1982).

As we analyze the results of the time-series models in this reading, we encounter complications. One common complication is significant **seasonality**, a case in which the series shows regular patterns of movement within the year. At first glance, seasonality might appear to rule out using autoregressive time-series models. After all, autocorrelations will differ by season. This problem can often be solved, however, by using seasonal lags in an autoregressive model.

A seasonal lag is usually the value of the time series one year before the current period, included as an extra term in an autoregressive model. Suppose, for example, that we model a particular quarterly time series using an AR(1) model, $x_t = b_0 + b_1x_{t-1} + \varepsilon_t$. If the time series had significant seasonality, this model would not be correctly specified. The seasonality would be easy to detect because the seasonal autocorrelation (in the case of quarterly data, the fourth autocorrelation) of the error term would differ significantly from 0. Suppose this quarterly model has significant seasonality. In this case, we might include a seasonal lag in the autoregressive model and estimate

$$x_t = b_0 + b_1x_{t-1} + b_2x_{t-4} + \varepsilon_t \tag{15}$$

to test whether including the seasonal lag would eliminate statistically significant autocorrelation in the error term.

In Examples 15 and 16, we illustrate how to test and adjust for seasonality in a time-series model. We also illustrate how to compute a forecast using an autoregressive model with a seasonal lag.

EXAMPLE 15

Seasonality in Sales at Starbucks

Earlier, we concluded that we could not model the log of Starbucks' quarterly sales using only a time-trend line (as shown in Example 3) because the Durbin–Watson statistic from the regression provided evidence of positive serial correlation in the error term. Based on methods presented in this reading, we might next investigate using the first difference of log sales to remove an exponential trend from the data to obtain a covariance-stationary time series.

Using quarterly data from the last quarter of 2001 to the second quarter of 2019, we estimate the following AR(1) model using ordinary least squares: (ln Sales_t − ln Sales_{t-1}) = $b_0 + b_1$(ln Sales_{t-1} − ln Sales_{t-2}) + ε_t. Exhibit 26 shows the results of the regression.

Exhibit 26 Log Differenced Sales: AR(1) Model—Starbucks, Quarterly Observations, 2001–2019

Regression Statistics	
R^2	0.2044
Standard error	0.0611
Observations	72
Durbin–Watson	1.9904

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0469	0.0080	5.8625
ln Sales_{t-1} − ln Sales_{t-2}	−0.4533	0.1069	−4.2404

Exhibit 26 (Continued)

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.0051	0.1179	−0.0433
2	−0.1676	0.1179	−1.4218
3	−0.0130	0.1179	−0.1099
4	0.7630	0.1179	6.4720

Source: Bloomberg.

The first thing to note in Exhibit 26 is the strong seasonal autocorrelation of the residuals. The bottom portion of the table shows that the fourth autocorrelation has a value of 0.7630 and a t-statistic of 6. With 72 observations and two parameters, this model has 70 degrees of freedom.[11] The critical value for a t-statistic is about 1.99 at the 0.05 significance level. Given this value of the t-statistic, we must reject the null hypothesis that the fourth autocorrelation is equal to 0 because the t-statistic is larger than the critical value of 1.99.

In this model, the fourth autocorrelation is the seasonal autocorrelation because this AR(1) model is estimated with quarterly data. Exhibit 26 shows the strong and statistically significant seasonal autocorrelation that occurs when a time series with strong seasonality is modeled without taking the seasonality into account. Therefore, the AR(1) model is misspecified, and we should not use it for forecasting.

Suppose we decide to use an autoregressive model with a seasonal lag because of the seasonal autocorrelation. We are modeling quarterly data, so we estimate Equation 15: $(\ln \text{Sales}_t - \ln \text{Sales}_{t-1}) = b_0 + b_1(\ln \text{Sales}_{t-1} - \ln \text{Sales}_{t-2}) + b_2(\ln \text{Sales}_{t-4} - \ln \text{Sales}_{t-5}) + \varepsilon_t$. Adding the seasonal difference $\ln \text{Sales}_{t-4} - \ln \text{Sales}_{t-5}$ is an attempt to remove a consistent quarterly pattern in the data and could also eliminate a seasonal nonstationarity if one existed. The estimates of this equation appear in Exhibit 27.

Exhibit 27 Log Differenced Sales: AR(1) Model with Seasonal Lag—Starbucks, Quarterly Observations, 2005–2019

Regression Statistics	
R^2	0.7032
Standard error	0.0373
Observations	69
Durbin–Watson	2.0392

(continued)

11 In this example, we restrict the start of the sample period to the beginning of 2001, and we do not use prior observations for the lags. Accordingly, the number of observations decreases with an increase in the number of lags. In Exhibit 26, the first observation is for the third quarter of 2001 because we use up to two lags. In Exhibit 27, the first observation is for the second quarter of 2002 because we use up to five lags.

Exhibit 27 (Continued)

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0107	0.0059	1.8136
ln $Sales_{t-1}$ − ln $Sales_{t-2}$	−0.1540	0.0729	−2.1125
ln $Sales_{t-4}$ − ln $Sales_{t-5}$	0.7549	0.0720	10.4847

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.0135	0.1204	0.1121
2	−0.0171	0.1204	−0.1420
3	0.1589	0.1204	1.3198
4	−0.1498	0.1204	−1.2442

Source: Compustat.

Note the autocorrelations of the residual shown at the bottom of Exhibit 27. None of the t-statistics on the first four autocorrelations is now significant. Because the overall regression is highly significant (an F-test, not shown in the exhibit, is significant at the 0.01 level), we can take an AR(1) model with a seasonal lag as a reasonable working model for Starbucks sales. (A model having only a seasonal lag term was investigated and not found to improve on this model.)

How can we interpret the coefficients in this model? To predict the current quarter's sales growth at Starbucks, we need to know two things: sales growth in the previous quarter and sales growth four quarters ago. If sales remained constant in each of those two quarters, the model in Exhibit 27 would predict that sales will grow by 0.0107 (1.07%) in the current quarter. If sales grew by 1% last quarter and by 2% four quarters ago, then the model would predict that sales growth this quarter will be 0.0107 − 0.0154(0.01) + 0.7549(0.02) = 0.0256, or 2.56%. Note that all of these growth rates are exponential growth rates. Notice also that the R^2 in the model with the seasonal lag (0.7032 in Exhibit 27) was more than three times higher than the R^2 in the model without the seasonal lag (0.2044 in Exhibit 26). Again, the seasonal lag model does a much better job of explaining the data.

EXAMPLE 16 (HISTORICAL EXAMPLE)

Retail Sales Growth

We want to predict the growth in monthly retail sales of Canadian furniture and home furnishing stores so that we can decide whether to recommend the shares of these stores. We decide to use non-seasonally adjusted data on retail sales. To begin with, we estimate an AR(1) model with observations on the annualized monthly growth in retail sales from January 1995 to December 2012. We estimate the following equation: Sales $\text{growth}_t = b_0 + b_1(\text{Sales growth}_{t-1}) + \varepsilon_t$. Exhibit 28 shows the results from this model.

The autocorrelations of the residuals from this model, shown at the bottom of Exhibit 28, indicate that seasonality is extremely significant in this model. With 216 observations and two parameters, this model has 214 degrees of freedom. At the 0.05 significance level, the critical value for a t-statistic is about 1.97. The 12th-lag autocorrelation (the seasonal autocorrelation, because we are using monthly data) has a value of 0.7620 and a t-statistic of 11.21. The t-statistic on this autocorrelation is larger than the critical value (1.97), implying that we can reject the null hypothesis that the 12th autocorrelation is 0. Note also that many of the other t-statistics for autocorrelations shown in the table differ significantly from 0. Consequently, the model shown in Exhibit 28 is misspecified, so we cannot rely on it to forecast sales growth.

Suppose we add the seasonal lag of sales growth (the 12th lag) to the AR(1) model to estimate the equation Sales growth$_t$ = b_0 + b_1(Sales growth$_{t-1}$) + b_2(Sales growth$_{t-12}$) + ε_t. In this example, although we state that the sample period begins in 1995, we use prior observations for the lags. This results in the same number of observations irrespective of the number of lags. Exhibit 29 presents the results of estimating this equation. The estimated value of the seasonal autocorrelation (the 12th autocorrelation) has fallen to −0.1168. None of the first 12 autocorrelations has a t-statistic with an absolute value greater than the critical value of 1.97 at the 0.05 significance level. We can conclude that there is no significant serial correlation in the residuals from this model. Because we can reasonably believe that the model is correctly specified, we can use it to predict retail sales growth. Note that the R^2 in Exhibit 29 is 0.6724, much larger than the R^2 in Exhibit 28 (computed by the model without the seasonal lag).

Exhibit 28 Monthly Retail Sales Growth of Canadian Furniture and Home Furnishing Stores: AR(1) Model, January 1995–December 2012

Regression Statistics	
R^2	0.0509
Standard error	1.8198
Observations	216
Durbin–Watson	2.0956

	Coefficient	Standard Error	t-Statistic
Intercept	1.0518	0.1365	7.7055
Sales growth$_{t-1}$	−0.2252	0.0665	−3.3865

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	t-Statistic
1	−0.0109	0.0680	−0.1603
2	−0.1949	0.0680	−2.8662
3	0.1173	0.0680	1.7250
4	−0.0756	0.0680	−1.1118
5	−0.1270	0.0680	−1.8676
6	−0.1384	0.0680	−2.0353
7	−0.1374	0.0680	−2.0206

(continued)

Exhibit 28 (Continued)

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
8	−0.0325	0.0680	−0.4779
9	0.1207	0.0680	1.7750
10	−0.2197	0.0680	−3.2309
11	−0.0342	0.0680	−0.5029
12	0.7620	0.0680	11.2059

Source: Statistics Canada (Government of Canada).

How can we interpret the coefficients in the model? To predict growth in retail sales in this month, we need to know last month's retail sales growth and retail sales growth 12 months ago. If retail sales remained constant both last month and 12 months ago, the model in Exhibit 29 would predict that retail sales will grow at an annual rate of about 23.7% this month. If retail sales grew at an annual rate of 10% last month and at an annual rate of 5% 12 months ago, the model in Exhibit 29 would predict that retail sales will grow in the current month at an annual rate of 0.2371 − 0.0792(0.10) + 0.7798(0.05) = 0.2682, or 26.8%.

Exhibit 29 Monthly Retail Sales Growth of Canadian Furniture and Home Furnishing Stores: AR(1) Model with Seasonal Lag, January 1995–December 2012

Regression Statistics

R^2	0.6724
Standard error	1.0717
Observations	216
Durbin–Watson	2.1784

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.2371	0.0900	2.6344
Sales growth$_{t-1}$	−0.0792	0.0398	−1.9899
Sales growth$_{t-12}$	0.7798	0.0388	20.0979

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	−0.0770	0.0680	−1.1324
2	−0.0374	0.0680	−0.5500
3	0.0292	0.0680	0.4294
4	−0.0358	0.0680	−0.5265
5	−0.0399	0.0680	−0.5868
6	0.0227	0.0680	0.3338
7	−0.0967	0.0680	−1.4221

Exhibit 29 (Continued)

Autocorrelations of the Residual			
Lag	Autocorrelation	Standard Error	*t*-Statistic
8	0.1241	0.0680	1.8250
9	0.0499	0.0680	0.7338
10	−0.0631	0.0680	−0.9279
11	0.0231	0.0680	0.3397
12	−0.1168	0.0680	−1.7176

Source: Statistics Canada (Government of Canada).

14 AUTOREGRESSIVE MOVING-AVERAGE MODELS AND AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY MODELS

m explain autoregressive conditional heteroskedasticity (ARCH) and describe how ARCH models can be applied to predict the variance of a time series;

So far, we have presented autoregressive and moving-average models as alternatives for modeling a time series. The time series we have considered in examples have usually been explained quite well with a simple autoregressive model (with or without seasonal lags).[12] Some statisticians, however, have advocated using a more general model, the autoregressive moving-average (ARMA) model. The advocates of ARMA models argue that these models may fit the data better and provide better forecasts than do plain autoregressive (AR) models. However, as we discuss later in this section, there are severe limitations to estimating and using these models. Because you may encounter ARMA models, we next provide a brief overview.

An ARMA model combines both autoregressive lags of the dependent variable and moving-average errors. The equation for such a model with p autoregressive terms and q moving-average terms, denoted ARMA(p, q), is

$$\begin{aligned} x_t &= b_0 + b_1 x_{t-1} + \cdots + b_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \\ E(\varepsilon_t) &= 0,\ E\left(\varepsilon_t^2\right) = \sigma^2,\ \operatorname{cov}(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) = 0 \text{ for } t \neq s, \end{aligned} \tag{16}$$

where $b_1, b_2, \ldots, b_p$ are the autoregressive parameters and $\theta_1, \theta_2, \ldots, \theta_q$ are the moving-average parameters.

Estimating and using ARMA models has several limitations. First, the parameters in ARMA models can be very unstable. In particular, slight changes in the data sample or the initial guesses for the values of the ARMA parameters can result in very different final estimates of the ARMA parameters. Second, choosing the right ARMA model is more of an art than a science. The criteria for deciding on p and q for a particular time series are far from perfect. Moreover, even after a model is selected, that model may not forecast well.

12 For the returns on the S&P BSE 100 (see Example 14), we chose a moving-average model over an autoregressive model.

To reiterate, ARMA models can be very unstable, depending on the data sample used and the particular ARMA model estimated. Therefore, you should be skeptical of claims that a particular ARMA model provides much better forecasts of a time series than any other ARMA model. In fact, in most cases, you can use an AR model to produce forecasts that are just as accurate as those from ARMA models without nearly as much complexity. Even some of the strongest advocates of ARMA models admit that these models should not be used with fewer than 80 observations, and they do not recommend using ARMA models for predicting quarterly sales or gross margins for a company using even 15 years of quarterly data.

14.1 Autoregressive Conditional Heteroskedasticity Models

Up to now, we have ignored any issues of heteroskedasticity in time-series models and have assumed homoskedasticity. **Heteroskedasticity** is the dependence of the error term variance on the independent variable; **homoskedasticity** is the independence of the error term variance from the independent variable. We have assumed that the error term's variance is constant and does not depend on the value of the time series itself or on the size of previous errors. At times, however, this assumption is violated and the variance of the error term is not constant. In such a situation, the standard errors of the regression coefficients in AR, MA, or ARMA models will be incorrect, and our hypothesis tests would be invalid. Consequently, we can make poor investment decisions based on those tests.

For example, suppose you are building an autoregressive model of a company's sales. If heteroskedasticity is present, then the standard errors of the regression coefficients of your model will be incorrect. It is likely that because of heteroskedasticity, one or more of the lagged sales terms may appear statistically significant when in fact they are not. Therefore, if you use this model for your decision making, you may make some suboptimal decisions.

In work responsible in part for his shared 2003 Nobel Prize in Economics, Robert F. Engle in 1982 first suggested a way of testing whether the variance of the error in a particular time-series model in one period depends on the variance of the error in previous periods. He called this type of heteroskedasticity "autoregressive conditional heteroskedasticity" (ARCH).

As an example, consider the ARCH(1) model

$$\varepsilon_t \sim N\left(0, a_0 + a_1\varepsilon_{t-1}^2\right), \tag{17}$$

where the distribution of ε_t, conditional on its value in the previous period, ε_{t-1}, is normal, with mean 0 and variance $a_0 + a_1\varepsilon_{t-1}^2$. If $a_1 = 0$, the variance of the error in every period is just a_0. The variance is constant over time and does not depend on past errors. Now suppose that $a_1 > 0$. Then the variance of the error in one period depends on how large the squared error was in the previous period. If a large error occurs in one period, the variance of the error in the next period will be even larger.

Engle showed that we can test whether a time series is ARCH(1) by regressing the squared residuals from a previously estimated time-series model (AR, MA, or ARMA) on a constant and one lag of the squared residuals. We can estimate the linear regression equation

$$\hat{\varepsilon}_t^2 = a_0 + a_1\hat{\varepsilon}_{t-1}^2 + u_t, \tag{18}$$

where u_t is an error term. If the estimate of a_1 is statistically significantly different from zero, we conclude that the time series is ARCH(1). If a time-series model has ARCH(1) errors, then the variance of the errors in period $t + 1$ can be predicted in period t using the formula $\hat{\sigma}_{t+1}^2 = \hat{a}_0 + \hat{a}_1\hat{\varepsilon}_t^2$.

EXAMPLE 17

Testing for ARCH(1) in Monthly Inflation

Analyst Lisette Miller wants to test whether monthly data on CPI inflation contain autoregressive conditional heteroskedasticity. She could estimate Equation 18 using the residuals from the time-series model. Based on the analyses in Examples 6 through 9, she has concluded that if she modeled monthly CPI inflation from 1995 to 2018, there would not be much difference in the performance of AR(1) and AR(2) models in forecasting inflation. The AR(1) model is clearly better for the period 2008–2018. She decides to further explore the AR(1) model for the entire period 1995 to 2018. Exhibit 30 shows the results of testing whether the errors in that model are ARCH(1). Because the test involves the first lag of residuals of the estimated time-series model, the number of observations in the test is one less than that in the model.

The *t*-statistic for the coefficient on the previous period's squared residuals is greater than 4.8. Therefore, Miller easily rejects the null hypothesis that the variance of the error does not depend on the variance of previous errors. Consequently, the test statistics she computed in Exhibit 30 are not valid, and she should not use them in deciding her investment strategy.

Exhibit 30 Test for ARCH(1) in an AR(1) Model: Residuals from Monthly CPI Inflation at an Annual Rate, March 1995–December 2018

Regression Statistics	
R^2	0.0759
Standard error	23.7841
Observations	286
Durbin–Watson	2.0569

	Coefficient	Standard Error	*t*-Statistic
Intercept	6.3626	1.4928	4.2622
$\hat{\varepsilon}_{t-1}^2$	0.2754	0.0570	4.8316

Source: US Bureau of Labor Statistics.

It is possible Miller's conclusion—that the AR(1) model for monthly inflation has ARCH in the errors—may have been due to the sample period used (1995–2018). In Example 9, she used a shorter sample period, 2008–2018, and concluded that monthly CPI inflation follows an AR(1) process. (These results were shown in Exhibit 16.) Exhibit 30 shows that errors for a time-series model of inflation for the entire sample (1995–2018) have ARCH errors. Do the errors estimated with a shorter sample period (2008–2018) also display ARCH? For the shorter sample period, Miller estimated an AR(1) model using monthly inflation data. Now she tests to see whether the errors display ARCH. Exhibit 31 shows the results.

In this sample, the coefficient on the previous period's squared residual has a t-statistic of 4.0229. Consequently, Miller rejects the null hypothesis that the errors in this regression have no autoregressive conditional heteroskedasticity. The error variance appears to be heteroskedastic, and Miller cannot rely on the t-statistics.

Exhibit 31 Test for ARCH(1) in an AR(1) Model: Monthly CPI Inflation at an Annual Rate, February 2008–December 2018

Regression Statistics	
R^2	0.1113
Standard error	24.64
Observations	131
Durbin–Watson	2.0385

	Coefficient	Standard Error	t-Statistic
Intercept	6.2082	2.2873	2.7142
$\hat{\varepsilon}^2_{t-1}$	0.3336	0.0830	4.0229

Source: US Bureau of Labor Statistics.

Suppose a model contains ARCH(1) errors. What are the consequences of that fact? First, if ARCH exists, the standard errors for the regression parameters will not be correct. We will need to use generalized least squares[13] or other methods that correct for heteroskedasticity to correctly estimate the standard error of the parameters in the time-series model. Second, if ARCH exists and we have it modeled—for example, as ARCH(1)—we can predict the variance of the errors. Suppose, for instance, that we want to predict the variance of the error in inflation using the estimated parameters from Exhibit 30: $\hat{\sigma}^2_t = 6.3626 + 0.2754\hat{\varepsilon}^2_{t-1}$. If the error in one period were 0%, the predicted variance of the error in the next period would be 6.3626 + 0.2754(0) = 6.3626. If the error in one period were 1%, the predicted variance of the error in the next period would be 6.3626 + 0.2754(1^2) = 6.6380.

Engle and other researchers have suggested many generalizations of the ARCH(1) model, including ARCH(p) and generalized autoregressive conditional heteroskedasticity (GARCH) models. In an ARCH(p) model, the variance of the error term in the current period depends linearly on the squared errors from the previous p periods: $\sigma^2_t = a_0 + a_1\varepsilon^2_{t-1} + \cdots + a_p\varepsilon^2_{t-p}$. GARCH models are similar to ARMA models of the error variance in a time series. Just like ARMA models, GARCH models can be finicky and unstable: Their results can depend greatly on the sample period and the initial guesses of the parameters in the GARCH model. Financial analysts who use GARCH models should be well aware of how delicate these models can be, and they should examine whether GARCH estimates are robust to changes in the sample and the initial guesses about the parameters.[14]

13 See Greene (2018).

14 For more on ARCH, GARCH, and other models of time-series variance, see Hamilton (1994).

REGRESSIONS WITH MORE THAN ONE TIME SERIES

15

- explain how time-series variables should be analyzed for nonstationarity and/or cointegration before use in a linear regression; and

Up to now, we have discussed time-series models only for one time series. Although in the readings on correlation and regression and on multiple regression we used linear regression to analyze the relationship among different time series, in those readings we completely ignored unit roots. A time series that contains a unit root is not covariance stationary. If any time series in a linear regression contains a unit root, ordinary least squares estimates of regression test statistics may be invalid.

To determine whether we can use linear regression to model more than one time series, let us start with a single independent variable; that is, there are two time series, one corresponding to the dependent variable and one corresponding to the independent variable. We will then extend our discussion to multiple independent variables.

We first use a unit root test, such as the Dickey–Fuller test, for each of the two time series to determine whether either of them has a unit root.[15] There are several possible scenarios related to the outcome of these tests. One possible scenario is that we find that neither of the time series has a unit root. Then we can safely use linear regression to test the relations between the two time series. Otherwise, we may have to use additional tests, as we discuss later in this section.

EXAMPLE 18

Unit Roots and the Fisher Effect

Researchers at an asset management firm examined the Fisher effect by estimating the regression relation between expected inflation and US Treasury bill (T-bill) returns. They used 181 quarterly observations on expected inflation rates and T-bill returns from the sample period extending from the fourth quarter of 1968 through the fourth quarter of 2013. They used linear regression to analyze the relationship between the two time series. The results of this regression would be valid if both time series are covariance stationary; that is, neither of the two time series has a unit root. So, if they compute the Dickey–Fuller *t*-test statistic of the hypothesis of a unit root separately for each time series and find that they can reject the null hypothesis that the T-bill return series has a unit root and the null hypothesis that the expected inflation time series has a unit root, then they can use linear regression to analyze the relation between the two series. In that case, the results of their analysis of the Fisher effect would be valid.

A second possible scenario is that we reject the hypothesis of a unit root for the independent variable but fail to reject the hypothesis of a unit root for the dependent variable. In this case, the error term in the regression would not be covariance stationary. Therefore, one or more of the following linear regression assumptions would be violated: (1) that the expected value of the error term is 0, (2) that the variance of the error term is constant for all observations, and (3) that the error term is uncorrelated across observations. Consequently, the estimated regression coefficients and standard

15 For theoretical details of unit root tests, see Greene (2018) or Tsay (2010). Unit root tests are available in some econometric software packages, such as EViews.

errors would be inconsistent. The regression coefficients might appear significant, but those results would be spurious.[16] Thus we should not use linear regression to analyze the relation between the two time series in this scenario.

A third possible scenario is the reverse of the second scenario: We reject the hypothesis of a unit root for the dependent variable but fail to reject the hypothesis of a unit root for the independent variable. In this case also, like the second scenario, the error term in the regression would not be covariance stationary, and we cannot use linear regression to analyze the relation between the two time series.

EXAMPLE 19 (HISTORICAL EXAMPLE)

Unit Roots and Predictability of Stock Market Returns by Price-to-Earnings Ratio

Johann de Vries is analyzing the performance of the South African stock market. He examines whether the percentage change in the Johannesburg Stock Exchange (JSE) All Share Index can be predicted by the price-to-earnings ratio (P/E) for the index. Using monthly data from January 1994 to December 2013, he runs a regression using $(P_t - P_{t-1})/P_{t-1}$ as the dependent variable and P_{t-1}/E_{t-2} as the independent variable, where P_t is the value of the JSE index at time t and E_t is the earnings on the index. De Vries finds that the regression coefficient is negative and statistically significant and the value of the R^2 for the regression is quite high. What additional analysis should he perform before accepting the regression as valid?

De Vries needs to perform unit root tests for each of the two time series. If one of the two time series has a unit root, implying that it is not stationary, the results of the linear regression are not meaningful and cannot be used to conclude that stock market returns are predictable by P/E.[17]

The next possibility is that both time series have a unit root. In this case, we need to establish whether the two time series are **cointegrated** before we can rely on regression analysis.[18] Two time series are cointegrated if a long-term financial or economic relationship exists between them such that they do not diverge from each other without bound in the long run. For example, two time series are cointegrated if they share a common trend.

In the fourth scenario, both time series have a unit root but are not cointegrated. In this scenario, as in the second and third scenarios, the error term in the linear regression will not be covariance stationary, some regression assumptions will be violated, the regression coefficients and standard errors will not be consistent, and we cannot use them for hypothesis tests. Consequently, linear regression of one variable on the other would be meaningless.

Finally, the fifth possible scenario is that both time series have a unit root but they are cointegrated. In this case, the error term in the linear regression of one time series on the other will be covariance stationary. Accordingly, the regression coefficients and standard errors will be consistent, and we can use them for hypothesis tests. However, we should be very cautious in interpreting the results of a regression with cointegrated variables. The cointegrated regression estimates the long-term relation between the

16 The problem of spurious regression for nonstationary time series was first discussed by Granger and Newbold (1974).

17 Barr and Kantor (1999) contains evidence that the P/E time series is nonstationary.

18 Engle and Granger (1987) first discussed cointegration.

two series but may not be the best model of the short-term relation between the two series. Short-term models of cointegrated series (error correction models) are discussed in Engle and Granger (1987) and Tsay (2010), but these are specialist topics.

Now let us look at how we can test for cointegration between two time series that each have a unit root, as in the fourth and fifth scenarios.[19] Engle and Granger suggested the following test. If y_t and x_t are both time series with a unit root, we should do the following:

1. Estimate the regression $y_t = b_0 + b_1 x_t + \varepsilon_t$.
2. Test whether the error term from the regression in Step 1 has a unit root using a Dickey–Fuller test. Because the residuals are based on the estimated coefficients of the regression, we cannot use the standard critical values for the Dickey–Fuller test. Instead, we must use the critical values computed by Engle and Granger, which take into account the effect of uncertainty about the regression parameters on the distribution of the Dickey–Fuller test.
3. If the (Engle–Granger) Dickey–Fuller test fails to reject the null hypothesis that the error term has a unit root, then we conclude that the error term in the regression is not covariance stationary. Therefore, the two time series are not cointegrated. In this case, any regression relation between the two series is spurious.
4. If the (Engle–Granger) Dickey–Fuller test rejects the null hypothesis that the error term has a unit root, then we may assume that the error term in the regression is covariance stationary and that the two time series are cointegrated. The parameters and standard errors from linear regression will be consistent and will let us test hypotheses about the long-term relation between the two series.

EXAMPLE 20

Testing for Cointegration between Intel Sales and Nominal GDP

Suppose we want to test whether the natural log of Intel's sales and the natural log of GDP are cointegrated (that is, whether there is a long-term relation between GDP and Intel sales). We want to test this hypothesis using quarterly data from the first quarter of 1995 through the fourth quarter of 2019. Here are the steps:

1. Test whether the two series each have a unit root. If we cannot reject the null hypothesis of a unit root for both series, implying that both series are nonstationary, we must then test whether the two series are cointegrated.
2. Having established that each series has a unit root, we estimate the regression ln Intel sales$_t = b_0 + b_1(\ln \text{GDP}_t) + \varepsilon_t$, then conduct the (Engle–Granger) Dickey–Fuller test of the hypothesis that there is a unit root in the error term of this regression using the residuals from the estimated regression. If we reject the null hypothesis of a unit root in the error term of the regression, we reject the null hypothesis of no cointegration. That

19 Consider a time series, x_t, that has a unit root. For many such financial and economic time series, the first difference of the series, $x_t - x_{t-1}$, is stationary. We say that such a series, whose first difference is stationary, has a *single* unit root. However, for some time series, even the first difference may not be stationary and further differencing may be needed to achieve stationarity. Such a time series is said to have *multiple* unit roots. In this section, we consider only the case in which each nonstationary series has a single unit root (which is quite common).

is, the two series would be cointegrated. If the two series are cointegrated, we can use linear regression to estimate the long-term relation between the natural log of Intel sales and the natural log of GDP.

We have so far discussed models with a single independent variable. We now extend the discussion to a model with two or more independent variables, so that there are three or more time series. The simplest possibility is that none of the time series in the model has a unit root. Then, we can safely use multiple regression to test the relation among the time series.

EXAMPLE 21

Unit Roots and Returns to the Fidelity Select Technology Fund

In earlier coverage of multiple regression, we used a multiple linear regression model to examine whether returns to either the S&P 500 Growth Index or the S&P 500 Value Index explain returns to the Fidelity Select Technology Portfolio using monthly observations between October 2015 and August 2019. Of course, if any of the three time series has a unit root, then the results of our regression analysis may be invalid. Therefore, we could use a Dickey–Fuller test to determine whether any of these series has a unit root.

If we reject the hypothesis of unit roots for all three series, we can use linear regression to analyze the relation among the series. In that case, the results of our analysis of the factors affecting returns to the Fidelity Select Technology Portfolio would be valid.

If at least one time series (the dependent variable or one of the independent variables) has a unit root while at least one time series (the dependent variable or one of the independent variables) does not, the error term in the regression cannot be covariance stationary. Consequently, we should not use multiple linear regression to analyze the relation among the time series in this scenario.

Another possibility is that each time series, including the dependent variable and each of the independent variables, has a unit root. If this is the case, we need to establish whether the time series are cointegrated. To test for cointegration, the procedure is similar to that for a model with a single independent variable. First, estimate the regression $y_t = b_0 + b_1x_{1t} + b_2x_{2t} + \ldots + b_kx_{kt} + \varepsilon_t$. Then conduct the (Engle–Granger) Dickey–Fuller test of the hypothesis that there is a unit root in the errors of this regression using the residuals from the estimated regression.

If we cannot reject the null hypothesis of a unit root in the error term of the regression, we cannot reject the null hypothesis of no cointegration. In this scenario, the error term in the multiple regression will not be covariance stationary, so we cannot use multiple regression to analyze the relationship among the time series.

If we can reject the null hypothesis of a unit root in the error term of the regression, we can reject the null hypothesis of no cointegration. However, modeling three or more time series that are cointegrated may be difficult. For example, an analyst may want to predict a retirement services company's sales based on the country's GDP and the total population over age 65. Although the company's sales, GDP, and the population over 65 may each have a unit root and be cointegrated, modeling the cointegration of the three series may be difficult, and doing so is beyond the scope of this volume.

Analysts who have not mastered all these complex issues should avoid forecasting models with multiple time series that have unit roots; the regression coefficients may be inconsistent and may produce incorrect forecasts.

OTHER ISSUES IN TIME SERIES AND SUGGESTED STEPS IN TIME-SERIES FORECASTING 16

o determine an appropriate time-series model to analyze a given investment problem and justify that choice.

Time-series analysis is an extensive topic and includes many highly complex issues. Our objective in this reading has been to present those issues in time series that are the most important for financial analysts and can also be handled with relative ease. In this section, we briefly discuss some of the issues that we have not covered but could be useful for analysts.

In this reading, we have shown how to use time-series models to make forecasts. We have also introduced the RMSE as a criterion for comparing forecasting models. However, we have not discussed measuring the uncertainty associated with forecasts made using time-series models. The uncertainty of these forecasts can be very large, and should be taken into account when making investment decisions. Fortunately, the same techniques apply to evaluating the uncertainty of time-series forecasts as apply to evaluating the uncertainty about forecasts from linear regression models. To accurately evaluate forecast uncertainty, we need to consider both the uncertainty about the error term and the uncertainty about the estimated parameters in the time-series model. Evaluating this uncertainty is fairly complicated when using regressions with more than one independent variable.

In this reading, we used the US CPI inflation series to illustrate some of the practical challenges analysts face in using time-series models. We used information on US Federal Reserve policy to explore the consequences of splitting the inflation series in two. In financial time-series work, we may suspect that a time series has more than one regime but lack the information to attempt to sort the data into different regimes. If you face such a problem, you may want to investigate other methods, especially switching regression models, to identify multiple regimes using only the time series itself.

If you are interested in these and other advanced time-series topics, you can learn more from Diebold (2008) and Tsay (2010).

16.1 Suggested Steps in Time-Series Forecasting

The following is a step-by-step guide to building a model to predict a time series.

1 Understand the investment problem you have, and make an initial choice of model. One alternative is a regression model that predicts the future behavior of a variable based on hypothesized causal relationships with other variables. Another is a time-series model that attempts to predict the future behavior of a variable based on the past behavior of the same variable.

2 If you have decided to use a time-series model, compile the time series and plot it to see whether it looks covariance stationary. The plot might show important deviations from covariance stationarity, including the following:

- a linear trend,
- an exponential trend,

- seasonality, or
- a significant shift in the time series during the sample period (for example, a change in mean or variance).

3 If you find no significant seasonality or shift in the time series, then perhaps either a linear trend or an exponential trend will be sufficient to model the time series. In that case, take the following steps:
   - Determine whether a linear or exponential trend seems most reasonable (usually by plotting the series).
   - Estimate the trend.
   - Compute the residuals.
   - Use the Durbin–Watson statistic to determine whether the residuals have significant serial correlation. If you find no significant serial correlation in the residuals, then the trend model is sufficient to capture the dynamics of the time series and you can use that model for forecasting.

4 If you find significant serial correlation in the residuals from the trend model, use a more complex model, such as an autoregressive model. First, however, reexamine whether the time series is covariance stationary. The following is a list of violations of stationarity, along with potential methods to adjust the time series to make it covariance stationary:
   - If the time series has a linear trend, first-difference the time series.
   - If the time series has an exponential trend, take the natural log of the time series and then first-difference it.
   - If the time series shifts significantly during the sample period, estimate different time-series models before and after the shift.
   - If the time series has significant seasonality, include seasonal lags (discussed in Step 7).

5 After you have successfully transformed a raw time series into a covariance-stationary time series, you can usually model the transformed series with a short autoregression.[20] To decide which autoregressive model to use, take the following steps:
   - Estimate an AR(1) model.
   - Test to see whether the residuals from this model have significant serial correlation.
   - If you find no significant serial correlation in the residuals, you can use the AR(1) model to forecast.

6 If you find significant serial correlation in the residuals, use an AR(2) model and test for significant serial correlation of the residuals of the AR(2) model.
   - If you find no significant serial correlation, use the AR(2) model.
   - If you find significant serial correlation of the residuals, keep increasing the order of the AR model until the residual serial correlation is no longer significant.

7 Your next move is to check for seasonality. You can use one of two approaches:
   - Graph the data and check for regular seasonal patterns.

20 Most financial time series can be modeled using an autoregressive process. For a few time series, a moving-average model may fit better. To see whether this is the case, examine the first five or six autocorrelations of the time series. If the autocorrelations suddenly drop to 0 after the first q autocorrelations, a moving-average model (of order q) is appropriate. If the autocorrelations start large and decline gradually, an autoregressive model is appropriate.

- Examine the data to see whether the seasonal autocorrelations of the residuals from an AR model are significant (for example, the fourth autocorrelation for quarterly data) and whether the autocorrelations before and after the seasonal autocorrelations are significant. To correct for seasonality, add seasonal lags to your AR model. For example, if you are using quarterly data, you might add the fourth lag of a time series as an additional variable in an AR(1) or an AR(2) model.

8 Next, test whether the residuals have autoregressive conditional heteroskedasticity. To test for ARCH(1), for example, do the following:

- Regress the squared residual from your time-series model on a lagged value of the squared residual.
- Test whether the coefficient on the squared lagged residual differs significantly from 0.
- If the coefficient on the squared lagged residual does not differ significantly from 0, the residuals do not display ARCH and you can rely on the standard errors from your time-series estimates.
- If the coefficient on the squared lagged residual does differ significantly from 0, use generalized least squares or other methods to correct for ARCH.

9 Finally, you may also want to perform tests of the model's out-of-sample forecasting performance to see how the model's out-of-sample performance compares to its in-sample performance.

Using these steps in sequence, you can be reasonably sure that your model is correctly specified.

SUMMARY

- The predicted trend value of a time series in period t is $\hat{b}_0 + \hat{b}_1 t$ in a linear trend model; the predicted trend value of a time series in a log-linear trend model is $e^{\hat{b}_0 + \hat{b}_1 t}$.
- Time series that tend to grow by a constant amount from period to period should be modeled by linear trend models, whereas time series that tend to grow at a constant rate should be modeled by log-linear trend models.
- Trend models often do not completely capture the behavior of a time series, as indicated by serial correlation of the error term. If the Durbin–Watson statistic from a trend model differs significantly from 2, indicating serial correlation, we need to build a different kind of model.
- An autoregressive model of order p, denoted AR(p), uses p lags of a time series to predict its current value: $x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + b_p x_{t-p} + \varepsilon_t$.
- A time series is covariance stationary if the following three conditions are satisfied: First, the expected value of the time series must be constant and finite in all periods. Second, the variance of the time series must be constant and finite in all periods. Third, the covariance of the time series with itself for a fixed number of periods in the past or future must be constant and finite in all periods. Inspection of a nonstationary time-series plot may reveal an upward or downward trend (nonconstant mean) and/or nonconstant variance. The use of linear regression to estimate an autoregressive time-series model is not valid unless the time series is covariance stationary.

- For a specific autoregressive model to be a good fit to the data, the autocorrelations of the error term should be 0 at all lags.
- A time series is mean reverting if it tends to fall when its level is above its long-run mean and rise when its level is below its long-run mean. If a time series is covariance stationary, then it will be mean reverting.
- The one-period-ahead forecast of a variable x_t from an AR(1) model made in period t for period $t + 1$ is $\hat{x}_{t+1} = \hat{b}_0 + \hat{b}_1 x_t$. This forecast can be used to create the two-period-ahead forecast from the model made in period t, $\hat{x}_{t+2} = \hat{b}_0 + \hat{b}_1 x_{t+1}$. Similar results hold for AR(p) models.
- In-sample forecasts are the in-sample predicted values from the estimated time-series model. Out-of-sample forecasts are the forecasts made from the estimated time-series model for a time period different from the one for which the model was estimated. Out-of-sample forecasts are usually more valuable in evaluating the forecasting performance of a time-series model than are in-sample forecasts. The root mean squared error (RMSE), defined as the square root of the average squared forecast error, is a criterion for comparing the forecast accuracy of different time-series models; a smaller RMSE implies greater forecast accuracy.
- Just as in regression models, the coefficients in time-series models are often unstable across different sample periods. In selecting a sample period for estimating a time-series model, we should seek to assure ourselves that the time series was stationary in the sample period.
- A random walk is a time series in which the value of the series in one period is the value of the series in the previous period plus an unpredictable random error. If the time series is a random walk, it is not covariance stationary. A random walk with drift is a random walk with a nonzero intercept term. All random walks have unit roots. If a time series has a unit root, then it will not be covariance stationary.
- If a time series has a unit root, we can sometimes transform the time series into one that is covariance stationary by first-differencing the time series; we may then be able to estimate an autoregressive model for the first-differenced series.
- An n-period moving average of the current and past $(n - 1)$ values of a time series, x_t, is calculated as $[x_t + x_{t-1} + \ldots + x_{t-(n-1)}]/n$.
- A moving-average model of order q, denoted MA(q), uses q lags of a random error term to predict its current value.
- The order q of a moving-average model can be determined using the fact that if a time series is a moving-average time series of order q, its first q autocorrelations are nonzero while autocorrelations beyond the first q are zero.
- The autocorrelations of most autoregressive time series start large and decline gradually, whereas the autocorrelations of an MA(q) time series suddenly drop to 0 after the first q autocorrelations. This helps in distinguishing between autoregressive and moving-average time series.
- If the error term of a time-series model shows significant serial correlation at seasonal lags, the time series has significant seasonality. This seasonality can often be modeled by including a seasonal lag in the model, such as adding a term lagged four quarters to an AR(1) model on quarterly observations.
- The forecast made in time t for time $t + 1$ using a quarterly AR(1) model with a seasonal lag would be $x_{t+1} = \hat{b}_0 + \hat{b}_1 x_t + \hat{b}_2 x_{t-3}$.

- ARMA models have several limitations: The parameters in ARMA models can be very unstable; determining the AR and MA order of the model can be difficult; and even with their additional complexity, ARMA models may not forecast well.
- The variance of the error in a time-series model sometimes depends on the variance of previous errors, representing autoregressive conditional heteroskedasticity (ARCH). Analysts can test for first-order ARCH in a time-series model by regressing the squared residual on the squared residual from the previous period. If the coefficient on the squared residual is statistically significant, the time-series model has ARCH(1) errors.
- If a time-series model has ARCH(1) errors, then the variance of the errors in period $t + 1$ can be predicted in period t using the formula $\hat{\sigma}_{t+1}^2 = \hat{a}_0 + \hat{a}_1 \hat{\varepsilon}_t^2$.
- If linear regression is used to model the relationship between two time series, a test should be performed to determine whether either time series has a unit root:
  - If neither of the time series has a unit root, then we can safely use linear regression.
  - If one of the two time series has a unit root, then we should not use linear regression.
  - If both time series have a unit root and the time series are cointegrated, we may safely use linear regression; however, if they are not cointegrated, we should not use linear regression. The (Engle–Granger) Dickey–Fuller test can be used to determine whether time series are cointegrated.

REFERENCES

Barr, G. D. I., and B. S. Kantor. 1999. "Price–Earnings Ratios on the Johannesburg Stock Exchange—Are They a Good Value?" *SA Journal of Accounting Research* 13 (1): 1–23.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74 (366): 427–31.

Diebold, Francis X. 2008. *Elements of Forecasting*, 4th ed. Cincinnati: South-Western.

Engle, Robert F., and Clive W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55 (2): 251–76.

Garbade, Kenneth. 1982. *Securities Markets*. New York: McGraw-Hill.

Granger, Clive W. J., and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2 (2): 111–20.

Greene, William H., Econometric Analysis, New York. 2018. NY: Pearson.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Tsay, Ruey S. 2010. *Analysis of Financial Time Series*, 3rd ed. New York: Wiley.

PRACTICE PROBLEMS

Note: In the problems and solutions for this reading, we use the hat (ˆ) to indicate an estimate if we are trying to differentiate between an estimated and an actual value. However, we suppress the hat when we are clearly showing regression output.

1 The civilian unemployment rate (UER) is an important component of many economic models. Exhibit 1 gives regression statistics from estimating a linear trend model of the unemployment rate: $UER_t = b_0 + b_1 t + \varepsilon_t$.

Exhibit 1 Estimating a Linear Trend in the Civilian Unemployment Rate: Monthly Observations, January 2013–August 2019

Regression Statistics	
R^2	0.9316
Standard error	0.3227
Observations	80
Durbin–Watson	0.1878

	Coefficient	Standard Error	*t*-Statistic
Intercept	7.2237	0.0728	99.1704
Trend	–0.0510	0.0016	–32.6136

A Using the regression output in the previous table, what is the model's prediction of the unemployment rate for July 2013?

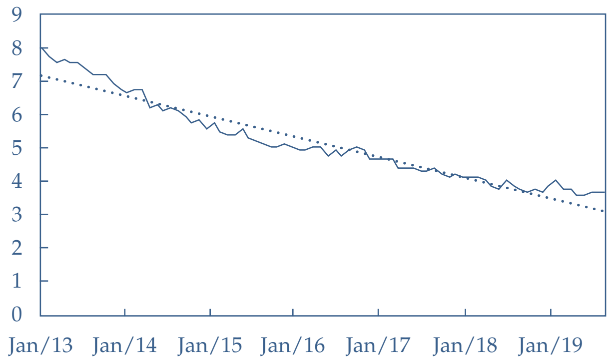
B How should we interpret the Durbin–Watson (DW) statistic for this regression? What does the value of the DW statistic say about the validity of a *t*-test on the coefficient estimates?

2 Exhibit 2 compares the predicted civilian unemployment rate (PRED) with the actual civilian unemployment rate (UER) from January 2013 to August 2019. The predicted results come from estimating the linear time trend model $UER_t = b_0 + b_1 t + \varepsilon_t$.

What can we conclude about the appropriateness of this model?

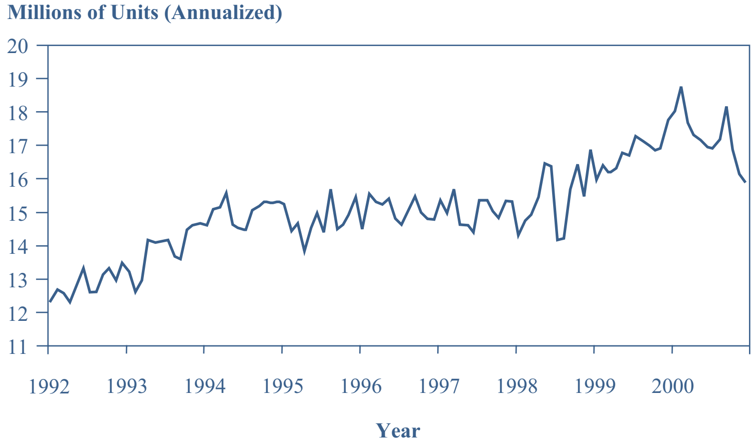
© 2017 CFA Institute. All rights reserved.

Exhibit 2 Predicted and Actual Civilian Unemployment Rates



3 You have been assigned to analyze automobile manufacturers, and as a first step in your analysis, you decide to model monthly sales of lightweight vehicles to determine sales growth in that part of the industry. Exhibit 3 gives lightweight vehicle monthly sales (annualized) from January 1992 to December 2000.

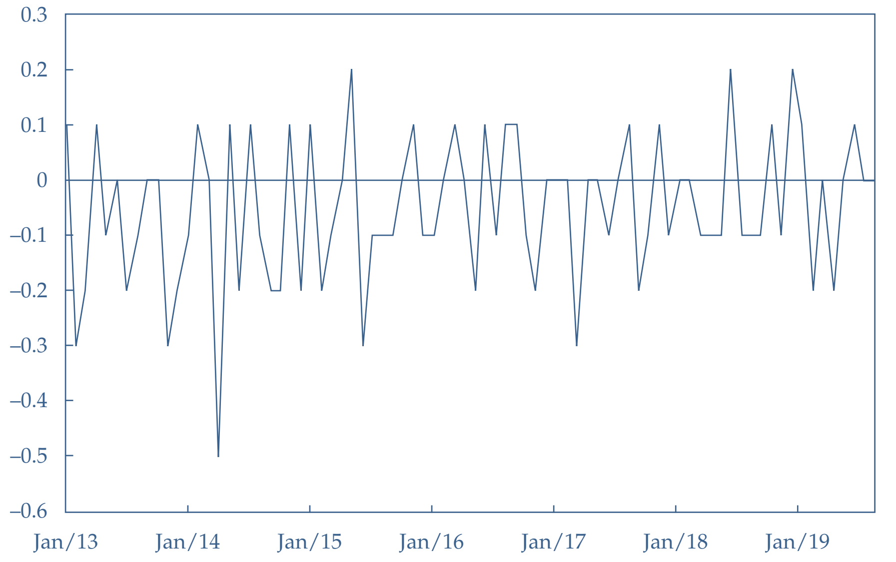
Exhibit 3 Lightweight Vehicle Sales



Monthly sales in the lightweight vehicle sector, Sales_t, have been increasing over time, but you suspect that the growth rate of monthly sales is relatively constant. Write the simplest time-series model for Sales_t that is consistent with your perception.

4 Exhibit 4 shows a plot of the first differences in the civilian unemployment rate (UER) between January 2013 and August 2019, $\Delta\text{UER}_t = \text{UER}_t - \text{UER}_{t-1}$.

Exhibit 4 Change in Civilian Unemployment Rate



A Has differencing the data made the new series, ΔUER_t, covariance stationary? Explain your answer.

B Given the graph of the change in the unemployment rate shown in the figure, describe the steps we should take to determine the appropriate autoregressive time-series model specification for the series ΔUER_t.

5 Exhibit 5 gives the regression output of an AR(1) model on first differences in the unemployment rate. Describe how to interpret the DW statistic for this regression.

Exhibit 5 Estimating an AR(1) Model of Changes in the Civilian Unemployment Rate: Monthly Observations, February 2013–August 2019

Regression Statistics	
R^2	0.0546
Standard error	0.1309
Observations	79
Durbin–Watson	2.0756

	Coefficient	Standard Error	*t*-Statistic
Intercept	–0.0668	0.0158	–4.2278
ΔUER_{t-1}	–0.2320	0.1100	–2.191

6 Assume that changes in the civilian unemployment rate are covariance stationary and that an AR(1) model is a good description for the time series of changes in the unemployment rate. Specifically, we have $\Delta UER_t = -0.0668$

$- 0.2320\Delta UER_{t-1}$ (using the coefficient estimates given in the previous problem). Given this equation, what is the mean-reverting level to which changes in the unemployment rate converge?

7 Suppose the following model describes changes in the civilian unemployment rate: $\Delta UER_t = -0.0668 - 0.2320\Delta UER_{t-1}$. The current change (first difference) in the unemployment rate is 0.0300. Assume that the mean-reverting level for changes in the unemployment rate is −0.0542.

A What is the best prediction of the next change?

B What is the prediction of the change following the next change?

C Explain your answer to Part B in terms of equilibrium.

8 Exhibit 6 gives the actual sales, log of sales, and changes in the log of sales of Cisco Systems for the period 1Q 2019 to 4Q 2019.

Exhibit 6

Date	Actual Sales ($ Millions)	Log of Sales	Changes in Log of Sales $\Delta\ln(\text{Sales}_t)$
1Q 2019	13,072	9.4782	0.0176
2Q 2019	12,446	9.4292	−0.0491
3Q 2019	12,958	9.4695	0.403
4Q 2019	13,428	9.5051	0.0356
1Q 2020			
2Q 2020			

Forecast the first- and second-quarter sales of Cisco Systems for 2020 using the regression $\Delta\ln(\text{Sales}_t) = 0.0068 + 0.2633\Delta\ln(\text{Sales}_{t-1})$.

9 Exhibit 7 gives the actual change in the log of sales of Cisco Systems from 1Q 2019 to 4Q 2019, along with the forecasts from the regression model $\Delta\ln(\text{Sales}_t) = 0.0068 + 0.2633\Delta\ln(\text{Sales}_{t-1})$ estimated using data from 1Q 2001 to 4Q 2018. (Note that the observations after the fourth quarter of 2018 are out of sample.)

Exhibit 7

Date	Actual Value of Changes in the Log of Sales $\Delta\ln(\text{Sales}_t)$	Forecast Value of Changes in the Log of Sales $\Delta\ln(\text{Sales}_t)$
1Q 2019	0.0176	0.0147
2Q 2019	−0.0491	0.0107
3Q 2019	0.4030	0.0096
4Q 2019	0.0356	0.0093

A Calculate the RMSE for the out-of-sample forecast errors.

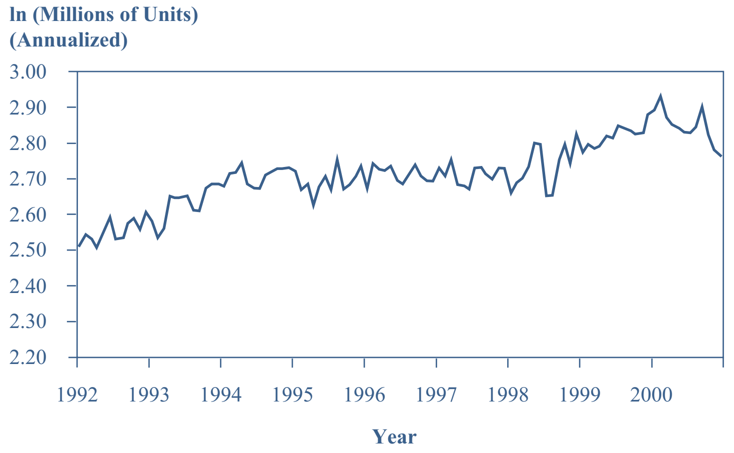
B Compare the forecasting performance of the model given with that of another model having an out-of-sample RMSE of 2%.

10 **A** The AR(1) model for the civilian unemployment rate, $\Delta UER_t = -0.0405 - 0.4674\Delta UER_{t-1}$, was developed with five years of data. What would be the drawback to using the AR(1) model to predict changes in the civilian unemployment rate 12 months or more ahead, as compared with 1 month ahead?

B For purposes of estimating a predictive equation, what would be the drawback to using 30 years of civilian unemployment data rather than only 5 years?

11 Exhibit 8 shows monthly observations on the natural log of lightweight vehicle sales, $\ln(\text{Sales}_t)$, for January 1992 to December 2000.

Exhibit 8 Lightweight Vehicle Sales

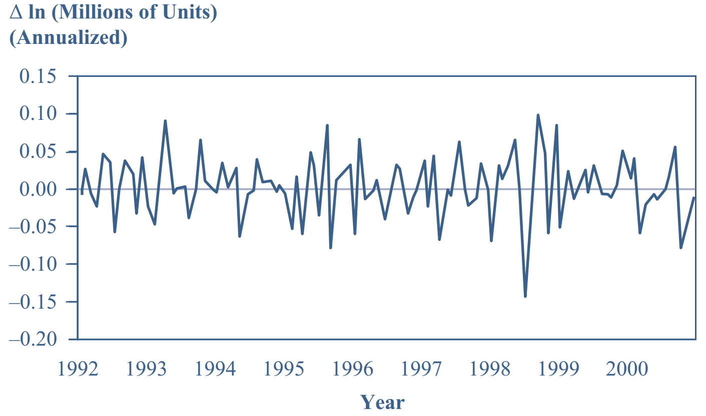


A Using the figure, comment on whether the specification $\ln(\text{Sales}_t) = b_0 + b_1[\ln(\text{Sales}_{t-1})] + \varepsilon_t$ is appropriate.

B State an appropriate transformation of the time series.

12 Exhibit 9 shows a plot of first differences in the log of monthly lightweight vehicle sales over the same period as in Problem 11. Has differencing the data made the resulting series, $\Delta\ln(\text{Sales}_t) = \ln(\text{Sales}_t) - \ln(\text{Sales}_{t-1})$, covariance stationary?

Exhibit 9 Change in Natural Log of Lightweight Vehicle Sales



13 Using monthly data from January 1992 to December 2000, we estimate the following equation for lightweight vehicle sales: $\Delta\ln(\text{Sales}_t) = 2.7108 + 0.3987\Delta\ln(\text{Sales}_{t-1}) + \varepsilon_t$. Exhibit 10 gives sample autocorrelations of the errors from this model.

Exhibit 10 Different Order Autocorrelations of Differences in the Logs of Vehicle Sales

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.9358	0.0962	9.7247
2	0.8565	0.0962	8.9005
3	0.8083	0.0962	8.4001
4	0.7723	0.0962	8.0257
5	0.7476	0.0962	7.7696
6	0.7326	0.0962	7.6137
7	0.6941	0.0962	7.2138
8	0.6353	0.0962	6.6025
9	0.5867	0.0962	6.0968
10	0.5378	0.0962	5.5892
11	0.4745	0.0962	4.9315
12	0.4217	0.0962	4.3827

A Use the information in the table to assess the appropriateness of the specification given by the equation.

B If the residuals from the AR(1) model above violate a regression assumption, how would you modify the AR(1) specification?

14 Exhibit 11 shows the quarterly sales of Cisco Systems from 3Q 2001 to 2Q 2019.

Exhibit 11 Quarterly Sales at Cisco

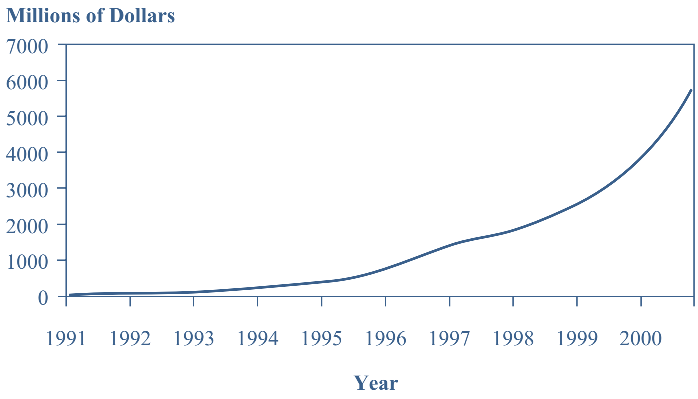


Exhibit 12 gives the regression statistics from estimating the model $\Delta\ln(\text{Sales}_t) = b_0 + b_1\Delta\ln(\text{Sales}_{t-1}) + \varepsilon_t$.

Exhibit 12 Change in the Natural Log of Sales for Cisco Quarterly Observations, 3Q 1991–4Q 2000

Regression Statistics	
R^2	0.2899
Standard error	0.0408
Observations	38
Durbin–Watson	1.5707

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0661	0.0175	3.7840
$\Delta\ln(\text{Sales}_{t-1})$	0.4698	0.1225	3.8339

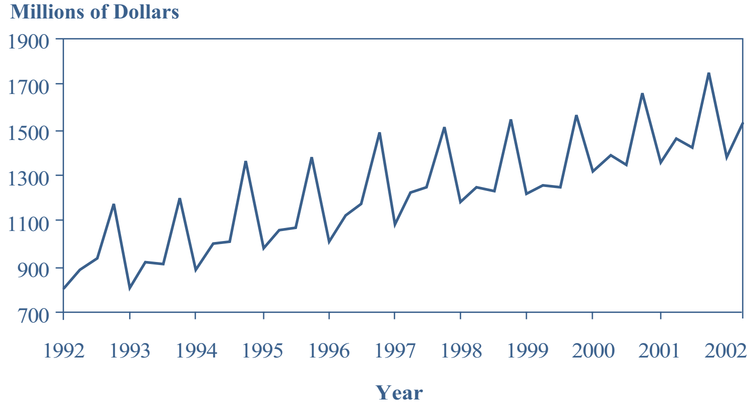
A Describe the salient features of the quarterly sales series.

B Describe the procedures we should use to determine whether the AR(1) specification is correct.

C Assuming the model is correctly specified, what is the long-run change in the log of sales toward which the series will tend to converge?

15 Exhibit 13 shows the quarterly sales of Avon Products from 1Q 1992 to 2Q 2002. Describe the salient features of the data shown.

Exhibit 13 Quarterly Sales at Avon



16 Exhibit 14 shows the autocorrelations of the residuals from an AR(1) model fit to the changes in the gross profit margin (GPM) of the Home Depot, Inc.

Exhibit 14 Autocorrelations of the Residuals from Estimating the Regression $\Delta GPM_t = 0.0006 - 0.3330\Delta GPM_{t-1} + \varepsilon_t$, 1Q 1992–4Q 2001 (40 Observations)

Lag	Autocorrelation
1	−0.1106
2	−0.5981
3	−0.1525
4	0.8496
5	−0.1099

Exhibit 15 shows the output from a regression on changes in the GPM for Home Depot, where we have changed the specification of the AR regression.

Exhibit 15 Change in Gross Profit Margin for Home Depot, 1Q 1992–4Q 2001

Regression Statistics	
R^2	0.9155
Standard error	0.0057
Observations	40
Durbin–Watson	2.6464

	Coefficient	Standard Error	*t*-Statistic
Intercept	−0.0001	0.0009	−0.0610
ΔGPM_{t-1}	−0.0608	0.0687	−0.8850
ΔGPM_{t-4}	0.8720	0.0678	12.8683

A Identify the change that was made to the regression model.

B Discuss the rationale for changing the regression specification.

17 Suppose we decide to use an autoregressive model with a seasonal lag because of the seasonal autocorrelation in the previous problem. We are modeling quarterly data, so we estimate Equation 15: $(\ln \text{Sales}_t - \ln \text{Sales}_{t-1}) = b_0 + b_1(\ln \text{Sales}_{t-1} - \ln \text{Sales}_{t-2}) + b_2(\ln \text{Sales}_{t-4} - \ln \text{Sales}_{t-5}) + \varepsilon_t$. Exhibit 16 shows the regression statistics from this equation.

Exhibit 16 Log Differenced Sales: AR(1) Model with Seasonal Lag Johnson & Johnson Quarterly Observations, January 1985–December 2001

Regression Statistics	
R^2	0.4220
Standard error	0.0318
Observations	68
Durbin–Watson	1.8784

(continued)

Exhibit 16 (Continued)

	Coefficient	Standard Error	t-Statistic
Intercept	0.0121	0.0053	2.3055
Lag 1	−0.0839	0.0958	−0.8757
Lag 4	0.6292	0.0958	6.5693

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	t-Statistic
1	0.0572	0.1213	0.4720
2	−0.0700	0.1213	−0.5771
3	0.0065	0.1213	−0.0532
4	−0.0368	0.1213	−0.3033

A Using the information in Exhibit 16, determine whether the model is correctly specified.

B If sales grew by 1% last quarter and by 2% four quarters ago, use the model to predict the sales growth for this quarter.

18 Describe how to test for autoregressive conditional heteroskedasticity (ARCH) in the residuals from the AR(1) regression on first differences in the civilian unemployment rate, $\Delta UER_t = b_0 + b_1\Delta UER_{t-1} + \varepsilon_t$.

19 Suppose we want to predict the annualized return of the five-year T-bill using the annualized return of the three-month T-bill with monthly observations from January 1993 to December 2002. Our analysis produces the data shown in Exhibit 17.

Exhibit 17 Regression with Three-Month T-Bill as the Independent Variable and the Five-Year T-Bill as the Dependent Variable: Monthly Observations, January 1993–December 2002

Regression Statistics

R^2	0.5829
Standard error	0.6598
Observations	120
Durbin–Watson	0.1130

	Coefficient	Standard Error	t-Statistic
Intercept	3.0530	0.2060	14.8181
Three-month	0.5722	0.0446	12.8408

Can we rely on the regression model in Exhibit 17 to produce meaningful predictions? Specify what problem might be a concern with this regression.

The following information relates to Questions 20–26

Angela Martinez, an energy sector analyst at an investment bank, is concerned about the future level of oil prices and how it might affect portfolio values. She is considering whether to recommend a hedge for the bank portfolio's exposure to changes in oil prices. Martinez examines West Texas Intermediate (WTI) monthly crude oil price data, expressed in US dollars per barrel, for the 181-month period from August 2000 through August 2015. The end-of-month WTI oil price was $51.16 in July 2015 and $42.86 in August 2015 (Month 181).

After reviewing the time-series data, Martinez determines that the mean and variance of the time series of oil prices are not constant over time. She then runs the following four regressions using the WTI time-series data.

- Linear trend model: Oil $\text{price}_t = b_0 + \text{b}_1 t + e_t$.
- Log-linear trend model: ln Oil $\text{price}_t = b_0 + b_1 t + e_t$.
- AR(1) model: Oil $\text{price}_t = b_0 + b_1$Oil $\text{price}_{t-1} + e_t$.
- AR(2) model: Oil $\text{price}_t = b_0 + b_1$Oil $\text{price}_{t-1} + b_2$Oil $\text{price}_{t-2} + e_t$.

Exhibit 1 presents selected data from all four regressions, and Exhibit 2 presents selected autocorrelation data from the AR(1) models.

Exhibit 1 Crude Oil Price per Barrel, August 2000–August 2015

	Regression Statistics (*t*-statistics for coefficients are reported in parentheses)			
	Linear	**Log-Linear**	**AR(1)**	**AR(2)**
R^2	0.5703	0.6255	0.9583	0.9656
Standard error	18.6327	0.3034	5.7977	5.2799
Observations	181	181	180	179
Durbin–Watson	0.10	0.08	1.16	2.08
RMSE			2.0787	2.0530
Coefficients:				
Intercept	28.3278	3.3929	1.5948	2.0017
	(10.1846)	(74.9091)	(1.4610)	(1.9957)
t (Trend)	0.4086	0.0075		
	(15.4148)	(17.2898)		
Oil price_{t-1}			0.9767	1.3946
			(63.9535)	(20.2999)
Oil price_{t-2}				−0.4249
				(−6.2064)

In Exhibit 1, at the 5% significance level, the lower critical value for the Durbin–Watson test statistic is 1.75 for both the linear and log-linear regressions.

Exhibit 2 Autocorrelations of the Residual from AR(1) Model

Lag	Autocorrelation	*t*-Statistic
1	0.4157	5.5768
2	0.2388	3.2045
3	0.0336	0.4512
4	−0.0426	−0.5712

Note: At the 5% significance level, the critical value for a *t*-statistic is 1.97.

After reviewing the data and regression results, Martinez draws the following conclusions.

Conclusion 1 The time series for WTI oil prices is covariance stationary.

Conclusion 2 Out-of-sample forecasting using the AR(1) model appears to be more accurate than that of the AR(2) model.

20 Based on Exhibit 1, the predicted WTI oil price for October 2015 using the linear trend model is *closest* to:

A $29.15.

B $74.77.

C $103.10.

21 Based on Exhibit 1, the predicted WTI oil price for September 2015 using the log-linear trend model is *closest* to:

A $29.75.

B $29.98.

C $116.50.

22 Based on the regression output in Exhibit 1, there is evidence of positive serial correlation in the errors in:

A the linear trend model but not the log-linear trend model.

B both the linear trend model and the log-linear trend model.

C neither the linear trend model nor the log-linear trend model.

23 Martinez's Conclusion 1 is:

A correct.

B incorrect because the mean and variance of WTI oil prices are not constant over time.

C incorrect because the Durbin–Watson statistic of the AR(2) model is greater than 1.75.

24 Based on Exhibit 1, the forecasted oil price in September 2015 based on the AR(2) model is *closest* to:

A $38.03.

B $40.04.

C $61.77.

25 Based on the data for the AR(1) model in Exhibits 1 and 2, Martinez can conclude that the:

A residuals are not serially correlated.

B autocorrelations do not differ significantly from zero.

C standard error for each of the autocorrelations is 0.0745.

26 Based on the mean-reverting level implied by the AR(1) model regression output in Exhibit 1, the forecasted oil price for September 2015 is *most likely* to be:

A less than \$42.86.

B equal to \$42.86.

C greater than \$42.86.

The following information relates to Question 27–35

Max Busse is an analyst in the research department of a large hedge fund. He was recently asked to develop a model to predict the future exchange rate between two currencies. Busse gathers monthly exchange rate data from the most recent 10-year period and runs a regression based on the following AR(1) model specification:

Regression 1: $x_t = b_0 + b_1 x_{t-1} + \varepsilon_t$, where x_t is the exchange rate at time t.

Based on his analysis of the time series and the regression results, Busse reaches the following conclusions:

Conclusion 1 The variance of x_t increases over time.

Conclusion 2 The mean-reverting level is undefined.

Conclusion 3 b_0 does not appear to be significantly different from 0.

Busse decides to do additional analysis by first-differencing the data and running a new regression.

Regression 2: $y_t = b_0 + b_1 y_{t-1} + \varepsilon_t$, where $y_t = x_t - x_{t-1}$.

Exhibit 1 shows the regression results.

Exhibit 1 First-Differenced Exchange Rate AR(1) Model: Month-End Observations, Last 10 Years

Regression Statistics	
R^2	0.0017
Standard error	7.3336
Observations	118
Durbin–Watson	1.9937

	Coefficient	Standard Error	*t*-Statistic
Intercept	−0.8803	0.6792	−1.2960
$x_{t-1} - x_{t-2}$	0.0412	0.0915	0.4504

(continued)

Exhibit 1 (Continued)

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	t-Statistic
1	0.0028	0.0921	0.0300
2	0.0205	0.0921	0.2223
3	0.0707	0.0921	0.7684
4	0.0485	0.0921	0.5271

Note: The critical t-statistic at the 5% significance level is 1.98.

Busse decides that he will need to test the data for nonstationarity using a Dickey–Fuller test. To do so, he knows he must model a transformed version of Regression 1.

Busse's next assignment is to develop a model to predict future quarterly sales for PoweredUP, Inc., a major electronics retailer. He begins by running the following regression:

Regression 3: $\ln \text{Sales}_t - \ln \text{Sales}_{t-1} = b_0 + b_1(\ln \text{Sales}_{t-1} - \ln \text{Sales}_{t-2}) + \varepsilon_t$.

Exhibit 2 presents the results of this regression.

Exhibit 2 Log Differenced Sales AR(1) Model: PoweredUP, Inc., Last 10 Years of Quarterly Sales

Regression Statistics

R^2	0.2011
Standard error	0.0651
Observations	38
Durbin–Watson	1.9677

	Coefficient	Standard Error	t-Statistic
Intercept	0.0408	0.0112	3.6406
$\ln \text{Sales}_{t-1} - \ln \text{Sales}_{t-2}$	−0.4311	0.1432	−3.0099

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	t-Statistic
1	0.0146	0.1622	0.0903
2	−0.1317	0.1622	−0.8119
3	−0.1123	0.1622	−0.6922
4	0.6994	0.1622	4.3111

Note: The critical t-statistic at the 5% significance level is 2.02.

Because the regression output from Exhibit 2 raises some concerns, Busse runs a different regression. These regression results, along with quarterly sales data for the past five quarters, are presented in Exhibits 3 and 4, respectively.

Exhibit 3 Log Differenced Sales AR(1) Model with Seasonal Lag: PoweredUP, Inc., Last 10 Years of Quarterly Sales

Regression Statistics	
R^2	0.6788
Standard error	0.0424
Observations	35
Durbin–Watson	1.8799

	Coefficient	Standard Error	*t*-Statistic
Intercept	0.0092	0.0087	1.0582
$\ln \text{Sales}_{t-1} - \ln \text{Sales}_{t-2}$	−0.1279	0.1137	−1.1252
$\ln \text{Sales}_{t-4} - \ln \text{Sales}_{t-5}$	0.7239	0.1093	6.6209

Autocorrelations of the Residual

Lag	Autocorrelation	Standard Error	*t*-Statistic
1	0.0574	0.1690	0.3396
2	0.0440	0.1690	0.2604
3	0.1923	0.1690	1.1379
4	−0.1054	0.1690	−0.6237

Note: The critical *t*-statistic at the 5% significance level is 2.03.

Exhibit 4 Most Recent Quarterly Sales Data (in billions)

Dec 2015 (Sales_{t-1})	$3.868
Sep 2015 (Sales_{t-2})	$3.780
June 2015 (Sales_{t-3})	$3.692
Mar 2015 (Sales_{t-4})	$3.836
Dec 2014 (Sales_{t-5})	$3.418

After completing his work on PoweredUP, Busse is asked to analyze the relationship of oil prices and the stock prices of three transportation companies. His firm wants to know whether the stock prices can be predicted by the price of oil. Exhibit 5 shows selected information from the results of his analysis.

Exhibit 5 Analysis Summary of Stock Prices for Three Transportation Stocks and the Price of Oil

	Unit Root?	Linear or Exponential Trend?	Serial Correlation of Residuals in Trend Model?	ARCH(1)?	Comments
Company 1	Yes	Exponential	Yes	Yes	Not cointegrated with oil price
Company 2	Yes	Linear	Yes	No	Cointegrated with oil price

(continued)

Exhibit 5 (Continued)

	Unit Root?	Linear or Exponential Trend?	Serial Correlation of Residuals in Trend Model?	ARCH(1)?	Comments
Company 3	No	Exponential	Yes	No	Not cointegrated with oil price
Oil Price	Yes				

To assess the relationship between oil prices and stock prices, Busse runs three regressions using the time series of each company's stock prices as the dependent variable and the time series of oil prices as the independent variable.

27 Which of Busse's conclusions regarding the exchange rate time series is consistent with both the properties of a covariance-stationary time series and the properties of a random walk?

A Conclusion 1

B Conclusion 2

C Conclusion 3

28 Based on the regression output in Exhibit 1, the first-differenced series used to run Regression 2 is consistent with:

A a random walk.

B covariance stationarity.

C a random walk with drift.

29 Based on the regression results in Exhibit 1, the *original* time series of exchange rates:

A has a unit root.

B exhibits stationarity.

C can be modeled using linear regression.

30 In order to perform the nonstationarity test, Busse should transform the Regression 1 equation by:

A adding the second lag to the equation.

B changing the regression's independent variable.

C subtracting the independent variable from both sides of the equation.

31 Based on the regression output in Exhibit 2, what should lead Busse to conclude that the Regression 3 equation is not correctly specified?

A The Durbin–Watson statistic

B The t-statistic for the slope coefficient

C The t-statistics for the autocorrelations of the residual

32 Based on the regression output in Exhibit 3 and sales data in Exhibit 4, the forecasted value of quarterly sales for March 2016 for PoweredUP is *closest* to:

A \$4.193 billion.

B \$4.205 billion.

C \$4.231 billion.

33 Based on Exhibit 5, Busse should conclude that the variance of the error terms for Company 1:

A is constant.

B can be predicted.

C is homoskedastic.

34 Based on Exhibit 5, for which company would the regression of stock prices on oil prices be expected to yield valid coefficients that could be used to estimate the long-term relationship between stock price and oil price?

A Company 1

B Company 2

C Company 3

35 Based on Exhibit 5, which single time-series model would *most likely* be appropriate for Busse to use in predicting the future stock price of Company 3?

A Log-linear trend model

B First-differenced AR(2) model

C First-differenced log AR(1) model

SOLUTIONS

1 **A** The estimated forecasting equation is $UER_t = 5.5098 - 0.0294(t)$. The data begin in January 2013, and July 2013 is Period 7. Thus the linear trend model predicts the unemployment rate to be $UER_7 = 7.2237 - 0.0510(7) = 6.8667$, or approximately 6.9%.

B The DW statistic is designed to detect positive serial correlation of the errors of a regression equation. Under the null hypothesis of no positive serial correlation, the DW statistic is 2.0. Positive serial correlation will lead to a DW statistic that is less than 2.0. From the table in Problem 1, we see that the DW statistic is 0.1878. To see whether this result is significantly less than 2.0, refer to the Durbin–Watson table in Appendix E at the end of this volume, in the column marked $k = 1$ (one independent variable) and the row corresponding to 60 observations. We see that $d_l = 1.61$. Because our DW statistic is clearly less than d_l, we reject the null hypothesis of no serial correlation at the 0.05 significance level.

The presence of serial correlation in the error term violates one of the regression assumptions. The standard errors of the estimated coefficients will be biased downward, so we cannot conduct hypothesis testing on the coefficients.

2 The difference between UER and its forecast value, PRED, is the forecast error. In an appropriately specified regression model, the forecast errors are randomly distributed around the regression line and have a constant variance. We can see that the errors from this model specification are persistent. The errors tend first to be above the regression line, and then, starting in 2014, they tend to be below the regression line until 2017, when they again are persistently above the regression line. This persistence suggests that the errors are positively serially correlated. Therefore, we conclude that the model is not appropriate for making estimates.

3 A log-linear model captures growth at a constant rate. The log-linear model $\ln(\text{Sales}_t) = b_0 + b_1 t + \varepsilon_t$ would be the simplest model consistent with a constant growth rate for monthly sales. Note that we would need to confirm that the regression assumptions are satisfied before accepting the model as valid.

4 **A** The plot of the series ΔUER_t seems to fluctuate around a constant mean; its volatility appears to be constant throughout the period. Our initial judgment is that the differenced series is covariance stationary.

B The change in the unemployment rate seems covariance stationary, so we should first estimate an AR(1) model and test to see whether the residuals from this model have significant serial correlation. If the residuals do not display significant serial correlation, we should use the AR(1) model. If the residuals do display significant serial correlation, we should try an AR(2) model and test for serial correlation of the residuals of the AR(2) model. We should continue this procedure until the errors from the final AR(p) model are serially uncorrelated.

5 The DW statistic cannot be appropriately used for a regression that has a lagged value of the dependent variable as one of the explanatory variables. To test for serial correlation, we need to examine the autocorrelations.

6 When a covariance-stationary series is at its mean-reverting level, the series will tend not to change until it receives a shock (ε_t). So, if the series ΔUER_t is at the mean-reverting level, $\Delta UER_t = \Delta UER_{t-1}$. This implies that $\Delta UER_t = -0.0668$

$-\ 0.2320\Delta UER_t$, so that $(1 + 0.2320)\Delta UER_t = -0.0668$ and $\Delta UER_t = -0.0668/(1 + 0.2320) = -0.0542$. The mean-reverting level is −0.0542. In an AR(1) model, the general expression for the mean-reverting level is $b_0/(1 - b_1)$.

7 **A** The predicted change in the unemployment rate for next period is −7.38%, found by substituting 0.0300 into the forecasting model: $-0.0668 - 0.2320(0.03) = -0.0738$.

B If we substitute our one-period-ahead forecast of −0.0738 into the model (using the chain rule of forecasting), we get a two-period-ahead forecast of −0.0497, or −4.97%.

C The answer to Part B is quite close to the mean-reverting level of −0.0542. A stationary time series may need many periods to return to its equilibrium, mean-reverting level.

8 The forecast of sales is $13,647 million for the first quarter of 2020 and $13,800 million for the second quarter of 2002, as the following table shows.

Date	Sales ($ Millions)	Log of Sales	Actual Value of Changes in the Log of Sales $\Delta\ln(\text{Sales}_t)$	Forecast Value of Changes in the Log of Sales $\Delta\ln(\text{Sales}_t)$
1Q 2019	13,072	9.4782	0.0176	
2Q 2019	12,446	9.4292	−0.0491	
3Q 2019	12,958	9.4695	0.4030	
4Q 2019	13,428	9.5051	0.0356	
1Q 2020	13,647	9.5213		0.0162
2Q 2020	13,800	9.5324		0.0111

We find the forecasted change in the log of sales for the first quarter of 2020 by inputting the value for the change in the log of sales from the previous quarter into the equation $\Delta\ln(\text{Sales}_t) = 0.0068 + 0.2633\Delta\ln(\text{Sales}_{t-1})$. Specifically, $\Delta\ln(\text{Sales}_t) = 0.0068 + 0.2633(0.0356) = 0.0162$, which means that we forecast the log of sales in the first quarter of 2020 to be $9.5051 + 0.0162 = 9.5213$.

Next, we forecast the change in the log of sales for the second quarter of 2020 as $\Delta\ln(\text{Sales}_t) = 0.0068 + 0.2633(0.0162) = 0.0111$. Note that we have to use our first-quarter 2020 estimated value of the change in the log of sales as our input for $\Delta\ln(\text{Sales}_{t-1})$ because we are forecasting past the period for which we have actual data.

With a forecasted change of 0.0111, we forecast the log of sales in the second quarter of 2020 to be $9.5213 + 0.0111 = 9.5324$.

We have forecasted the log of sales in the first and second quarters of 2020 to be 9.5213 and 9.5324, respectively. Finally, we take the antilog of our estimates of the log of sales in the first and second quarters of 2020 to get our estimates of the level of sales: $e^{9.5213} = 13{,}647$ and $e^{9.5324} = 13{,}800$, respectively, for sales of $13,647 million and $13,800 million.

9 **A** The RMSE of the out-of-sample forecast errors is approximately 3.6%. Out-of-sample error refers to the difference between the realized value and the forecasted value of $\Delta\ln(\text{Sales}_t)$ for dates beyond the estimation period. In this case, the out-of-sample period is 1Q 2019 to 4Q 2019. These are the four quarters for which we have data that we did not use to obtain the estimated model $\Delta\ln(\text{Sales}_t) = 0.0068 + 0.2633\Delta\ln(\text{Sales}_{t-1})$.

The steps to calculate RMSE are as follows:

i. Take the difference between the actual and the forecast values. This is the error.

ii. Square the error.

iii. Sum the squared errors.

iv. Divide by the number of forecasts.

v. Take the square root of the average.

We show the calculations for RMSE in the following table.

Actual Values of Changes in the Log of Sales $\Delta\ln(\text{Sales}_t)$	Forecast Values of Changes in the Log of Sales $\Delta\ln(\text{Sales}_t)$	Error (Column 1 – Column 2)	Squared Error (Column 3 Squared)
0.0176	0.0147	0.0029	0.0000
−0.0491	0.0107	−0.0598	0.0036
0.0403	0.0096	0.0307	0.0009
0.0356	0.0093	0.0263	0.0007
		Sum	0.0052
		Mean	0.0013
		RMSE	0.036

B The lower the RMSE, the more accurate the forecasts of a model in forecasting. Therefore, the model with the RMSE of 2% has greater accuracy in forecasting than the model in Part A, which has an RMSE of 3.6%.

10 **A** Predictions too far ahead can be nonsensical. For example, the AR(1) model we have been examining, $\Delta\text{UER}_t = -0.0405 - 0.4674\Delta\text{UER}_{t-1}$, taken at face value, predicts declining civilian unemployment into the indefinite future. Because the civilian unemployment rate will probably not go below 3% frictional unemployment and cannot go below 0% unemployment, this model's long-range forecasts are implausible. The model is designed for short-term forecasting, as are many time-series models.

B Using more years of data for estimation may lead to nonstationarity even in the series of first differences in the civilian unemployment rate. As we go further back in time, we increase the risk that the underlying civilian unemployment rate series has more than one regime (or true model). If the series has more than one regime, fitting one model to the entire period would not be correct. Note that when we have good reason to believe that a time series is stationary, a longer series of data is generally desirable.

11 **A** The graph of $\ln(\text{Sales}_t)$ appears to trend upward over time. A series that trends upward or downward over time often has a unit root and is thus not covariance stationary. Therefore, using an AR(1) regression on the undifferenced series is probably not correct. In practice, we need to examine regression statistics to confirm such visual impressions.

B The most common way to transform a time series with a unit root into a covariance-stationary time series is to difference the data—that is, to create a new series: $\Delta\ln(\text{Sales}_t) = \ln(\text{Sales}_t) - \ln(\text{Sales}_{t-1})$.

12 The plot of the series $\Delta\ln(\text{Sales}_t)$ appears to fluctuate around a constant mean; its volatility seems constant throughout the period. Differencing the data appears to have made the time series covariance stationary.

13 A In a correctly specified regression, the residuals must be serially uncorrelated. We have 108 observations, so the standard error of the autocorrelation is $1/\sqrt{T}$, or in this case $1/\sqrt{108} = 0.0962$. The t-statistic for each lag is significant at the 0.01 level. We would have to modify the model specification before continuing with the analysis.

B Because the residuals from the AR(1) specification display significant serial correlation, we should estimate an AR(2) model and test for serial correlation of the residuals of the AR(2) model. If the residuals from the AR(2) model are serially uncorrelated, we should then test for seasonality and ARCH behavior. If any serial correlation remains in the residuals, we should estimate an AR(3) process and test the residuals from that specification for serial correlation. We should continue this procedure until the errors from the final AR(p) model are serially uncorrelated. When serial correlation is eliminated, we should test for seasonality and ARCH behavior.

14 A The series has a steady upward trend of growth, suggesting an exponential growth rate. This finding suggests transforming the series by taking the natural log and differencing the data.

B First, we should determine whether the residuals from the AR(1) specification are serially uncorrelated. If the residuals are serially correlated, then we should try an AR(2) specification and then test the residuals from the AR(2) model for serial correlation. We should continue in this fashion until the residuals are serially uncorrelated and then look for seasonality in the residuals. If seasonality is present, we should add a seasonal lag. If no seasonality is present, we should test for ARCH. If ARCH is not present, we can conclude that the model is correctly specified.

C If the model $\Delta\ln(\text{Sales}_t) = b_0 + b_1[\Delta\ln(\text{Sales}_{t-1})] + \varepsilon_t$ is correctly specified, then the series $\Delta\ln(\text{Sales}_t)$ is covariance stationary. So, this series tends to its mean-reverting level, which is $b_0/(1 - b_1)$, or $0.0661/(1 - 0.4698) = 0.1247$.

15 The quarterly sales of Avon show an upward trend and a clear seasonal pattern, as indicated by the repeated regular cycle.

16 A A second explanatory variable, the change in the gross profit margin lagged four quarters, ΔGPM_{t-4}, was added.

B The model was augmented to account for seasonality in the time series (with quarterly data, significant autocorrelation at the fourth lag indicates seasonality). The standard error of the autocorrelation coefficient equals 1 divided by the square root of the number of observations: $1/\sqrt{40}$, or 0.1581. The autocorrelation at the fourth lag (0.8496) is significant: $t = 0.8496/0.1581 = 5.37$. This indicates seasonality, and accordingly we added ΔGPM_{t-4}. Note that in the augmented regression, the coefficient on ΔGPM_{t-4} is highly significant. (Although the autocorrelation at second lag is also significant, the fourth lag is more important because of the rationale of seasonality. Once the fourth lag is introduced as an independent variable, we might expect that the second lag in the residuals would not be significant.)

17 A In order to determine whether this model is correctly specified, we need to test for serial correlation among the residuals. We want to test whether we can reject the null hypothesis that the value of each autocorrelation is 0 against the alternative hypothesis that each is not equal to 0. At the 0.05 significance level, with 68 observations and three parameters, this model has 65 degrees of freedom. The critical value of the t-statistic needed to reject the null hypothesis is thus about 2.0. The absolute value of the t-statistic for each autocorrelation is below 0.60 (less than 2.0), so we cannot reject the null hypothesis that each autocorrelation is not significantly different from 0. We have determined that the model is correctly specified.

B If sales grew by 1% last quarter and by 2% four quarters ago, then the model predicts that sales growth this quarter will be 0.0121 – 0.0839[ln(1.01)] + 0.6292[ln(1.02)] = $e^{0.02372}$ – 1 = 2.40%.

18 We should estimate the regression $\Delta\text{UER}_t = b_0 + b_1\Delta\text{UER}_{t-1} + \varepsilon_t$ and save the residuals from the regression. Then we should create a new variable, $\hat{\varepsilon}_t^2$, by squaring the residuals. Finally, we should estimate $\hat{\varepsilon}_t^2 = a_0 + a_1\hat{\varepsilon}_{t-1}^2 + u_t$ and test to see whether a_1 is statistically different from 0.

19 To determine whether we can use linear regression to model more than one time series, we should first determine whether any of the time series has a unit root. If none of the time series has a unit root, then we can safely use linear regression to test the relations between the two time series. Note that if one of the two variables has a unit root, then our analysis would not provide valid results; if both of the variables have unit roots, then we would need to evaluate whether the variables are cointegrated.

20 C is correct. The predicted value for period t from a linear trend is calculated as $\hat{y}_t = \hat{b}_0 + \hat{b}_1(t)$.

October 2015 is the second month out of sample, or t = 183. So, the predicted value for October 2015 is calculated as

$$\hat{y}_t = 28.3278 + 0.4086(183) = \$103.10.$$

Therefore, the predicted WTI oil price for October 2015 based on the linear trend model is $103.10.

21 C is correct. The predicted value for period t from a log-linear trend is calculated as $\ln\hat{y}_t = \hat{b}_0 + \hat{b}_1(t)$.

September 2015 is the first month out of sample, or t = 182. So, the predicted value for September 2015 is calculated as follows:

$$\ln \hat{y}_t = 3.3929 + 0.0075(182).$$

$$\ln \hat{y}_t = 4.7579.$$

$$\hat{y}_t = e^{4.7579} = \$116.50.$$

Therefore, the predicted WTI oil price for September 2015, based on the log-linear trend model, is $116.50.

22 B is correct. The Durbin–Watson statistic for the linear trend model is 0.10 and for the log-linear trend model is 0.08. Both of these values are below the critical value of 1.75. Therefore, we can reject the hypothesis of no positive serial correlation in the regression errors in both the linear trend model and the log-linear trend model.

23 B is correct. There are three requirements for a time series to be covariance stationary. First, the expected value of the time series must be constant and finite in all periods. Second, the variance of the time series must be constant and finite in all periods. Third, the covariance of the time series with itself for a fixed number of periods in the past or future must be constant and finite in all periods. Martinez concludes that the mean and variance of the time series of WTI oil prices are not constant over time. Therefore, the time series is not covariance stationary.

24 B is correct. The last two observations in the WTI time series are July and August 2015, when the WTI oil price was \$51.16 and \$42.86, respectively. Therefore, September 2015 represents a one-period-ahead forecast. The one-period-ahead forecast from an AR(2) model is calculated as

$$\hat{x}_{t+1} = \hat{b}_0 + \hat{b}_1 x_t + \hat{b}_2 x_{t-1}.$$

So, the one-period-ahead (September 2015) forecast is calculated as

$$\hat{x}_{t+1} = 2.0017 + 1.3946(\$42.86) - 0.4249(\$51.16) = \$40.04.$$

Therefore, the September 2015 forecast based on the AR(2) model is \$40.04.

25 C is correct. The standard error of the autocorrelations is calculated as $\frac{1}{\sqrt{T}}$, where T represents the number of observations used in the regression.

Therefore, the standard error for each of the autocorrelations is $\frac{1}{\sqrt{180}} = 0.0745$.

Martinez can conclude that the residuals are serially correlated and are significantly different from zero because two of the four autocorrelations in Exhibit 2 have a t-statistic in absolute value that is greater than the critical value of 1.97.

Choices A and B are incorrect because two of the four autocorrelations have a t-statistic in absolute value that is greater than the critical value of the t-statistic of 1.97.

26 C is correct. The mean-reverting level from the AR(1) model is calculated as

$$\hat{x}_t = \frac{b_0}{1 - b_1} = \frac{1.5948}{1 - 0.9767} = \$68.45.$$

Therefore, the mean-reverting WTI oil price from the AR(1) model is \$68.45. The forecasted oil price in September 2015 will likely be greater than \$42.86 because the model predicts that the price will rise in the next period from the August 2015 price of \$42.86.

27 C is correct. A random walk can be described by the equation $x_t = b_0 + b_1 x_{t-1} + \varepsilon_t$, where $b_0 = 0$ and $b_1 = 1$. So $b_0 = 0$ is a characteristic of a random walk time series. A covariance-stationary series must satisfy the following three requirements:

1. The expected value of the time series must be constant and finite in all periods.
2. The variance of the time series must be constant and finite in all periods.
3. The covariance of the time series with itself for a fixed number of periods in the past or future must be constant and finite in all periods.

$b_0 = 0$ does not violate any of these three requirements and is thus consistent with the properties of a covariance-stationary time series.

28 B is correct. The critical t-statistic at a 5% confidence level is 1.98. As a result, neither the intercept nor the coefficient on the first lag of the first-differenced exchange rate in Regression 2 differs significantly from zero. Also, the residual autocorrelations do not differ significantly from zero. As a result, Regression 2 can be reduced to $y_t = \varepsilon_t$, with a mean-reverting level of $b_0/(1 - b_1) = 0/1 = 0$. Therefore, the variance of y_t in each period is $\text{var}(\varepsilon_t) = \sigma^2$. The fact that the residuals are not autocorrelated is consistent with the covariance of the times series with itself being constant and finite at different lags. Because the variance and the mean of y_t are constant and finite in each period, we can also conclude that y_t is covariance stationary.

29 A is correct. If the exchange rate series is a random walk, then the first-differenced series will yield $b_0 = 0$ and $b_1 = 0$ and the error terms will not be serially correlated. The data in Exhibit 1 show that this is the case: Neither the intercept nor the coefficient on the first lag of the first-differenced exchange rate in Regression 2 differs significantly from zero because the t-statistics of both coefficients are less than the critical t-statistic of 1.98. Also, the residual autocorrelations do not differ significantly from zero because the t-statistics of all autocorrelations are less than the critical t-statistic of 1.98. Therefore, because all random walks have unit roots, the exchange rate time series used to run Regression 1 has a unit root.

30 C is correct. To conduct the Dickey–Fuller test, one must subtract the independent variable, x_{t-1}, from both sides of the original AR(1) model. This results in a change of the dependent variable (from x_t to $x_t - x_{t-1}$) and a change in the regression's slope coefficient (from b_1 to $b_1 - 1$) but not a change in the independent variable.

31 C is correct. The regression output in Exhibit 2 suggests there is serial correlation in the residual errors. The fourth autocorrelation of the residual has a value of 0.6994 and a t-statistic of 4.3111, which is greater than the t-statistic critical value of 2.02. Therefore, the null hypothesis that the fourth autocorrelation is equal to zero can be rejected. This indicates strong and significant seasonal autocorrelation, which means the Regression 3 equation is misspecified.

32 C is correct. The quarterly sales for March 2016 are calculated as follows:

$$\begin{aligned}
\ln \text{Sales}_t - \ln \text{Sales}_{t-1} &= b_0 + b_1(\ln \text{Sales}_{t-1} - \ln \text{Sales}_{t-2}) + b_2(\ln \text{Sales}_{t-4} - \\
&\quad \ln \text{Sales}_{t-5}).\\
\ln \text{Sales}_t - \ln 3.868 &= 0.0092 - 0.1279(\ln 3.868 - \ln 3.780) + 0.7239(\ln \\
&\quad 3.836 - \ln 3.418).\\
\ln \text{Sales}_t - 1.35274 &= 0.0092 - 0.1279(1.35274 - 1.32972) + \\
&\quad 0.7239(1.34443 - 1.22906).\\
\ln \text{Sales}_t &= 1.35274 + 0.0092 - 0.1279(0.02301) + \\
&\quad 0.7239(0.11538).\\
\ln \text{Sales}_t &= 1.44251.\\
\text{Sales}_t &= e^{1.44251} = 4.231.
\end{aligned}$$

33 B is correct. Exhibit 3 shows that the time series of the stock prices of Company 1 exhibits heteroskedasticity, as evidenced by the fact that the time series is ARCH(1). If a time series is ARCH(1), then the variance of the error in one period depends on the variance of the error in previous periods. Therefore, the variance of the errors in period $t + 1$ can be predicted in period t using the formula

$$\hat{\sigma}_{t+1}^2 = \hat{a}_0 + \hat{a}_1\hat{\varepsilon}_t^2.$$

34 B is correct. When two time series have a unit root but are cointegrated, the error term in the linear regression of one time series on the other will be covariance stationary. Exhibit 5 shows that the series of stock prices of Company 2 and the oil prices both contain a unit root and the two time series are cointegrated. As a result, the regression coefficients and standard errors are consistent and can be used for hypothesis tests. Although the cointegrated regression estimates the long-term relation between the two series, it may not be the best model of the short-term relationship.

35 C is correct. As a result of the exponential trend in the time series of stock prices for Company 3, Busse would want to take the natural log of the series and then first-difference it. Because the time series also has serial correlation in the residuals from the trend model, Busse should use a more complex model, such as an autoregressive (AR) model.

Quantitative Methods (2)

This study session provides coverage on techniques that underlie how financial technology (fintech) is affecting areas within the investment industry, such as investment analysis, automated advice, and risk management. The first reading introduces techniques in machine learning (ML) that involve clustering, simplifying, classifying, and predicting relationships in the large datasets that are often found in finance. The session concludes with a reading that examines how data projects involving large datasets are structured with an application to sentiment analysis in investment analysis using machine learning techniques for natural language processing (NLP).

READING ASSIGNMENTS

Reading 4	Machine Learning by Kathleen DeRose, CFA, Matthew Dixon, PhD, FRM, and Christophe Le Lannou
Reading 5	Big Data Projects by Sreekanth Mallikarjun, PhD, and Ahmed Abbasi, PhD

© 2021 CFA Institute. All rights reserved.

READING

4

Machine Learning

by Kathleen DeRose, CFA, Matthew Dixon, PhD, FRM, and Christophe Le Lannou

Kathleen DeRose, CFA, is at New York University, Stern School of Business (USA). Matthew Dixon, PhD, FRM, is at Illinois Institute of Technology, Stuart School of Business (USA). Christophe Le Lannou is at dataLearning (United Kingdom).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe supervised machine learning, unsupervised machine learning, and deep learning;
☐	**b.** describe overfitting and identify methods of addressing it;
☐	**c.** describe supervised machine learning algorithms—including penalized regression, support vector machine, k-nearest neighbor, classification and regression tree, ensemble learning, and random forest—and determine the problems for which they are best suited;
☐	**d.** describe unsupervised machine learning algorithms—including principal components analysis, k-means clustering, and hierarchical clustering—and determine the problems for which they are best suited;
☐	**e.** describe neural networks, deep learning nets, and reinforcement learning.

INTRODUCTION

1

Investment firms are increasingly using technology at every step of the investment management value chain—from improving their understanding of clients to uncovering new sources of alpha and executing trades more efficiently. Machine learning techniques, a central part of that technology, are the subject of this reading. These techniques first appeared in finance in the 1990s and have since flourished with the explosion of data and cheap computing power.

This reading provides a high-level view of machine learning (ML). It covers a selection of key ML algorithms and their investment applications. Investment practitioners should be equipped with a basic understanding of the types of investment

© 2021 CFA Institute. All rights reserved.

problems that machine learning can address, an idea of how the algorithms work, and the vocabulary to interact with machine learning and data science experts. While investment practitioners need not master the details and mathematics of machine learning, as domain experts in investments they can play an important role in the implementation of these techniques by being able to source appropriate model inputs, interpret model outputs, and translate outputs into appropriate investment actions.

Section 1 gives an overview of machine learning in investment management. Section 2 defines machine learning and the types of problems that can be addressed by supervised and unsupervised learning. Section 3 describes evaluating machine learning algorithm performance. Key supervised machine learning algorithms are covered in Sections 4–8, and Sections 9–12 describe key unsupervised machine learning algorithms. Neural networks, deep learning nets, and reinforcement learning are covered in Sections 13 and 14. Section 15 provides a decision flowchart for selecting the appropriate ML algorithm. The reading concludes with a summary.

1.1 Machine Learning and Investment Management

a describe supervised machine learning, unsupervised machine learning, and deep learning

The growing volume and exploding diversity of data, as well as the perceived increasing economic value of insights extracted from these data, have inspired rapid growth in data science. This newly emerging field combines mathematics, computer science, and business analytics. It also strikes out in a new direction that relies on learning—from basic learning functions that map relationships between variables to advanced neural networks that mimic physical processes that absorb, order, and adapt to information.

Machine learning has theoretical and practical implications for investment management. For example, machine learning could potentially reshape accepted wisdom about asset risk premiums and reconfigure investment management business processes. Large datasets and learning models are already affecting investment management practices—from client profiling to asset allocation, stock selection, portfolio construction and risk management, and trading.

Machine learning applications are at each step of the asset and wealth management value chain. Chatbots answer basic retirement savings questions, learning from their interactions with investors. Machine learning methods can be used to generate alpha signals used in security selection by creating a non-linear forecast for a single time series, by deriving a forecast from a suite of predefined factors, or even by choosing input signals from existing or newly found data. For example, researchers using textual analysis have found that year-over-year changes in annual (10-K) and quarterly (10-Q) filings, particularly negative changes in the management discussion and risk sections, can strongly predict equity returns.

Machine learning methods can help calculate target portfolio weights that incorporate client restrictions and then dynamically weight them to maximize a Sharpe ratio. Another use of machine learning methods is better estimation of the variance–covariance matrix via principal components analysis, which reduces the number of variables needed to explain the variation in the data. Research suggests that machine learning solutions outperform mean–variance optimization in portfolio construction. Machine learning techniques are already creating better order flow management tools with non-linear trading algorithms that reduce the costs of implementing portfolio decisions. These developments have caused an evolution in the automation of tools, processes, and businesses (such as robo-advising).

WHAT IS MACHINE LEARNING

2

a describe supervised machine learning, unsupervised machine learning, and deep learning

We now discuss some fundamental concepts of machine learning, including a definition and an overview of key types of machine learning, such as supervised and unsupervised ML.

2.1 Defining Machine Learning

Statistical approaches and machine learning techniques both analyze observations to reveal some underlying process; however, they diverge in their assumptions, terminology, and techniques. Statistical approaches rely on foundational assumptions and explicit models of structure, such as observed samples that are assumed to be drawn from a specified underlying probability distribution. These a priori restrictive assumptions can fail in reality.

In contrast, machine learning seeks to extract knowledge from large amounts of data with fewer such restrictions. The goal of machine learning algorithms is to automate decision-making processes by generalizing (i.e., "learning") from known examples to determine an underlying structure in the data. The emphasis is on the ability of the algorithm to generate structure or predictions from data without any human help. An elementary way to think of ML algorithms is to "find the pattern, apply the pattern."

Machine learning techniques are better able than statistical approaches (such as linear regression) to handle problems with many variables (high dimensionality) or with a high degree of non-linearity. ML algorithms are particularly good at detecting change, even in highly non-linear systems, because they can detect the preconditions of a model's break or anticipate the probability of a regime switch.

Machine learning is broadly divided into three distinct classes of techniques: supervised learning, unsupervised learning, and deep learning/reinforcement learning.

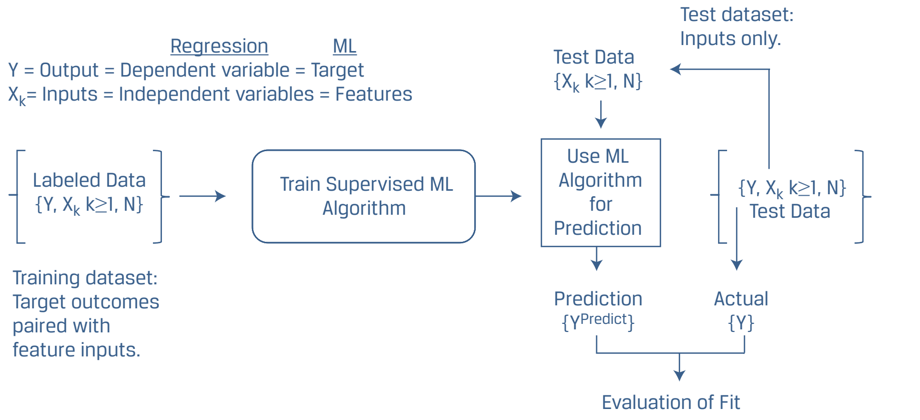
2.2 Supervised Learning

Supervised learning involves ML algorithms that infer patterns between a set of inputs (the *X*'s) and the desired output (*Y*). The inferred pattern is then used to map a given input set into a predicted output. Supervised learning requires a **labeled dataset**, one that contains matched sets of observed inputs and the associated output. Applying the ML algorithm to this dataset to infer the pattern between the inputs and output is called "training" the algorithm. Once the algorithm has been trained, the inferred pattern can be used to predict output values based on new inputs (i.e., ones not in the training dataset).

Multiple regression is an example of supervised learning. A regression model takes matched data (*X*'s, *Y*) and uses it to estimate parameters that characterize the relationship between *Y* and the *X*'s. The estimated parameters can then be used to predict *Y* on a new, different set of *X*'s. The difference between the predicted and actual *Y* is used to evaluate how well the regression model predicts out-of-sample (i.e., using new data).

The terminology used with ML algorithms differs from that used in regression. Exhibit 1 provides a visual of the supervised learning model training process and a translation between regression and ML terminologies.

Exhibit 1 Overview of Supervised Learning



In supervised machine learning, the dependent variable (Y) is the **target** and the independent variables (X's) are known as **features**. The labeled data (training dataset) is used to train the supervised ML algorithm to infer a pattern-based prediction rule. The fit of the ML model is evaluated using labeled test data in which the predicted targets ($Y^{Predict}$) are compared to the actual targets (Y^{Actual}).

An example of supervised learning is the case in which ML algorithms are used to predict whether credit card transactions are fraudulent or legitimate. In the credit card example, the target is a binary variable with a value of 1 for "fraudulent" or 0 for "non-fraudulent." The features are the transaction characteristics. The chosen ML algorithm uses these data elements to train a model to predict the likelihood of fraud more accurately in new transactions. The ML program "learns from experience" if the percentage of correctly predicted credit card transactions increases as the amount of input from a growing credit card database increases. One possible ML algorithm to use would be to fit a logistic regression model to the data to provide an estimate of the probability a transaction is fraudulent.

Supervised learning can be divided into two categories of problems—regression and classification—with the distinction between them being determined by the nature of the target (Y) variable. If the target variable is continuous, then the task is one of regression (even if the ML technique used is not "regression"; note this nuance of ML terminology). If the target variable is categorical or ordinal (i.e., a ranked category), then it is a classification problem. Regression and classification use different ML techniques.

Regression focuses on making predictions of continuous target variables. Most readers are already familiar with multiple linear regression (e.g., ordinary least squares) models, but other supervised learning techniques exist, including non-linear models. These non-linear models are useful for problems involving large datasets with large numbers of features, many of which may be correlated. Some examples of problems belonging to the regression category are using historical stock market returns to forecast stock price performance or using historical corporate financial ratios to forecast the probability of bond default.

Classification focuses on sorting observations into distinct categories. In a regression problem, when the dependent variable (target) is categorical, the model relating the outcome to the independent variables (features) is called a "classifier." You should already be familiar with logistic regression as a type of classifier. Many classification models are binary classifiers, as in the case of fraud detection for credit card transactions. Multi-category classification is not uncommon, as in the case of classifying firms into multiple credit rating categories. In assigning ratings, the outcome variable is ordinal, meaning the categories have a distinct order or ranking (e.g., from low to high creditworthiness). Ordinal variables are intermediate between categorical variables and continuous variables on a scale of measurement.

2.3 Unsupervised Learning

Unsupervised learning is machine learning that does not make use of labeled data. More formally, in unsupervised learning, we have inputs (X's) that are used for analysis without any target (Y) being supplied. In unsupervised learning, because the ML algorithm is not given labeled training data, the algorithm seeks to discover structure within the data themselves. As such, unsupervised learning is useful for exploring new datasets because it can provide human experts with insights into a dataset too big or too complex to visualize.

Two important types of problems that are well suited to unsupervised machine learning are reducing the dimension of data and sorting data into clusters, known as dimension reduction and clustering, respectively.

Dimension reduction focuses on reducing the number of features while retaining variation across observations to preserve the information contained in that variation. Dimension reduction may have several purposes. It may be applied to data with a large number of features to produce a lower dimensional representation (i.e., with fewer features) that can fit, for example, on a computer screen. Dimension reduction is also used in many quantitative investment and risk management applications where it is critical to identify the most predictive factors underlying asset price movements.

Clustering focuses on sorting observations into groups (clusters) such that observations in the same cluster are more similar to each other than they are to observations in other clusters. Groups are formed based on a set of criteria that may or may not be prespecified (such as the number of groups). Clustering has been used by asset managers to sort companies into groupings driven by data (e.g., based on their financial statement data or corporate characteristics) rather than conventional groupings (e.g., based on sectors or countries).

2.4 Deep Learning and Reinforcement Learning

More broadly in the field of artificial intelligence, additional categories of machine learning algorithms are distinguished. In **deep learning**, sophisticated algorithms address complex tasks, such as image classification, face recognition, speech recognition, and natural language processing. Deep learning is based on **neural networks** (NNs), also called artificial neural networks (ANNs)—highly flexible ML algorithms that have been successfully applied to a variety of supervised and unsupervised tasks characterized by large datasets, non-linearities, and interactions among features. In **reinforcement learning**, a computer learns from interacting with itself or data generated by the same algorithm. Deep learning and reinforcement learning principles have been combined to create efficient algorithms for solving a range of highly complex problems in robotics, health care, and finance.

2.5 Summary of ML Algorithms and How to Choose among Them

Exhibit 2 is a guide to the various machine learning algorithms organized by algorithm type (supervised or unsupervised) and by type of variables (continuous, categorical, or both). We will not cover linear or logistic regression since they are covered elsewhere in readings on quantitative methods. The extensions of linear regression, such as penalized regression and least absolute shrinkage and selection operator (LASSO), as well as the other ML algorithms shown in Exhibit 2, will be covered in this reading.

Exhibit 2 Guide to ML Algorithms

	ML Algorithm Type	
Variables	**Supervised (Target Variable)**	**Unsupervised (No Target Variable)**
Continuous	**Regression**	**Dimension Reduction**
	• Linear; Penalized Regression/LASSO	• Principal Components Analysis (PCA)
	• Logistic	**Clustering**
	• Classification and Regression Tree (CART)	• *K*-Means
	• Random Forest	• Hierarchical
Categorical	**Classification**	**Dimension Reduction**
	• Logistic	• Principal Components Analysis (PCA)
	• Support Vector Machine (SVM)	**Clustering**
	• *K*-Nearest Neighbor (KNN)	• *K*-Means
	• Classification and Regression Tree (CART)	• Hierarchical
Continuous or Categorical	Neural Networks	Neural Networks
	Deep Learning	Deep Learning
	Reinforcement Learning	Reinforcement Learning

EXAMPLE 1

Machine Learning Overview

1. Which of the following *best* describes machine learning? Machine learning:
   - **A** is a type of computer algorithm used just for linear regression.
   - **B** is a set of algorithmic approaches aimed at generating structure or predictions from data without human intervention by finding a pattern and then applying the pattern.
   - **C** is a set of computer-driven approaches adapted to extracting information from linear, labeled datasets.
2. Which of the following statements is *most* accurate? When attempting to discover groupings of data without any target (Y) variable:
   - **A** an unsupervised ML algorithm is used.
   - **B** an ML algorithm that is given labeled training data is used.
   - **C** a supervised ML algorithm is used.
3. Which of the following statements concerning supervised learning *best* distinguishes it from unsupervised learning? Supervised learning involves:
   - **A** training on labeled data to infer a pattern-based prediction rule.
   - **B** training on unlabeled data to infer a pattern-based prediction rule.
   - **C** learning from unlabeled data by discovering underlying structure in the data themselves.

4 Which of the following *best* describes dimension reduction? Dimension reduction:

A focuses on classifying observations in a dataset into known groups using labeled training data.

B focuses on clustering observations in a dataset into unknown groups using unlabeled data.

C focuses on reducing the number of features in a dataset while retaining variation across observations to preserve the information in that variation.

Solution to 1:

B is correct. A is incorrect because machine learning algorithms are typically not used for linear regression. C is incorrect because machine learning is not limited to extracting information from linear, labeled datasets.

Solution to 2:

A is correct. B is incorrect because the term "labeled training data" means the target (Y) is provided. C is incorrect because a supervised ML algorithm is meant to predict a target (Y) variable.

Solution to 3:

A is correct. B is incorrect because supervised learning uses labeled training data. C is incorrect because it describes unsupervised learning.

Solution to 4:

C is correct. A is incorrect because it describes classification, not dimension reduction. B is incorrect because it describes clustering, not dimension reduction.

OVERVIEW OF EVALUATING ML ALGORITHM PERFORMANCE 3

b describe overfitting and identify methods of addressing it

Machine learning algorithms promise several advantages relative to a structured statistical approach in exploring and analyzing the structure of very large datasets. ML algorithms have the ability to uncover complex interactions between feature variables and the target variable, and they can process massive amounts of data quickly. Moreover, many ML algorithms can easily capture non-linear relationships and may be able to recognize and predict structural changes between features and the target. These advantages mainly derive from the non-parametric and non-linear models that allow more flexibility when inferring relationships.

The flexibility of ML algorithms comes with a price, however. ML algorithms can produce overly complex models with results that are difficult to interpret, may be sensitive to noise or particulars of the data, and may fit the training data too well. An ML algorithm that fits the training data too well will typically not predict well using new data. This problem is known as **overfitting**, and it means that the fitted algorithm does not **generalize** well to new data. A model that generalizes well is a model that retains its explanatory power when predicting using out-of-sample (i.e., new) data. An overfit model has incorporated the noise or random fluctuations in the training data into its learned relationship. The problem is that these aspects often do not apply to new data the algorithm receives and so will negatively impact the model's ability to generalize, therefore reducing its overall predictive value. The evaluation of any ML algorithm thus focuses on its prediction error on new data rather than on its goodness of fit on the data with which the algorithm was fitted (i.e., trained).

Generalization is an objective in model building, so the problem of overfitting is a challenge to attaining that objective. These two concepts are the focus of the discussion below.

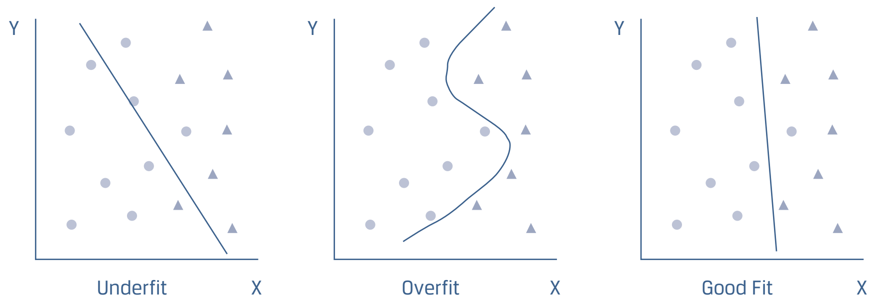
3.1 Generalization and Overfitting

To properly describe generalization and overfitting of an ML model, it is important to note the partitioning of the dataset to which the model will be applied. The dataset is typically divided into three non-overlapping samples: (1) **training sample** used to train the model, (2) **validation sample** for validating and tuning the model, and (3) **test sample** for testing the model's ability to predict well on new data. The training and validation samples are often referred to as being "in-sample," and the test sample is commonly referred to as being "out-of-sample." We will return shortly to the topic of partitioning the dataset.

To be valid and useful, any supervised machine learning model must generalize well beyond the training data. The model should retain its explanatory power when tested out-of-sample. As mentioned, one common reason for failure to generalize is overfitting. Think of overfitting as tailoring a custom suit that fits only one person. Continuing the analogy, underfitting is similar to making a baggy suit that fits no one, whereas robust fitting, the desired result, is similar to fashioning a universal suit that fits all people of similar dimensions.

The concepts of underfitting, overfitting, and good (or robust) fitting are illustrated in Exhibit 3. Underfitting means the model does not capture the relationships in the data. The left graph shows four errors in this underfit model (three misclassified circles and one misclassified triangle). Overfitting means training a model to such a degree of specificity to the training data that the model begins to incorporate noise coming from quirks or spurious correlations; it mistakes randomness for patterns and relationships. The algorithm may have memorized the data, rather than learned from it, so it has perfect hindsight but no foresight. The main contributors to overfitting are thus high noise levels in the data and too much complexity in the model. The middle graph shows no errors in this overfit model. **Complexity** refers to the number of features, terms, or branches in the model and to whether the model is linear or non-linear (non-linear is more complex). As models become more complex, overfitting risk increases. A good fit/robust model fits the training (in-sample) data well and generalizes well to out-of-sample data, both within acceptable degrees of error. The right graph shows that the good fitting model has only one error, the misclassified circle.

Exhibit 3 Underfitting, Overfitting, and Good Fitting



3.2 Errors and Overfitting

To capture these effects and calibrate degree of fit, data scientists compare error rates in- and out-of-sample as a function of both the data and the algorithm. Total in-sample errors (E_{in}) are generated by the predictions of the fitted relationship relative to actual target outcomes on the training sample. Total out-of-sample errors (E_{out}) are from either the validation or test samples. Low or no in-sample error but large out-of-sample error are indicative of poor generalization. Data scientists decompose the total out-of-sample error into three sources:

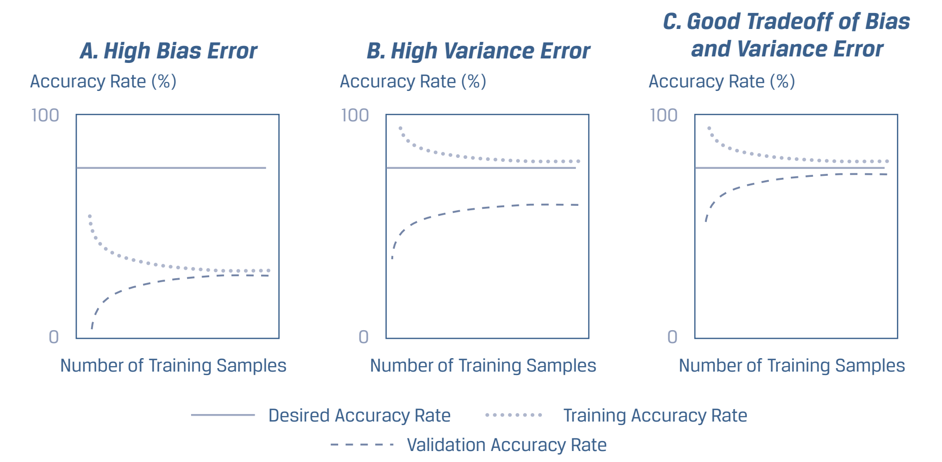
1 **Bias error**, or the degree to which a model fits the training data. Algorithms with erroneous assumptions produce high bias with poor approximation, causing underfitting and high in-sample error.

2 **Variance error**, or how much the model's results change in response to new data from validation and test samples. Unstable models pick up noise and produce high variance, causing overfitting and high out-of-sample error.

3 **Base error** due to randomness in the data.

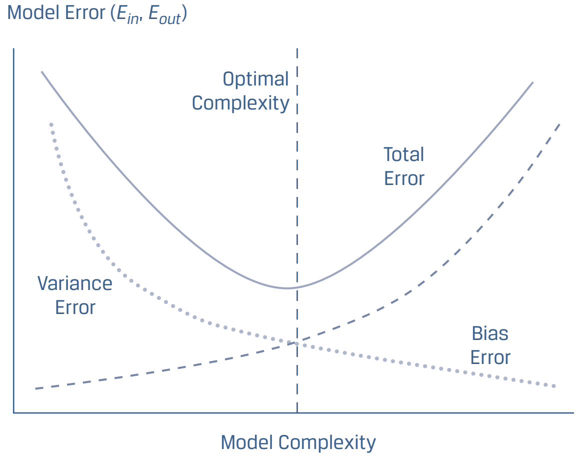
A **learning curve** plots the accuracy rate (= 1 – error rate) in the validation or test samples (i.e., out-of-sample) against the amount of data in the training sample, so it is useful for describing under- and overfitting as a function of bias and variance errors. If the model is robust, out-of-sample accuracy increases as the training sample size increases. This implies that error rates experienced in the validation or test samples (E_{out}) and in the training sample (E_{in}) converge toward each other and toward a desired error rate (or, alternatively, the base error). In an underfitted model with high bias error, shown in the left panel of Exhibit 4, high error rates cause convergence below the desired accuracy rate. Adding more training samples will not improve the model to the desired performance level. In an overfitted model with high variance error, shown in the middle panel of Exhibit 4, the validation sample and training sample error rates fail to converge. In building models, data scientists try to simultaneously minimize both bias and variance errors while selecting an algorithm with good predictive or classifying power, as seen in the right panel of Exhibit 4.

Exhibit 4 Learning Curves: Accuracy in Validation and Training Samples



Out-of-sample error rates are also a function of model complexity. As complexity increases in the training set, error rates (E_{in}) fall and bias error shrinks. As complexity increases in the test set, however, error rates (E_{out}) rise and variance error rises. Typically, linear functions are more susceptible to bias error and underfitting, while non-linear functions are more prone to variance error and overfitting. Therefore, an optimal point of model complexity exists where the bias and variance error curves intersect and in- and out-of-sample error rates are minimized. A **fitting curve**, which shows in- and out-of-sample error rates (E_{in} and E_{out}) on the y-axis plotted against model complexity on the x-axis, is presented in Exhibit 5 and illustrates this trade-off.

Exhibit 5 Fitting Curve Shows Trade-Off between Bias and Variance Errors and Model Complexity



Finding the optimal point (managing overfitting risk)—the point just before the total error rate starts to rise (due to increasing variance error)—is a core part of the machine learning process and the key to successful generalization. Data scientists express the trade-off between overfitting and generalization as a trade-off between *cost* (the difference between in- and out-of-sample error rates) and *complexity*. They use the trade-off between cost and complexity to calibrate and visualize under- and overfitting and to optimize their models.

3.3 Preventing Overfitting in Supervised Machine Learning

We have seen that overfitting impairs generalization, but overfitting potential is endemic to the supervised machine learning process due to the presence of noise. So, how do data scientists combat this risk? Two common methods are used to reduce overfitting: (1) preventing the algorithm from getting too complex during selection and training, which requires estimating an overfitting penalty, and (2) proper data sampling achieved by using **cross-validation**, a technique for estimating out-of-sample error directly by determining the error in validation samples.

The first strategy comes from Occam's razor, the problem-solving principle that the simplest solution tends to be the correct one. In supervised machine learning, it means limiting the number of features and penalizing algorithms that are too complex or too flexible by constraining them to include only parameters that reduce out-of-sample error.

The second strategy comes from the principle of avoiding sampling bias. But sampling bias can creep into machine learning in many ways. The challenge is having a large enough dataset to make both training and testing possible on representative samples. An unrepresentative sample or reducing the training sample size too much could obscure its true patterns, thereby increasing bias. In supervised machine learning, the technique for reducing sampling bias is through careful partitioning of the dataset into three groups: (1) training sample, the set of labeled training data where the target variable (Y) is known;(2) validation sample, the set of data used for making structural choices on the degree of model complexity, comparing various solutions, and tuning the selected model, thereby validating the model; and (3) test sample, the set of data held aside for testing to confirm the model's predictive or classifying power. The goal, of course, is to deploy the tested model on fresh data from the same domain.

To mitigate the problem of such **holdout samples** (i.e., data samples not used to train the model) reducing the training set size too much, modelers use special cross-validation techniques. One such technique is ***k*-fold cross-validation**, in which the data (excluding test sample and fresh data) are shuffled randomly and then are divided into k equal sub-samples, with $k - 1$ samples used as training samples and one sample, the kth, used as a validation sample. Note that k is typically set at 5 or 10. This process is then repeated k times, which helps minimize both bias and variance by insuring that each data point is used in the training set $k - 1$ times and in the validation set once. The average of the k validation errors (mean E_{val}) is then taken as a reasonable estimate of the model's out-of-sample error (E_{out}). A limitation of k-fold cross-validation is that it cannot be used with time-series data, where only the most recent data can reasonably be used for model validation.

In sum, mitigating overfitting risk by avoiding excessive out-of-sample error is critical to creating a supervised machine learning model that generalizes well to fresh datasets drawn from the same distribution. The main techniques used to mitigate overfitting risk in model construction are complexity reduction (or regularization) and cross-validation.

EXAMPLE 2

Evaluating ML Algorithm Performance

Shreya Anand is a portfolio manager based in the Mumbai headquarters office of an investment firm, where she runs a high-dividend-yield fund for wealthy clients. Anand has some knowledge of data science from her university studies. She is interested in classifying companies in the NIFTY 200 Index—an index of large- and mid-cap companies listed on the National Stock Exchange of India—into two categories: dividend increase and no dividend increase. She assembles data for training, validating, and testing an ML-based model that consists of 1,000 observations of NIFTY 200 companies, each consisting of 25 features (fundamental and technical) and the labeled target (dividend increase or no dividend increase).

After training her model, Anand discovers that while it is good at correctly classifying using the training sample, it does not perform well on new data. In consulting her colleagues about this issue, Anand hears conflicting explanations about what constitutes good generalization in an ML model:

Statement 1 The model retains its explanatory power when predicting using new data (i.e., out-of-sample).

Statement 2 The model shows low explanatory power after training using in-sample data (i.e., training data).

Statement 3 The model loses its explanatory power when predicting using new data (i.e., out-of-sample).

1 Which statement made to Anand is *most* accurate?

A Statement 1

B Statement 2

C Statement 3

2 Anand's model is *most likely* being impaired by which of the following?

A Underfitting and bias error

B Overfitting and variance error

C Overfitting and bias error

3 By implementing which one of the following actions can Anand address the problem?

A Estimate and incorporate into the model a penalty that decreases in size with the number of included features.

B Use the *k*-fold cross-validation technique to estimate the model's out-of-sample error, and then adjust the model accordingly.

C Use an unsupervised learning model.

Solution to 1:

A, Statement 1, is correct. B, Statement 2, is incorrect because it describes a poorly fitting model with high bias. C, Statement 3, is incorrect because it describes an overfitted model with poor generalization.

Solution to 2:

B is correct. Anand's model is good at correctly classifying using the training sample, but it does not perform well using new data. The model is overfitted, so it has high variance error.

Solution to 3:

B is correct. A is incorrect because the penalty should increase in size with the number of included features. C is incorrect because Anand is using labeled data for classification, and unsupervised learning models do not use labeled data.

4 SUPERVISED MACHINE LEARNING ALGORITHMS: PENALIZED REGRESSION

c describe supervised machine learning algorithms—including penalized regression, support vector machine, K-nearest neighbor, classification and regression tree, ensemble learning, and random forest—and determine the problems for which they are best suited

Supervised machine learning models are trained using labeled data, and depending on the nature of the target (Y) variable, they can be divided into two types: regression for a continuous target variable and classification for a categorical or ordinal target variable. As shown in Exhibit 2 under regression, we will now cover penalized regression and LASSO. Then, as shown under classification, we will introduce support vector machine (SVM), *k*-nearest neighbor (KNN), and classification and regression tree (CART) algorithms. Note that CART, as its name implies, can be used for both classification and regression problems.

In the following discussion, assume we have a number of observations of a target variable, Y, and n real valued features, $X_1, \ldots, X_n$, that we may use to establish a relationship (regression or classification) between X (a vector of the X_i) and Y for each observation in our dataset.

4.1 Penalized Regression

Penalized regression is a computationally efficient technique used in prediction problems. In practice, penalized regression has been useful for reducing a large number of features to a manageable set and for making good predictions in a variety of large datasets, especially where features are correlated (i.e., when classical linear regression breaks down).

In a large dataset context, we may have many features that potentially could be used to explain Y. When a model is fit to training data, the model may so closely reflect the characteristics of the specific training data that the model does not perform well on new data. Features may be included that reflect noise or randomness in the training dataset that will not be present in new or future data used for making predictions. That is the problem of overfitting, and penalized regression can be described as a technique to avoid overfitting. In prediction, out-of-sample performance is key, so relatively parsimonious models (that is, models in which each variable plays an essential role) tend to work well because they are less subject to overfitting.

Let us suppose that we standardize our data so the features have a mean of 0 and a variance of 1. Standardization of features will allow us to compare the magnitudes of regression coefficients for the feature variables. In ordinary linear regression (i.e., ordinary least squares, or OLS), the regression coefficients $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_K$ are chosen to *minimize* the sum of the squared residuals (i.e., the sum of the squared difference between the actual values, Y_i, and the predicted values, $\hat{Y}_i$), or

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2.$$

Penalized regression includes a constraint such that the regression coefficients are chosen to minimize the sum of squared residuals *plus* a penalty term that increases in size with the number of included features. So, in a penalized regression, a feature must make a sufficient contribution to model fit to offset the penalty from including it. Therefore, only the more important features for explaining Y will remain in the penalized regression model.

In one popular type of penalized regression, **LASSO**, or least absolute shrinkage and selection operator, the penalty term has the following form, with $\lambda > 0$:

$$\text{Penalty term} = \lambda\sum_{k=1}^{K}\left|\hat{b}_k\right|.$$

In addition to minimizing the sum of the squared residuals, LASSO involves minimizing the sum of the absolute values of the regression coefficients (see the following expression). The greater the number of included features (i.e., variables with non-zero coefficients), the larger the penalty term. Therefore, penalized regression ensures that a feature is included only if the sum of squared residuals declines by more than the penalty term increases. All types of penalized regression involve a trade-off of this type. Also, since LASSO eliminates the least important features from the model, it automatically performs a type of feature selection.

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \lambda\sum_{k=1}^{K}\left|\hat{b}_k\right|.$$

Lambda (λ) is a **hyperparameter**—a parameter whose value must be set by the researcher before learning begins—of the regression model and will determine the balance between fitting the model versus keeping the model parsimonious. In practice, a hyperparameter is set by reviewing model performance repeatedly at different settings on the validation set, and hence the test set is also essential to avoid overfitting of hyperparameters to the validation data.

Note that in the case where λ = 0, the LASSO penalized regression is equivalent to an OLS regression. When using LASSO or other penalized regression techniques, the penalty term is added only during the model building process (i.e., when fitting the model to the training data). Once the model has been built, the penalty term is no longer needed, and the model is then evaluated by the sum of the squared residuals generated using the test dataset.

With today's availability of fast computation algorithms, investment analysts are increasingly using LASSO and other regularization techniques to remove less pertinent features and build parsimonious models. **Regularization** describes methods that reduce statistical variability in high-dimensional data estimation problems—in this case, reducing regression coefficient estimates toward zero and thereby avoiding complex models and the risk of overfitting. LASSO has been used, for example, for forecasting default probabilities in industrial sectors where scores of potential features, many collinear, have been reduced to fewer than 10 variables, which is important given the relatively small number (about 100) of observations of default.

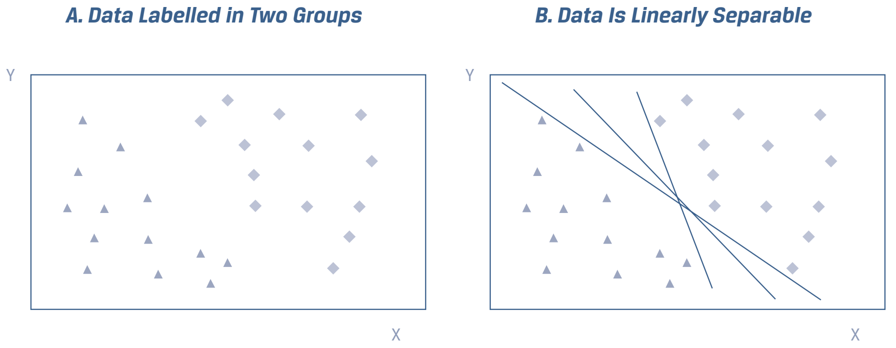
Regularization methods can also be applied to non-linear models. A long-term challenge of the asset management industry in applying mean–variance optimization has been the estimation of stable covariance matrixes and asset weights for large portfolios. Asset returns typically exhibit strong multi-collinearity, making the estimation of the covariance matrix highly sensitive to noise and outliers, so the resulting optimized asset weights are highly unstable. Regularization methods have been used to address this problem. The relatively parsimonious models produced by applying penalized regression methods, such as LASSO, tend to work well because they are less subject to overfitting.

5 SUPPORT VECTOR MACHINE

c describe supervised machine learning algorithms—including penalized regression, support vector machine, k-nearest neighbor, classification and regression tree, ensemble learning, and random forest—and determine the problems for which they are best suited

Support vector machine (SVM) is one of the most popular algorithms in machine learning. It is a powerful supervised algorithm used for classification, regression, and outlier detection. Despite its complicated-sounding name, the notion is relatively straightforward and best explained with a few pictures. The left panel in Exhibit 6 presents a simple dataset with two features (x and y coordinates) labeled in two groups (triangles and crosses). These binary labeled data are noticeably separated into two distinct regions, which could represent stocks with positive and negative returns in a given year. These two regions can be easily separated by an infinite number of straight lines; three of them are shown in the right panel of Exhibit 6. The data are thus linearly separable, and any of the straight lines shown would be called a **linear classifier**—a binary classifier that makes its classification decision based on a linear combination of the features of each data point.

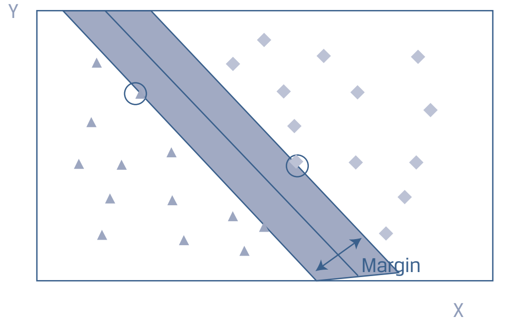
Exhibit 6 Scatterplots and Linear Separation of Labeled Data



With two dimensions or features (x and y), linear classifiers can be represented as straight lines. Observations with n features can be represented in an n-dimension space, and the dataset would be linearly separable if the observations can be separated into two distinct regions by a linear space boundary. The general term for such a space boundary is an n-dimensional hyperplane, which with n = 1 is called a line and with n = 2 is called a plane.

Support vector machine is a linear classifier that determines the hyperplane that optimally separates the observations into two sets of data points. The intuitive idea behind the SVM algorithm is maximizing the probability of making a correct prediction (here, that an observation is a triangle or a cross) by determining the boundary that is the furthest away from all the observations. In Exhibit 7, SVM separates the data by the maximum margin, where the margin is the shaded strip that divides the observations into two groups. The straight line in the middle of the shaded strip is the discriminant boundary, or boundary, for short. We can see that the SVM algorithm produces the widest shaded strip (i.e., the one with the maximum margin on either side of the boundary). The margin is determined by the observations closest to the boundary (the circled points) in each set, and these observations are called support vectors. Adding more training data away from the support vectors will not affect the boundary. In our training datasets, however, adding data points which are close to the hyperplane may move the margin by changing the set of support vectors.

Exhibit 7 Linear Support Vector Machine Classifier



In Exhibit 7, SVM is classifying all observations perfectly. Most real-world datasets, however, are not linearly separable. Some observations may fall on the wrong side of the boundary and be misclassified by the SVM algorithm. The SVM algorithm handles this problem by an adaptation called **soft margin classification**, which adds a penalty to the objective function for observations in the training set that are misclassified. In essence, the SVM algorithm will choose a discriminant boundary that optimizes the trade-off between a wider margin and a lower total error penalty.

As an alternative to soft margin classification, a non-linear SVM algorithm can be run by introducing more advanced, non-linear separation boundaries. These algorithms may reduce the number of misclassified instances in the training datasets but are more complex and, so, are prone to overfitting.

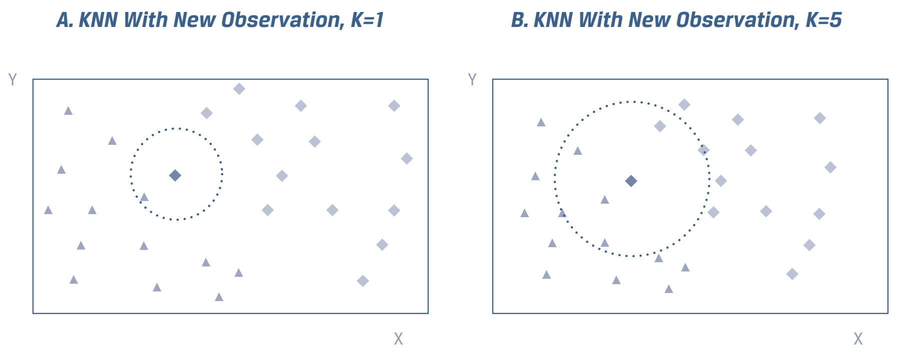
SVM has many applications in investment management. It is particularly suited for small to medium-size but complex high-dimensional datasets, such as corporate financial statements or bankruptcy databases. Investors seek to predict company failures for identifying stocks to avoid or to short sell, and SVM can generate a binary classification (e.g., bankruptcy likely versus bankruptcy unlikely) using many fundamental and technical feature variables. SVM can effectively capture the characteristics of such data with many features while being resilient to outliers and correlated features. SVM can also be used to classify text from documents (e.g., news articles, company announcements, and company annual reports) into useful categories for investors (e.g., positive sentiment and negative sentiment).

6 K-NEAREST NEIGHBOR

c describe supervised machine learning algorithms—including penalized regression, support vector machine, k-nearest neighbor, classification and regression tree, ensemble learning, and random forest—and determine the problems for which they are best suited

***K*-nearest neighbor** (KNN) is a supervised learning technique used most often for classification and sometimes for regression. The idea is to classify a new observation by finding similarities ("nearness") between this new observation and the existing data. Going back to the scatterplot in Exhibit 6, let us assume we have a new observation: The diamond in Exhibit 8 needs to be classified as belonging to either the cross or the triangle category. If $k = 1$, the diamond will be classified into the same category as its nearest neighbor (i.e., the triangle in the left panel). The right panel in Exhibit 8 presents the case where $k = 5$, so the algorithm will look at the diamond's five nearest neighbors, which are three triangles and two crosses. The decision rule is to choose the classification with the largest number of nearest neighbors out of the five being considered. So, the diamond is again classified as belonging to the triangle category.

Exhibit 8 *K*-Nearest Neighbor Algorithm



Let us suppose we have a database of corporate bonds classified by credit rating that also contains detailed information on the characteristics of these bonds. Such features would include those of the issuing company (e.g., asset size, industry, leverage ratios, cash flow ratios) and of the bond issue itself (e.g., tenor, fixed/floating coupon, embedded options). Now, assume a new bond is about to be issued with no credit rating. By nature, corporate bonds with similar issuer and issue characteristics should be given a similar credit rating. So, by using KNN, we can predict the implied credit rating of the new bond based on the similarities of its characteristics to those of the bonds in our database.

KNN is a straightforward, intuitive model that is still very powerful because it is non-parametric; the model makes no assumptions about the distribution of the data. Moreover, it can be used directly for multi-class classification. A critical challenge of KNN, however, is defining what it means to be "similar" (or near). Besides the selection of features, an important decision relates to the distance metric used to model similarity because an inappropriate measure will generate poorly performing models. The choice of a correct distance measure may be even more subjective for ordinal or categorical data. For example, if an analyst is looking at the similarities in market performance of various equities, he or she may consider using the correlation between the stocks' historical returns as an appropriate measure of similarity.

Knowledge of the data and understanding of the business objectives of the analysis are critical aspects in the process of defining similarity. KNN results can be sensitive to inclusion of irrelevant or correlated features, so it may be necessary to select features manually. By doing so, the analyst removes less valuable information to keep the most relevant and pertinent information. If done correctly, this process should generate a more representative distance measure. KNN algorithms tend to work better with a small number of features.

Finally, the number *k*, the hyperparameter of the model, must be chosen with the understanding that different values of *k* can lead to different conclusions. For predicting the credit rating of an unrated bond, for example, should *k* be the 3, 15, or 50 bonds most similar to the unrated bond? If *k* is an even number, there may be ties and no clear classification. Choosing a value for *k* that is too small would result in a high error rate and sensitivity to local outliers, but choosing a value for *k* that is too large would dilute the concept of nearest neighbors by averaging too many outcomes. In practice, several different techniques can be used to determine an optimal value for *k*, taking into account the number of categories and their partitioning of the feature space.

The KNN algorithm has many applications in the investment industry, including bankruptcy prediction, stock price prediction, corporate bond credit rating assignment, and customized equity and bond index creation. For example, KNN is useful for determining bonds that are similar and those that are dissimilar, which is critical information for creating a custom, diversified bond index.

CLASSIFICATION AND REGRESSION TREE 7

c describe supervised machine learning algorithms—including penalized regression, support vector machine, k-nearest neighbor, classification and regression tree, ensemble learning, and random forest—and determine the problems for which they are best suited

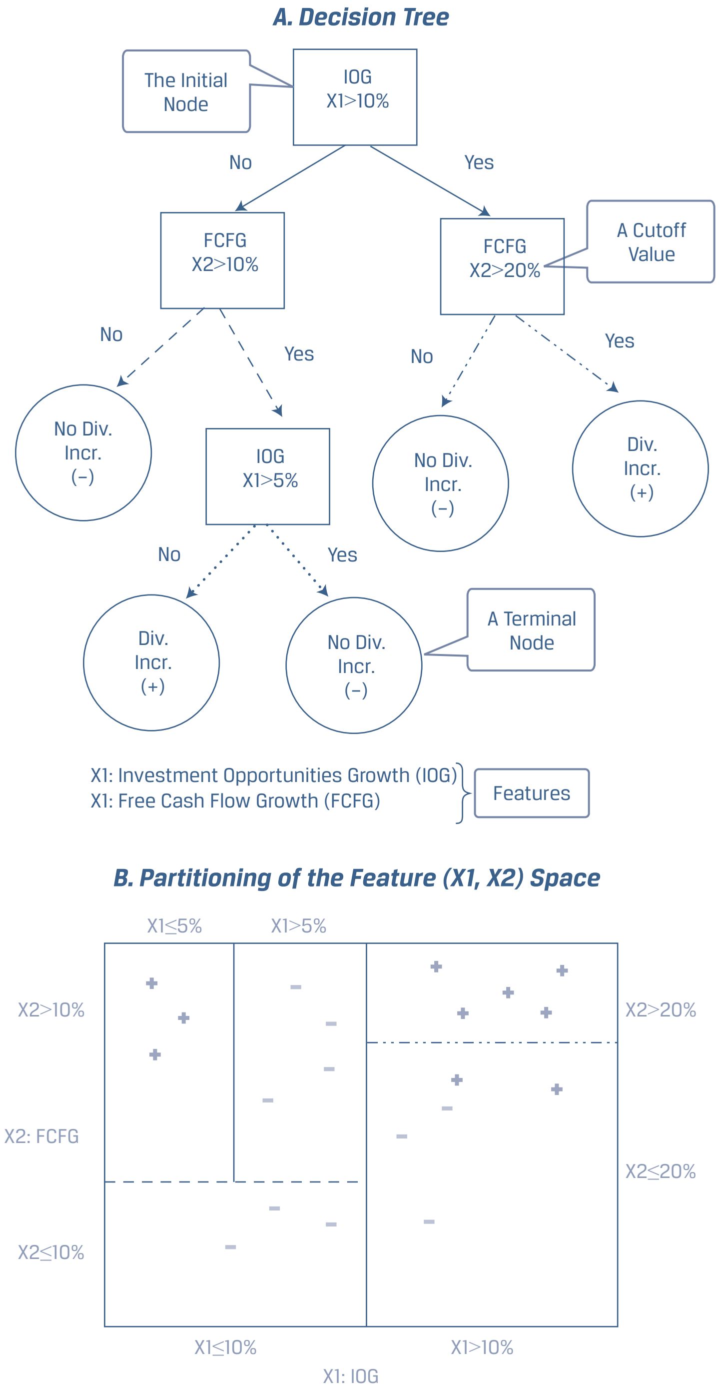
Classification and regression tree (CART) is another common supervised machine learning technique that can be applied to predict either a categorical target variable, producing a classification tree, or a continuous target variable, producing a regression tree. CART is commonly applied to binary classification or regression.

CART will be discussed in the context of a simplified model for classifying companies by whether they are likely to increase their dividends to shareholders. Such a classification requires a binary tree: a combination of an initial root node, decision nodes, and terminal nodes. The root node and each decision node represent a single feature (*f*) and a cutoff value (*c*) for that feature. As shown in Panel A of Exhibit 9, we start at the initial root node for a new data point. In this case, the initial root node represents the feature investment opportunities growth (IOG), designated as X1, with a cutoff value of 10%. From the initial root node, the data are partitioned at decision nodes into smaller and smaller subgroups until terminal nodes that contain the predicted labels are formed. In this case, the predicted labels are either dividend increase (the cross) or no dividend increase (the dash).

Also shown in Panel A of Exhibit 9, if the value of feature IOG (X1) is greater than 10% (Yes), then we proceed to the decision node for free cash flow growth (FCFG), designated as X2, which has a cutoff value of 20%. Now, if the value of FCFG is not greater than 20% (No), then CART will predict that that data point belongs to the no dividend increase (dash) category, which represents a terminal node. Conversely, if the value of X2 is greater than 20% (Yes), then CART will predict that that data point belongs to the dividend increase (cross) category, which represents another terminal node.

It is important to note that the same feature can appear several times in a tree in combination with other features. Moreover, some features may be relevant only if other conditions have been met. For example, going back to the initial root node, if IOG is not greater than 10% (X1 ≤ 10%) and FCFG is greater than 10%, then IOG appears again as another decision node, but this time it is lower down in the tree and has a cutoff value of 5%.

Exhibit 9 Classification and Regression Tree—Decision Tree and Partitioning of the Feature Space



We now turn to how the CART algorithm selects features and cutoff values for them. Initially, the classification model is trained from the labeled data, which in this hypothetical case are 10 instances of companies having a dividend increase (the crosses) and 10 instances of companies with no dividend increase (the dashes). As shown in Panel B of Exhibit 9, at the initial root node and at each decision node, the feature space (i.e., the plane defined by X1 and X2) is split into two rectangles for values above and below the cutoff value for the particular feature represented at that node. This can be seen by noting the distinct patterns of the lines that emanate from the decision nodes in Panel A. These same distinct patterns are used for partitioning the feature space in Panel B.

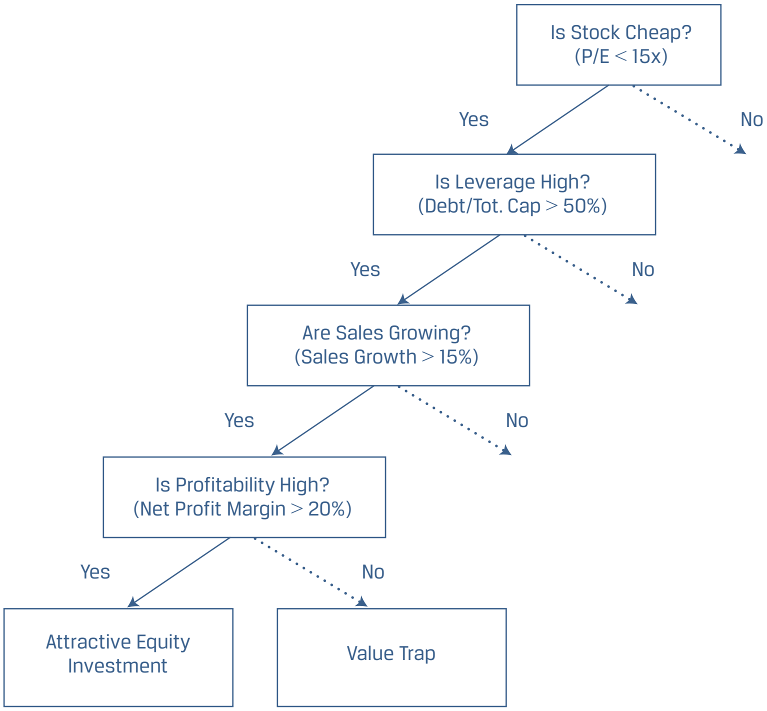
The CART algorithm chooses the feature and the cutoff value at each node that generates the widest separation of the labeled data to minimize classification error (e.g., by a criterion, such as mean-squared error). After each decision node, the partition of the feature space becomes smaller and smaller, so observations in each group have lower within-group error than before. At any level of the tree, when the classification error does not diminish much more from another split (bifurcation), the process stops, the node is a terminal node, and the category that is in the majority at that node is assigned to it. If the objective of the model is classification, then the prediction of the algorithm at each terminal node will be the category with the majority of data points. For example, in Panel B of Exhibit 9, the top right rectangle of the feature space, representing IOG (X1) > 10% and FCFG (X2)> 20%, contains five crosses, the most data points of any of the partitions. So, CART would predict that a new data point (i.e., a company) with such features belongs to the dividend increase (cross) category. However, if instead the new data point had IOG (X1) > 10% and FCFG (X2) ≤ 20%, then it would be predicted to belong to the no dividend increase (dash) category—represented by the lower right rectangle, with two crosses but with three dashes. Finally, if the goal is regression, then the prediction at each terminal node is the mean of the labeled values.

CART makes no assumptions about the characteristics of the training data, so if left unconstrained, it potentially can perfectly learn the training data. To avoid such overfitting, regularization parameters can be added, such as the maximum depth of the tree, the minimum population at a node, or the maximum number of decision nodes. The iterative process of building the tree is stopped once the regularization criterion has been reached. For example, in Panel B of Exhibit 9, the upper left rectangle of the feature space (determined by X1 ≤ 10%, X2 > 10%, and X1 ≤ 5% with three crosses) might represent a terminal node resulting from a regularization criterion with minimum population equal to 3. Alternatively, regularization can occur via a **pruning** technique that can be used afterward to reduce the size of the tree. Sections of the tree that provide little classifying power are pruned (i.e., cut back or removed).

By its iterative structure, CART can uncover complex dependencies between features that other models cannot reveal. As demonstrated in Exhibit 9, the same feature can appear several times in combination with other features and some features may be relevant only if other conditions have been met.

As shown in Exhibit 10, high profitability is a critical feature for predicting whether a stock is an attractive investment or a value trap (i.e., an investment that, although apparently priced cheaply, is likely to be unprofitable). This feature is relevant only if the stock is cheap: For example, in this hypothetical case, if P/E is less than 15, leverage is high (debt to total capital > 50%) and sales are expanding (sales growth > 15%). Said another way, high profitability is irrelevant in this context if the stock is not cheap *and* if leverage is not high *and* if sales are not expanding. Multiple linear regression typically fails in such situations where the relationship between the features and the outcome is non-linear.

Exhibit 10 Stylized Decision Tree—Attractive Investment or Value Trap?



CART models are popular supervised machine learning models because the tree provides a visual explanation for the prediction. This contrasts favorably with other algorithms that are often considered to be "black boxes" because it may be difficult to understand the reasoning behind their outcomes and thus to place trust in them. CART is a powerful tool to build expert systems for decision-making processes. It can induce robust rules despite noisy data and complex relationships between high numbers of features. Typical applications of CART in investment management include, among others, enhancing detection of fraud in financial statements, generating consistent decision processes in equity and fixed-income selection, and simplifying communication of investment strategies to clients.

ENSEMBLE LEARNING AND RANDOM FOREST

8

c describe supervised machine learning algorithms—including penalized regression, support vector machine, k-nearest neighbor, classification and regression tree, ensemble learning, and random forest—and determine the problems for which they are best suited

Instead of basing predictions on the results of a single model as in the previous discussion, why not use the predictions of a group—or an ensemble—of models? Each single model will have a certain error rate and will make noisy predictions. But by taking the average result of many predictions from many models, we can expect to achieve a reduction in noise as the average result converges toward a more accurate prediction. This technique of combining the predictions from a collection of models is called **ensemble learning**, and the combination of multiple learning algorithms is known as the **ensemble method**. Ensemble learning typically produces more accurate and more stable predictions than the best single model. In fact, in many prestigious machine learning competitions, an ensemble method is often the winning solution.

Ensemble learning can be divided into two main categories: (1) aggregation of heterogeneous learners (i.e., different types of algorithms combined with a voting classifier) or (2) aggregation of homogeneous learners (i.e., a combination of the same algorithm using different training data that are based, for example, on a bootstrap aggregating, or bagging, technique, as discussed later).

8.1 Voting Classifiers

Suppose you have been working on a machine learning project for some time and have trained and compared the results of several algorithms, such as SVM, KNN, and CART. A **majority-vote classifier** will assign to a new data point the predicted label with the most votes. For example, if the SVM and KNN models are both predicting the category "stock outperformance" and the CART model is predicting the category "stock underperformance," then the majority-vote classifier will choose stock outperformance." The more individual models you have trained, the higher the accuracy of the aggregated prediction up to a point. There is an optimal number of models beyond which performance would be expected to deteriorate from overfitting. The trick is to look for diversity in the choice of algorithms, modeling techniques, and hypotheses. The (extreme) assumption here is that if the predictions of the individual models are independent, then we can use the law of large numbers to achieve a more accurate prediction.

8.2 Bootstrap Aggregating (Bagging)

Alternatively, one can use the same machine learning algorithm but with different training data. **Bootstrap aggregating (or bagging)** is a technique whereby the original training dataset is used to generate n new training datasets or bags of data. Each new bag of data is generated by random sampling with replacement from the initial training set. The algorithm can now be trained on n independent datasets that will generate n new models. Then, for each new observation, we can aggregate the n predictions using a majority-vote classifier for a classification or an average for a regression. Bagging is a very useful technique because it helps to improve the stability of predictions and protects against overfitting the model.

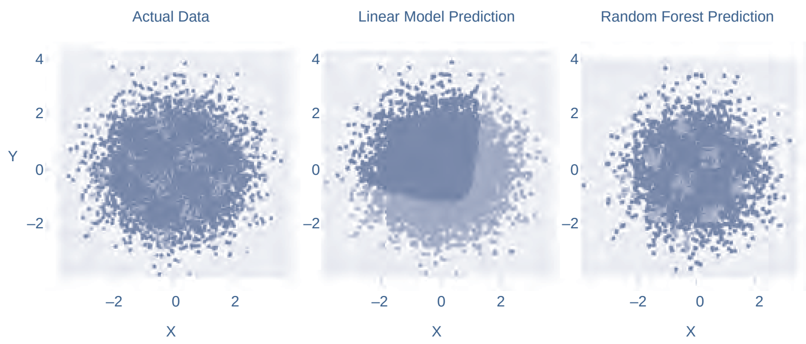
8.3 Random Forest

A **random forest classifier** is a collection of a large number of decision trees trained via a bagging method. For example, a CART algorithm would be trained using each of the n independent datasets (from the bagging process) to generate the multitude of different decision trees that make up the random forest classifier.

To derive even more individual predictions, added diversity can be generated in the trees by randomly reducing the number of features available during training. So, if each observation has n features, one can randomly select a subset of m features (where $m < n$) that will then be considered by the CART algorithm for splitting the dataset at each of the decision nodes. The number of subset features (m), the number of trees to use, the minimum size (population) of each node (or leaf), and the maximum depth of each tree are all hyperparameters that can be tuned to improve overall model prediction accuracy. For any new observation, we let all the classifier trees (the "random forest") undertake classification by majority vote—implementing a machine learning version of the "wisdom of crowds." The process involved in random forest construction tends to reduce variance and protect against overfitting on the training data. It also reduces the ratio of noise to signal because errors cancel out across the collection of slightly different classification trees. However, an important drawback of random forest is that it lacks the ease of interpretability of individual trees; as a result, it is considered a relatively black box type of algorithm.

Exhibit 11 presents three scatterplots of actual and predicted defaults by small and medium-sized businesses with respect to two features, X and Y—for example, firm profitability and leverage, respectively. The left plot shows the actual cases of default in light shade and no default in dark shade, while the middle and right plots present the predicted defaults and no defaults (also in light and dark shades, respectively). It is clear from the middle plot, which is based on a traditional linear regression model, that the model fails to predict the complex non-linear relationship between the features. Conversely, the right plot, which presents the prediction results of a random forest model, shows that this model performs very well in matching the actual distribution of the data.

Exhibit 11 Credit Defaults of Small- and Medium-Sized Borrowers



Source: Bacham and Zhao (2017).

Ensemble Learning with Random Forest

In making use of voting across classifier trees, random forest is an example of ensemble learning: Incorporating the output of a collection of models produces classifications that have better signal-to-noise ratios than the individual classifiers. A good example is a credit card fraud detection problem that comes from an open source dataset on Kaggle.[1] Here, the data contained several anonymized features that might be used to explain which transactions were fraudulent. The difficulty in the analysis arises from the fact that the rate of fraudulent transactions is very low; in a sample of 284,807 transactions, only 492 were fraudulent (0.17%). This is akin to finding a needle in a haystack. Applying a random forest classification algorithm with an oversampling technique—which involves increasing the proportional representation of fraudulent data in the training set—does

1 See www.kaggle.com/mlg-ulb/creditcardfraud (accessed 1 October 2018).

extremely well. Despite the lopsided sample, it delivers **precision** (the ratio of correctly predicted fraudulent cases to all predicted fraudulent cases) of 89% and **recall** (the ratio of correctly predicted fraudulent cases to all actual fraudulent cases) of 82%.

Despite its relative simplicity, random forest is a powerful algorithm with many investment applications. These include, for example, use in factor-based investment strategies for asset allocation and investment selection or use in predicting whether an IPO will be successful (e.g., percent oversubscribed, first trading day close/IPO price) given the attributes of the IPO offering and the corporate issuer. Later, in a mini-case study, Deep Neural Network–Based Equity Factor Model, we present further details of how supervised machine learning is used for fundamental factor modeling.

EXAMPLE 3

Support Vector Machine and *K*-Nearest Neighbor

Rachel Lee is a fixed-income portfolio manager with Zeta Investment Management Company. Zeta manages an investment-grade bond portfolio for small, conservative institutions and a non-investment-grade (i.e., high-yield) bond portfolio for yield-seeking, high-net-worth individuals. Both portfolios can hold unrated bonds if the characteristics of the unrated bonds closely match those of the respective portfolio's average holding.

Lee is discussing an upcoming straight, 10-year fixed-coupon bond issue with senior credit analyst Marc Watson. Watson comments that although the bond's issuer, Biotron Corporation, has not had this issue rated, his analysis of the company's profitability, cash flow, leverage, and coverage ratios places the issue near the borderline between low investment-grade (Baa3/BBB–) and high non-investment-grade (Ba1/BB+) bonds.

Lee decides to use machine learning methods to confirm the implied credit rating of Biotron Corporation.

1 State the type of problem being addressed by Lee.

2 State two ML algorithms that Lee could use to explore the implied credit rating of Biotron Corporation, and then describe how each algorithm could be applied.

Lee decides to apply the two identified ML algorithms. Both algorithms clearly support a high non-investment-grade rating. Watson states that because both ML algorithms agree on the rating, he has confidence in relying on the rating.

3 State one argument in support of Watson's viewpoint.

Solution to 1:

Lee is addressing a supervised learning classification problem because she must determine whether Biotron's upcoming bond issue would be classified as investment grade or non-investment grade.

Solution to 2:

One suitable ML algorithm is SVM. The SVM algorithm is a linear classifier that aims to find the optimal hyperplane—the one that separates observations into two distinct sets by the maximum margin. So, SVM is well suited to binary classification problems, such as the one facing Lee (investment grade versus non-investment grade). In this case, Lee could train the SVM algorithm on data—characteristics (features) and rating (target)—of low investment-grade (Baa3/BBB–) and high non-investment-grade (Ba1/BB+) bonds. Lee would then note on which side of the margin the new data point (Biotron's new bonds) lies.

The KNN algorithm is also well suited for classification problems because it classifies a new observation by finding similarities (or nearness) between the new observation and the existing data. Training the algorithm with data as for SVM, the decision rule for classifying Biotron's new bonds is which classification is in the majority among its *k*-nearest neighbors. Note that *k* (a hyperparameter) must be pre-specified by Lee.

Solution to 3:

If the ML algorithms disagreed on the classification, the classification would be more likely to be sensitive to the algorithm's approach to classifying data. Because the classification of Biotron's new issue appears robust to the choice of ML algorithm (i.e., both algorithms agree on the rating), the resulting classification will more likely be correct.

EXAMPLE 4

CART and Ensemble Learning

Laurie Kim is a portfolio manager at Hilux LLC, a high-yield bond investment firm. The economy has been in recession for several months, and high-yield bond prices have declined precipitously as credit spreads have widened in response to the weak macroeconomic environment. Kim, however, believes this is a good time to buy because she expects to profit as credit spreads narrow and high-yield bond prices rise in anticipation of economic recovery.

Based on her analysis, Kim believes that corporate high-yield bonds in the credit quality range of B/B2 to CCC/Caa2 are the most attractive. However, she must carefully select which bonds to buy and which bonds to avoid because of the elevated default risk caused by the currently weak economy.

To help with her bond selection, Kim turns to Hilux's data analytics team. Kim has supplied them with historical data consisting of 19 fundamental and 5 technical factors for several thousand high-yield bond issuers and issues labeled to indicate default or no default. Kim requests that the team develop an ML-based model using all the factors provided that will make accurate classifications in two categories: default and no default. Exploratory data analysis suggests considerable non-linearities among the feature set.

1 State the type of problem being addressed by Kim.
2 Describe the dimensionality of the model that Kim requests her analytics team to develop.
3 Evaluate whether a CART model is appropriate for addressing her problem.
4 Describe how a CART model operates at each node of the tree.
5 Describe how the team might avoid overfitting and improve the predictive power of a CART model.
6 Describe how ensemble learning might be used by the team to develop even better predictions for Kim's selection of corporate high-yield bonds.

Solution to 1:

Kim is addressing a classification problem because she must determine whether bonds that she is considering purchasing in the credit quality range of B/B2 to CCC/Caa2 will default or not default.

Solution to 2:

With 19 fundamental and 5 technical factors (i.e., the features), the dimensionality of the model is 24.

Solution to 3:

The CART model is an algorithm for addressing classification problems. Its ability to handle complex, non-linear relationships makes it a good choice to address the modeling problem at hand. An important advantage of CART is that its results are relatively straightforward to visualize and interpret, which should help Kim explain her recommendations based on the model to Hilux's investment committee and the firm's clients.

Solution to 4:

At each node in the decision tree, the algorithm will choose the feature and the cutoff value for the selected feature that generates the widest separation of the labeled data to minimize classification error.

Solution to 5:

The team can avoid overfitting and improve the predictive power of the CART model by adding regularization parameters. For example, the team could specify the maximum depth of the tree, the minimum population at a node, or the maximum number of decision nodes. The iterative process of building nodes will be stopped once the regularization criterion has been reached. Alternatively, a pruning technique can be used afterward to remove parts of the CART model that provide little power to correctly classify instances into default or no default categories.

Solution to 6:

The analytics team might use ensemble learning to combine the predictions from a collection of models, where the average result of many predictions leads to a reduction in noise and thus more accurate predictions. Ensemble learning can be achieved by an aggregation of either heterogeneous learners—different types of algorithms combined with a voting classifier—or homogeneous learners—a combination of the same algorithm but using different training data based on the bootstrap aggregating (i.e., bagging) technique. The team may also consider developing a random forest classifier (i.e., a collection of many decision trees) trained via a bagging method.

CASE STUDY: CLASSIFICATION OF WINNING AND LOSING FUNDS

The following case study was developed and written by Matthew Dixon, PhD, FRM.

A research analyst for a fund of funds has been tasked with identifying a set of attractive exchange-traded funds (ETFs) and mutual funds (MFs) in which to invest. She decides to use machine learning to identify the best (i.e., winners) and worst (i.e., losers) performing funds and the features which are most important in such an identification. Her aim is to train a model to correctly classify the winners and losers and then to use it to predict future outperformers. She is unsure of which type of machine learning classification model (i.e., classifier) would work best, so she reports and cross-compares her findings using several different well-known machine learning algorithms.

The goal of this case is to demonstrate the application of machine learning classification to fund selection. Therefore, the analyst will use the following classifiers to identify the best and worst performing funds:

- classification and regression tree (CART),
- support vector machine (SVM),
- *k*-nearest neighbors (KNN), and
- random forests.

Data Description

In the following experiments, the performance of each fund is learned by the machine learning algorithms based on fund type and size, asset class composition, fundamentals (i.e., valuation multiples), and sector composition characteristics. To form a cross-sectional classifier, the sector composition and fund size reported on 15 February 2019 are assumed to be representative of the latest month over which the fund return is reported. Exhibit 12 presents a description of the dataset.

Exhibit 12 Dataset Description

Dataset: MF and ETF Data

There are two separate datasets, one for MFs and one for ETFs, consisting of fund type, size, asset class composition, fundamental financial ratios, sector weights, and monthly total return labeled to indicate the fund as being a winner, a loser, or neither. Number of observations: 6,085 MFs and 1,594 ETFs.

Features: Up to 21, as shown below:

General (six features):

1 cat investment*: Fund type, either "blend," "growth," or "value"

2 net assets: Total net assets in US dollars

3 cat size: Investment category size, either "small," "medium," or "large" market capitalization stocks

4 portfolio cash**: The ratio of cash to total assets in the fund

5 portfolio stocks: The ratio of stocks to total assets in the fund

6 portfolio bonds: The ratio of bonds to total assets in the fund

Fundamentals (four features):

1 price earnings: The ratio of price per share to earnings per share

2 price book: The ratio of price per share to book value per share

3 price sales: The ratio of price per share to sales per share

4 price cashflow: The ratio of price per share to cash flow per share

Sector weights (for 11 sectors) provided as percentages:

1 basic materials

2 consumer cyclical

3 financial services

(continued)

Exhibit 12 (Continued)

4 real estate
5 consumer defensive
6 healthcare
7 utilities
8 communication services
9 energy
10 industrials
11 technology

Labels:

Winning and losing ETFs or MFs are determined based on whether their returns are one standard deviation or more above or below the distribution of one-month fund returns across all ETFs or across all MFs, respectively. More precisely, the labels are:

1, if fund_return_1 month ≥ mean(fund_return_1 month) + one std.dev(-fund_return_1 month), indicating a winning fund;

-1, if fund_return_1 month ≤ mean(fund_return_1 month) – one std.dev(-fund_return_1 month), indicating a losing fund; and

0, otherwise.

*Feature appears in the ETF dataset only.
**Feature appears in the MF dataset only.
Data sources: Kaggle, Yahoo Finance on 15 February 2019.

Methodology

The classification model is trained to determine whether a fund's performance is one standard deviation or more above the mean return (Label 1), within one standard deviation of the mean return (Label 0), or one standard deviation or more below the mean return (Label -1), where the mean return and standard deviation are either for all ETFs or all MFs, depending on the particular fund's type (ETF or MF). Performance is based on the one-month return of each fund as of 15 February 2019.

This procedure results in most of the funds being labeled as "0" (or average). After removing missing values in the dataset, there are 1,594 and 6,085 observations in the ETF and MF datasets, respectively. The data table is a 7,679 × 22 matrix, with 7,679 rows for each fund observation (1,594 for ETFs and 6,085 for MFs) and 22 columns for the 21 features plus the return label, and all data are recorded as of 15 February 2019.

The aim of the experiment is to identify not only winning and losing funds but also the features which are useful for distinguishing winners from losers. An important caveat, however, is that no claim is made that such features are causal.

A separate multi-classifier, with three classes, is run for each dataset. Four types of machine learning algorithms are used to build each classifier: (i) CART, (ii) SVM, (iii) KNN, and (iv) random forest. Random forest is an example of an ensemble method (based on bagging), whereas the other three algorithms do not use bagging.

A typical experimental design would involve using 70% of the data for training and holding 15% for tuning model hyperparameters and the remaining 15% of the data for testing. For simplicity, we shall not tune the hyperparameters but simply use the default settings without attempting to fine tune each one for best performance. So, in this case, we do not withhold 15% of the data for validation but instead train the classifier on a random split of 70% of the dataset, with the remaining 30% of the dataset used for testing. Crucially, for fairness of evaluation, each algorithm is trained and tested on identical data: The same 70% of observations are used for training each algorithm, and the same 30% are used for testing each one. The most important hyperparameters and settings for the algorithms are shown in Exhibit 13.

Exhibit 13 Parameter Settings for the Four Machine Learning Classifiers

1 CART: maximum tree depth: 5 levels
2 SVM: cost parameter: 1.0
3 KNN: number of nearest neighbors: 4
4 Random forest: number of trees: 100; maximum tree depth: 20 levels

The choices of hyperparameter values for the four machine learning classifiers are supported by theory, academic research, practice, and experimentation to yield a satisfactory bias–variance trade-off. For SVM, the cost parameter is a penalty on the margin of the decision boundary. A large cost parameter forces the SVM to use a thin margin, whereas a smaller cost parameter widens the margin. For random forests, recall that this is an ensemble method which uses multiple decision trees to classify, typically by majority vote. Importantly, no claim is made that these choices of hyperparameters are universally optimal for any dataset.

Results

The results of each classifier are evaluated separately on the test portion of the ETF and MF datasets. The evaluation metrics used are based on Type I and Type II classification errors, where a Type I error is a false positive (FP) and a Type II error is a false negative (FN). Correct classifications are true positive (TP) and true negative (TN).

- The first evaluation metric is **accuracy**, the percentage of correctly predicted classes out of total predictions. So, high accuracy implies low Type I and Type II errors.
- **F1 score**, the second evaluation metric, is the weighted average of precision and recall. Precision is the ratio of correctly predicted positive classes to all predicted positive classes, and recall is the ratio of correctly predicted positive classes to all actual positive classes.

F1 score is a more appropriate evaluation metric to use than accuracy when there is unequal class distribution ("class imbalance") in the dataset, as is the case here. As mentioned, most of the funds in the ETF and MF datasets are designated as "0," indicating average performers.

Exhibit 14 shows the comparative performance results for each algorithm applied to the ETF dataset. These results show the random forest model is the most accurate (0.812), but once class imbalance is accounted for using F1 score (0.770), random forest is about as good as CART. Generally, ensemble methods, such as random forest, are expected to be at least as good as their single-model counterparts because ensemble forecasts generalize better out-of-sample. Importantly, while the relative accuracies and F1 scores across the different methods provide a basis for comparison, they do not speak to the absolute performance. In this regard, values approaching 1 suggest an excellent model, whereas values of approximately 1/3 would indicate the model is useless: 1/3 is premised on three (+1, 0, -1) equally distributed labels. However, because the distribution of classes is often not balanced, this ratio typically requires some adjustment.

Exhibit 14 Comparison of Accuracy and F1 Score for Each Classifier Applied to the ETF Dataset

	CART	SVM	KNN	Random Forest
Accuracy	0.770	0.774	0.724	0.812
F1 score	0.769	0.693	0.683	0.770

Exhibit 15 shows that the random forest model outperforms all the other classifiers under both metrics when applied to the MF dataset. Overall, the accuracy and F1 score for the SVM and KNN methods are similar for each dataset, and these algorithms are dominated by CART and random forest, especially in the larger MF dataset. The difference in performance between the two datasets for all the algorithms is to be expected, since the MF dataset is approximately four times larger than the ETF dataset and a larger sample set generally leads to better model performance. Moreover, the precise explanation of why random forest and CART outperform SVM and KNN is beyond the scope of this case. Suffice it to say that random forests are well known to be more robust to noise than most other classifiers.

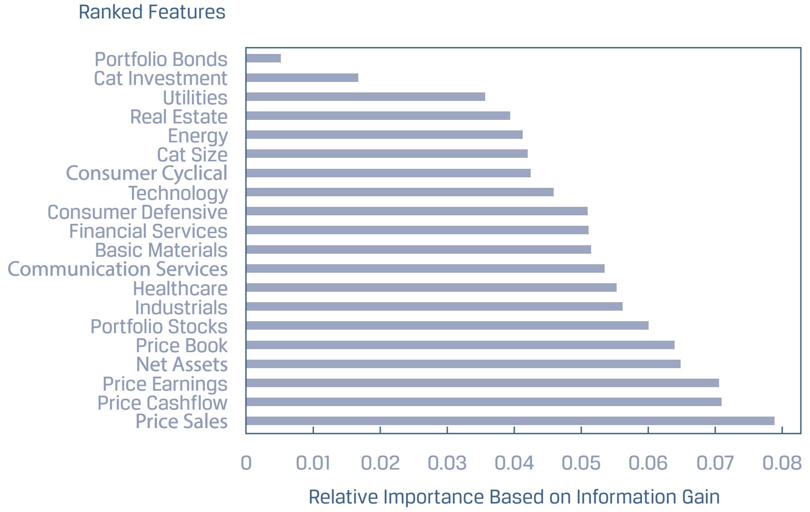
Exhibit 15 Comparison of Accuracy and F1 Score for Each Classifier Applied to the Mutual Fund Dataset

	CART	SVM	KNN	Random Forest
Accuracy	0.959	0.859	0.856	0.969
F1 score	0.959	0.847	0.855	0.969

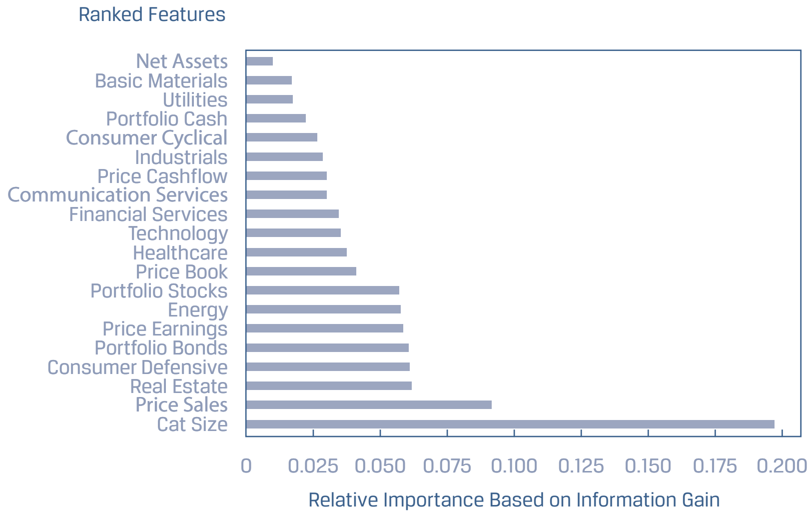
Exhibit 16 presents results on the relative importance of the features in the random forest model for both the ETF (Panel A) and MF (Panel B) datasets. Relative importance is determined by **information gain**, which quantifies the amount of information that the feature holds about the response. Information gain can be regarded as a form of non-linear correlation between Y and X. Note the horizontal scale of Panel B (MF dataset) is more than twice as large as that of Panel A (ETF dataset), and the bar colors represent the feature rankings, not the features themselves.

Exhibit 16 Relative Importance of Features in the Random Forest Model

A. ETF Dataset



B. Mutual Fund Dataset



The prices-to-sales (price_sales) and prices-to-earnings (price_earnings) ratios are observed to be important indicators of performance, at about 0.08–0.09 and 0.06–0.07, respectively, in the random forest models for each dataset. The ratio of stocks to total assets (portfolio_stocks), at 0.06, is another key feature. Moreover, the industrials, health care, and communication services sector weightings are relatively important in the ETF dataset, while the real estate, consumer defensive, and energy sector weightings are key features in the MF dataset for differentiating between winning and losing funds.

Another important observation is that the category of the fund size (cat_size) is by far the most important feature in the model's performance for the MF dataset (≈ 0.20), whereas it is of much less importance for model performance using the ETF dataset (≈ 0.04). Conversely, net assets is a relatively important feature for model performance using the ETF dataset (0.065), while it is the least important feature when the random forest model is applied to the MF dataset (0.01).

Conclusion

The research analyst has trained and tested machine learning–based models that she can use to identify potential winning and losing ETFs and MFs. Her classification models use input features based on fund type and size, asset class composition, fundamentals, and sector composition characteristics. She is more confident in her assessment of MFs than of ETFs, owing to the substantially larger sample size of the former. She is also confident that any imbalance in class has not led to misinterpretation of her models' results, since she uses F1 score as her primary model evaluation metric. Moreover, she determines that the best performing model using both datasets is an ensemble-type random forest model. Finally, she concludes that while fundamental ratios, asset class ratios, and sector composition are important features for both models, net assets and category size also figure prominently in discriminating between winning and losing ETFs and MFs.

EXAMPLE 5

Classification of Funds

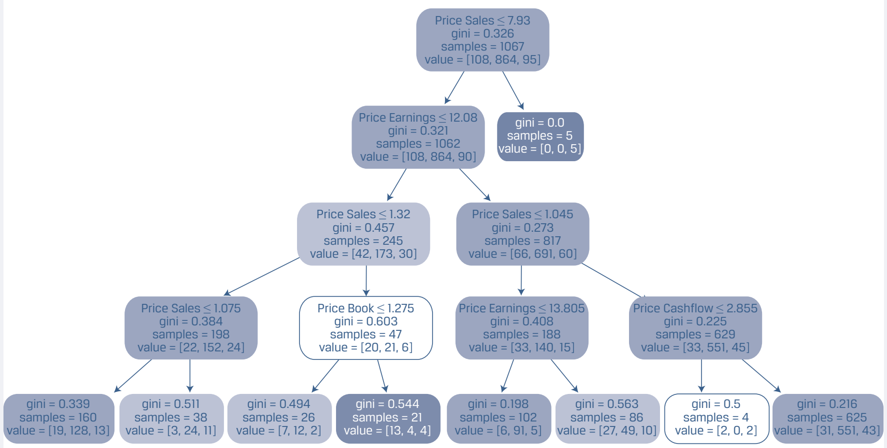
The research analyst from the previous case uses CART to generate the decision tree shown in Exhibit 17, which she will use to predict whether and explain why a new ETF is likely to be a winner (+1), an average performer (0), or a loser (-1). This ETF's fundamental valuation ratios are as follows: Price-to-sales = 2.29, price-to-earnings = 7.20, price-to-book = 1.41, and price-to-cash flow = 2.65. Note that the sample size is 1,067 ETFs and the CART model uses just valuation ratios, because these are deemed the most important features for ETF performance classification.

1. Explain the CART model's prediction for performance of the new ETF: winner, loser, or average performer.

2. Calculate the probability that the fund will be in the class predicted by the CART model.

3. Explain why the analyst should be cautious in basing the ETF's predicted performance solely on the CART-generated decision tree.

Exhibit 17 CART-Based Decision Tree for EFT Performance Classification



Legend:
• Darkest shade, 5th (last) level: Winner (Class = +1)

Exhibit 17 (Continued)

- Light to medium shade: Average Performer (Class = 0); note that the medium shade indicates more confidence in the classification.
- Darkest shade, 2nd level: Loser (Class = -1)
- White: Inconclusive, either because there is a tie with multiple categories or there are too few samples
- Value: The number of sample cases in each of the three classes: Winner, Average Performer, or Loser
- Path: Left path is True and right path is False.

Solution to 1:

Based on its valuation ratios (P/S = 2.29; P/E = 7.20; P/B = 1.41), the new ETF is predicted to be a winner because the decision path leads to the dark shaded, 5th level ("winner") terminal node. The split criteria and decisions are as follows:

Initial node: P/S ≤ 7.93 and EFT P/S = 2.29, so True.
2nd-level node: P/E ≤ 12.08 and EFT P/E = 7.20, so True.
3rd-level node: P/S ≤ 1.32 and EFT P/S = 2.29, so False.
4th-level node: P/B ≤ 1.275 and EFT P/B = 1.41, so False.
5th-level (terminal) node: darkest shaded terminal node indicates "winner."

Solution to 2:

The output from the CART model in the darkest shaded, 5th level (winner) terminal node is [13, 4, 4], which indicates it includes 13 funds of Class +1 (winners), 4 funds of Class 0 (average performers), and 4 funds of Class -1 (losers). Thus, the probability predicted by the CART model that this ETF will be in the "winner" class is 13/21, or 62%. There are also equal probabilities of it being an average performer (19%) or a loser (19%).

Solution to 3:

There are several reasons why the analyst should be cautious in basing the ETF's predicted performance solely on the CART-generated decision tree. First, this CART model had a maximum depth of just five levels. Truncating at five levels facilitates visualization, but a more realistic decision path is likely to be nuanced and so would require greater depth. Second, only some of the important variables (from Exhibit 16) were used in generating this tree, again for simplicity of visualization. A CART model using additional features, including fund asset class ratios, sector composition, and, especially, net assets would be expected to generate a more accurate (using F1 score) model. Finally, the number of funds reaching the darkest shaded, 5th level ("winner") terminal node (21) is small compared to the total sample size (1,067), so there may be too few clear winners (13) under this decision path from which to draw a statistically significant conclusion. Besides increasing the maximum tree depth and adding more features, another approach the analyst might take in this case for achieving a more accurate model is random forest; being an ensemble classifier, a random forest model would generalize out-of-sample better than any single CART model.

ESG Data as Alternative Data and ML/AI for Integrating ESG Data into Investment Decisions

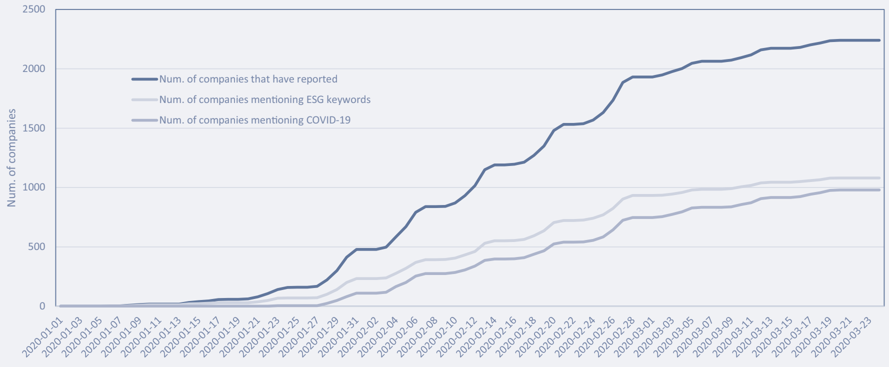
As an investment professional, how might you set about measuring the potential impact of climate change on a company's future prospects? Negative climate outcomes in coming years may include higher temperatures, more intense storms, melting glaciers, rising sea levels, shifting agricultural patterns, pressure on food and water, and new threats to human health. Assessing the likely severity of these future events and then quantifying the impact on companies is no easy task. Big Data techniques could be pivotal in generating usable information that could help investment professionals unlock long-term shareholder value.

Some fund managers, influenced by evolving investor preferences and increasing disclosure by companies on non-financial issues, have already incorporated ESG analysis into their investment processes. Governance ("G") data are generally objective: Investors are able to observe and measure corporate board actions, making governance comparable across companies and regions. Data on Environmental ("E") and Social ("S") impacts on listed companies, on the other hand, are more subjective, less reliable, and less comparable.

ESG data resemble alternative data in the sense that they have generally been poorly defined, are complex and unstructured, and need considerable due diligence before being used in investment decision making. Applying Machine Learning (ML) and Artificial Intelligence (AI) techniques can transform ESG data into meaningful information that is more useful for investment analysis.

Corporate sustainability reports often suffer from haphazard data collection and missing values. Equally, when data vendors acquire and combine raw ESG data into aggregate ESG scores, potential signals may be lost. ESG data and scoring across companies and data vendors can lack consistency and comparability; as a result, using simple summary scores in investment analysis is potentially flawed. Data analysts can apply data-science methods, such as data cleansing and data wrangling, to raw ESG data to create a structured dataset. Then, ML/AI techniques, such as natural language processing (NLP), can be applied to text-based, video, or audio ESG data. The foundation of NLP consists of supervised machine learning algorithms that typically include logistic regression, SVM, CART, random forests, or neural networks.

NLP can, for instance, search for key ESG words in corporate earnings calls. An increase in the number of mentions of, say, "human capital," employee "health and safety," or "flexible working" arrangements may indicate an increased focus on the "S" pillar of ESG. This would potentially raise the overall ESG score of a particular company. The results of such an application of NLP to corporate earnings calls are illustrated in the following exhibit:



Source: "GS SUSTAIN: ESG—Neither Gone Nor Forgotten" by Evan Tylenda, Sharmini Chetwode, and Derek R. Bingham, Goldman Sachs Global Investment Research (2 April 2020).

ML/AI can help fund managers apply only those ESG factors that are relevant to a company and its sector. For example, "E" factors are important for mining and utility companies but less so for clothing manufacturers. Likewise, "S" factors are important for the global clothing manufacturing sector but less so for mining and utility companies.

ML/AI techniques are not used in isolation. ESG scoring systems tend to rely on cross-functional teams, with data scientists operating in tandem with economists, fundamental analysts, and portfolio managers to identify strengths and weaknesses of companies and sectors. Fundamental analysts, for instance, typically do not need to know the details of ML algorithms to make valuable contributions to the ESG investment workflow. The industry-specific knowledge of fundamental analysts can provide nuanced viewpoints that help to: 1) identify relevant raw data; 2) enable data scientists to incorporate ESG data into appropriate investment models; and 3) interpret model outputs and investment implications.

UNSUPERVISED MACHINE LEARNING ALGORITHMS AND PRINCIPAL COMPONENT ANALYSIS

9

d describe unsupervised machine learning algorithms—including principal components analysis, K-means clustering, and hierarchical clustering—and determine the problems for which they are best suited

Unsupervised learning is machine learning that does not use labeled data (i.e., no target variable); thus, the algorithms are tasked with finding patterns within the data themselves. The two main types of unsupervised ML algorithms shown in Exhibit 2 are dimension reduction, using principal components analysis, and clustering, which includes *k*-means and hierarchical clustering. These will now be described in turn.

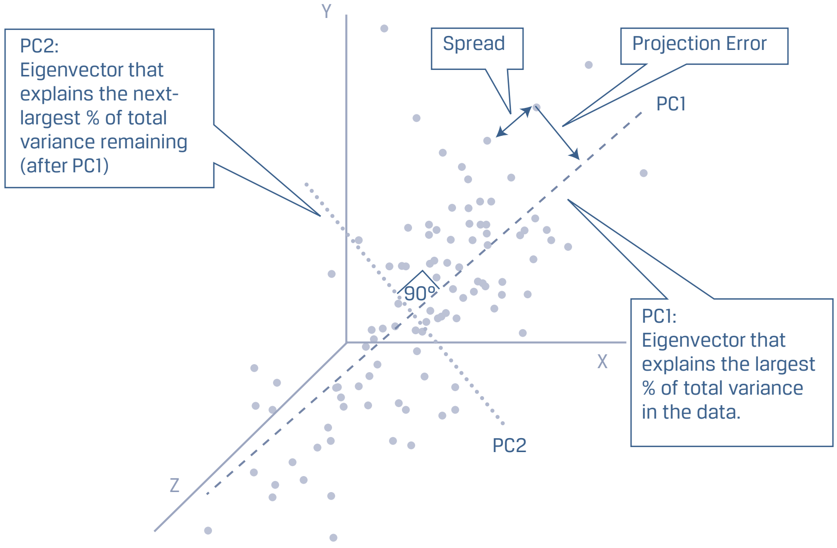
9.1 Principal Components Analysis

Dimension reduction is an important type of unsupervised learning that is used widely in practice. When many features are in a dataset, representing the data visually or fitting models to the data may become extremely complex and "noisy" in the sense of reflecting random influences specific to a dataset. In such cases, dimension reduction may be necessary. Dimension reduction aims to represent a dataset with many typically correlated features by a smaller set of features that still does well in describing the data.

A long-established statistical method for dimension reduction is **principal components analysis (PCA)**. PCA is used to summarize or transform highly correlated features of data into a few main, uncorrelated composite variables. A **composite variable** is a variable that combines two or more variables that are statistically strongly related to each other. Informally, PCA involves transforming the covariance matrix of the features and involves two key concepts: eigenvectors and eigenvalues. In the context of PCA, **eigenvectors** define new, mutually uncorrelated composite variables that are linear combinations of the original features. As a vector, an eigenvector also represents a direction. Associated with each eigenvector is an eigenvalue. An **eigenvalue** gives the proportion of total variance in the initial data that is explained by each eigenvector. The PCA algorithm orders the eigenvectors from highest to lowest according to their eigenvalues—that is, in terms of their usefulness in explaining the total variance in the initial data (this will be shown shortly using a scree plot). PCA selects as the first principal component the eigenvector that explains the largest proportion of variation in the dataset (the eigenvector with the largest eigenvalue). The second principal component explains the next-largest proportion of variation remaining after the first principal component; this process continues for the third, fourth, and subsequent principal components. Because the principal components are linear combinations of the initial feature set, only a few principal components are typically required to explain most of the total variance in the initial feature covariance matrix.

Exhibit 18 shows a hypothetical dataset with three features, so it is plotted in three dimensions along the *x*-, *y*-, and *z*-axes. Each data point has a measurement (*x*, *y*, *z*), and the data should be standardized so that the mean of each series (*x*'s, *y*'s, and *z*'s) is 0 and the standard deviation is 1. Assume PCA has been applied, revealing the first two principal components, PC1 and PC2. With respect to PC1, a perpendicular line dropped from each data point to PC1 shows the vertical distance between the data point and PC1, representing **projection error**. Moreover, the distance between each data point in the direction that is parallel to PC1 represents the spread or variation of the data along PC1. The PCA algorithm operates in such a way that it finds PC1 by selecting the line for which the sum of the projection errors for all data points is minimized and for which the sum of the spread between all the data is maximized. As a consequence of these selection criteria, PC1 is the unique vector that accounts for the largest proportion of the variance in the initial data. The next-largest portion of the remaining variance is best explained by PC2, which is at right angles to PC1 and thus is uncorrelated with PC1. The data points can now be represented by the first two principal components. This example demonstrates the effectiveness of the PCA algorithm in summarizing the variability of the data and the resulting dimension reduction.

Exhibit 18 First and Second Principal Components of a Hypothetical Three-Dimensional Dataset



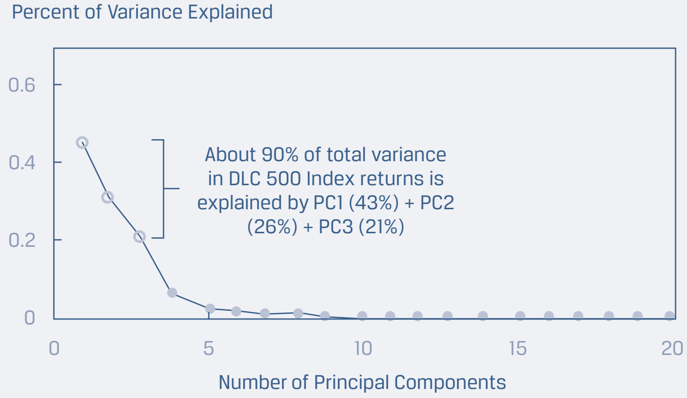
It is important to know how many principal components to retain because there is a trade-off between a lower-dimensional, more manageable view of a complex dataset when a few are selected and some loss of information. **Scree plots**, which show the proportion of total variance in the data explained by each principal component, can be helpful in this regard (see the accompanying sidebar). In practice, the smallest number of principal components that should be retained is that which the scree plot shows as explaining a desired proportion of total variance in the initial dataset (often 85% to 95%).

Scree Plots for the Principal Components of Returns to the Hypothetical DLC 500 and VLC 30 Equity Indexes

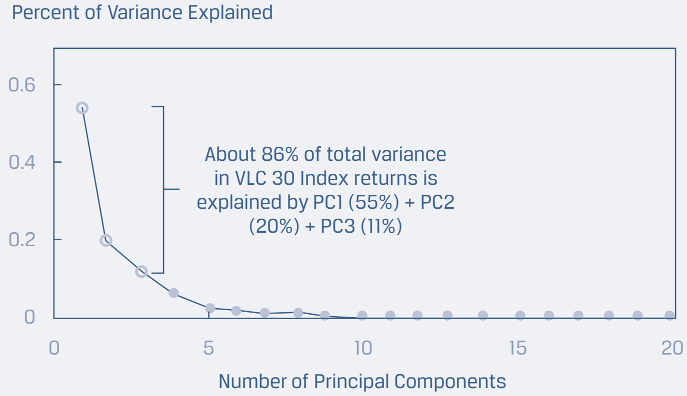
In this illustration, researchers use scree plots and decide that three principal components are sufficient for explaining the returns to the hypothetical Diversified Large Cap (DLC) 500 and Very Large Cap (VLC) 30 equity indexes over the last 10-year period. The DLC 500 can be thought of as a diversified index of large-cap companies covering all economic sectors, while the VLC 30 is a more concentrated index of the 30 largest publicly traded companies. The dataset consists of index prices and more than 2,000 fundamental and technical features. Multi-collinearity among the features is a typical problem because that many features or combinations of features tend to have overlaps. To mitigate the problem, PCA can be used to capture the information and variance in the data. The following scree plots show that of the 20 principal components generated, the first 3 together explain about 90% and 86% of the variance in the value of the DLC 500 and VLC 30 indexes, respectively. The scree plots indicate that for each of these indexes, the incremental contribution to explaining the variance structure of the data is quite small after about the fifth principal component. Therefore, these less useful principal components can be ignored without much loss of information.

Scree Plots of Percent of Total Variance Explained by Each Principal Component for Hypothetical DLC 500 and VLC 30 Equity Indexes

A. Diversified Large Cap 500 Index (DLC 500)



B. Very Large Cap 30 Index (VLC 30)



The main drawback of PCA is that since the principal components are combinations of the dataset's initial features, they typically cannot be easily labeled or directly interpreted by the analyst. Compared to modeling data with variables that represent well-defined concepts, the end user of PCA may perceive PCA as something of a "black box."

Reducing the number of features to the most relevant predictors is very useful, even when working with datasets having as few as 10 or so features. Notably, dimension reduction facilitates visually representing the data in two or three dimensions. It is typically performed as part of exploratory data analysis, before training another supervised or unsupervised learning model. Machine learning models are quicker to train, tend to reduce overfitting (by avoiding the curse of dimensionality), and are easier to interpret if provided with lower-dimensional datasets.

10 CLUSTERING

d describe unsupervised machine learning algorithms—including principal components analysis, k-means clustering, and hierarchical clustering—and determine the problems for which they are best suited

Clustering is another type of unsupervised machine learning, which is used to organize data points into similar groups called clusters. A **cluster** contains a subset of observations from the dataset such that all the observations within the same cluster are deemed "similar." The aim is to find a good clustering of the data—meaning that the observations inside each cluster are similar or close to each other (a property known as cohesion) and the observations in two different clusters are as far away from one another or are as dissimilar as possible (a property known as separation). Exhibit 19 depicts this intra-cluster cohesion and inter-cluster separation.

Exhibit 19 Evaluating Clustering—Intra-Cluster Cohesion and Inter-Cluster Separation

Clustering algorithms are particularly useful in the many investment problems and applications in which the concept of similarity is important. Applied to grouping companies, for example, clustering may uncover important similarities and differences among companies that are not captured by standard classifications of companies by industry and sector. In portfolio management, clustering methods have been used for improving portfolio diversification.

In practice, expert human judgment has a role in using clustering algorithms. In the first place, one must establish what it means to be "similar." Each company can be considered an observation with multiple features, including such financial statement items as total revenue and profit to shareholders, a wide array of financial ratios, or any other potential model inputs. Based on these features, a measure of similarity or "distance" between two observations (i.e., companies) can be defined. The smaller the distance, the more similar the observations; the larger the distance, the more dissimilar the observations.

A commonly used definition of distance is the Euclidian distance, the straight-line distance between two points. A closely related distance useful in portfolio diversification is correlation, which is the average Euclidian distance between a set of standardized points. Roughly a dozen different distance measures are used regularly in ML. In practice, the choice of the distance measures depends on the nature of the data (numerical or not) and the business problem being investigated. Once the relevant distance measure is defined, similar observations can be grouped together. We now introduce two of the more popular clustering approaches: *k*-means and hierarchical clustering.

K-MEANS CLUSTERING 11

d describe unsupervised machine learning algorithms—including principal components analysis, k-means clustering, and hierarchical clustering—and determine the problems for which they are best suited

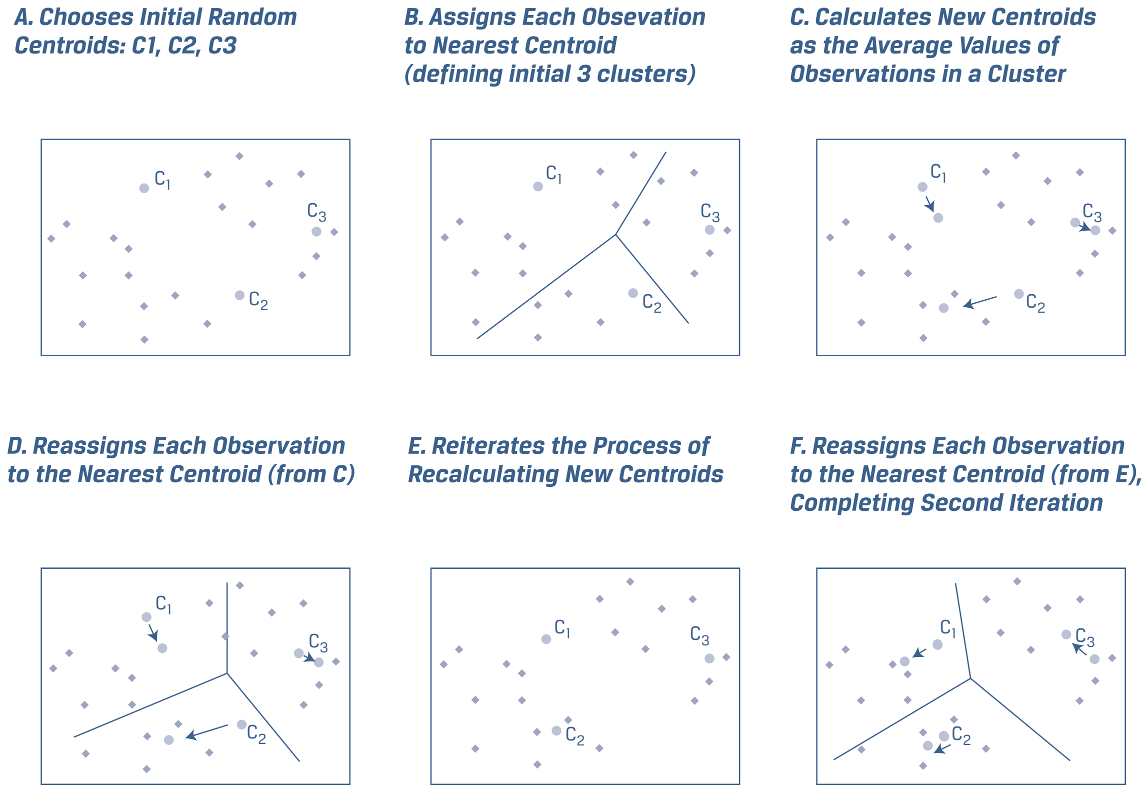
***K*-means** is an algorithm that repeatedly partitions observations into a fixed number, *k*, of non-overlapping clusters. The number of clusters, *k*, is a model hyperparameter. Each cluster is characterized by its **centroid** (i.e., center), and each observation is assigned by the algorithm to the cluster with the centroid to which that observation is closest. Notably, once the clusters are formed, there is no defined relationship between them.

The *k*-means algorithm follows an iterative process. It is illustrated in Exhibit 20 for $k = 3$ and a set of observations on a variable that can be described by two features. In Exhibit 20, the horizontal and vertical axes represent, respectively, the first and second features. For example, an investment analyst may want to group a set of firms into three groups according to two numerical measures of management quality. The algorithm groups the observations in the following steps:

1. *K*-means starts by determining the position of the *k* (here, 3) initial random centroids.
2. The algorithm then analyzes the features for each observation. Based on the distance measure that is used, *k*-means assigns each observation to its closest centroid, which defines a cluster.
3. Using the observations within each cluster, *k*-means then calculates the new (*k*) centroids for each cluster, where the centroid is the average value of their assigned observations.
4. *K*-means then reassigns the observations to the new centroids, redefining the clusters in terms of included and excluded observations.
5. The process of recalculating the new (*k*) centroids for each cluster is reiterated.
6. *K*-means then reassigns the observations to the revised centroids, again redefining the clusters in terms of observations that are included and excluded.

The *k*-means algorithm will continue to iterate until no observation is reassigned to a new cluster (i.e., no need to recalculate new centroids). The algorithm has then converged and reveals the final *k* clusters with their member observations. The *k*-means algorithm has minimized intra-cluster distance (thereby maximizing cohesion) and has maximized inter-cluster distance (thereby maximizing separation) under the constraint that $k = 3$.

Exhibit 20 Example of 3-Means Algorithm



The *k*-means algorithm is fast and works well on very large datasets, those with hundreds of millions of observations. However, the final assignment of observations to clusters can depend on the initial location of the centroids. To address this problem, the algorithm can be run several times using different sets of initial centroids, and then one can choose the clustering that is most useful given the business purpose.

One limitation of this technique is that the hyperparameter, *k*, the number of clusters in which to partition the data, must be decided before *k*-means can be run. So, one needs to have a sense of how many clusters are reasonable for the problem under investigation and the dataset being analyzed. Alternatively, one can run the algorithm using a range of values for *k* to find the optimal number of clusters—the *k* that minimizes intra-cluster distance and thereby maximizes intra-cluster similarity (i.e., cohesion) and that maximizes inter-cluster distance (i.e., separation). However, note that the final results can be subjective and dependent on the context of the problem and the particular training set. In practice, it is common to make the final choice of *k* based on face validity, such that the clusters feel sensible and are interpretable. This decision is greatly assisted by using summary information about the centroids and ranges of values and naming example items in each cluster.

For example, consider the Russell 3000 Index, which tracks the 3,000 highest market capitalization stocks in the United States. These 3,000 stocks can be grouped in 10, 50, or even more clusters based on their financial characteristics (e.g., total assets, total revenue, profitability, leverage) and operating characteristics (e.g., employee headcount, R&D intensity). Because companies in the same standard industry classification can have very different financial and operating characteristics, using *k*-means to derive different clusters can provide insights and understanding into the nature of "peer" groups. As mentioned, the exact choice of the *k*, the number of clusters, will depend on the level of precision or segmentation desired. In a similar vein, clustering can be used to classify collective investment vehicles or hedge funds as an alternative to standard classifications. Clustering analysis can also help visualize the data and facilitate detecting trends or outliers.

In sum, the *k*-means algorithm is among the most used algorithms in investment practice, particularly in data exploration for discovering patterns in high-dimensional data or as a method for deriving alternatives to existing static industry classifications.

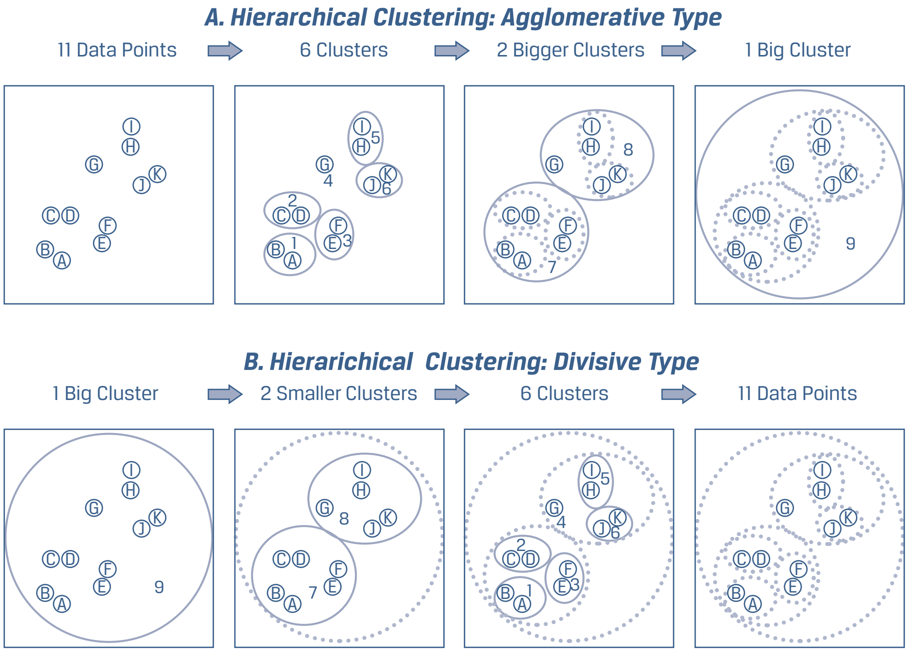
12 HIERARCHICAL CLUSTERING: AGGLOMERATIVE AND DENDROGRAMS

d describe unsupervised machine learning algorithms—including principal components analysis, k-means clustering, and hierarchical clustering—and determine the problems for which they are best suited

Hierarchical clustering is an iterative procedure used to build a hierarchy of clusters. In *k*-means clustering, the algorithm segments the data into a predetermined number of clusters; there is no defined relationship among the resulting clusters. In hierarchical clustering, however, the algorithms create intermediate rounds of clusters of increasing (in "agglomerative") or decreasing (in "divisive") size until a final clustering is reached. The process creates relationships among the rounds of clusters, as the word "hierarchical" suggests. Although more computationally intensive than *k*-means clustering, hierarchical clustering has the advantage of allowing the investment analyst to examine alternative segmentations of data of different granularity before deciding which one to use.

Agglomerative clustering (or bottom-up hierarchical clustering) begins with each observation being treated as its own cluster. Then, the algorithm finds the two closest clusters, defined by some measure of distance (similarity), and combines them into one new larger cluster. This process is repeated iteratively until all observations are clumped into a single cluster. A hypothetical example of how agglomerative clustering develops a hierarchical clustering scheme is depicted in the top part of Exhibit 21, where observations are lettered (A to K) and circles around observations denote clusters. The process begins with 11 individual clusters and then generates a sequence of groupings. The first sequence includes five clusters with two observations each and one cluster with a single observation, G, for a total of six clusters. It then generates two clusters—one cluster with six observations and the other with five observations. The final result is one large cluster containing all 11 observations. It is easily seen that this final large cluster includes the two main sub-clusters, with each containing three smaller sub-clusters.

Exhibit 21 Agglomerative and Divisive Hierarchical Clustering



By contrast, **divisive clustering** (or top-down hierarchical clustering) starts with all the observations belonging to a single cluster. The observations are then divided into two clusters based on some measure of distance (similarity). The algorithm then progressively partitions the intermediate clusters into smaller clusters until each cluster contains only one observation. Divisive clustering is depicted in the bottom part of Exhibit 21, which begins with all 11 observations in one large cluster. Next, the algorithm generates two smaller clusters, one with six observations and the other with five observations, and then six clusters, with two observations each except for observation G, which is its own cluster. Finally, 11 clusters are generated, with each cluster containing only one observation.

Although this is not a typical outcome (because the two methods generally use different algorithms), in this hypothetical illustration, the agglomerative and divisive clustering methods produced the same result: two main sub-clusters each having three smaller sub-clusters. The analyst could decide between using a six- or a two-cluster representation of the data. The agglomerative method is the approach typically used with large datasets because of the algorithm's fast computing speed. The agglomerative clustering algorithm makes clustering decisions based on local patterns without initially accounting for the global structure of the data. As such, the agglomerative method is well suited for identifying small clusters. However, because the divisive method starts with a holistic representation of the data, the divisive clustering algorithm is designed to account for the global structure of the data and thus is better suited for identifying large clusters.

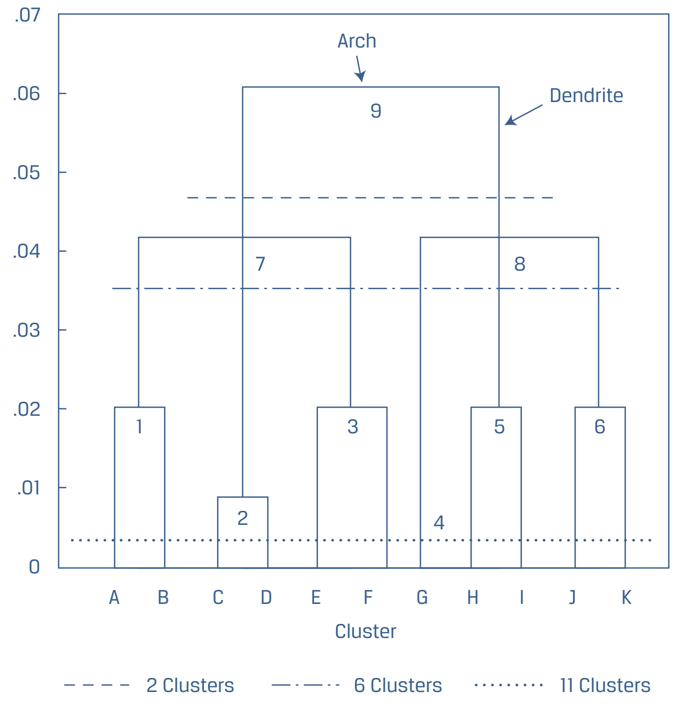
To decide on the closest clusters for combining in the agglomerative process or for dividing in the divisive process, an explicit definition for the distance between two clusters is required. Some commonly used definitions for the distance between two clusters involve finding the minimum, the maximum, or the average of the straight-line distances between all the pairs of observations in each cluster.

12.1 Dendrograms

A type of tree diagram for visualizing a hierarchical cluster analysis is known as a **dendrogram**, which highlights the hierarchical relationships among the clusters. Exhibit 22 shows a dendrogram representation for the clustering shown in Exhibit 21. First, a few technical points on dendrograms bear mentioning—although they may not all be apparent in Exhibit 22. The x-axis shows the clusters, and the y-axis indicates some distance measure. Clusters are represented by a horizontal line, the arch, which connects two vertical lines, called dendrites, where the height of each arch represents the distance between the two clusters being considered. Shorter dendrites represent a shorter distance (and greater similarity) between clusters. The horizontal dashed lines cutting across the dendrites show the number of clusters into which the data are split at each stage.

The agglomerative algorithm starts at the bottom of the dendrite, where each observation is its own cluster (A to K). Agglomerative clustering then generates the six larger clusters (1 to 6). For example, Clusters A and B combine to form Cluster 1, and Observation G remains its own cluster, now Cluster 4. Moving up the dendrogram, two larger clusters are formed, where, for example, Cluster 7 includes Clusters 1 to 3. Finally, at the top of the dendrogram is the single large cluster (9). The dendrogram readily shows how this largest cluster is composed of the two main sub-clusters (7 and 8), each having three smaller sub-clusters (1 to 3 and 4 to 6, respectively). The dendrogram also facilitates visualization of divisive clustering by starting at the top of the largest cluster and then working downward until the bottom is reached, where all 11 single-observation clusters are shown.

Exhibit 22 Dendrogram of Agglomerative Hierarchical Clustering



Clustering has many applications in investment management. For example, portfolio diversification can be approached as a clustering problem with the aim of optimally diversifying risks by investing in assets from multiple different clusters. Because the clusters have maximum inter-cluster separation, diversifying among them helps ensure that the portfolio reflects a wide diversity of characteristics with well-diversified risk. In contrast, information that investments are concentrated in a cluster indicates a high probability of concentrated risk. Finally, it is important to note that while the results of clustering algorithms are often difficult to evaluate (because the resulting clusters themselves are not explicitly defined), they are still very useful in practice for uncovering important underlying structure (namely, similarities among observations) in complex datasets.

EXAMPLE 6

Investment Uses of Clustering Algorithms

István Perényi is a portfolio manager of the Europe Diversified Equity Fund ("the Fund") within the Diversified Investment Management Company (DIMCO) fund family. The Fund is benchmarked to the STOXX Europe 600 Index, which spans 17 countries, 19 industry sectors, and three market capitalization groupings (large-, mid-, and small-cap).

Examining the Fund's most recent performance, Perényi is concerned that the Fund's holdings, although approximately aligned with the STOXX Europe 600 Index's country weights, may have unrecognized risk biases and concentrations. Perényi asks Elsa Lund, DIMCO's chief risk officer, to investigate the Fund's diversification. Lund asks her analysts for ideas on how Perényi's request can be addressed and receives three suggestions:

Suggestion 1 Estimate the country, industry, and market cap exposures of each Fund holding, aggregate them, and compare the aggregate exposures to the benchmark's exposures. Then, examine mismatches for evidence of unexpected biases or concentrations.

Suggestion 2 Identify inherent groupings among fund holdings based on a broad set of eight numerical (operating and financial) measures related to the holdings' characteristics. Then, examine the groupings for evidence of unexpected biases or concentrations.

Suggestion 3 Regress the return of the Fund on a set of country equity market indexes and sector indexes based on the Fund's benchmark. Then, examine the regression coefficients for evidence of unexpected biases or concentrations.

Lund has several questions for analyst Greg Kane about using one or more clustering machine learning algorithms in relation to addressing Perényi's request.

Lund asks whether any information needs to be specified for the ML clustering algorithms no matter which one is used. Kane replies that only the distance measure that the algorithm will use and the hyperparameter, k, for k-means clustering need to be specified.

Lund further asks whether there would be an advantage to using k-means clustering as opposed to hierarchical clustering. Kane replies that in his opinion, hierarchical clustering is the more appropriate algorithm.

1 Which analyst suggestion is *most likely* to be implemented using machine learning?

A Suggestion 1

B Suggestion 2

C Suggestion 3

2 Kane's reply to Lund's first question about specification of ML clustering models is:

A correct.

B not correct, because other hyperparameters must also be specified.

C not correct, because the feature set for describing the measure used to group holdings must also be specified.

3 The best justification for Kane's preference for hierarchical clustering in his reply to Lund's second question is that Kane is *most likely* giving consideration to:

A the speed of the algorithms.

B the dimensionality of the dataset.

C the need to specify the hyperparameter, k, in using a k-means algorithm.

Solution to 1:

B is correct. A machine learning clustering algorithm could be used to implement Suggestion 2. A and C are incorrect because Suggestions 1 and 3, respectively, can be addressed easily using traditional regression analysis.

Solution to 2:

C is correct. Beyond specifying a distance measure and the k for k-means, whichever clustering algorithm is selected, the feature set used to group holdings by similarities must also be specified. Operating and financial characteristics of the companies represented in the Fund's portfolio are examples of such features.

Solution to 3:

C is correct. The value of the hyperparameter, k, the number of distinct groups into which the STOXX Europe 600 Index can be segmented, is not known and needs to be specified in advance by the analyst. Using a hierarchical algorithm, the sorting of observations into clusters will occur without any prior input on the analyst's part.

CASE STUDY: CLUSTERING STOCKS BASED ON CO-MOVEMENT SIMILARITY

The following case study was developed and written by Matthew Dixon, PhD, FRM.

An endowment fund's Investment Committee is seeking three "buy" recommendations for the fund's large-cap equity portfolio. An analyst working for the Investment Committee is given a subset of eight stocks from the S&P 500 Index and asked to determine the co-movement similarity (i.e., correlation) of their returns. Specifically, for diversification purposes, the Investment Committee wants the correlation of returns between the recommended stocks to be low, so the analyst decides to use clustering to identify the most similar stocks and then choose one stock from each cluster. Although this case study focuses mainly on hierarchical agglomerative clustering, the analyst's results using other clustering algorithms (i.e., divisive clustering and k-means) are also briefly discussed. Exhibit 23 provides a description of the data used by the analyst.

Exhibit 23 Dataset of Eight Stocks from the S&P 500 Index

Description: Daily adjusted closing prices of eight S&P 500 member stocks

Trading Dates: 30 May 2017 to 24 May 2019

Number of Observations: 501

Stocks (Ticker Symbols): AAPL, F, FB, GM, GS, GOOG, JPM, and UBS

The following steps are taken by the analyst to perform the hierarchical agglomerative cluster analysis:

1. Collect panel data on adjusted closing prices for the stocks under investigation.
2. Calculate the daily log returns for each stock, where each time series of stock returns is an n-vector (n = 500).
3. Run the agglomerative hierarchical clustering algorithm.
   a. The algorithm calculates the pairwise distance (i.e., Euclidean distance) between vectors of any two stocks' returns. Each pairwise distance is an element of a distance matrix (i.e., dissimilarity matrix) with zero diagonals.
   b. The algorithm starts with each stock as its own cluster, finds the pair of clusters which are closest to each other, and then redefines them as a new cluster.
   c. The algorithm finds the distances from this new cluster to the remaining return clusters. Using a process called average (centroid) linkage, it determines the distances from the center of the new cluster to the centers of the remaining clusters. Note that there are several other linkage methods, but whichever method is selected, the algorithm proceeds in the same fashion: It combines the pair of clusters which are closest, redefines them as a new cluster, and recalculates the distances to the remaining clusters.
4. Repeat Step 3c until the data are aggregated into a single large cluster.
5. Plot the resulting dendrogram to visualize the hierarchical clusters and draw the highest horizontal line intersecting three (i.e., the desired number of clusters, since the Investment Committee wants three "buy" recommendations) vertical lines (or dendrites) to determine the appropriate cluster configuration.

Exhibit 24 shows for illustrative purposes a subset of the panel data on daily returns, calculated from the adjusted closing prices of the eight stocks collected in Step 1. The clustering is performed on the daily returns.

Exhibit 24 Subset of Stock Returns, Calculated from Adjusted Closing Prices, for Clustering

Date	JPM	UBS	GS	FB	AAPL	GOOG	GM	F
2017-05-31	−0.021	−0.007	−0.033	−0.006	−0.006	−0.011	0.012	0.004
2017-06-01	0.011	0.013	0.018	0.000	0.003	0.002	0.015	0.026
2017-06-02	−0.005	−0.002	−0.008	0.014	0.015	0.009	0.001	−0.005
2017-06-05	0.002	−0.007	0.003	0.000	−0.010	0.008	0.000	−0.009
2017-06-06	0.002	0.002	0.003	−0.005	0.003	−0.007	−0.001	−0.012

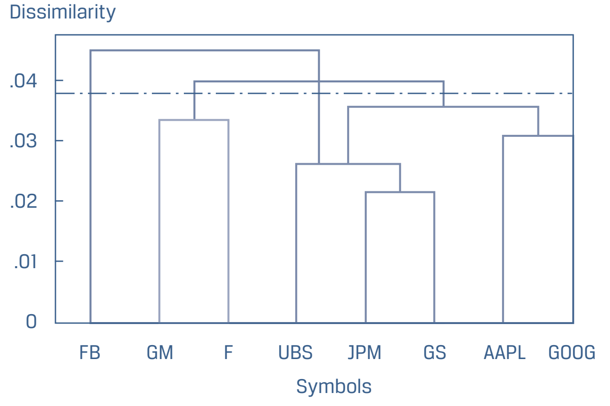
The results of the remaining steps are described using the distance matrix shown in Exhibit 25.

Exhibit 25 Distance Matrix for Hierarchical Agglomerative Clustering

	JPM	UBS	GS	FB	AAPL	GOOG	GM	F
JPM	0.000	0.243	0.215	0.456	0.364	0.332	0.358	0.348
UBS	0.243	0.000	0.281	0.460	0.380	0.338	0.384	0.385
GS	0.215	0.281	0.000	0.471	0.375	0.345	0.383	0.393
FB	0.456	0.460	0.471	0.000	0.437	0.357	0.491	0.480
AAPL	0.364	0.380	0.375	0.437	0.000	0.307	0.445	0.456
GOOG	0.332	0.338	0.345	0.357	0.307	0.000	0.405	0.422
GM	0.358	0.384	0.383	0.491	0.445	0.405	0.000	0.334
F	0.348	0.385	0.393	0.480	0.456	0.422	0.334	0.000

The distance matrix reveals the closest pair of stocks is JPM and GS, with a distance of 0.215. Therefore, this pair becomes the first combined cluster as shown in the dendrogram in Exhibit 26. Note that the vertical distance connecting the various clusters represents the Euclidean distance between clusters, so the arch between this pair has a height of 0.215. Now that JPM and GS are paired in a cluster (i.e., GS_JPM), we treat the mean of their two return vectors as a new point.

Exhibit 26 Dendrogram for Hierarchical Agglomerative Clustering



From the distance matrix, the average distance of UBS to the new cluster (i.e., GS_JPM) is the sum of the distance between UBS and JPM, 0.243, and the distance between UBS and GS, 0.281, divided by two, which is 0.262 (= (0.243 + 0.281)/2). Since this distance is smaller than the distance between any of the other unpaired stock clusters, UBS is merged with this cluster to create a new cluster (i.e., GS_JPM_UBS). The height of the arch in the dendrogram for this new cluster is 0.262, which is now observed to contain three banking sector stocks. Although not shown in the dendrogram, the cluster is identified by the return vector averaged over the three stocks.

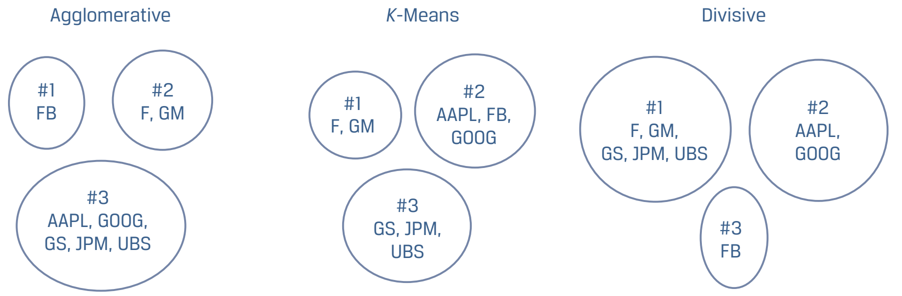
The next closest pair of points, whether stock to stock or stock to cluster, is AAPL and GOOG, with a distance of 0.307, so the algorithm merges these two points into a second cluster (i.e., AAPL_GOOG), with an arch height of 0.307. Next, GM and F are paired into a third cluster (i.e., F_GM), with an arch height of 0.334. Finally, the first two clusters are merged to form a five-stock cluster (i.e., GS_JPM_UBS_AAPL_GOOG), with an arch height of 0.356. Note that this value is determined by taking the average distance between the three banks and AAPL and GOOG: 0.356 = (0.364 + 0.380 + 0.375 + 0.332 + 0.338 + 0.345)/6. The result is three separate clusters: the five-stock cluster, F_GM, and FB by itself. Also, note the horizontal dashed line that cuts the dendrogram into three distinct clusters, with FB as its own cluster.

This agglomerative hierarchical clustering analysis reveals some interesting preliminary results—largely grouping the stocks by their sectors but also uncovering some anomalies. In particular, FB is found to behave quite differently, in terms of return co-movement similarity, from the other technology stocks (AAPL and GOOG). Also, AAPL and GOOG are found to behave more like the bank stocks and less like the auto stocks (F and GM), which appear in their own cluster.

In contrast to agglomerative clustering, the divisive clustering algorithm starts with all stocks assigned to one large cluster and then splits the cluster into sub-clusters recursively, until each stock occupies its own cluster. Determining how to split the first cluster requires searching over all combinations of possible splits, so it is too numerically intensive to cover the details here. However, results of the first two splits for divisive clustering, into three clusters, are shown in Exhibit 27. Results for *k*-means, with *k* = 3, and agglomerative clustering are also presented.

Exhibit 27 Comparison of Results of Different Clustering Algorithms

	Agglomerative	K-means	Divisive
AAPL	3	2	2
F	2	1	1
FB	1	2	3
GM	2	1	1
GOOG	3	2	2
GS	3	3	1
JPM	3	3	1
UBS	3	3	1



Whereas the assignment of the cluster number (1, 2, 3), shown in the upper panel, can be taken as arbitrary across each algorithm, the useful information is in the grouping of like stocks. As seen in the stylized clusters in the lower panel, all three clustering algorithms agree that bank stocks belong in the same cluster. Both hierarchical agglomerative and *k*-means algorithms also agree that auto stocks belong in their own separate cluster. *K*-means clusters the stocks precisely by industry sector, whereas hierarchical agglomerative and divisive clustering identify FB as an outlier and place it in its own cluster. In general, the most agreement is expected between the two hierarchical clustering algorithms, although their results are not guaranteed to match, even when using the same linkage process. *K*-means starts with three clusters (k = 3) and iteratively swaps points in and out of these clusters using a partitioning mechanism different from that of hierarchical clustering. Thus, *k*-means results are typically not expected to match those of hierarchical clustering.

In conclusion, based on the analyses of the co-movement similarity of returns among the eight stocks using the agglomerative clustering algorithm and the Investment Committee's requirement that the correlation of returns between the recommended stocks should be low, the analyst's recommendation should be as follows:

- buy FB,
- buy the most attractive of the two auto stocks (F or GM), and
- buy the most attractive of the three bank stocks (GS, JPM, or UBS).

EXAMPLE 7

Hierarchical Agglomerative Clustering

Assume the analyst is given the same set of stocks as previously excluding F and GM (i.e., no auto stocks)—so now, six stocks. Using the information from this mini-case study, answer the following questions:

1 Describe how the inputs to the hierarchical agglomerative clustering algorithm would differ from those in the mini-case study.

2 Describe the three clusters that would now result from running the hierarchical agglomerative clustering algorithm.

3 Explain why these results differ from the previous case, with eight stocks (including the two auto stocks).

4 Describe the analyst's new recommendation to the Investment Committee.

Solution to 1:

The panel data on closing prices and daily log returns would include the same stocks as before but without F and GM—so, AAPL, FB, GOOG, GS, JPM, and UBS. The distance matrix would also appear the same except without F, GM, or any of the pairwise distances between them and the remaining stocks.

Solution to 2:

The three clusters that would now result from running the agglomerative clustering algorithm are GS_JPM_UBS (i.e., one cluster of three bank stocks), AAPL_GOOG (i.e., one cluster of two technology stocks), and FB by itself.

Solution to 3:

The agglomerative clustering algorithm now combines GS and JPM and then UBS, as before, to form a bank cluster. Next, and as previously, the algorithm combines AAPL and GOOG into a cluster. However, without the auto stocks, there is no need to combine AAPL_GOOG with the bank cluster. There are now three distinct clusters, since (as before) the algorithm treats FB as its own cluster, given the high degree of return co-movement dissimilarity between FB and the other clusters (i.e., AAPL_GOOG, and GS_JPM_UBS).

Solution to 4:

The analyst's new recommendation to the Investment Committee would be to buy FB, buy the cheapest of AAPL or GOOG, and buy the most attractive of the three bank stocks (GS, JPM, or UBS).

13 NEURAL NETWORKS, DEEP LEARNING NETS AND REINFORCEMENT LEARNING AND NEURAL NETWORKS

e describe neural networks, deep learning nets, and reinforcement learning

The artificial intelligence revolution has been driven in large part by advances in neural networks, deep learning algorithms, and reinforcement learning. These sophisticated algorithms can address highly complex machine learning tasks, such as image classification, face recognition, speech recognition, and natural language processing. These complicated tasks are characterized by non-linearities and interactions between large numbers of feature inputs. We now provide an overview of these algorithms and their investment applications.

13.1 Neural Networks

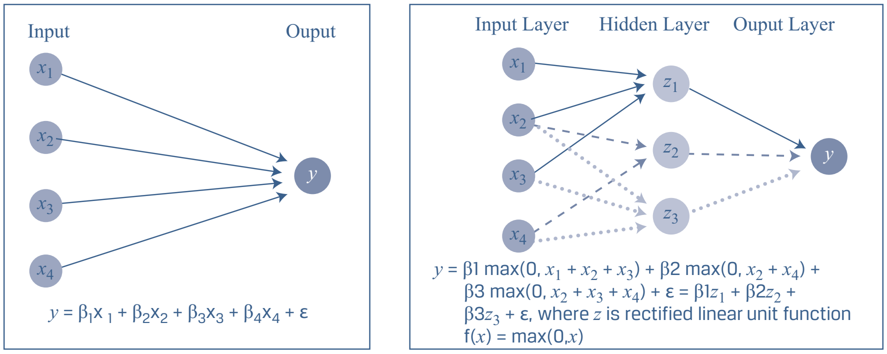
Neural networks (also called artificial neural networks, or ANNs) are a highly flexible type of ML algorithm that have been successfully applied to a variety of tasks characterized by non-linearities and complex interactions among features. Neural networks are commonly used for classification and regression in supervised learning but are also important in reinforcement learning, which does not require human-labeled training data.

Exhibit 28 shows the connection between multiple regression and neural networks. Panel A represents a hypothetical regression for data using four inputs, the features x_1 to x_4, and one output—the predicted value of the target variable y. Panel B shows a schematic representation of a basic neural network, which consists of nodes (circles) connected by links (arrows connecting nodes). Neural networks have three types of layers: an input layer (here with a node for each of the four features); hidden layers, where learning occurs in training and inputs are processed on trained nets; and an output layer (here consisting of a single node for the target variable y), which passes information outside the network.

Besides the network structure, another important difference between multiple regression and neural networks is that the nodes in the neural network's hidden layer transform the inputs in a non-linear fashion into new values that are then combined into the target value. For example, consider the popular rectified linear unit (ReLU) function, $f(x) = \max(0, x)$, which takes on a value of zero if there is a negative input and takes on the value of the input if it is positive. In this case, y will be equal to β_1 times z_1, where z_1 is the maximum of $(x_1 + x_2 + x_3)$ or 0, plus β_2 times z_2, the maximum of $(x_2 + x_4)$ or 0, plus β_3 times z_3, the maximum of $(x_2 + x_3 + x_4)$ or 0, plus an error term.

Exhibit 28 Regression and Neural Networks (Regression with Transformed Features)

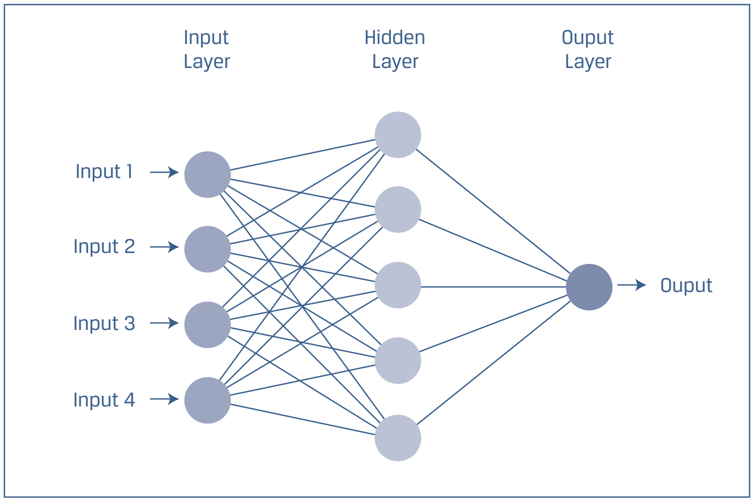
A. Conceptual Illustration of Regression *B. Conceptual Illustration of Hypothetical Neural Network*



Note that for neural networks, the feature inputs would be scaled (i.e., standardized) to account for differences in the units of the data. For example, if the inputs were positive numbers, each could be scaled by its maximum value so that their values lie between 0 and 1.

Exhibit 29 shows a more complex neural network, with an input layer consisting of four nodes (i.e., four features), one hidden layer consisting of five hidden nodes, and an output node. These three numbers—4, 5, and 1—for the neural network are hyperparameters that determine the structure of the neural network.

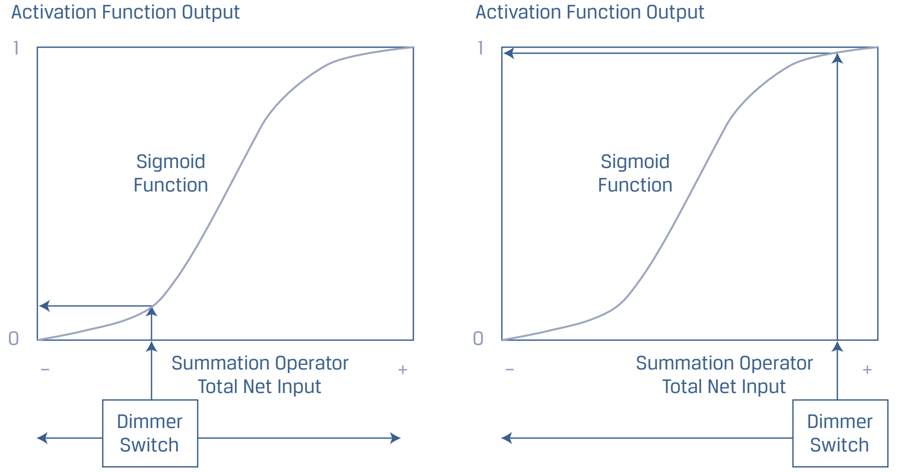
Exhibit 29 A More Complex Neural Network with One Hidden Layer



Now consider any of the nodes to the right of the input layer. These nodes are sometimes called "neurons" because they process information received. Take the topmost hidden node. Four links connect to that node from the inputs, so the node gets four values transmitted by the links. Each link has a weight meant to represent its importance (initially these weights may be assigned randomly). Each node has, conceptually, two functional parts: a summation operator and an activation function. Once the node receives the four input values, the **summation operator** multiplies each value by its respective weight and then sums the weighted values to form the total net input. The total net input is then passed to the **activation function**, which transforms this input into the final output of the node. Informally, the activation function operates like a light dimmer switch that decreases or increases the strength of the input. The activation function, which is chosen by the modeler (i.e., a hyperparameter), is characteristically non-linear, such as an S-shaped (sigmoidal) function (with output range of 0 to 1) or the rectified linear unit function shown in Panel B of Exhibit 28. Non-linearity implies that the rate of change of output differs at different levels of input.

This activation function is shown in Exhibit 30, where in the left graph a negative total net input is transformed via the S-shaped function into an output close to 0. This low output implies the node does not trigger, so there is nothing to pass to the next node. Conversely, in the right graph a positive total net input is transformed into an output close to 1, so the node does trigger. The output of the activation function is then transmitted to the next set of nodes if there is a second hidden layer or, as in this case, to the output layer node as the predicted value. The process of transmission just described (think of forward pointing arrows in Exhibit 29) is referred to as **forward propagation**.

Exhibit 30 Activation Function as "Light Dimmer Switch" at Each Node in a Neural Network



Starting with an initialized set of random network weights (i.e., the weights assigned to each of the links), training a neural network in a supervised learning context is an iterative process in which predictions are compared to actual values of labeled data and evaluated by a specified performance measure (e.g., mean squared error). Then, the network weights are adjusted to reduce total error of the network. (If the process of adjustment works backward through the layers of the network, this process is called **backward propagation**). Learning takes place through this process of adjustment to the network weights with the aim of reducing total error. Without proliferating notation relating to nodes, the gist of the updating can be expressed informally as

New weight = (Old weight) – (Learning rate) × (Partial derivative of the total error with respect to the old weight),

where partial derivative is a gradient or rate of change of the total error with respect to the change in the old weight and **learning rate** is a hyperparameter that affects the magnitude of adjustments. When learning is completed, all the network weights have assigned values; these are the parameters of the network.

The structure of a network in which all the features are interconnected with non-linear activation functions allows neural networks to uncover and approximate complex non-linear relationships among features. Broadly speaking, when more nodes and more hidden layers are specified, a neural network's ability to handle complexity tends to increase (but so does the risk of overfitting).

Asset pricing is a noisy, stochastic process with potentially unstable relationships that challenge modeling processes, so researchers are asking if machine learning can improve our understanding of how markets work. Research comparing statistical and machine learning methods' abilities to explain and predict equity prices so far indicates that simple neural networks produce models of equity returns at the individual stock and portfolio level that are superior to models built using traditional statistical methods due to their ability to capture dynamic and interacting variables. This suggests that ML-based models, such as neural networks, may simply be better able to cope with the non-linear relationships inherent in security prices. However, the trade-offs in using neural networks are their lack of interpretability (i.e., black box nature) and the large amounts of data and high computation intensity needed to train such models; thus, neural networks may not be a good choice in many investment applications.

DEEP LEARNING NETS, REINFORCEMENT AND LEARNING 14

e describe neural networks, deep learning nets, and reinforcement learning

The previous discussion of neural networks was limited to types of neural networks referred to as "shallow neural networks"—exhibiting just one hidden layer. Neural networks with many hidden layers—at least 2 but potentially more than 20—are known as **deep neural networks** (DNNs). DNNs are the foundation of deep learning and have proven to be successful across a wide range of artificial intelligence applications. Advances in DNNs have driven developments in many complex activities, such as image, pattern, and speech recognition. To state the operation of DNNs succinctly, they take a set of inputs x from a feature set (the input layer), which are then passed to a layer of non-linear mathematical functions (neurons) with weights w_{ij} (for neuron i and input j), each of which usually produces a scaled number in the range (0, 1) or (–1, 1). These numbers are then passed to another layer of functions and into another and so on until the final layer produces a set of probabilities of the observation being in any of the target categories (each represented by a node in the output layer). The DNN assigns the category based on the category with the highest probability. The DNN is trained on large datasets; during training, the weights, w_i, are determined to minimize a specified loss function.

In practice, while the number of nodes in the input and the output layers are typically determined by the characteristics of the features and predicted output, many model hyperparameters still must be decided, particularly the number of hidden layers, the number of nodes per hidden layer, and their connectivity and activation architecture. The objective is to choose them to achieve the best out-of-sample performance, but it is still a challenge with no simple solution. As such, a good starting point is a "reasonable" guess for hyperparameters based on experience and literature. The researcher can then observe the result and adjust the hyperparameters incrementally until the model performance goal is reached. In practice, DNNs require substantial time to train, and systematically varying the hyperparameters may not be feasible. So, for many problems with relatively small datasets, one can start with just two or three hidden layers and a few hundred nodes before tuning the parameters until a model with acceptable predictive power is achieved.

DNNs have been shown to be useful in general for pattern recognition problems (e.g., character and image recognition), credit card fraud detection, vision and control problems in autonomous cars, natural language processing (such as machine translation), and other applications. DNNs have become hugely successful because of a confluence of three developments: (1) the availability of large quantities of machine-readable data to train models, (2) advances in analytical methods for fitting these models, and (3) fast computers, especially new chips in the graphics processing unit (GPU) class, tailored for the type of calculations done on DNNs.

Several financial firms are experimenting with DNNs for trading as well as automating their internal processes. Culkin and Das (2017) described how they trained DNNs to price options, mimicking the Black–Scholes–Merton model. Their research used the same six input parameters for the model as input layer features—spot price, strike, time to maturity, dividend yield, risk-free interest rate, and volatility—with four hidden layers of 100 neurons each and one output layer. The predicted option prices out-of-sample were very close to the actual option prices: A regression of predicted option prices on actual prices had an R^2 of 99.8%.

14.1 Reinforcement Learning

Reinforcement learning (RL) made headlines in 2017 when DeepMind's AlphaGo program beat the reigning world champion at the ancient game of Go. The RL framework involves an agent that is designed to perform actions that will maximize its rewards over time, taking into consideration the constraints of its environment. In the case of AlphaGo, a virtual gamer (the agent) uses his or her console commands (the actions) with the information on the screen (the environment) to maximize his or her score (the reward). Unlike supervised learning, reinforcement learning has neither direct labeled data for each observation nor instantaneous feedback. With RL, the algorithm needs to observe its environment, learn by testing new actions (some of which may not be immediately optimal), and reuse its previous experiences. The learning subsequently occurs through millions of trials and errors. Academics and practitioners are applying RL in a similar way in investment strategies where the agent could be a virtual trader who follows certain trading rules (the actions) in a specific market (the environment) to maximize its profits (its reward). The success of RL in dealing with the complexities of financial markets is still an open question.

EXAMPLE 8

Deep Neural Networks

Glen Mitsui is the chief investment officer for a large Australian state's Public Employees' Pension Fund (PEPF), which currently has assets under management (AUM) of A$20 billion. The fund manages one-quarter of its assets internally, with A$5 billion mostly in domestic government and corporate fixed-income instruments and domestic equities. The remaining three-quarters of AUM, or A$15 billion, is managed by nearly 100 mostly active external asset managers and is invested in a wide range of asset classes, including foreign fixed income and equities, domestic and foreign hedge funds, REITs, commodities, and derivatives.

PEPF has a small staff of four investment professionals tasked with selecting and monitoring these external managers to whom it pays more than A$400 million in fees annually. Performance (compared to appropriate benchmarks) of many of PEPF's external managers has been lagging over the past several years.

After studying the situation, Mitsui concludes that style drift may be an important factor in explaining such underperformance, for which PEPF is not happy to pay. Mitsui believes that machine learning may help and consults with Frank Monroe, professor of data analysis at Epsilon University.

Monroe suggests using a deep neural network model that collects and analyzes the real-time trading data of PEPF's external managers and compares them to well-known investment styles (e.g., high dividend, minimum volatility, momentum, growth, value) to detect potential style drift. Mitsui arranges for Monroe to meet with PEPF's investment committee (IC) to discuss the matter. As a junior data analyst working with Monroe, you must help him satisfy the following requests from the IC:

1 Define a deep neural network.
2 Evaluate Monroe's opinion on the applicability of deep neural networks to Mitsui's problem.
3 Describe the functions of the three groups of layers of a deep neural network.

Solution to 1:

A deep neural network is a neural network (NN) with many hidden layers (at least 2 but often more than 20). NNs and DNNs have been successfully applied to a wide variety of complex tasks characterized by non-linearities and interactions among features, particularly pattern recognition problems.

Solution to 2:

Mitsui wants to detect patterns of potential style drift in the daily trading of nearly 100 external asset managers in many markets. This task will involve the processing of huge amounts of complicated data. Monroe is correct that a DNN is well suited to PEPF's needs.

Solution to 3:

The input layer, the hidden layers, and the output layer constitute the three groups of layers of DNNs. The input layer receives the inputs (i.e., features) and has as many nodes as there are dimensions of the feature set. The hidden layers consist of nodes, each comprising a summation operator and an activation function that are connected by links. These hidden layers are, in effect, where the model is learned. The final layer, the output layer, produces a set of probabilities of an observation being in any of the target style categories (each represented by a node in the output layer). For example, if there are three target style categories, then three nodes in the output layer are activated to produce outputs that sum to one. So, output (Style Category I, 0.7; Style Category II, 0.2; Style Category III, 0.1) would indicate that the model assigns the greatest probability to an observation being in Style Category I and the least probability to Style Category III. The DNN assigns the observation to the style category with the highest probability.

CASE STUDY: DEEP NEURAL NETWORK–BASED EQUITY FACTOR MODEL

The following case study was developed and written by Matthew Dixon, PhD, FRM.

An investment manager wants to select stocks based on their predicted performance using a fundamental equity factor model. She seeks to capture superior performance from stocks with the largest excess return using a non-linear factor model and so chooses a deep neural network to predict the stock returns. The goal of this mini-case study is to demonstrate the application of deep neural networks to fundamental equity factor modeling. We shall focus on using feed-forward (i.e., forward propagation) network regression in place of ordinary least squares linear regression. Since neural networks are prone to over-fitting, we shall use LASSO penalization, the same penalty score–based approach used previously with regression, to mitigate this issue.

Introduction

Cross-sectional fundamental factor models are used extensively by investment managers to capture the effects of company-specific factors on individual securities. A fixed universe of N assets is first chosen, together with a set of K fundamental factors. Each asset's sensitivity (i.e., exposure or loading) to a fundamental factor is represented by beta, B, and the factors are represented by factor returns (f_t). There are two standard approaches to estimating a factor model: (i) adopt time-series regression (TSR) to recover loadings if factors are known or (ii) use cross-sectional regression (CSR) to recover factor returns from known loadings. We shall follow the CSR approach; the factor exposures are used to predict a stock's return (r_t) by estimating the factor returns using multivariate linear regression (where ε_t is the model error at time t):

$$r_t = Bf_t + \varepsilon_t.$$

However, this CSR model is too simplistic to capture non-linear relationships between stock returns and fundamental factors. So, instead we use a deep neural network to learn the non-linear relationships between the betas (B) and asset returns (r_t) at each time t. The goal of deep learning is to find the network weights which minimize the out-of-sample mean squared error (MSE) between the predicted stock returns, $\hat{r}$, and the observed stock returns, r. We shall see that simply increasing the number of neurons in the network will increase predictive performance using the in-sample data but to the detriment of out-of-sample performance; this phenomenon is the bias–variance trade-off. To mitigate this effect, we add a LASSO penalty term to the loss function to automatically shrink the number of non-zero weights in the network. In doing so, we shall see that this leads to better out-of-sample predictive performance.

Note that each weight corresponds to a link between a node in the previous and current layer. Reducing the number of weights generally means that the number of connections—not the number of nodes—is reduced. The exception is when all weights from the neurons in the previous layer are set to zero—in which case the number of nodes in the current layer would be reduced. In the special case when the previous layer is the input layer, the number of features is also reduced.

We shall illustrate the data preparation and the neural network fitting using six fundamental equity factors. This choice of number and type of fundamental factor is arbitrary, and an investment manager may use many more factors in her or his model, often representing industry sectors and sub-sectors using dummy variables.

Data Description

A description of the stock price and fundamental equity factor data used for training and evaluating the neural network is shown in Exhibit 31.

Exhibit 31 Dataset of S&P 500 Stocks and Fundamental Factors

Description:

A subset of S&P 500 Index stocks, historical monthly adjusted closing prices, and corresponding monthly fundamental factor loadings.

Time period: June 2010 to November 2018

Number of periods: 101

Number of stocks (N): 218 stocks

Number of features (K): 6

Features: Fundamental equity factors:

1. Current enterprise value (i.e., market values of equity + preferred stock + debt − cash − short-term investments)
2. Current enterprise value to trailing 12-month EBITDA
3. Price-to-sales ratio
4. Price-to-earnings ratio
5. Price-to-book ratio
6. Log of stock's market capitalization (i.e., share price × number of shares outstanding)

Output: Monthly return for each stock over the following month.

We define the universe as the top 250 stocks from the S&P 500, ranked by market capitalization as of June 2010. All stock prices and factor loadings are sourced from Bloomberg. An illustrative extract of the data is given in Exhibit 32. Note that after removing stocks with missing factor loadings, we are left with 218 stocks.

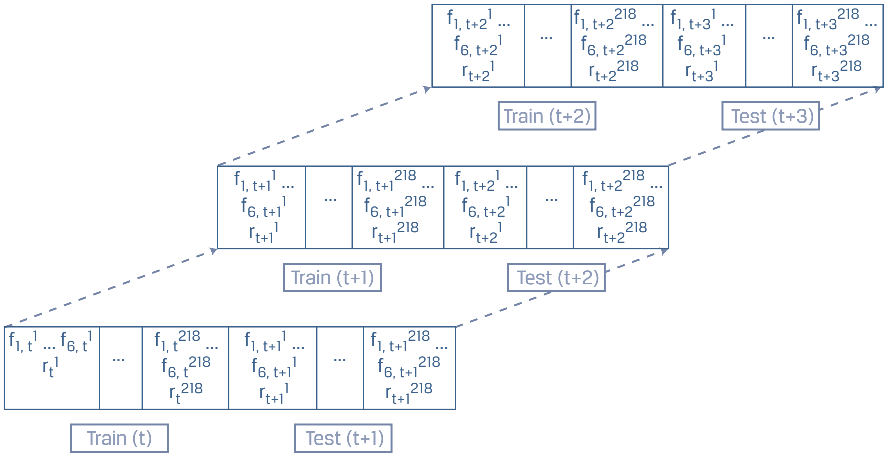
Exhibit 32 Extract of Six Factor Loadings and Return for Three Selected Stocks

TICKER	CURR_EV ($ Mil.)	CURR_EV_ TO_T12M_ EBITDA (X)	PX_ TO_ SALES (X)	PX_ TO_ EARN (X)	PX_ TO_ BOOK (X)	LOG_ CAP ($ Mil.)	RETURN (%)
SWK	10,775.676	30.328	1.138	16.985	1.346	9.082970	−0.132996
STZ	7,433.553	15.653	1.052	10.324	1.480	8.142253	−0.133333
SRE	19,587.124	10.497	1.286	10.597	1.223	9.314892	−0.109589

Experimental Design

The method used to train the deep neural network is time-series cross-validation (i.e., walk-forward optimization), as depicted in Exhibit 33. At each time period, the investment manager fits a new model; each factor (f_1 to f_6) is a feature in the network, and the loadings of the factors for each stock is a feature vector observation (i.e., the set of observations for each stock for each period), leading to N = 218 observations of pairs of feature vectors and output (monthly return, r_t) in the training set per period. The network is initially trained at period t, and then it is tested over the next period, t +1, which also has N = 218 observations of pairs of feature vectors and output. In the next iteration, the t + 1 data become the new training set and the revised model is tested on the $t + 2$ data. The walk-forward optimization of the neural network continues until the last iteration: model training with $t + 99$ data (from Period 100) and testing with $t + 100$ data (from the last period, 101).

Exhibit 33 Time-Series Cross-Validation on Asset Returns (Walk-Forward Optimization)—The First Three Iterations



We use a feed-forward neural network with six input nodes (i.e., neurons), two hidden layers, and one output neuron. There are 50 neurons in each hidden layer to intentionally over-specify the number of parameters needed in the model, meaning bias (variance) is substantially lower (higher) than optimal. LASSO penalization is then used to automatically shrink the parameter set. Additionally, it is important for the number of nodes in each hidden layer not to exceed the number of observations in the training set (50 nodes per layer versus 218 observations). The model training in period t involves finding the optimal bias-versus-variance trade-off. Once fitted, we record the in-sample MSE and the out-of-sample MSE in addition to the optimal regularization parameter. This procedure is then repeated sequentially over the horizon of 100 remaining periods, tuning the hyperparameters at each stage using cross-validation. The end result of this procedure is a fitted model, trained monthly on the current cross-sectional data and for which hyperparameters have been tuned at each step.

Results

Exhibit 34 presents the results from model evaluation; it compares the in-sample and out-of-sample MSEs of the deep neural network over all 101 months. Note that the out-of-sample error (dotted line) is typically significantly larger than the in-sample error (solid line). However, as the time periods pass and the model is repeatedly trained and tested, the difference between the out-of-sample and in-sample MSEs narrows dramatically.

Exhibit 34 In-Sample and Out-of-Sample MSE for Each Training and Testing Period

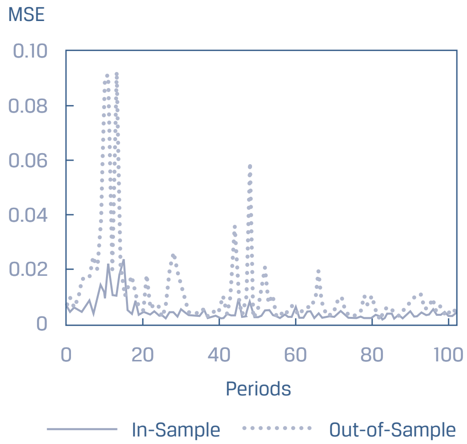
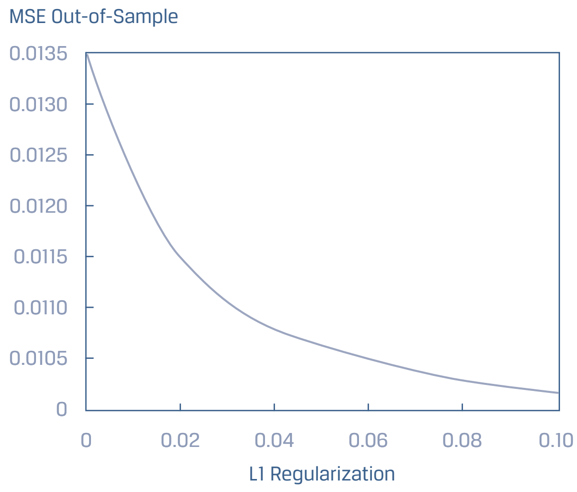


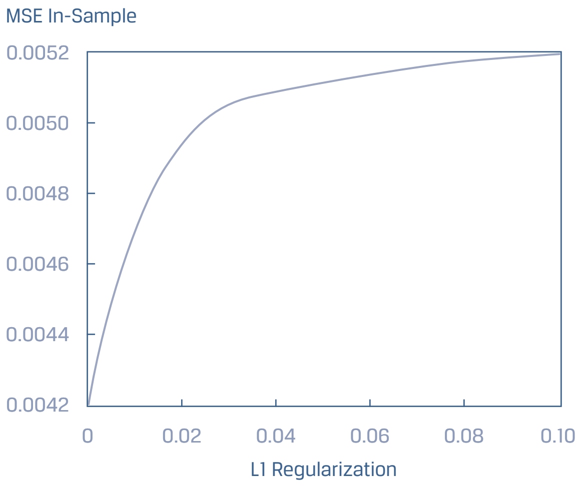
Exhibit 35 shows the effect of LASSO regularization on the in-sample MSE (lower panel, B) and the out-of-sample MSE (upper panel, A) for the first iteration of the time-series cross-validation (training with data from period t and testing with data from period t +1). The degree of LASSO regularization needed is found by cross-validation using 50 neurons in each hidden layer. Increasing the LASSO regularization, which reduces the number of non-zero weights in the model, introduces more bias and hence increases the in-sample error. Conversely, increasing the LASSO regularization reduces the model's variance and thereby reduces the out-of-sample error. Overall, the amount of LASSO regularization needed is significant, at 0.10; typically the regularization hyperparameter is between 0.001 and 1.0. Also, the out-of-sample and in-sample MSEs have not yet converged. There is still a substantial gap, of roughly 0.0051 (= 0.01025 – 0.0052), and the slope of the curves in each plot suggests the optimal value of the regularization hyperparameter is significantly more than 0.10. Note that the value of the regularization hyperparameter is not interpretable and does not correspond to the number of weights eliminated. Suffice it to say, the larger the value of the regularization hyperparameter, the more the loss is being penalized.

Exhibit 35 LASSO Regularization for Optimizing Bias–Variance Trade-Off (First Iteration)

A.



B.

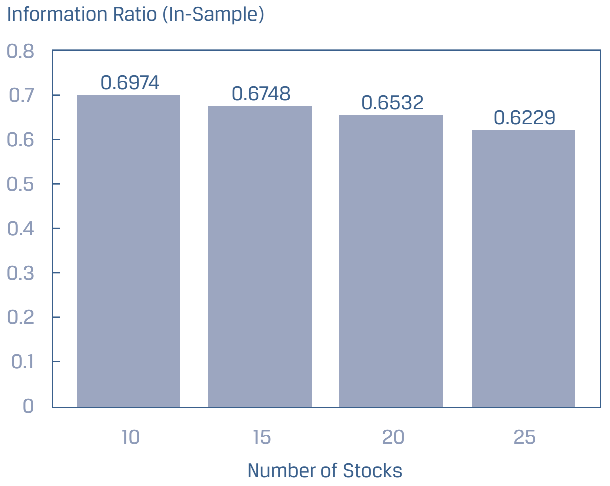


It is important to recognize that although the out-of-sample MSE of this deep learning neural network is key to characterizing its predictive performance, it does not necessarily follow that a stock selection strategy based on the neural network will be successful. This is because the neural network predicts the next month's expected (i.e., mean) asset returns and not the full distribution of returns. Hence a simple stock selection strategy—measured by information ratios (recall the information ratio, or IR, is alpha divided by nonsystematic risk, so it measures the abnormal return per unit of risk for a well-diversified portfolio) of the portfolio returns—that selects stocks ranked by predicted returns will not necessarily lead to positive information ratios.

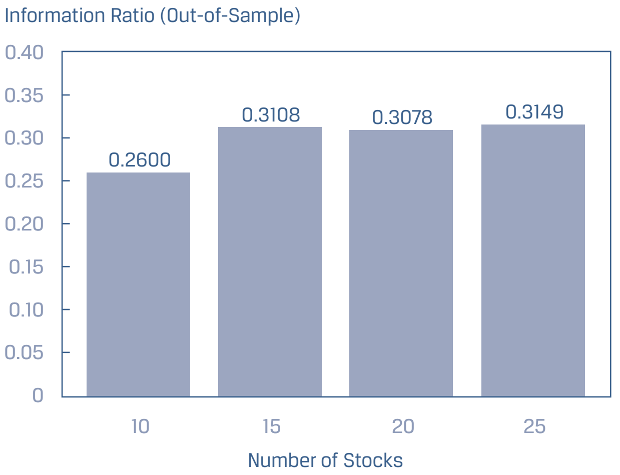
Exhibit 36 presents the information ratios found by back-testing a simple stock selection strategy that picks the top performing stocks determined by the neural network's forecasted returns realized in month t +1 using features observed in month t. Note these IRs do not account for transaction costs, interest rates, or any other fees. The upper panel (A) shows the best-case scenario; the neural network in-sample prediction is used to select the n (where n is 10, 15, 20, or 25) top performing stocks. The IRs are shown for each of the different-sized portfolios; they range from 0.697 to 0.623. Note that as a rule of thumb, IRs in the range of 0.40–0.60 are considered quite good. The lower panel (B) shows the IRs from back-test results for the same strategy applied to the out-of-sample data. The out-of-sample IRs range from 0.260 to 0.315 and so are substantially smaller than in-sample IRs.

Exhibit 36 Information Ratios from Back-Testing a Stock Selection Strategy Using Top Performers from the Neural Network

A.



B.



Importantly, the out-of-sample performance provides the most realistic assessment of the likely future investment performance from applying this deep learning neural network to stock selection. It is a baseline for further model refinements, including adding more fundamental and macroeconomic factors. With such refinements, it can be expected that the out-of-sample IRs should improve substantially.

EXAMPLE 9

Deep Learning–Based Fundamental Factor Model

A research analyst, Jane Hinton, has been tasked with further developing the deep learning–based fundamental factor model. She decides to refine the model by adding four more fundamental factors (such as debt leverage and R&D intensity) given by firm characteristics and by including dummy variables for 11 industrial sectors. Moreover, she additionally expands the universe of stocks to 420 from 218 by using a supplementary data source.

1 Describe how Jane would modify the inputs of the neural network architecture for this new dataset.
2 Describe the size of the new training and test datasets.
3 Describe any additional changes to the architecture and hyperparameters of the neural network that Jane would likely need to make to ensure good performance of the network.
4 Explain how Jane should evaluate whether the new model leads to improved portfolio performance.

Solution to 1:

Jane adds four more fundamental factors and 11 dummy variables, to represent each industrial sector, for a total of 21 (= 4 + 11 + 6) features. Therefore, the refined neural network will have 21 input neurons. The output layer will remain the same. Note that concerns of collinearity of the features through the dummy variables or high correlation, which are problematic for linear regression, are not an issue for a deep learning–based model.

Solution to 2:

There are now data on 420 stocks, for each of the 101 time periods, consisting of factor loadings for the 21 features and the monthly return for each stock. Per the time-series cross-validation method, the test dataset in the current iteration will become the training dataset in the next iteration.

Solution to 3:

Jane should find the new optimal LASSO regularization hyperparameter using time-series cross-validation. Alternatively, she may find the optimal bias–variance trade-off by first increasing the number of neurons in the hidden layers and then performing the cross-validation.

Solution to 4:

Once Jane has found the optimal LASSO hyperparameter and network architecture, she will use the model to forecast the out-of-sample monthly asset returns (i.e., the model forecasts from factor loadings which are not in the training set). She will then rank and select the top predicted performers and finally measure the realized monthly portfolio return. She will then repeat the experiment by moving forward one month in the dataset and repeating the out-of-sample forecast of the asset returns, until she has generated forecasts for all time periods. Finally, Jane will calculate the information ratios from the mean and standard deviation of the monthly portfolio excess returns.

EXAMPLE 10

Summing Up the Major Types of Machine Learning

1 As used in supervised machine learning, classification problems involve the following *except*:

A binary target variables.

B continuous target variables.

C categorical target variables.

2 Which of the following *best* describes penalized regression? Penalized regression:

A is unrelated to multiple linear regression.

B involves a penalty term that is added to the predicted target variable.

C is a category of general linear models used when the number of features and overfitting are concerns.

3 CART is *best* described as:

A an unsupervised ML algorithm.

B a clustering algorithm based on decision trees.

C a supervised ML algorithm that accounts for non-linear relationships among the features.

4 A neural network is *best* described as a technique for machine learning that is:

A exactly modeled on the human nervous system.

B based on layers of nodes connected by links when the relationships among the features are usually non-linear.

C based on a tree structure of nodes when the relationships among the features are linear.

5 Hierarchical clustering is *best* described as a technique in which:

A the grouping of observations is unsupervised.

B features are grouped into a pre-specified number, *k*, of clusters.

C observations are classified according to predetermined labels.

6 Dimension reduction techniques are *best* described as a means to reduce a set of features to a manageable size:

A without regard for the variation in the data.

B while increasing the variation in the data.

C while retaining as much of the variation in the data as possible.

Solution to 1:

B is correct. A and C are incorrect because when the target variable is binary or categorical (not continuous), the problem is a classification problem.

Solution to 2:

C is correct. A is incorrect because penalized regression is related to multiple linear regression. B is incorrect because penalized regression involves adding a penalty term to the sum of the squared regression residuals.

Solution to 3:

C is correct. A is incorrect because CART is a supervised ML algorithm. B is incorrect because CART is a classification and regression algorithm, not a clustering algorithm.

Solution to 4:

B is correct. A is incorrect because neural networks are not exactly modeled on the human nervous system. C is incorrect because neural networks are not based on a tree structure of nodes when the relationships among the features are linear.

Solution to 5:

A is correct. B is incorrect because it refers to *k*-means clustering. C is incorrect because it refers to classification, which involves supervised learning.

Solution to 6:

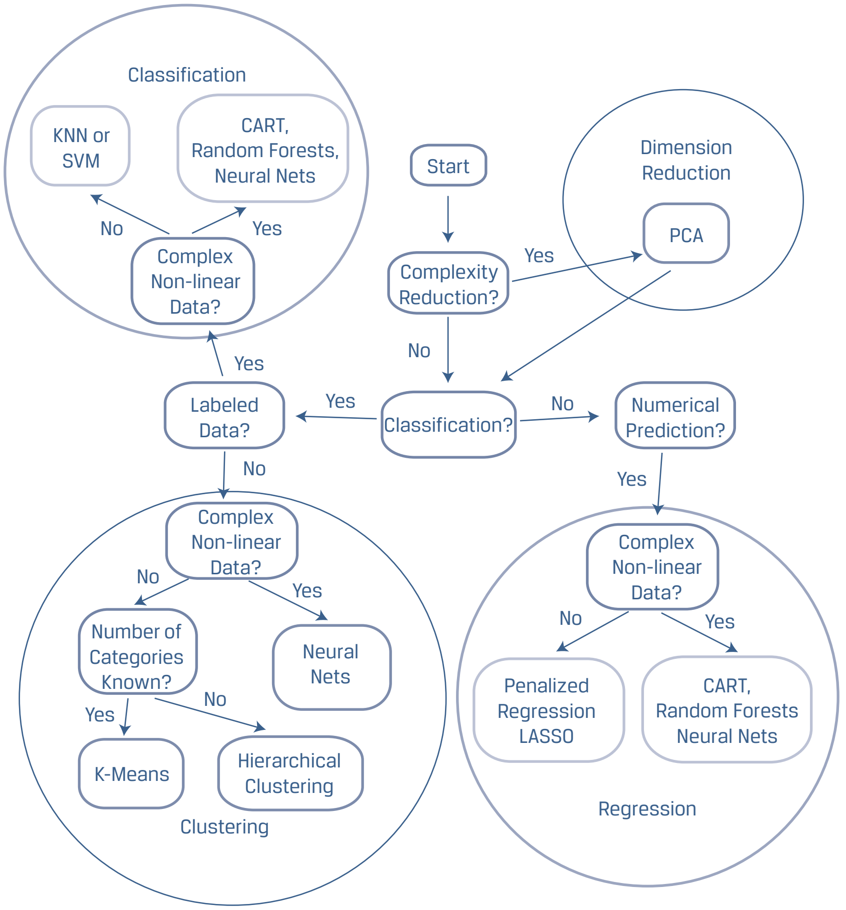
C is correct because dimension reduction techniques, such as PCA, are aimed at reducing the feature set to a manageable size while retaining as much of the variation in the data as possible.

15 CHOOSING AN APPROPRIATE ML ALGORITHM

e describe neural networks, deep learning nets, and reinforcement learning

Exhibit 37 presents a simplified decision flowchart for choosing among the machine learning algorithms which have been discussed. The dark-shaded ovals contain the supervised ML algorithms, the light-shaded ovals contain the unsupervised ML algorithms, and the key questions to consider are shown in the unshaded rounded rectangles.

Exhibit 37 Stylized Decision Flowchart for Choosing ML Algorithms



First, start by asking, Are the data complex, having many features that are highly correlated? If yes, then dimension reduction using principal components analysis is appropriate.

Next, is the problem one of classification or numerical prediction? If numerical prediction, then depending on whether the data have non-linear characteristics, the choice of ML algorithms is from a set of regression algorithms—either penalized regression/LASSO for linear data or CART, random forest, or neural networks for non-linear data.

If the problem is one of classification, then depending on whether the data are labeled, the choice is either from a set of classification algorithms using labeled data or from a set of clustering algorithms using unlabeled data.

If the data are labeled, then depending on whether the data have non-linear characteristics, the choice of classification algorithm would be *K*-nearest neighbor and support vector machine for linear data or CART, random forest, or neural networks (or deep neural networks) for non-linear data.

Finally, if the data are unlabeled, the choice of clustering algorithm depends on whether the data have non-linear characteristics. The choice of clustering algorithm would be neural networks (or deep neural networks) for non-linear data or for linear data, *K*-means with a known number of categories and hierarchical clustering with an unknown number of categories.

SUMMARY

Machine learning methods are gaining usage at many stages in the investment management value chain. Among the major points made are the following:

- Machine learning aims at extracting knowledge from large amounts of data by learning from known examples to determine an underlying structure in the data. The emphasis is on generating structure or predictions without human intervention. An elementary way to think of ML algorithms is to "find the pattern, apply the pattern."
- Supervised learning depends on having labeled training data as well as matched sets of observed inputs (*X*'s, or features) and the associated output (*Y*, or target). Supervised learning can be divided into two categories: regression and classification. If the target variable to be predicted is continuous, then the task is one of regression. If the target variable is categorical or ordinal (e.g., determining a firm's rating), then it is a classification problem.
- With unsupervised learning, algorithms are trained with no labeled data, so they must infer relations between features, summarize them, or present underlying structure in their distributions that has not been explicitly provided. Two important types of problems well suited to unsupervised ML are dimension reduction and clustering.
- In deep learning, sophisticated algorithms address complex tasks (e.g., image classification, natural language processing). Deep learning is based on neural networks, highly flexible ML algorithms for solving a variety of supervised and unsupervised tasks characterized by large datasets, non-linearities, and interactions among features. In reinforcement learning, a computer learns from interacting with itself or data generated by the same algorithm.

- Generalization describes the degree to which an ML model retains its explanatory power when predicting out-of-sample. Overfitting, a primary reason for lack of generalization, is the tendency of ML algorithms to tailor models to the training data at the expense of generalization to new data points.
- Bias error is the degree to which a model fits the training data. Variance error describes how much a model's results change in response to new data from validation and test samples. Base error is due to randomness in the data. Out-of-sample error equals bias error plus variance error plus base error.
- *K*-fold cross-validation is a technique for mitigating the holdout sample problem (excessive reduction of the training set size). The data (excluding test sample and fresh data) are shuffled randomly and then divided into k equal sub-samples, with $k - 1$ samples used as training samples and one sample, the kth, used as a validation sample.
- Regularization describes methods that reduce statistical variability in high-dimensional data estimation or prediction problems via reducing model complexity.
- LASSO (least absolute shrinkage and selection operator) is a popular type of penalized regression where the penalty term involves summing the absolute values of the regression coefficients. The greater the number of included features, the larger the penalty. So, a feature must make a sufficient contribution to model fit to offset the penalty from including it.
- Support vector machine (SVM) is a classifier that aims to seek the optimal hyperplane—the one that separates the two sets of data points by the maximum margin (and thus is typically used for classification).
- *K*-nearest neighbor (KNN) is a supervised learning technique most often used for classification. The idea is to classify a new observation by finding similarities ("nearness") between it and its k-nearest neighbors in the existing dataset.
- Classification and regression tree (CART) can be applied to predict either a categorical target variable, producing a classification tree, or a continuous target variable, producing a regression tree.
- A binary CART is a combination of an initial root node, decision nodes, and terminal nodes. The root node and each decision node represent a single feature (f) and a cutoff value (c) for that feature. The CART algorithm iteratively partitions the data into sub-groups until terminal nodes are formed that contain the predicted label.
- Ensemble learning is a technique of combining the predictions from a collection of models. It typically produces more accurate and more stable predictions than any single model.
- A random forest classifier is a collection of many different decision trees generated by a bagging method or by randomly reducing the number of features available during training.
- Principal components analysis (PCA) is an unsupervised ML algorithm that reduces highly correlated features into fewer uncorrelated composite variables by transforming the feature covariance matrix. PCA produces eigenvectors that define the principal components (i.e., the new uncorrelated composite variables) and eigenvalues, which give the proportion of total variance in the initial data that is explained by each eigenvector and its associated principal component.

- *K*-means is an unsupervised ML algorithm that partitions observations into a fixed number (*k*) of non-overlapping clusters. Each cluster is characterized by its centroid, and each observation belongs to the cluster with the centroid to which that observation is closest.
- Hierarchical clustering is an unsupervised iterative algorithm that is used to build a hierarchy of clusters. Two main strategies are used to define the intermediary clusters (i.e., those clusters between the initial dataset and the final set of clustered data).
- Agglomerative (bottom-up) hierarchical clustering begins with each observation being its own cluster. Then, the algorithm finds the two closest clusters, defined by some measure of distance, and combines them into a new, larger cluster. This process is repeated until all observations are clumped into a single cluster.
- Divisive (top-down) hierarchical clustering starts with all observations belonging to a single cluster. The observations are then divided into two clusters based on some measure of distance. The algorithm then progressively partitions the intermediate clusters into smaller clusters until each cluster contains only one observation.
- Neural networks consist of nodes connected by links. They have three types of layers: an input layer, hidden layers, and an output layer. Learning takes place in the hidden layer nodes, each of which consists of a summation operator and an activation function. Neural networks have been successfully applied to a variety of investment tasks characterized by non-linearities and complex interactions among variables.
- Neural networks with many hidden layers (at least 2 but often more than 20) are known as deep neural networks (DNNs) and are the backbone of the artificial intelligence revolution.
- Reinforcement learning (RL) involves an agent that should perform actions that will maximize its rewards over time, taking into consideration the constraints of its environment.

REFERENCES

Bacham, Dinesh, and Janet Zhao. 2017. “Machine Learning: Challenges, Lessons, and Opportunities in Credit Risk Modeling.” *Moody’s Analytics Risk Perspectives* 9:28–35.

Culkin, Robert, and Sanjiv R. Das. 2017. “Machine Learning in Finance: The Case of Deep Learning for Option Pricing.” *Journal of Investment Management* 15 (4): 92–100.

PRACTICE PROBLEMS

The following information relates to Questions 1–10

Alef Associates manages a long-only fund specializing in global smallcap equities. Since its founding a decade ago, Alef maintains a portfolio of 100 stocks (out of an eligible universe of about 10,000 stocks). Some of these holdings are the result of screening the universe for attractive stocks based on several ratios that use readily available market and accounting data; others are the result of investment ideas generated by Alef's professional staff of five securities analysts and two portfolio managers.

Although Alef's investment performance has been good, its Chief Investment Officer, Paul Moresanu, is contemplating a change in the investment process aimed at achieving even better returns. After attending multiple workshops and being approached by data vendors, Moresanu feels that data science should play a role in the way Alef selects its investments. He has also noticed that much of Alef's past outperformance is due to stocks that became takeover targets. After some research and reflection, Moresanu writes the following email to the Alef's CEO.

Subject: Investment Process Reorganization

I have been thinking about modernizing the way we select stock investments. Given that our past success has put Alef Associates in an excellent financial position, now seems to be a good time to invest in our future. What I propose is that we continue managing a portfolio of 100 global small-cap stocks but restructure our process to benefit from machine learning (ML). Importantly, the new process will still allow a role for human insight, for example, in providing domain knowledge. In addition, I think we should make a special effort to identify companies that are likely to be acquired. Specifically, I suggest following the four steps which would be repeated every quarter.

Step 1 We apply ML techniques to a model including fundamental and technical variables (features) to predict next quarter's return for each of the 100 stocks currently in our portfolio. Then, the 20 stocks with the lowest estimated return are identified for replacement.

Step 2 We utilize ML techniques to divide our investable universe of about 10,000 stocks into 20 different groups, based on a wide variety of the most relevant financial and non-financial characteristics. The idea is to prevent unintended portfolio concentration by selecting stocks from each of these distinct groups.

Step 3 For each of the 20 different groups, we use labeled data to train a model that will predict the five stocks (in any given group) that are most likely to become acquisition targets in the next one year.

© 2020 CFA Institute. All rights reserved.

(Continued)

Step 4 Our five experienced securities analysts are each assigned four of the groups, and then each analyst selects their one best stock pick from each of their assigned groups. These 20 "high-conviction" stocks will be added to our portfolio (in replacement of the 20 relatively underperforming stocks to be sold in Step 1).

A couple of additional comments related to the above:

Comment 1 The ML algorithms will require large amounts of data. We would first need to explore using free or inexpensive historical datasets and then evaluate their usefulness for the ML-based stock selection processes before deciding on using data that requires subscription.

Comment 2 As time passes, we expect to find additional ways to apply ML techniques to refine Alef's investment processes.

What do you think?
Paul Moresanu

---

1 The machine learning techniques appropriate for executing Step 1 are *most* likely to be based on:

A regression

B classification

C clustering

2 Assuming regularization is utilized in the machine learning technique used for executing Step 1, which of the following ML models would be *least* appropriate:

A Regression tree with pruning.

B LASSO with lambda (λ) equal to 0.

C LASSO with lambda (λ) between 0.5 and 1.

3 Which of the following machine learning techniques is *most* appropriate for executing Step 2:

A K-Means Clustering

B Principal Components Analysis (PCA)

C Classification and Regression Trees (CART)

4 The hyperparameter in the ML model to be used for accomplishing Step 2 is?

A 100, the number of small-cap stocks in Alef's portfolio.

B 10,000, the eligible universe of small-cap stocks in which Alef can potentially invest.

C 20, the number of different groups (i.e. clusters) into which the eligible universe of small-cap stocks will be divided.

5 The target variable for the labelled training data to be used in Step 3 is *most* likely which one of the following?

A A continuous target variable.

B A categorical target variable.

C An ordinal target variable.

6 Comparing two ML models that could be used to accomplish Step 3, which statement(s) *best* describe(s) the advantages of using Classification and Regression Trees (CART) instead of K-Nearest Neighbor (KNN)?

Statement I For CART there is no requirement to specify an initial hyperparameter (like K).

Statement II For CART there is no requirement to specify a similarity (or distance) measure.

Statement III For CART the output provides a visual explanation for the prediction.

A Statement I only.

B Statement III only.

C Statements I, II and III.

7 Assuming a Classification and Regression Tree (CART) model is used to accomplish Step 3, which of the following is *most* likely to result in model overfitting?

A Using the k-fold cross validation method

B Including an overfitting penalty (i.e., regularization term).

C Using a fitting curve to select a model with low bias error and high variance error.

8 Assuming a Classification and Regression Tree (CART) model is initially used to accomplish Step 3, as a further step which of the following techniques is most likely to result in more accurate predictions?

A Discarding CART and using the predictions of a Support Vector Machine (SVM) model instead.

B Discarding CART and using the predictions of a K-Nearest Neighbor (KNN) model instead.

C Combining the predictions of the CART model with the predictions of other models – such as logistic regression, SVM, and KNN – via ensemble learning.

9 Regarding Comment #2, Moresanu has been thinking about the applications of neural networks (NNs) and deep learning (DL) to investment management. Which statement(s) *best* describe(s) the tasks for which NNs and DL are well-suited?

Statement I NNs and DL are well-suited for image and speech recognition, and natural language processing.

Statement II NNs and DL are well-suited for developing single variable ordinary least squares regression models.

Statement III NNs and DL are well-suited for modelling non-linearities and complex interactions among many features.

A Statement II only.

B Statements I and III.

C Statements I, II and III.

10 Regarding neural networks (NNs) that Alef might potentially implement, which of the following statements is *least* accurate?

A NNs must have at least 10 hidden layers to be considered deep learning nets.

B The activation function in a node operates like a light dimmer switch since it decreases or increases the strength of the total net input.

C The summation operator receives input values, multiplies each by a weight, sums up the weighted values into the total net input, and passes it to the activation function.

SOLUTIONS

1 A is correct. The target variable (quarterly return) is continuous, hence this calls for a supervised machine learning based regression model.

B is incorrect, since classification uses categorical or ordinal target variables, while in Step 1 the target variable (quarterly return) is continuous.

C is incorrect, since clustering involves unsupervised machine learning so does not have a target variable.

2 B is correct. It is least appropriate because with LASSO, when $\lambda = 0$ the penalty (i.e., regularization) term reduces to zero, so there is no regularization and the regression is equivalent to an ordinary least squares (OLS) regression.

A is incorrect. With Classification and Regression Trees (CART), one way that regularization can be implemented is via pruning which will reduce the size of the regression tree—sections that provide little explanatory power are pruned (i.e., removed).

C is incorrect. With LASSO, when λ is between 0.5 and 1 the relatively large penalty (i.e., regularization) term requires that a feature makes a sufficient contribution to model fit to offset the penalty from including it in the model.

3 A is correct. K-Means clustering is an unsupervised machine learning algorithm which repeatedly partitions observations into a fixed number, *k*, of non-overlapping clusters (i.e., groups).

B is incorrect. Principal Components Analysis is a long-established statistical method for dimension reduction, not clustering. PCA aims to summarize or reduce highly correlated features of data into a few main, uncorrelated composite variables.

C is incorrect. CART is a supervised machine learning technique that is most commonly applied to binary classification or regression.

4 C is correct. Here, 20 is a hyperparameter (in the K-Means algorithm), which is a parameter whose value must be set by the researcher before learning begins.

A is incorrect, because it is not a hyperparameter. It is just the size (number of stocks) of Alef's portfolio.

B is incorrect, because it is not a hyperparameter. It is just the size (number of stocks) of Alef's eligible universe.

5 B is correct. To predict which stocks are likely to become acquisition targets, the ML model would need to be trained on categorical labelled data having the following two categories: "0" for "not acquisition target", and "1" for "acquisition target".

A is incorrect, because the target variable is categorical, not continuous.

C is incorrect, because the target variable is categorical, not ordinal (i.e., 1st, 2nd, 3rd, etc.).

6 C is correct. The advantages of using CART over KNN to classify companies into two categories ("not acquisition target" and "acquisition target"), include all of the following: For CART there are no requirements to specify an initial hyperparameter (like K) or a similarity (or distance) measure as with KNN, and CART provides a visual explanation for the prediction (i.e., the feature variables and their cut-off values at each node).

A is incorrect, because CART provides all of the advantages indicated in Statements I, II and III.

B is incorrect, because CART provides all of the advantages indicated in Statements I, II and III.

7 C is correct. A fitting curve shows the trade-off between bias error and variance error for various potential models. A model with low bias error and high variance error is, by definition, overfitted.

A is incorrect, because there are two common methods to reduce overfitting, one of which is proper data sampling and cross-validation. K-fold cross validation is such a method for estimating out-of-sample error directly by determining the error in validation samples.

B is incorrect, because there are two common methods to reduce overfitting, one of which is preventing the algorithm from getting too complex during selection and training, which requires estimating an overfitting penalty.

8 C is correct. Ensemble learning is the technique of combining the predictions from a collection of models, and it typically produces more accurate and more stable predictions than the best single model.

A is incorrect, because a single model will have a certain error rate and will make noisy predictions. By taking the average result of many predictions from many models (i.e., ensemble learning) one can expect to achieve a reduction in noise as the average result converges towards a more accurate prediction.

B is incorrect, because a single model will have a certain error rate and will make noisy predictions. By taking the average result of many predictions from many models (i.e., ensemble learning) one can expect to achieve a reduction in noise as the average result converges towards a more accurate prediction.

9 B is correct. NNs and DL are well-suited for addressing highly complex machine learning tasks, such as image classification, face recognition, speech recognition and natural language processing. These complicated tasks are characterized by non-linearities and complex interactions between large numbers of feature inputs.

A is incorrect, because NNs and DL are well-suited for addressing highly complex machine learning tasks, not simple single variable OLS regression models.

C is incorrect, because NNs and DL are well-suited for addressing highly complex machine learning tasks, not simple single variable OLS regression models.

10 A is correct. It is the least accurate answer because neural networks with many hidden layers—at least 3, but often more than 20 hidden layers—are known as deep learning nets.

B is incorrect, because the node's activation function operates like a light dimmer switch which decreases or increases the strength of the (total net) input.

C is incorrect, because the node's summation operator multiplies each (input) value by a weight and sums up the weighted values to form the total net input. The total net input is then passed to the activation function.

READING

5

Big Data Projects

by Sreekanth Mallikarjun, PhD, and Ahmed Abbasi, PhD

Sreekanth Mallikarjun, PhD, is at Reorg (USA) and the University of Virginia, McIntire School of Commerce (USA). Ahmed Abbasi, PhD, is at the University of Virginia, McIntire School of Commerce (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** identify and explain steps in a data analysis project;
☐	**b.** describe objectives, steps, and examples of preparing and wrangling data;
☐	**c.** describe objectives, methods, and examples of data exploration;
☐	**d.** describe objectives, steps, and techniques in model training;
☐	**e.** describe preparing, wrangling, and exploring text-based data for financial forecasting;
☐	**f.** describe methods for extracting, selecting and engineering features from textual data;
☐	**g.** evaluate the fit of a machine learning algorithm.

INTRODUCTION AND BIG DATA IN INVESTMENT MANAGEMENT

1

Big data (also referred to as alternative data) encompasses data generated by financial markets (e.g., stock and bond prices), businesses (e.g., company financials, production volumes), governments (e.g., economic and trade data), individuals (e.g., credit card purchases, social media posts), sensors (e.g., satellite imagery, traffic patterns), and the Internet of Things, or IoT, (i.e., the network of interrelated digital devices that can transfer data among themselves without human interaction). A veritable explosion in big data has occurred over the past decade or so, especially in unstructured data generated from social media (e.g., posts, tweets, blogs), email and text communications, web traffic, online news sites, electronic images, and other electronic information sources. The prospects are for exponential growth in big data to continue.

© 2019 CFA Institute. All rights reserved.

Investment managers are increasingly using big data in their investment processes as they strive to discover signals embedded in such data that can provide them with an information edge. They seek to augment structured data with a plethora of unstructured data to develop improved forecasts of trends in asset prices, detect anomalies, etc. A typical example involves a fund manager using financial text data from 10-K reports for forecasting stock sentiment (i.e., positive or negative), which can then be used as an input to a more comprehensive forecasting model that includes corporate financial data.

Unlike structured data (numbers and values) that can be readily organized into data tables to be read and analyzed by computers, unstructured data typically require specific methods of preparation and refinement before being usable by machines (i.e., computers) and useful to investment professionals. Given the volume, variety, and velocity of available big data, it is important for portfolio managers and investment analysts to have a basic understanding of how unstructured data can be transformed into structured data suitable as inputs to machine learning (ML) methods (in fact, for any type of modeling methods) that can potentially improve their financial forecasts.

This reading describes the steps in using big data, both structured and unstructured, in financial forecasting. The concepts and methods are then demonstrated in a case study of an actual big data project. The project uses text-based data derived from financial documents to train an ML model to classify text into positive or negative sentiment classes for the respective stocks and then to predict sentiment.

Section 1 of the reading covers a description of the key characteristics of big data. Section 2 provides an overview of the steps in executing a financial forecasting project using big data. We then describe in Sections 3–9 key aspects of data preparation and wrangling, data exploration, and model training using structured data and unstructured (textual) data. In Section 10–13, we bring these pieces together by covering the execution of an actual big data project.

1.1 Big Data in Investment Management

Big data differs from traditional data sources based on the presence of a set of characteristics commonly referred to as the 3Vs: volume, variety, and velocity.

Volume refers to the quantity of data. The US Library of Congress, which is tasked with archiving both digital and physical information artifacts in the United States, has collected hundreds of terabytes of data (one terabyte equals 1,024 gigabytes, which are equal to 1,048,576 megabytes). Several years ago, one of the authors managed an archival project for the Library of Congress in which many terabytes of online content were collected—a copious amount of data at the time. However, in most US industry sectors today, the average company collects more data than the Library of Congress! In big data conversations, terabytes have been replaced with petabytes and exabytes (one exabyte equals 1,024 petabytes, which are equal to 1,048,576 terabytes). The classic grains of sand analogy puts these volumes into perspective: If a megabyte is a tablespoon of sand, then a petabyte is a 1.6-kilometer-long beach and an exabyte is a beach extending about 1,600 kilometers.

Variety pertains to the array of available data sources. Organizations are now dealing with structured, semi-structured, and unstructured data from within and outside the enterprise. Variety includes traditional transactional data; user-generated text, images, and videos; social media; sensor-based data; web and mobile clickstreams; and spatial-temporal data. Effectively leveraging the variety of available data presents both opportunities and challenges, including such legal and ethical issues as data privacy.

Velocity is the speed at which data are created. Many large organizations collect several petabytes of data every hour. With respect to unstructured data, more than one billion new tweets (i.e., a message of 280 characters or less posted on the social media website Twitter) are generated every three days; five billion search queries occur daily. Such information has important implications for real-time predictive analytics in various financial applications. Analyzing such "data-in-motion" poses challenges since relevant patterns and insights might be moving targets relative to situations of "data-at-rest."

When using big data for inference or prediction, there is a "fourth V": *Veracity relates to the credibility and reliability of different data sources.* Determining the credibility and reliability of data sources is an important part of any empirical investigation. The issue of veracity becomes critically important for big data, however, because of the varied sources of these large datasets. Big data amplifies the age-old challenge of disentangling quality from quantity. Social media, including blogs, forums, and social networking sites, are plagued with spam; by some estimates, as much as 10%–15% of such content is completely fake. Similarly, according to our research, web spam accounts for more than 20% of all content on the worldwide web. Clickstreams from website and mobile traffic are equally susceptible to noise. Furthermore, deriving deep semantic knowledge from text remains challenging in certain instances despite significant advances in natural language processing (NLP).

These Vs have numerous implications for financial technology (commonly referred to as "fintech") pertaining to investment management. Machine learning assessments of creditworthiness, which have traditionally relied on structured financial metrics, are being enhanced by incorporating text derived from financial statements, news articles, and call transcripts. Customers in the financial industry are being segmented based not only on their transactional data but also on their views and preferences expressed on social media (to the degree permissible under applicable privacy agreements). Big data also affords opportunities for enhanced fraud detection and risk management.

STEPS IN EXECUTING A DATA ANALYSIS PROJECT: FINANCIAL FORECASTING WITH BIG DATA 2

a identify and explain steps in a data analysis project;

In the era of big data, firms treat data like they do important assets. However, effective big data analytics are critical to allow appropriate data monetization. Let us take financial forecasting as an application area. Numerous forecasting tasks in this domain can benefit from predictive analytics models built using machine learning methods. One common example is predicting whether stock prices (for an individual stock or a portfolio) will go up or down in value at some specific point in the future. Traditionally, financial forecasting relied on various financial and accounting numbers, ratios, and metrics coupled with statistical or mathematical models. More recently, machine learning models have been commonly utilized. However, with the proliferation of textual big data (e.g., online news articles, internet financial forums, social networking platforms), such unstructured data have been shown to offer insights faster (as they are real-time) and have enhanced predictive power.

Textual big data provides several valuable types of information, including topics and sentiment. Topics are what people are talking about (e.g., a firm, an industry, a particular event). Sentiment is how people feel about what they are discussing. For instance, they might express positive, negative, or neutral views (i.e., sentiments) toward a topic of discussion. One study conducted in the United States found that positive sentiment on Twitter could predict the trend for the Dow Jones Industrial Average up to three days later with nearly 87% accuracy.

Deriving such insights requires supplementing traditional data with textual big data. As depicted in Exhibit 1, the inclusion of big data has immediate implications for building the machine learning model as well as downstream implications for financial forecasting and analysis. We begin with the top half of Exhibit 1, which shows the traditional (i.e., with structured data) *ML Model Building Steps:*

1 *Conceptualization of the modeling task*. This crucial first step entails determining what the output of the model should be (e.g., whether the price of a stock will go up/down one week from now), how this model will be used and by whom, and how it will be embedded in existing or new business processes.

2 *Data collection*. The data traditionally used for financial forecasting tasks are mostly numeric data derived from internal and external sources. Such data are typically already in a structured tabular format, with columns of features, rows of instances, and each cell representing a particular value.

3 *Data preparation and wrangling*. This step involves cleansing and preprocessing of the raw data. Cleansing may entail resolving missing values, out-of-range values, and the like. Preprocessing may involve extracting, aggregating, filtering, and selecting relevant data columns.

4 *Data exploration.* This step encompasses exploratory data analysis, feature selection, and feature engineering.

5 *Model training*. This step involves selecting the appropriate ML method (or methods), evaluating performance of the trained model, and tuning the model accordingly.

Note that these steps are iterative because model building is an iterative process. The insights gained from one iteration may inform the next iteration, beginning with reconceptualization. In contrast with structured data sources, textual big data originating in online news articles, social media, internal/external documents (such as public financial statements), and other openly available data sources are unstructured.

The *Text ML Model Building Steps* used for the unstructured data sources of big data are shown in the bottom half of Exhibit 1. They differ from those used for traditional data sources and are typically intended to create output information that is structured. The differences in steps between the text model and traditional model account for the characteristics of big data: volume, velocity, variety, and veracity. In this reading, we mostly focus on the variety and veracity dimensions of big data as they manifest themselves in text. The major differences in the *Text ML Model Building Steps* are in the first four steps:

1 *Text problem formulation*. Analysts begin by determining how to formulate the text classification problem, identifying the exact inputs and outputs for the model. Perhaps we are interested in computing sentiment scores (structured output) from text (unstructured input). Analysts must also decide how the text ML model's classification output will be utilized.

2 *Data (text) curation*. This step involves gathering relevant external text data via web services or **web spidering (scraping or crawling) programs** that extract raw content from a source, typically web pages. Annotation of the text data with high-quality, reliable target (dependent) variable labels might also be necessary for supervised learning and performance evaluation purposes. For instance, experts might need to label whether a given expert assessment of a stock is bearish or bullish.

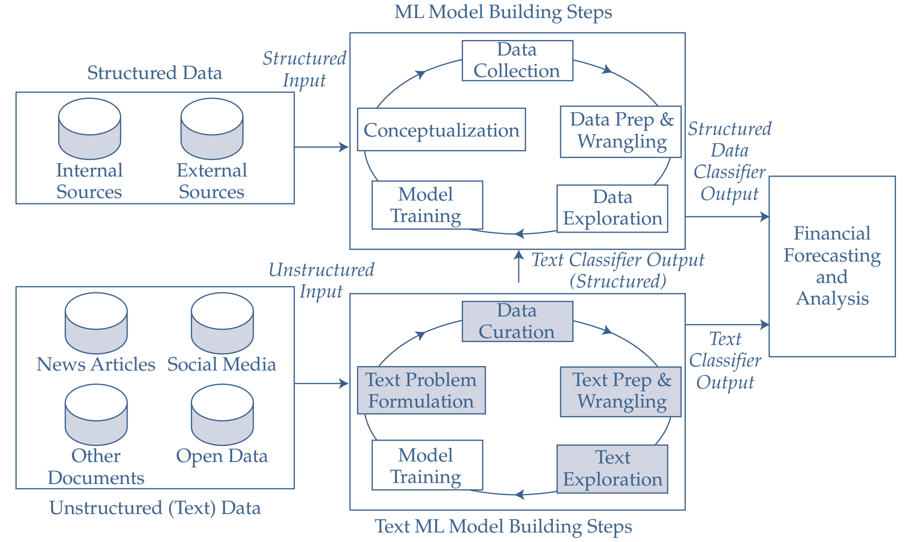
3 *Text preparation and wrangling.* This step involves critical cleansing and preprocessing tasks necessary to convert streams of unstructured data into a format that is usable by traditional modeling methods designed for structured inputs.

4 *Text exploration.* This step encompasses text visualization through techniques, such as word clouds, and text feature selection and engineering.

The resulting output (e.g., sentiment prediction scores) can either be combined with other structured variables or used directly for forecasting and/or analysis.

Next, we describe two key steps from the *ML Model Building Steps* depicted in Exhibit 1 that typically differ for structured data versus textual big data: data/text preparation and wrangling and data/text exploration. We then discuss model training. Finally, we focus on applying these steps to a case study related to classifying and predicting stock sentiment from financial texts.

Exhibit 1 Model Building for Financial Forecasting Using Big Data: Structured (Traditional) vs. Unstructured (Text)



EXAMPLE 1

Steps in ML Model Building

LendALot Corporation is a B2C (business-to-consumer) lender that has traditionally outsourced potential customers' creditworthiness scoring to a third-party firm. Given the recent advances in machine learning (ML)-based "fintech" that goes beyond traditional "repayment history" and "ability to repay" assessments derived from structured data, LendALot would like to develop in-house, ML-based credit scoring capabilities to enhance borrower risk assessment and differentiate itself in the B2C lending market. LendALot would like to follow a phased approach beginning with traditional (structured) data sources and then eventually incorporating textual (unstructured) big data sources. Paul Wang has been asked to lead a new analytics team at LendALot tasked with developing the ML-based creditworthiness scoring model. In the context of machine learning using structured data sources, address the following questions.

1 State and explain one decision Wang will need to make related to:
   A conceptualizing the modeling task.
   B data collection.
   C data preparation and wrangling.
   D data exploration.
   E model training.

In a later phase of the project, LendALot attempts to improve its credit scoring processes by incorporating textual data in credit scoring. Wang tells his team, "Enhance the creditworthiness scoring model by incorporating insights from text provided by the prospective borrowers in the loan application free response fields."

2 Identify the process step that Wang's statement addresses.
3 State two potential needs of the LendAlot team in relation to text curation.
4 State two potential needs of the LendAlot team in relation to text preparation and wrangling.

Solution to 1:

A In the conceptualization step, Wang will need to decide how the output of the ML model will be specified (e.g., a binary classification of creditworthiness), how the model will be used and by whom, and how it will be embedded in LendALot's business processes.
B In the data collection phase, Wang must decide on what data—internal, external, or both—to use for credit scoring.
C In the data preparation and wrangling step, Wang will need to decide on data cleansing and preprocessing needs. Cleansing may entail resolving missing values, extreme values, etc. Preprocessing may involve extracting, aggregating, filtering, and selecting relevant data columns.
D In the data exploration phase, Wang will need to decide which exploratory data analysis methods are appropriate, which features to use in building a credit scoring model, and which features may need to be engineered.
E In the model training step, Wang must decide which ML algorithm(s) to use. Assuming labeled training data are available, the choice will be among supervised learning algorithms. Decisions will need to be made on how model fit is measured and how the model is validated and tuned.

Solution to 2:

Wang's statement relates to the initial step of text problem formulation.

Solution to 3:

Related to text curation, the team will be using internal data (from loan applications). They will need to ensure that the text comment fields on the loan applications have been correctly implemented and enabled. If these fields are not required, they need to ensure there is a sufficient response rate to analyze.

Solution to 4:

Related to text preparation and wrangling, the team will need to carry out the critical tasks of text cleansing and text preprocessing. These two tasks are necessary to convert an unstructured stream of data into structured values for use by traditional modeling methods.

DATA PREPARATION AND WRANGLING; STRUCTURED DATA

3

b describe objectives, steps, and examples of preparing and wrangling data;

Data preparation and wrangling involve cleansing and organizing raw data into a consolidated format. The resulting dataset is suitable to use for further analyses and training a machine learning (ML) model. This is a critical stage, the foundation, in big data projects. Most of the project time is spent on this step, and the quality of the data affects the training of the selected ML model. Domain knowledge—that is, the involvement of specialists in the particular field in which the data are obtained and used—is beneficial and often necessary to successfully execute this step. Data preparation is preceded by data collection, so we discuss the data collection process first.

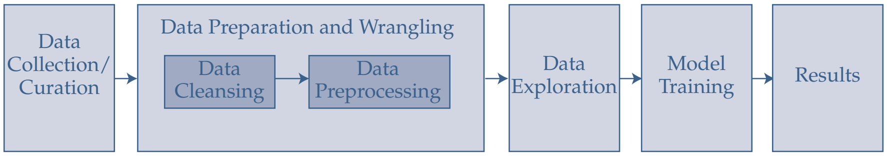
Before the data collection process even begins, it is important to state the problem, define objectives, identify useful data points, and conceptualize the model. Conceptualization is like a blueprint on a drawing board, a modifiable plan that is necessary to initiate the model building process. A project overview is established by determining the ML model type—supervised or unsupervised—and data sources/collection plans with respect to the needs of the project.

Data collection involves searching for and downloading the raw data from one or multiple sources. Data can be stored in different formats, sources, and locations. As databases are the most common primary sources, building necessary queries with the help of database administrators is critical. Database schemas are built with certain assumptions and exceptions, and it is safest to clarify the database architecture with an administrator or database architect before downloading the necessary data. Data also exist in the form of spreadsheets, comma-separated values (csv) files, text files, and other formats. Care must be taken before using such data, and documentation (often referred to as "Readme" files) must be referred to, if available. **Readme files** are text files provided with the raw data that contain information related to a data file. They are useful for understanding the data and how they can be interpreted correctly.

Alternatively, third-party data vendors can be sources of clean data. External data usually can be accessed through an **application programming interface (API)**—a set of well-defined methods of communication between various software components—or the vendors can deliver the required data in the form of csv files or other formats (as previously mentioned). Using external data can save time and resources that would otherwise go into data preparation and wrangling; however, vendor contracts come with a price. Depending on the big data project constraints, a decision must be made regarding the use of internal or external data based on the trade-offs between time, financial costs, and accuracy. For projects using internal user data, external data might not be suitable. For example, to understand user traffic on a company website, internally recorded site visits and click frequency may be captured and stored in the internal databases. External data are advantageous when a project requires generic data, such as demographics of a geographic area or traffic data of a public service. Another consideration in using external vendor provided data is that during the cleansing process, underlying trends in the data that are important for particular end-uses may be masked or even lost. This is where "alpha" is often found; so by simply buying a dataset from a vendor, you may lose your information edge. Of course, application of the data (e.g., merging and combining, putting through different types of models) will be different for everyone who uses it; there are always different ways to extract value.

Once the data are collected, the data preparation and wrangling stage begins. This stage involves two important tasks: cleansing and preprocessing, respectively. Exhibit 2 outlines data preparation and wrangling and defines the two component tasks. These tasks are explained in detail under the structured and unstructured sub-sections because the steps vary by the nature of data.

Exhibit 2 Data Preparation and Wrangling Stage



Data Preparation (Cleansing): This is the initial and most common task in data preparation that is performed on raw data. Data cleansing is the process of examining, identifying, and mitigating errors in raw data. Normally, the raw data are neither sufficiently complete nor sufficiently clean to directly train the ML model. Manually entered data can have incomplete, duplicated, erroneous, or inaccurate values. Automated data (recorded by systems) can have similar problems due to server failures and software bugs.

Data Wrangling (Preprocessing): This task performs transformations and critical processing steps on the cleansed data to make the data ready for ML model training. Raw data most commonly are not present in the appropriate format for model consumption. After the cleansing step, data need to be processed by dealing with outliers, extracting useful variables from existing data points, and scaling the data.

3.1 Structured Data

Data Preparation (Cleansing)

Structured data are organized in a systematic format that is readily searchable and readable by computer operations for processing and analyzing. In structured data, data errors can be in the form of incomplete, invalid, inaccurate, inconsistent, non-uniform, and duplicate data observations. The data cleansing process mainly deals with identifying and mitigating all such errors. Exhibit 3 shows a raw dataset before cleansing. The data have been collected from different sources and are organized in a data matrix (or data table) format. Each row contains observations of each customer of a US-based bank. Each column represents a variable (or feature) corresponding to each customer.

Exhibit 3 Raw Data Before Cleansing

1	ID	Name	Gender	Date of Birth	Salary	Other Income	State	Credit Card
2	1	Mr. ABC	M	12/5/1970	$50,200	$5,000	VA	Y
3	2	Ms. XYZ	M	15 Jan, 1975	$60,500	$0	NY	Y

Exhibit 3 (Continued)

1	ID	Name	Gender	Date of Birth	Salary	Other Income	State	Credit Card
4	3	EFG		1/13/1979	$65,000	$1,000	CA	N
5	4	Ms. MNO	F	1/1/1900	—	—	FL	Don't Know
6	5	Ms. XYZ	F	15/1/1975	$60,500	$0		Y
7	6	Mr. GHI	M	9/10/1942	NA	$55,000	TX	N
8	7	Mr. TUV	M	2/27/1956	$300,000	$50,000	CT	Y
9	8	Ms. DEF	F	4/4/1980	$55,000	$0	British Columbia	N

The possible errors in a raw dataset include the following:

1. *Incompleteness error* is where the data are not present, resulting in missing data. This can be corrected by investigating alternate data sources. Missing values and NAs (not applicable or not available values) must be either omitted or replaced with "NA" for deletion or substitution with imputed values during the data exploration stage. The most common imputations are mean, median, or mode of the variable or simply assuming zero. In Exhibit 3, rows 4 (ID 3), 5 (ID 4), 6 (ID 5), and 7 (ID 6) are incomplete due to missing values in either Gender, Salary, Other Income, Name (Salutation), and State columns.
2. *Invalidity error* is where the data are outside of a meaningful range, resulting in invalid data. This can be corrected by verifying other administrative data records. In Exhibit 3, row 5 likely contains invalid data as the date of birth is out of the range of the expected human life span.
3. *Inaccuracy error* is where the data are not a measure of true value. This can be rectified with the help of business records and administrators. In Exhibit 3, row 5 is inaccurate (it shows "Don't Know"); in reality, every person either has a credit card or does not.
4. *Inconsistency error* is where the data conflict with the corresponding data points or reality. This contradiction should be eliminated by clarifying with another source. In Exhibit 3, row 3 (ID 2) is likely to be inconsistent as the Name column contains a female title and the Gender column contains male.
5. *Non-uniformity error* is where the data are not present in an identical format. This can be resolved by converting the data points into a preferable standard format. In Exhibit 3, the data under the Date of Birth column is present in various formats. The data under the Salary column may also be non-uniform as the monetary units are ambiguous; the dollar symbol can represent US dollar, Canadian dollar, or others.
6. *Duplication error* is where duplicate observations are present. This can be corrected by removing the duplicate entries. In Exhibit 3, row 6 is a duplicate as the data under Name and Date of Birth columns are identical to the ones in row 3, referring to the same customer.

Exhibit 4 shows the dataset after completion of the cleansing process.

Exhibit 4 Data After Cleansing

1	ID	Name	Gender	Date of Birth	Salary	Other Income	State	Credit Card
2	1	Mr. ABC	M	12/5/1970	USD 50200	USD 5000	VA	Y
3	2	Ms. XYZ	F	1/15/1975	USD 60500	USD 0	NY	Y
4	3	Mr. EFG	M	1/13/1979	USD 65000	USD 1000	CA	N
5	6	Mr. GHI	M	9/10/1942	USD 0	USD 55000	TX	N
6	7	Mr. TUV	M	2/27/1956	USD 300000	USD 50000	CT	Y
7	8	Ms. DEF	F	4/4/1980	CAD 55000	CAD 0	British Columbia	N

Data cleansing can be expensive and cumbersome because it involves the use of automated, rule-based, and pattern recognition tools coupled with manual human inspection to sequentially check for the aforementioned types of errors row by row and column by column. The process involves a detailed data analysis as an initial step in identifying various errors that are present in the data. In addition to a manual inspection and verification of the data, analysis software, such as SPSS, can be used to understand **metadata** (data that describes and gives information about other data) about the data properties to use as a starting point to investigate any errors in the data. The business value of the project determines the necessary quality of data cleansing and subsequently the amount of resources used in the cleansing process. In case the errors cannot be resolved due to lack of available resources, the data points with errors can simply be omitted depending on the size of the dataset. For instance, if a dataset is large with more than 10,000 rows, removing a few rows (approximately 100) may not have a significant impact on the project. If a dataset is small with less than 1,000 rows, every row might be important and deleting many rows thus harmful to the project.

Data Wrangling (Preprocessing)

To make structured data ready for analyses, the data should be preprocessed. Data preprocessing primarily includes transformations and scaling of the data. These processes are exercised on the cleansed dataset. The following transformations are common in practice:

1. *Extraction:* A new variable can be extracted from the current variable for ease of analyzing and using for training the ML model. In Exhibit 4, the Date of Birth column consists of dates that are not directly suitable for analyses. Thus, an additional variable called "Age" can be extracted by calculating the number of years between the present day and date of birth.
2. *Aggregation:* Two or more variables can be aggregated into one variable to consolidate similar variables. In Exhibit 4, the two forms of income, Salary and Other Income, can be summed into a single variable called Total Income.
3. *Filtration:* The data rows that are not needed for the project must be identified and filtered. In Exhibit 4, row 7 (ID 8) has a non-US state; however, this dataset is for the US-based bank customers where it is required to have a US address.

4 *Selection:* The data columns that are intuitively not needed for the project can be removed. This should not be confused with feature selection, which is explained later. In Exhibit 4, Name and Date of Birth columns are not required for training the ML model. The ID column is sufficient to identify the observations, and the new extracted variable Age replaces the Date of Birth column.

5 *Conversion:* The variables can be of different types: nominal, ordinal, continuous, and categorical. The variables in the dataset must be converted into appropriate types to further process and analyze them correctly. This is critical for ML model training. Before converting, values must be stripped out with prefixes and suffixes, such as currency symbols. In Exhibit 4, Name is nominal, Salary and Income are continuous, Gender and Credit Card are categorical with 2 classes, and State is nominal. In case row 7 is not excluded, the Salary in row 7 must be converted into US dollars. Also, the conversion task applies to adjusting time value of money, time zones, and others when present.

Outliers may be present in the data, and domain knowledge is needed to deal with them. Any outliers that are present must first be identified. The outliers then should be examined and a decision made to either remove or replace them with values imputed using statistical techniques. In Exhibit 4, row 6 (ID 7) is an outlier because the Salary value is far above the upper quartile. Row 5 (ID 6) is also an outlier because the Salary value is far below the lower quartile. However, after the aggregation and formation of a new variable Total Income, as shown in Exhibit 5, row 5 (ID 6), it is no longer an outlier.

In practice, several techniques can be used to detect outliers in the data. Standard deviation can be used to identify outliers in normally distributed data. In general, a data value that is outside of 3 standard deviations from the mean may be considered an outlier. The interquartile range (IQR) can be used to identify outliers in data with any form of distribution. IQR is the difference between the 75th and the 25th percentile values of the data. In general, data values outside of the following are considered as outliers: +1.5 x IQR + 3rd Quartile Upper Bound; and -1.5 x IQR + 2nd Quartile Lower Bound. Using a multiple of 3.0 (instead of 1.5) times IQR would indicate extreme values.

There are several practical methods for handling outliers. When extreme values and outliers are simply removed from the dataset, it is known as **trimming** (also called truncation). For example, a 5% trimmed dataset is one for which the 5% highest and the 5% lowest values have been removed. When extreme values and outliers are replaced with the maximum (for large value outliers) and minimum (for small value outliers) values of data points that are not outliers, the process is known as **winsorization**.

Exhibit 5 Data After Applying Transformations

1	ID	Gender	Age	Total Income	State	Credit Card
2	1	M	48	55200	VA	Y
3	2	F	43	60500	NY	Y
4	3	M	39	66000	CA	N
5	6	M	76	55000	TX	N

Scaling is a process of adjusting the range of a feature by shifting and changing the scale of data. Variables, such as age and income, can have a diversity of ranges that result in a heterogeneous training dataset. For better ML model training when using such methods as support vector machines (SVMs) and artificial neural networks

(ANNs), all variables should have values in the same range to make the dataset homogeneous. It is important to remove outliers before scaling is performed. Here are two of the most common ways of scaling:

1. *Normalization* is the process of rescaling numeric variables in the range of [0, 1]. To normalize variable X, the minimum value (X_{min}) is subtracted from each observation (X_i), and then this value is divided by the difference between the maximum and minimum values of X ($X_{max} - X_{min}$) as follows:

$$X_{i\ (\text{normalized})} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

2. *Standardization* is the process of both centering and scaling the variables. Centering involves subtracting the mean (μ) of the variable from each observation (X_i) so the new mean is 0. Scaling adjusts the range of the data by dividing the centered values (X_i − μ) by the standard deviation (σ) of feature X. The resultant standardized variable will have an arithmetic mean of 0 and standard deviation of 1.

$$X_{i\ (\text{standardized})} = \frac{X_i - \mu}{\sigma} \tag{2}$$

Normalization is sensitive to outliers, so treatment of outliers is necessary before normalization is performed. Normalization can be used when the distribution of the data is not known. Standardization is relatively less sensitive to outliers as it depends on the mean and standard deviation of the data. However, the data must be normally distributed to use standardization.

EXAMPLE 2

Preparing and Wrangling Structured Data

Paul Wang's analytics team at LendALot Corporation is working to develop its first ML model for classifying prospective borrowers' creditworthiness. Wang has asked one of his data scientists, Lynn Lee, to perform a preliminary assessment of the data cleansing and preprocessing tasks the team will need to perform. As part of this assessment, Lee pulled the following sample of data for manual examination, which she brings to Wang to discuss.

1	ID	Name	Loan Outcome	Income (USD)	Loan Amount (USD)	Credit Score	Loan Type
2	1	Mr. Alpha	No Default	34,000	10,000	685	Mortgage
3	2	Ms. Beta	No Default	−63,050	49,000	770	Student Loan
4	3	Mr. Gamma	Defaulted	20,565	35,000	730	
5	4	Ms. Delta	No Default	50,021	unknown	664	Mortgage
6	5	Mr. Epsilon	Defaulted	100,350	129,000	705	Car Loan
7	6	Mr. Zeta	No Default	800,000	300,000	800	Boat Loan
8	6	Mr. Zeta	No Default	800,000	300,000	800	Boat Loan

After sharing a concern that the data should be thoroughly cleansed, Wang makes the following statements:

Statement 1 "Let's keep the ID column and remove the column for Name from the dataset."

Statement 2 "Let's create a new feature, "Loan Amount as a Percent of Income," to use as an additional feature."

1 The data shown for Ms. Beta contain what is *best described* as an:

A invalidity error.

B inaccuracy error.

C incompleteness error.

2 The data shown for Mr. Gamma contain what is *best described* as an:

A invalidity error.

B duplication error.

C incompleteness error.

3 The data shown for Ms. Delta contain what is *best described* as an:

A invalidity error.

B inaccuracy error.

C duplication error.

4 The data shown for Mr. Zeta contain what is *best described* as an:

A invalidity error.

B inaccuracy error.

C duplication error.

5 The process mentioned in Wang's first statement is *best described* as:

A feature selection.

B feature extraction.

C feature engineering

6 Wang's second statement is *best described* as:

A feature selection.

B feature extraction.

C feature engineering.

Solution to 1:

A is correct. This is an invalidity error because the data are outside of a meaningful range. Income cannot be negative.

Solution to 2:

C is correct. This is an incompleteness error as the loan type is missing.

Solution to 3:

B is correct. This is an inaccuracy error because LendALot must know how much they have lent to that particular borrower (who eventually repaid the loan as indicated by the loan outcome of no default).

Solution to 4:

C is correct. Row 8 duplicates row 7: This is a duplication error.

Solution to 5:

A is correct. The process mentioned involves selecting the features to use. The proposal makes sense; with "ID," "Name" is not needed to identify an observation.

Solution to 6:

B is correct. The proposed feature is a ratio of two existing features. *Feature extraction* is the process of creating (i.e., extracting) new variables from existing ones in the data.

4 UNSTRUCTURED (TEXT) DATA

b describe objectives, steps, and examples of preparing and wrangling data;

Unstructured data are not organized into any systematic format that can be processed by computers directly. They are available in formats meant for human usage rather than computer processing. Unstructured data constitute approximately 80% of the total data available today. They can be in the form of text, images, videos, and audio files. Unlike in structured data, preparing and wrangling unstructured data are both more challenging. For analysis and use to train the ML model, the unstructured data must be transformed into structured data. In this section, text data will be used to demonstrate unstructured data preparation and wrangling. The cleansing and preprocessing of text data is called *text processing*. Text processing is essentially cleansing and transforming the unstructured text data into a structured format. Text processing can be divided into two tasks: cleansing and preprocessing. The following content is related to text data in the English language.

Text Preparation (Cleansing)

Raw text data are a sequence of characters and contain other non-useful elements, including html tags, punctuations, and white spaces (including tabs, line breaks, and new lines). It is important to clean the text data before preprocessing. Exhibit 6 shows a sample text from the home page for the hypothetical company Robots Are Us website. The text appears to be clean visually and is designed for human readability.

Exhibit 6 Sample Text from Robots Are Us Home Page

Robots Are Us

Every home and business should have a robot.

However, the source text that can be downloaded is not as clean. The raw text contains html tags and formatting elements along with the actual text. Exhibit 7 shows the raw text from the source.

Exhibit 7 Raw Text from the Source

<h1 class="text-left mb-3">Robots Are Us</h1>
<h2> Every home and business shoudl have a robot </h2>

The initial step in text processing is cleansing, which involves basic operations to clean the text by removing unnecessary elements from the raw text. Text operations often use regular expressions. A **regular expression (regex)** is a series that contains characters in a particular order. Regex is used to search for patterns of interest in a given text. For example, a regex "<.*?>" can be used to find all the html tags that are present in the form of <…> in text.[1] GREP (global regular expression print) is a commonly available utility in programming languages for searching patterns using regex. Once a pattern is found, it can be removed or replaced. Additionally, advanced html parsers and packages are available in the popular programming languages, such as R and Python, to deal with this task.

The following steps describe the basic operations in the text cleansing process.

1 *Remove html tags*: Most of the text data are acquired from web pages, and the text inherits html markup tags with the actual content. The initial task is to remove (or strip) the html tags that are not part of the actual text using programming functions or using regular expressions. In Exhibit 7, </h2> is an html tag that can be identified by a regex and be removed. Note that it is not uncommon to keep some generic html tags to maintain certain formatting meaning in the text.

2 *Remove Punctuations*: Most punctuations are not necessary for text analysis and should be removed. However, some punctuations, such as percentage signs, currency symbols, and question marks, may be useful for ML model training. These punctuations should be substituted with such annotations as /percentSign/, /dollarSign/, and /questionMark/ to preserve their grammatical meaning in the text. Such annotations preserve the semantic meaning of important characters in the text for further text processing and analysis stages. It is important to note that periods (dots) in the text need to be processed carefully. There are different circumstances for periods to be present in text—characteristically used for abbreviations, sentence boundaries, and decimal points. The periods and the context in which they are used need to be identified and must be appropriately replaced or removed. In general, periods after abbreviations can be removed, but the periods separating sentences should be replaced by the annotation /endSentence/. Some punctuations, such as hyphens and underscores, can be kept in the text to keep the consecutive words intact as a single term (e.g., e-mail). Regex are often used to remove or replace punctuations.

3 *Remove Numbers*: When numbers (or digits) are present in the text, they should be removed or substituted with an annotation /number/. This helps inform the computer that a number is present, but the actual value of the number itself is not helpful for categorizing/analyzing the text. Such operations are critical for ML model training. Otherwise, the computers will treat each number as a separate word, which may complicate the analyses or add noise. Regex are often used to remove or replace numbers. However, the number and any decimals must be retained where the outputs of interest are the actual values of the number. One such text application is information extraction (IE), where the goal is to extract relevant information from a given text. An IE task could be extracting monetary values from financial reports, where the actual number values are critical.

4 *Remove white spaces*: It is possible to have extra white spaces, tab spaces, and leading and ending spaces in the text. The extra white spaces may be introduced after executing the previously mentioned operations. These should be identified

1 A regex of the form "<.*?>" will identify all html tags with anything (*) of any length (?) between the brackets (< >).

and removed to keep the text intact and clean. Certain functions in programming languages can be used to remove unnecessary white spaces from the text. For example, the text mining package in R offers a *stripwhitespace* function.

Exhibit 8 uses a sample financial text to show the transformations occurring after applying each operation of the text cleansing process. The four steps are applied on a mock financial text after scraping from a source. As noted previously, scraping (or web scraping) is a technique to extract raw content from a source, typically web pages. It is important to note that the sequence and choice of cleansing operations does matter. For instance, after removing punctuation, the "1.2 million" becomes "12 million." This is acceptable here since a subsequent operation replaces all numbers with a "/number/" tag. However, if numbers were not replaced with such tags, the punctuation removal operation could affect the data.

Exhibit 8 Text Cleansing Process Example

Original text from a financial statement as shown on a webpage

CapEx on the normal operations remained stable on historicallylow levels, $800,000 compared to $1.2 million last year.

Quarter 3, so far, is 5% sales growth quarter-to-date, and year-to-date, we have a 4% local currency sales development.

Raw text after scraping from the source

<p> CapEx on the normal operations remained stable on historically low levels, $800,000 compared to $1.2 million last year. <b/><b/> Quarter 3, so far, is 5% sales growth quarter-to-date, and year-to-date, we have a 4% local currency sales development.</p>

Text after removing html tags

CapEx on the normal operations remained stable on historically low levels, $800,000 compared to $1.2 million last year.
Quarter 3, so far, is 5% sales growth quarter-to-date, and year-to-date, we have a 4% local currency sales development.

Text after removing and replacing punctuations

CapEx on the normal operations remained stable on historically low levels /dollarSign/800000 compared to /dollarSign/12 million last year /endSentence/ Quarter 3 so far is 5 /percentSign/ sales growth quarter-to-date and year-to-date we have a 4 /percentSign/ local currency sales development /endSentence/

Text after replacing numbers

CapEx on the normal operations remained stable on historically low levels /dollarSign//number/ compared to/dollarSign//number/ million last year /endSentence/ Quarter/number/ so far is /number/ /percentSign/sales growth quarter-to-date and year-to-date we have a /number/ /percentSign/ local currency sales development /endSentence/

Text after removing extra white spaces

CapEx on the normal operations remained stable on historically low levels/dollarSign//number/compared to/dollarSign//number/million last year/endSentence/ Quarter/number/so far is /number//percentSign/sales growth quarter-to-date and year-to-date we have a/number//percentSign/local currency sales development/endSentence/

Text Wrangling (Preprocessing)

To further understand text processing, tokens and tokenization need to be defined. A **token** is equivalent to a word, and **tokenization** is the process of splitting a given text into separate tokens. In other words, a text is considered to be a collection of tokens. Tokenization can be performed at word or character level, but it is most commonly performed at word level. Exhibit 9 shows a sample dataset of four cleansed texts and their word tokens.

Exhibit 9 Tokenization of Four Texts

	Cleaned Texts	Tokens
Text 1	The man went to the market today	The · man · went · to · the · market · today
Text 2	Market values are increasing	Market · values · are · increasing
Text 3	Increased marketing is needed	Increased · marketing · is · needed
Text 4	There is no market for the product	There · is · no · market · for · the · product

Similar to structured data, text data also require normalization. The normalization process in text processing involves the following:

1 *Lowercasing* the alphabet removes distinctions among the same words due to upper and lower cases. This action helps the computers to process the same words appropriately (e.g., "The" and "the").

2 *Stop words* are such commonly used words as "the," "is," and "a." Stop words do not carry a semantic meaning for the purpose of text analyses and ML training. However, depending on the end-use of text processing, for advance text applications it may be critical to keep the stop words in the text in order to understand the context of adjacent words. For ML training purposes, stop words typically are removed to reduce the number of tokens involved in the training set. A predefined list of stop words is available in programming languages to help with this task. In some cases, additional stop words can be added to the list based on the content. For example, the word "exhibit" may occur often in financial filings, which in general is not a stop word but in the context of the filings can be treated as a stop word.

3 *Stemming* is the process of converting inflected forms of a word into its base word (known as stem). Stemming is a rule-based approach, and the results need not necessarily be linguistically sensible. Stems may not be the same as the morphological root of the word. Porter's algorithm is the most popular method for stemming. For example, the stem of the words "analyzed" and "analyzing" is "analyz." Similarly, the British English variant "analysing" would become "analys." Stemming is available in R and Python. The text mining package in R provides a *stemDocument* function that uses this algorithm.

4 *Lemmatization* is the process of converting inflected forms of a word into its morphological root (known as lemma). Lemmatization is an algorithmic approach and depends on the knowledge of the word and language structure. For example, the lemma of the words "analyzed" and "analyzing" is "analyze." Lemmatization is computationally more expensive and advanced.

Stemming or lemmatization will reduce the repetition of words occurring in various forms and maintain the semantic structure of the text data. Stemming is more common than lemmatization in the English language since it is simpler to perform. In text data, data sparseness refers to words that appear very infrequently, resulting in data consisting of many unique, low frequency tokens. Both techniques decrease data sparseness by aggregating many sparsely occurring words in relatively less sparse stems or lemmas, thereby aiding in training less complex ML models.

After the cleansed text is normalized, a bag-of-words is created. **Bag-of-words (BOW)** representation is a basic procedure used to analyze text. It is essentially a collection of a distinct set of tokens from all the texts in a sample dataset. BOW is simply a set of words and does not capture the position or sequence of words present in the text. However, it is memory efficient and easy to handle for text analyses.

Exhibit 10 shows the BOW and transformations occurring in each step of normalization on the cleansed texts from Exhibit 9. Note that the number of words decreases as the normalizing steps are applied, making the resulting BOW smaller and simpler.

Exhibit 10 Bag-of-Words Representation of Four Texts Before and After Normalization Process

BOW before normalizing

"The"	"man"	"went"	"to"	"the"	"market"
"today"	"Market"	"values"	"are"	"increasing"	"Increased"
"marketing"	"is"	"needed"	"There"	"no"	"for"
"product"					

BOW after removing uppercase letters

"the"	"man"	"went"	"to"	"market"	"today"
"values"	"are"	"increasing"	"increased"	"marketing"	"is"
"needed"	"there"	"no"	"for"	"product"	

BOW after removing stop words

"man"	"went"	"market"	"today"	"values"	"increasing"
"increased"	"marketing"	"needed"	"product"		

BOW after stemming

"man"	"went"	"market"	"today"	"valu"	"increas"	"need"	"product"

The last step of text preprocessing is using the final BOW after normalizing to build a **document term matrix (DTM)**. DTM is a matrix that is similar to a data table for structured data and is widely used for text data. Each row of the matrix belongs to a document (or text file), and each column represents a token (or term). The number of rows of DTM is equal to the number of documents (or text files) in a sample dataset. The number of columns is equal to the number of tokens from the BOW that is built using all the documents in a sample dataset. The cells can contain the counts of the number of times a token is present in each document. The matrix cells can be filled with other values that will be explained in the financial forecasting project section of this reading; a large dataset is helpful in understanding the concepts. At this point, the unstructured text data are converted to structured data that can be processed further and used to train the ML model. Exhibit 11 shows a DTM constructed from the resultant BOW of the four texts from Exhibit 10.

Exhibit 11 DTM of Four Texts and Using Normalized BOW Filled with Counts of Occurrence

	man	went	market	today	valu	increas	need	product
Text 1	1	1	1	1	0	0	0	0
Text 2	0	0	1	0	1	1	0	0
Text 3	0	0	1	0	0	1	1	0
Text 4	0	0	1	0	0	0	0	1

As seen in Exhibit 10, BOW does not represent the word sequences or positions, which limits its use for some advanced ML training applications. In the example, the word "no" is treated as a single token and has been removed during the normalization because it is a stop word. Consequently, this fails to signify the negative meaning ("no market") of the text (i.e., Text 4). To overcome such problems, a technique called n-grams can be employed. **N-grams** is a representation of word sequences. The length of a sequence can vary from 1 to n. When one word is used, it is a unigram; a two-word sequence is a bigram; and a 3-word sequence is a trigram; and so on. Exhibit 10, for example, shows a unigram (n = 1) BOW. The advantage of n-grams is that they can be used in the same way as unigrams to build a BOW. In practice, different n-grams can be combined to form a BOW and eventually be used to build a DTM. Exhibit 12 shows unigrams, bigrams, and trigrams. Exhibit 12 also shows a combined unigram-to-trigram BOW for the particular text. Stemming can be applied on the cleansed text before building n-grams and BOW (not shown in Exhibit 12).

Exhibit 12 N-Grams and N-Grams BOW

Clean text

The man went to the market today

Unigrams

"The" "man" "went" "to" "the" "market" "today"

Bigrams

"The_man" "man_went" "went_to" "to_the" "the_market" "market_today"

Trigrams

"The_man_went" "man_went_to" "went_to_the" "to_the_market" "the_market_today"

BOW before normalizing

"The"	"man"	"went"	"to"	"the"	"market"	"today"
"The_man"	"man_went"	"went_to"	"to_the"	"the_market"	"market_today"	"The_man_went"
"man_went_to"	"went_to_the"	"to_the_market"	"the_market_today"			

BOW after removing upper case letters

"the"	"man"	"went"	"to"	"market"	"today"	"the_man"
"man_went"	"went_to"	"to_the"	"the_market"	"market_today"	"the_man_went"	"man_went_to"
"went_to_the"	"to_the_market"	"the_market_today"				

BOW after removing stop words

"man"	"went"	"market"	"today"	"the_man"	"man_went"	"went_to"
"to_the"	"the_market"	"market_today"	"the_man_went"	"man_went_to"	"went_to_the"	"to_the_market"
"the_market_today"						

The n-grams implementation will vary the impact of normalization on the BOW. Even after removing isolated stop words, stop words tend to persist when they are attached to their adjacent words. For instance, "to_the" (Exhibit 12) is a single bigram token consisting of stop words and will not be removed by the predetermined list of stop words.

EXAMPLE 3

Unstructured Data Preparation and Wrangling

1 The output produced by preparing and wrangling textual data is best described as a:

A data table.

B confusion matrix.

C document term matrix.

2 In text cleansing, situations in which one may need to add an annotation include the removal of:

A html tags.

B white spaces.

C punctuations.

3 A column of a document term matrix is *best* described as representing:

A a token.

B a regularization term.

C an instance.

4 A cell of a document term matrix is *best* described as containing:

A a token.

B a count of tokens.

C a count of instances.

5 Points to cover in normalizing textual data include:

A removing numbers.

B removing white spaces.

C lowercasing the alphabet.

6 When some words appear very infrequently in a textual dataset, techniques that may address the risk of training highly complex models include:

A stemming.

B scaling.

C data cleansing.

7 Which of the following statements concerning tokenization is *most* accurate?

A Tokenization is part of the text cleansing process.

B Tokenization is most commonly performed at the character level.

C Tokenization is the process of splitting a given text into separate tokens.

Solution to 1:

C is correct. The objective of data preparation and wrangling of textual data is to transform the unstructured data into structured data. The output of these processes is a document term matrix that can be read by computers. The document term matrix is similar to a data table for structured data.

Solution to 2:

C is correct. Some punctuations, such as percentage signs, currency symbols, and question marks, may be useful for ML model training, so when such punctuations are removed annotations should be added.

Solution to 3:

A is correct. Each column of a document term matrix represents a token from the bag-of-words that is built using all the documents in a sample dataset.

Solution to 4:

B is correct. A cell in a document term matrix contains a count of the number of tokens of the kind indicated in the column heading.

Solution to 5:

C is correct. The other choices are related to text cleansing.

Solution to 6:

A is correct. Stemming, the process of converting inflected word forms into a base word (or stem), is one technique that can address the problem described.

Solution to 7:

C is correct, by definition. The other choices are not true.

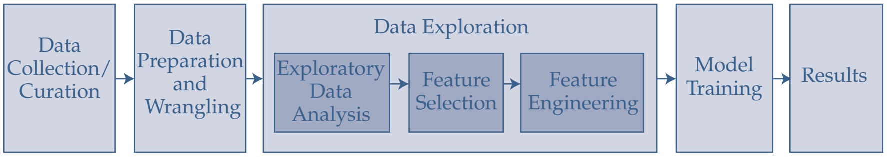
5 DATA EXPLORATION OBJECTIVES AND METHODS AND STRUCTURED DATA

c describe objectives, methods, and examples of data exploration;

Data exploration is a crucial part of big data projects. The prepared data are explored to investigate and comprehend data distributions and relationships. The knowledge that is gained about the data in this stage is used throughout the project. The outcome and quality of exploration strongly affects ML model training results. Domain knowledge plays a vital role in exploratory analysis as this stage should involve cooperation between analysts, model designers, and experts in the particular data domain. Data exploration without domain knowledge can result in ascertaining spurious relationships among the variables in the data that can mislead the analyses. The data exploration stage follows the data preparation stage and leads to the model training stage.

Data exploration involves three important tasks: exploratory data analysis, feature selection, and feature engineering. These three tasks are outlined in Exhibit 13 and are defined and further explained under the structured and unstructured data subsections.

Exhibit 13 Data Exploration Stage



Exploratory data analysis (EDA) is the preliminary step in data exploration. Exploratory graphs, charts, and other visualizations, such as heat maps and word clouds, are designed to summarize and observe data. In practice, many exploratory graphs are made for investigation and can be made swiftly using statistical programming and generic spreadsheet software tools. Data can also be summarized and examined using quantitative methods, such as descriptive statistics and central tendency measures. An important objective of EDA is to serve as a communication medium among project stakeholders, including business users, domain experts, and analysts. Relatively quick and easy exploratory visualizations help stakeholders connect and ensure the prepared data are sensible. Other objectives of EDA include:

- understanding data properties,
- finding patterns and relationships in data,
- inspecting basic questions and hypotheses,
- documenting data distributions and other characteristics, and
- planning modeling strategies for the next steps.

Feature selection is a process whereby only pertinent features from the dataset are selected for ML model training. Selecting fewer features decreases ML model complexity and training time. **Feature engineering** is a process of creating new features by changing or transforming existing features. Model performance heavily depends on feature selection and engineering.

5.1 Structured Data

Exploratory Data Analysis

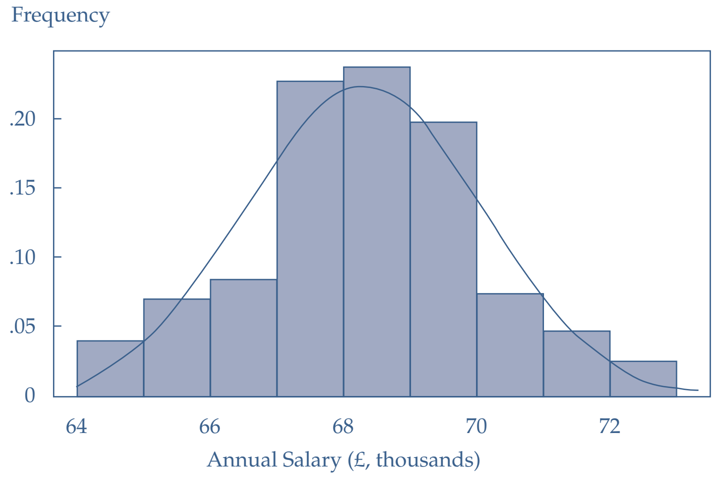
For structured data, each data table row contains an observation and each column contains a feature. EDA can be performed on a single feature (one-dimension) or on multiple features (multi-dimension). For high-dimension data with many features, EDA can be facilitated by using a dimension reduction technique, such as principal components analysis (PCA). Based on the number of dimensions, the exploratory techniques will vary.

For one-dimensional data, summary statistics, such as mean, median, quartiles, ranges, standard deviations, skewness, and kurtosis, of a feature can be computed. One-dimension visualization summarizes each feature in the dataset. The basic one-dimension exploratory visualizations are as follows:

- Histograms
- Bar charts
- Box plots
- Density plots

Histograms represent equal bins of data and their respective frequencies. They can be used to understand the high-level distribution of the data. Bar charts summarize the frequencies of categorical variables. Box plots show the distribution of continuous data by highlighting the median, quartiles, and outliers of a feature that is normally distributed. Density plots are another effective way to understand the distribution of continuous data. Density plots are smoothed histograms and are commonly laid on top of histograms, as shown in Exhibit 14. This histogram shows a hypothetical annual salary distribution (in £) of entry-level analyst positions at UK banks. The data represent a normal distribution with an approximate mean of £68,500.

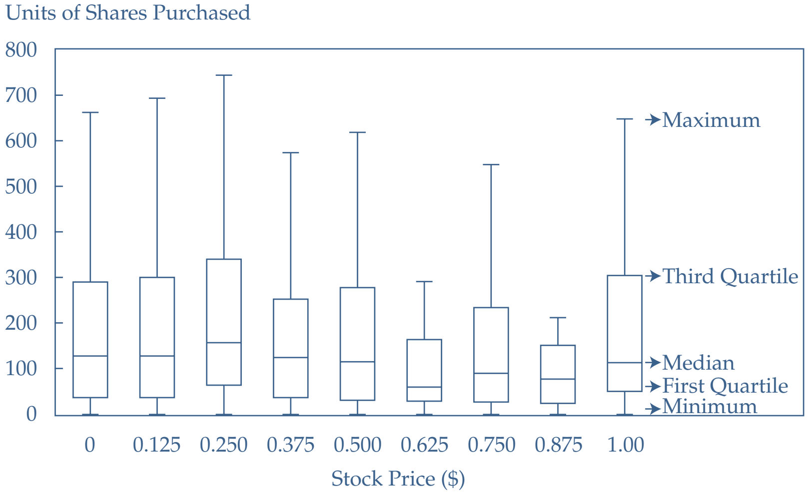
Exhibit 14 Histogram with Superimposed Density Plot



For data with two or more dimensions, summary statistics of relationships, such as a correlation matrix, can be calculated. Two- or more-dimensional visualization explores interactions between different features in the dataset. Common methods include scatterplots and line graphs. In multi-dimensional visualization, one-dimensional plots are overlaid to summarize each feature, thus enabling comparison between features. Additionally, attributes (e.g., color, shape, and size) and legends can be used creatively to pack more information about the data into fewer graphs.

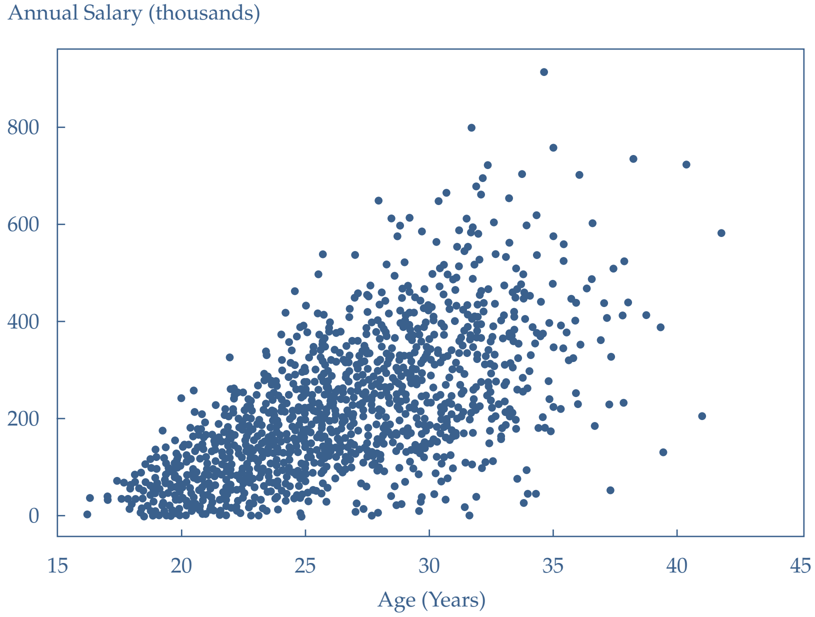
For multivariate data, commonly utilized exploratory visualization designs include stacked bar and line charts, multiple box plots, and scatterplots showing multivariate data that use different colors or shapes for each feature. Multiple box plots can be arranged in a single chart, where each individual box plot represents a feature. Such a multi-box plot chart assesses the relationship between each feature (x-axis) in the dataset and the target variable of interest (y-axis). The multi-box plot chart in Exhibit 15 represents units of shares purchased versus stock price for a hypothetical stock. The x-axis shows the stock price in increments of $0.125, and the y-axis shows units of shares purchased. The individual box plots indicate the distribution of shares purchased at the different stock prices. When the stock price is $0.25, the median number of shares purchased is the highest; when the stock price is $0.625, the median number of shares purchased is the lowest. However, visually it appears that the number of shares purchased at different stock prices is not significantly different.

Exhibit 15 Multiple Box Plots in One Chart



Two-dimensional charts can summarize and approximately measure relationships between two or more features. An example scatterplot in Exhibit 16 shows the interaction of two hypothetical features: age (x-axis) and annual salary (y-axis). The feature on the y-axis tends to increase as the feature on the x-axis increases. This pattern appears true visually; however, it may not be a statistically significant relationship. A scatterplot provides a starting point where relationships can be examined visually. These potential relationships should be tested further using statistical tests. Common parametric statistical tests include ANOVA, *t*-test, and Pearson correlation. Common non-parametric statistical tests include chi-square and the Spearman rank-order correlation.

Exhibit 16 Scatterplot Showing a Linear Relationship Between Two Features



In addition to visualization, descriptive statistics are a good means to summarize data. Central tendency measures as well as minimum and maximum values for continuous data are useful. Counts and frequencies for categorical data are commonly employed to gain insight regarding the distribution of possible values.

EDA is not only useful for revealing possible relationships among features or general trends in the data; it is also beneficial during the feature selection and engineering stages. These possible relationships and trends in the data may be used to suggest new features that, when incorporated into a model, may improve model training.

Feature Selection

Structured data consist of features, represented by different columns of data in a table or matrix. After using EDA to discover relevant patterns in the data, it is essential to identify and remove unneeded, irrelevant, and redundant features. Basic diagnostic testing should also be performed on features to identify redundancy, heteroscedasticity, and multi-collinearity. The objective of the feature selection process is to assist in identifying significant features that when used in a model retain the important patterns and complexities of the larger dataset while requiring fewer data overall. This last point is important since computing power is not free (i.e., explicit costs and processing time).

Typically, structured data even after the data preparation step can contain features that do not contribute to the accuracy of an ML model or that negatively affect the quality of ML training. The most desirable outcome is a parsimonious model with fewer features that provides the maximum predictive power out-of-sample.

Feature selection must not be confused with the data preprocessing steps during data preparation. Good feature selection requires an understanding of the data and statistics, and comprehensive EDA must be performed to assist with this step. Data preprocessing needs clarification only from data administrators and basic intuition (e.g., salary vs. income) during data preparation.

Feature selection on structured data is a methodical and iterative process. Statistical measures can be used to assign a score gauging the importance of each feature. The features can then be ranked using this score and either retained or eliminated from the dataset. The statistical methods utilized for this task are usually univariate and consider each feature independently or with regard to the target variable. Methods include chi-square test, correlation coefficients, and information-gain measures (i.e., R-squared values from regression analysis). All of these statistical methods can be combined in a manner that uses each method individually on each feature, automatically performing backward and forward passes over features to improve feature selection. Prebuilt feature selection functions are available in popular programming languages used to build and train ML models.

Dimensionality reduction assists in identifying the features in the data that account for the greatest variance between observations and allows for the processing of a reduced volume of data. Dimensionality reduction may be implemented to reduce a large number of features, which helps reduce the memory needed and speed up learning algorithms. Feature selection is different from dimensionality reduction, but both methods seek to reduce the number of features in the dataset. The dimensionality reduction method creates new combinations of features that are uncorrelated, whereas feature selection includes and excludes features present in the data without altering them.

Feature Engineering

After the appropriate features are selected, feature engineering helps further optimize and improve the features. The success of ML model training depends on how well the data are presented to the model. The feature engineering process attempts to produce good features that describe the structures inherent in the dataset. This process depends on the context of the project, domain of the data, and nature of the problem. Structured data are likely to contain quantities, which can be engineered to better present relevant patterns in the dataset. This action involves engineering an existing feature into a new feature or decomposing it into multiple features.

For continuous data, a new feature may be created—for example, by taking the logarithm of the product of two or more features. As another example, when considering a salary or income feature, it may be important to recognize that different salary brackets impose a different taxation rate. Domain knowledge can be used to decompose an income feature into different tax brackets, resulting in a new feature: "income_above_100k," with possible values 0 and 1. The value 1 under the new feature captures the fact that a subject has an annual salary of more than $100,000. By grouping subjects into income categories, assumptions about income tax can be made and utilized in a model that uses the income tax implications of higher and lower salaries to make financial predictions.

For categorical data, for example, a new feature can be a combination (e.g., sum or product) of two features or a decomposition of one feature into many. If a single categorical feature represents education level with five possible values—high school, associates, bachelor's, master's, and doctorate—then these values can be decomposed into five new features, one for each possible value (e.g., is_highSchool, is_doctorate) filled with 0s (for false) and 1s (for true). The process in which categorical variables are converted into binary form (0 or 1) for machine reading is called **one hot encoding**. It is one of the most common methods for handling categorical features in text data. When date-time is present in the data, such features as "second of the hour," "hour of the day," and "day of the date" can be engineered to capture critical information about temporal data attributes—which are important, for example, in modeling trading algorithms.

Feature engineering techniques systemically alter, decompose, or combine existing features to produce more meaningful features. More meaningful features allow an ML model to train more swiftly and easily. Different feature engineering strategies can lead to the generation of dramatically different results from the same ML model. The impact of feature selection and engineering on ML training is discussed further in the next section.

6 UNSTRUCTURED DATA - TEXT EXPLORATION

c describe objectives, methods, and examples of data exploration;

Exploratory Data Analysis

Just like with structured data, it is important to gain insight into existing patterns in the unstructured data for further analysis. In this section, text data will be discussed. Text analytics has various applications. The most common applications are text classification, topic modeling, fraud detection, and sentiment analysis. Text classification uses supervised ML approaches to classify texts into different classes. Topic modeling uses unsupervised ML approaches to group the texts in the dataset into topic clusters. Sentiment analysis predicts sentiment (negative, neutral, or positive) of the texts in a dataset using both supervised and unsupervised approaches.

Various statistics are used to explore, summarize, and analyze text data. Text data include a collection of texts (also known as a corpus) that are sequences of tokens. It is useful to perform EDA of text data by computing on the tokens such basic text statistics as **term frequency (TF)**, the ratio of the number of times a given token occurs in all the texts in the dataset to the total number of tokens in the dataset (e.g., word associations, average word and sentence length, and word and syllable counts).

Text statistics reveal patterns in the co-occurrence of words. There are many applications of text analytics, and necessary text statistics vary according to the context of the application. Topic modeling is a text data application in which the words that are most informative are identified by calculating the TF of each word. For example, the word "soccer" can be informative for the topic "sports." The words with high TF values are eliminated as they are likely to be stop words or other common vocabulary words, making the resulting BOW compact and more likely to be relevant to topics within the texts. In sentiment analysis and text classification applications, the chi-square measure of word association can be useful for understanding the significant word appearances in negative and positive sentences in the text or in different documents. The chi-square measure is further explained under feature selection. Such EDA plays a vital role in executing the feature selection step.

Text statistics can be visually comprehended by using the same methods as explained in the structured data section. For example, bar charts can be used to show word counts or frequency. Words clouds are common visualizations when working with text data as they can be made to visualize the most informative words and their TF values. The most commonly occurring words in the dataset can be shown by varying font size, and color is used to add more dimensions, such as frequency and length of words. Exhibit 17 shows a word cloud constructed from a sample dataset of generic financial news wires after text processing. Word cloud building functions and packages are available in several popular programming languages. A detailed demonstration of text data EDA will be presented in Section 10, where we work with actual text data in a financial forecasting project.

Exhibit 17 Word Cloud of Generic Financial Newsfeed Data Sample

Feature Selection

For text data, feature selection involves selecting a subset of the terms or tokens occurring in the dataset. The tokens serve as features for ML model training. Feature selection in text data effectively decreases the size of the vocabulary or BOW. This helps the ML model be more efficient and less complex. Another benefit is to eliminate noisy features from the dataset. Noisy features are tokens that do not contribute to ML model training and actually might detract from the ML model accuracy.

Noisy features are both the most frequent and most sparse (or rare) tokens in the dataset. On one end, noisy features can be stop words that are typically present frequently in all the texts across the dataset. On the other end, noisy features can be sparse terms that are present in only a few text cases. Text classification involves dividing text documents into assigned classes (a class is a category; examples include "relevant" and "irrelevant" text documents or "bearish" and "bullish" sentences). The *frequent* tokens strain the ML model to choose a decision boundary among the texts as the terms are present across all the texts, an example of model *underfitting*. The *rare* tokens mislead the ML model into classifying texts containing the rare terms into a specific class, an example of model *overfitting*. Identifying and removing noise features is very critical for text classification applications. The general feature selection methods in text data are as follows:

1 *Frequency* measures can be used for vocabulary pruning to remove noise features by filtering the tokens with very high and low TF values across all the texts. **Document frequency (DF)** is another frequency measure that helps to discard the noise features that carry no specific information about the text class and are present across all texts. The DF of a token is defined as the number of documents (texts) that contain the respective token divided by the total number of documents. It is the simplest feature selection method and often performs well when many thousands of tokens are present.

2 *Chi-square* test can be useful for feature selection in text data. The chi-square test is applied to test the independence of two events: occurrence of the token and occurrence of the class. The test ranks the tokens by their usefulness to each class in text classification problems. Tokens with the highest chi-square test statistic values occur more frequently in texts associated with a particular class and therefore can be selected for use as features for ML model training due to higher discriminatory potential.

3 *Mutual information* (MI) measures how much information is contributed by a token to a class of texts. The **mutual information** value will be equal to 0 if the token's distribution in all text classes is the same. The MI value approaches 1 as the token in any one class tends to occur more often in only that particular class of text. Exhibit 18 shows a simple depiction of some tokens with high MI scores for their corresponding text classes. Note how the tokens (or features) with the highest MI values narrowly relate to their corresponding text class name.

Exhibit 18 Tokens with Mutual Information (MI) Values for Two Given Text Classes

Text Classes: Sports or Politics			
Sports		Politics	
Token	MI Value	Token	MI Value
soccer	0.0781	election	0.0612
cup	0.0525	president	0.0511
match	0.0456	polls	0.0341
play	0.0387	vote	0.0288
game	0.0299	party	0.0202

Exhibit 18 (Continued)

Text Classes: Sports or Politics			
Sports		Politics	
Token	MI Value	Token	MI Value
team	0.0265	candidate	0.0201
win	0.0189	campaign	0.0201

Feature Engineering

As with structured data, feature engineering can greatly improve ML model training and remains a combination of art and science. The following are some techniques for feature engineering, which may overlap with text processing techniques.

1 *Numbers*: In text processing, numbers are converted into a token, such as "/number/." However, numbers can be of different lengths of digits representing different kinds of numbers, so it may be useful to convert different numbers into different tokens. For example, numbers with four digits may indicate years, and numbers with many digits could be an identification number. Four-digit numbers can be replaced with "/number4/," 10-digit numbers with "/number10/," and so forth.

2 *N-grams*: Multi-word patterns that are particularly discriminative can be identified and their connection kept intact. For example, "market" is a common word that can be indicative of many subjects or classes; the words "stock market" are used in a particular context and may be helpful to distinguish general texts from finance-related texts. Here, a bigram would be useful as it treats the two adjacent words as a single token (e.g., stock_market).

3 *Name entity recognition (NER)*: NER is an extensive procedure available as a library or package in many programming languages. The **name entity recognition** algorithm analyzes the individual tokens and their surrounding semantics while referring to its dictionary to tag an object class to the token. Exhibit 19 shows the NER tags of the text "*CFA Institute was formed in 1947 and is headquartered in Virginia.*" Additional object classes are, for example, MONEY, TIME, and PERCENT, which are not present in the example text. The NER tags, when applicable, can be used as features for ML model training for better model performance. NER tags can also help identify critical tokens on which such operations as lowercasing and stemming then can be avoided (e.g., Institute here refers to an organization rather than a verb). Such techniques make the features more discriminative.

Exhibit 19 Name Entity Recognition and Parts of Speech (POS) on Example Text

Token	NER Tag	POS Tag	POS Description
CFA	ORGANIZATION	NNP	Proper noun
Institute	ORGANIZATION	NNP	Proper noun
was		VBD	Verb, past tense
formed		VBN	Verb, past participle
in		IN	Preposition

(continued)

Exhibit 19 (Continued)

Token	NER Tag	POS Tag	POS Description
1947	DATE	CD	Cardinal number
and		CC	Coordinating conjunction
is		VBZ	Verb, 3rd person singular present
headquartered		VBN	Verb, past participle
in		IN	Preposition
Virginia	LOCATION	NNP	Proper noun

4 *Parts of speech (POS)*: Similar to NER, **parts of speech** uses language structure and dictionaries to tag every token in the text with a corresponding part of speech. Some common POS tags are noun, verb, adjective, and proper noun. Exhibit 19 shows the POS tags and descriptions of tags for the example text. POS tags can be used as features for ML model training and to identify the number of tokens that belong to each POS tag. If a given text contains many proper nouns, it means that it may be related to people and organizations and may be a business topic. POS tags can be useful for separating verbs and nouns for text analytics. For example, the word "market" can be a verb when used as "to market …" or noun when used as "in the market." Differentiating such tokens can help further clarify the meaning of the text. The use of "market" as a verb could indicate that the text relates to the topic of marketing and might discuss marketing a product or service. The use of "market" as a noun could suggest that the text relates to a physical or stock market and might discuss stock trading. Also for POS tagging, such compound nouns as "CFA Institute" can be treated as a single token. POS tagging can be performed using libraries or packages in programming languages.

In addition, many more creative techniques convey text information in a structured way to the ML training process. The goal of feature engineering is to maintain the semantic essence of the text while simplifying and converting it into structured data for ML.

EXAMPLE 4

Data Exploration

Paul Wang's analytics team at LendALot Corporation has completed its initial data preparation and wrangling related to their creditworthiness classification ML model building efforts. As a next step, Wang has asked one of the team members, Eric Kim, to examine the available structured data sources to see what types of exploratory data analysis might make sense. Kim has been tasked with reporting to the team on high-level patterns and trends in the data and which variables seem interesting. Greater situational awareness about the data can inform the team's decisions regarding model training and whether (and how) to incorporate textual big data in conjunction with the structured data inputs. Use the following sample of columns and rows Kim pulled for manual examination to answer the next questions.

1	ID	Loan Outcome	Income (USD)	Loan Amount (USD)	Credit Score	Loan Type	Free Responses to "Explain Credit Score" (excerpts from full text)
2	1	No Default	34,000	10,000	685	Mortgage	I am embarrassed that my score is below 700, but it was due to mitigating circumstances. I have developed a plan to improve my score.
3	2	No Default	63,050	49,000	770	Student Loan	I have a good credit score and am constantly looking to further improve it…
4	3	Defaulted	20,565	35,000	730	Student Loan	I think I have great credit. I don't think there are any issues. Having to provide a written response to these questions is kind of annoying…
5	4	No Default	50,021	10,000	664	Mortgage	I have a decent credit score. I regret not being as responsible in the past but feel I have worked hard to improve my score recently…
6	5	Defaulted	100,350	129,000	705	Car Loan	Honestly, my score probably would have been higher if I had worked harder. But it is probably good enough…
7	6	No Default	800,000	300,000	800	Boat Loan	I have worked hard to maintain a good credit rating. I am very responsible. I maintain a payment schedule and always stick to the payment plan…

1. Evaluate whether data visualization techniques, such as histograms, box plots, and scatterplots, could be relevant to exploratory data analysis.
2. State one visualization technique that could be used in relation to the free responses.
3. Describe how ranking methods can be used to select potentially interesting features to report back to the team.
4. State an example of a bigram from the free response texts that could be used to discriminate among loan outcomes.

Solution to 1:

The data provided include structured features (ID, Loan Outcome, Income, Loan Amount, Credit Score) and unstructured data. Histograms, box plots, and scatterplots are relevant visualization methods for structured data features. Histograms and box plots could be used by Kim to see how income, loan amount, and credit score are distributed. Moreover, these visualizations can be performed across all historical borrowing instances in the dataset as well as within the sets of defaulted loans versus non-defaulted loans. Scatterplots of income versus loan amount, income versus credit score, and loan amount versus credit score, both overall and within defaulted and non-defaulted datasets, can shed light on relationships between potentially important continuous variables.

Solution to 2:

For the text in the free response field, word clouds offer an appropriate starting point for exploratory analysis. A word cloud can enable a quick glimpse into the most frequently occurring words (i.e., term frequency). While some obvious words (e.g., "credit" and "score") may be valuable, other frequently occurring words (e.g., "worked," "hard," "probably," "embarrassed," "regret," "good," "decent," and "great") might have potential use for creditworthiness prediction.

Solution to 3:

Kim can use feature selection methods to rank all features. Since the target variable of interest (loan outcome) is discrete in this case, such techniques as chi-square and information gain would be well suited. These are univariate techniques that can score feature variables individually. In addition to the structured features, these univariate ranking methods can also be applied to word count-related features, such as term frequency and document frequency, that are derived from the text using frequently occurring words. Such frequently occurring words (e.g., "worked" and "hard") can be identified from the word cloud.

Solution to 4:

The bigrams "credit_score" and "worked_hard" from the text in the free response section may have potential to discriminate among loan outcomes.

EXAMPLE 5

Textual Feature Representations for ML Model Building

Having completed their exploration of the data, Paul Wang's analytics team at LendALot Corporation recognizes the importance of incorporating features derived from text data in their ML models for classifying creditworthiness. Wang has asked his colleagues, Lynn Lee and Eric Kim, to propose textual feature representations that might be well suited to constructing features for their task. As a starting point, Lee and Kim review the following sample of data:

1	ID	Loan Outcome	Income (USD)	Loan Amount (USD)	Credit Score	Loan Type	Free Responses to "Explain Credit Score" (excerpts from full text)
2	1	No Default	34,000	10,000	685	Mortgage	I am embarrassed that my score is below 700, but it was due to mitigating circumstances. I have developed a plan to improve my score.
3	2	No Default	63,050	49,000	770	Student Loan	I have a good credit score and am constantly looking to further improve it…
4	3	Defaulted	20,565	35,000	730	Student Loan	I think I have great credit. I don't think there are any issues. Having to provide a written response to these questions is kind of annoying…
5	4	No Default	50,021	10,000	664	Mortgage	I have a decent credit score. I regret not being as responsible in the past but feel I have worked hard to improve my score recently…
6	5	Defaulted	100,350	129,000	705	Car Loan	Honestly, my score probably would have been higher if I had worked harder. But it is probably good enough…
7	6	No Default	800,000	300,000	800	Boat Loan	I have worked hard to maintain a good credit rating. I am very responsible. I maintain a payment schedule and always stick to the payment plan…

Based on the information given, address the following questions.

1 Describe three textual feature representations that Lee and Kim should consider for their text data.

2 Describe a rationale for adopting each of the three textual feature representations identified in Question 1.

Solution 1:

Lee and Kim should consider bag-of-words (BOW), n-grams, and parts-of-speech (POS) as key textual feature representations for their text data. Conversely, name entity recognition (NER) might not be as applicable in this context because the data on prospective borrowers does not include any explicit references to people, locations, dates, or organizations.

Solution 2:

All three textual feature representations have the potential to add value.

Bag-of-words (BOW) is typically applicable in most contexts involving text features derived from languages where token boundaries are explicitly present (e.g., English) or can be inferred through processing (e.g., a different language, such as Spanish). BOW is generally the best starting point for most projects exploring text feature representations.

N-grams, representations of word or token sequences, are also applicable. N-grams can offer invaluable contextual information that can complement and enrich a BOW. In this specific credit-worthiness context, we examine the BOW token "worked." It appears three times (rows 5–7), twice in no-default loan texts and once in a defaulted loan text. This finding suggests that "worked" is being used to refer to the borrower's work ethic and may be a good predictor of credit worthiness. Digging deeper and looking at several trigrams (i.e., three-token sequences) involving "worked," we see that "have_worked_hard" appears in the two no-default loan related texts (referring to borrower accomplishments and plans) and "had_worked_harder" appears in the defaulted loan text (referring to what could have been done). This example illustrates how n-grams can provide richer contextualization capabilities for the creditworthiness prediction ML models.

Parts-of-speech tags can add value because they identify the composition of the texts. For example, POS provides information on whether the prospective borrowers are including many action words (verbs) or descriptors (adjectives) and whether this is being done differently in instances of no-default versus instances of defaulted loans.

MODEL TRAINING, STRUCTURED AND UNSTRUCTURED DATA, METHOD SELECTION 7

d describe objectives, steps, and techniques in model training;

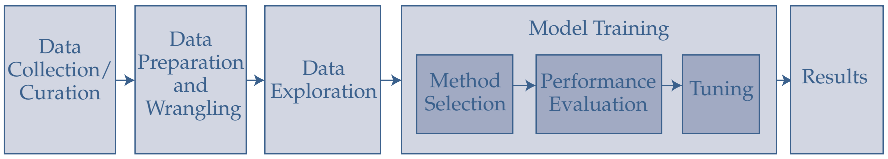
Machine learning model training is a systematic, iterative, and recursive process. The number of iterations required to reach optimum results depends on:

- the nature of the problem and input data and
- the level of model performance needed for practical application.

Machine learning models combine multiple principles and operations to provide predictions. As seen in the last two sections, typical ML model building requires data preparation and wrangling (cleansing and preprocessing) and data exploration (exploratory data analysis as well as feature selection and engineering). In addition, domain knowledge related to the nature of the data is required for good model building and training. For instance, knowledge of investment management and securities trading is important when using financial data to train a model for predicting costs of trading stocks. It is crucial for ML engineers and domain experts to work together in building and training robust ML models.

The three tasks of ML model training are method selection, performance evaluation, and tuning. Exhibit 20 outlines model training and its three component tasks. Method selection is the art and science of deciding which ML method(s) to incorporate and is guided by such considerations as the classification task, type of data, and size of data. Performance evaluation entails using an array of complementary techniques and measures to quantify and understand a model's performance. Tuning is the process of undertaking decisions and actions to improve model performance. These steps may be repeated multiple times until the desired level of ML model performance is attained. Although no standard rulebook for training an ML model exists, having a fundamental understanding of domain-specific training data and ML algorithm principles plays a vital role in good model training.

Exhibit 20 Model Training Stage



Before training a model, it is important to state the problem, define objectives, identify useful data points, and conceptualize the model. Conceptualization is like a blueprint on a drawing board, a modifiable plan that is necessary to initiate the model training process. Because modeling is an iterative process, many changes and refinements will be made to the model plan as the process evolves.

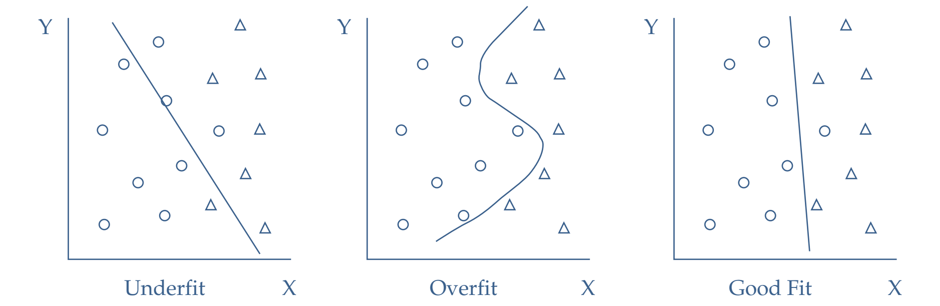
7.1 Structured and Unstructured Data

The ML model training process for structured and unstructured data is typically the same. Most ML models are intended to train on structured data, so unstructured data in the data preparation stage are processed and organized into a structured format. The systematic processing of unstructured text data so that they can be structured in the form of a data matrix has been previously covered. Similarly, other forms of unstructured data can also be prepared and formed into data matrixes or tables for ML training.

The fundamental idea of ML model training is fitting a system of rules on a training dataset to reveal a pattern in the data. In other words, fitting describes the degree to which (or how well) an ML model can be generalized to new data. A good model fit results in good model performance and can be validated using new data outside of the training dataset (i.e., out-of-sample). Exhibit 21 shows model decision boundaries in three possible model fitting scenarios for a classification task comprising two different classes of data (i.e., circles and triangles). The model on the left is underfit; it does not fit the training data well enough since it results in four misclassification errors (three circles and one triangle). Although the center model that generates the

"S"-shaped line has the best accuracy (no errors) on the training data, it is overfit (i.e., fits the training data too well) and thus unlikely to perform well on future test cases. The model on the right (with one classification error, a circle) is a model with good fit (i.e., it fits the training data well but not so well that it cannot be generalized to out-of-sample data).

Exhibit 21 Model Fitting Scenarios: Underfit, Overfit, and Good Fit



Model fitting errors are caused by several factors—the main ones being dataset size and number of features in the dataset.

- *Dataset Size*: Small datasets can lead to underfitting of the model since small datasets often are not sufficient to expose patterns in the data. Restricted by a small dataset, an ML model may not recognize important patterns.
- *Number of Features*: A dataset with a small number of features can lead to underfitting, and a dataset with a large number of features can lead to overfitting. As with small dataset size, a small number of features may not carry all the characteristics that explain relationships between the target variable and the features. Conversely, a large number of features can complicate the model and potentially distort patterns in the data due to low degrees of freedom, causing overfitting. Therefore, appropriate feature selection using the types of techniques described earlier (e.g., chi-square, mutual information) is a key factor in minimizing such model overfitting.

Feature engineering tends to prevent underfitting in the training of the model. New features, when engineered properly, can elevate the underlying data points that better explain the interactions of features. Thus, feature engineering can be critical to overcome underfitting. Method-related factors that affect model fitting are explained shortly under tuning.

Method Selection

ML model training is a craft (part art and part science); it has no strict guidelines. Selecting and applying a method or an algorithm is the first step of the training process. Method selection is governed by the following factors:

1. *Supervised or unsupervised learning.* The data for training and testing supervised ML models contain **ground truth**, the known outcome (i.e., target variable) of each observation in these datasets. Unsupervised ML modeling is relatively challenging because of the absence of ground truth (i.e., no target variable). Supervised models bring a structure that may or may not be supported by the data. Unsupervised models bring no structure beyond that which arises from the given data. For supervised learning (with labeled training data), typical methods of choice are regression, ensemble trees, support vector machines (SVMs), and neural networks (NNs). Supervised learning would be

used, for example, for default prediction based on high-yield corporate bond issuer data. For unsupervised learning, common methods are dimensionality reduction, clustering, and anomaly detection. Unsupervised learning, for example, would be used for clustering financial institutions into different groups based on their financial attributes.

2 *Type of data*. For numerical data (e.g., predicting stock prices using historical stock market values), classification and regression tree (CART) methods may be suitable. For text data (for example, predicting the topic of a financial news article by reading the headline of the article), such methods as generalized linear models (GLMs) and SVMs are commonly used. For image data (e.g., identifying objects in a satellite image, such as tanker ships moving in and out of port), NNs and deep learning methods tend to perform better than others. For speech data (e.g., predicting financial sentiment from quarterly earnings' conference call recordings), deep learning methods can offer promising results.

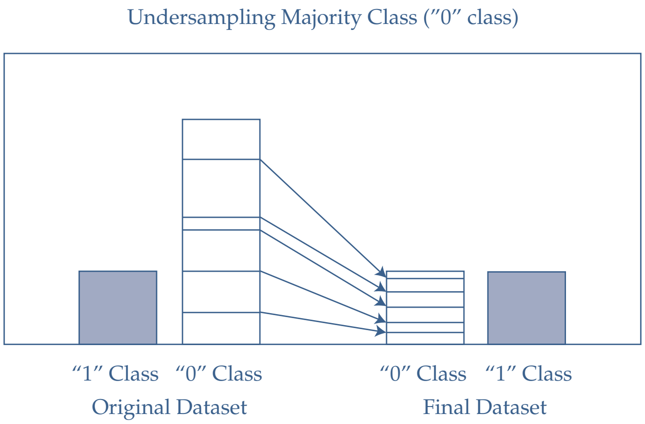
3 *Size of data.* A typical dataset has two basic characteristics: number of instances (i.e., observations) and number of features. The combination of these two characteristics can govern which method is most suitable for model training. For instance, SVMs have been found to work well on "wider" datasets with 10,000 to 100,000 features and with fewer instances. Conversely, NNs often work better on "longer" datasets, where the number of instances is much larger than the number of features.

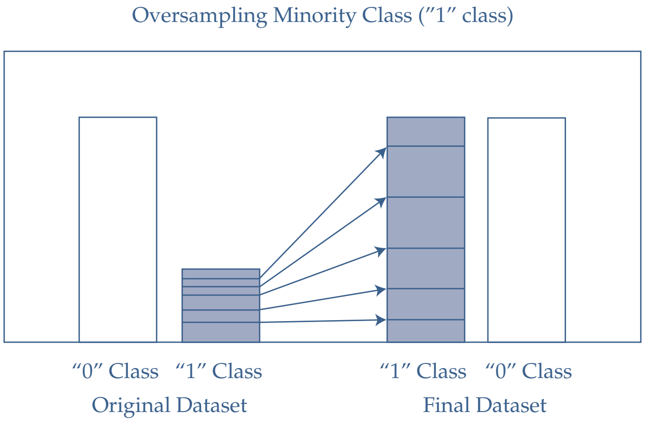
Once a method is selected, certain method-related decisions (e.g., on hyperparameters) need to be made. These decisions include the number of hidden layers in a neural network and the number of trees in ensemble methods (discussed later in the sub-section on tuning). In practice, datasets can be a combination of numerical and text data. To deal with mixed data, the results from more than one method can be combined. Sometimes, the predictions from one method can be used as predictors (features) by another. For example, unstructured financial text data can be used with logistic regression to classify stock sentiment as either positive or negative. Then, this sentiment classification cam be used as a predictor in a larger model, say CART, that also uses structured financial data as predictors for the purpose of stock selection. Finally, more than one method can be used and the results combined with quantitative or subjective weighing to exploit the advantages of each method.

Before model training begins, in the case of supervised learning the master dataset is split into three subsets used for model training and testing purposes. The first subset, a training set used to train the model, should constitute approximately 60% of the master dataset. The second subset, a cross-validation set (or validation set) used to tune and validate the model, should constitute approximately 20% of the master dataset. The third subset is a test set for testing the model and uses the remaining data. The data are split using a random sampling technique, such as the k-fold method. A commonly recommended split ratio is 60:20:20, as detailed above; however, the split percentages can vary. For unsupervised learning, no splitting is needed due to the absence of labeled training data.

Class imbalance, where the number of instances for a particular class is significantly larger than for other classes, may be a problem for data used in supervised learning because the ML classification method's objective is to train a high-accuracy model. In a high-yield bond default prediction example, say for corporate issuers in the BB+/Ba1 to B+/B1 credit quality range, issuers who defaulted (positive or "1" class) would be very few compared to issuers who did not default (negative or "0" class). Hence, on such training data, a naive model that simply assumes no corporate issuer will default may achieve good accuracy—albeit with all default cases misclassified. Balancing the training data can help alleviate such problems. In cases of unbalanced data, the "0" class (majority class) can be randomly undersampled or the "1" class (minority class) randomly oversampled. The random sampling can be done with or without replacement because they both work the same in general probability theory. Exhibit 22 depicts the idea of undersampling of the majority class and oversampling of the minority class. In practice, the choice of whether to undersample or oversample depends on the specific problem context. Advanced techniques can also reproduce synthetic observations from the existing data, and the new observations can be added to the dataset to balance the minority class.

Exhibit 22 Undersampling and Oversampling





8 PERFORMANCE EVALUATION

d describe objectives, steps, and techniques in model training;

It is important to measure the model training performance or goodness of fit for validation of the model. We shall cover several techniques to measure model performance that are well suited specifically for binary classification models.

1 *Error analysis.* For classification problems, error analysis involves computing four basic evaluation metrics: true positive (TP), false positive (FP), true negative (TN), and false negative (FN) metrics. FP is also called a Type I error, and FN is also called a Type II error. Exhibit 23 shows a **confusion matrix**, a grid that is used to summarize values of these four metrics.

Exhibit 23 Confusion Matrix for Error Analysis

		Actual Training Labels	
		Class "1"	Class "0"
Predicted Results	Class "1"	True Positives (TP)	False Positives (FP) Type I Error
	Class "0"	False Negatives (FN) Type II Error	True Negatives (TN)

Additional metrics, such as precision and recall, can be computed. Assume in the following explanation that Class "0" is "not defective" and Class "1" is "defective." **Precision** is the ratio of correctly predicted positive classes to all predicted positive classes. Precision is useful in situations where the cost of FP, or Type I error, is high—or example, when an expensive product fails quality inspection (predicted Class "1") and is scrapped, but it is actually perfectly good (actual Class "0"). **Recall** (also known as *sensitivity*) is the ratio of correctly predicted positive classes to all actual positive classes. Recall is useful in situations where the cost of FN or Type II error is high—for example, when an expensive product passes quality inspection (predicted Class "0") and is sent to the valued customer, but it is actually quite defective (actual Class "1"). The formulas for precision and recall are:

$$\text{Precision (P)} = \text{TP}/(\text{TP} + \text{FP}). \quad (3)$$

$$\text{Recall (R)} = \text{TP}/(\text{TP} + \text{FN}). \quad (4)$$

Trading off precision and recall is subject to business decisions and model application. Therefore, additional evaluation metrics that provide the overall performance of the model are generally used. The two overall performance metrics are accuracy and F1 score. **Accuracy** is the percentage of correctly predicted classes out of total predictions. **F1 score** is the harmonic mean of precision and recall. F1 score is more appropriate (than accuracy) when unequal class distribution is in the dataset and it is necessary to measure the equilibrium of precision and recall. High scores on both of these metrics suggest good model performance. The formulas for accuracy and F1 score are as follows:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}). \quad (5)$$

$$\text{F1 score} = (2 * \text{P} * \text{R})/(\text{P} + \text{R}). \quad (6)$$

Exhibit 24 illustrates computations of model evaluation metrics and performance scores on a sample dataset.

Exhibit 24 Performance Metrics and Scores Computation

Sample Dataset with Classification Results

Observation	Actual Training Labels	Predicted Results	Classification
1	1	1	TP
2	0	0	TN
3	1	1	TP
4	1	0	FN
5	1	1	TP
6	1	0	FN
7	0	0	TN
8	0	0	TN
9	0	0	TN
10	0	1	FP

Confusion Matrix

		Actual Training Labels	
		Class "1"	Class "0"
Predicted Results	Class "1"	3 (TP)	1 (FP)
	Class "0"	2 (FN)	4 (TN)

Performance Metrics

TP = 3, FP = 1, FN = 2, TN = 4

P = 3 / (3+1) = 0.75

R = 3 / (3+2) = 0.60

F1 Score = (2 × 0.75 × 0.60) / (0.75 + 0.60) = 0.67

Accuracy = (3 + 4) / (3 + 1 + 4 + 2) = 0.70

In Exhibit 24, if all "1" classes were predicted correctly (no FPs), the precision would have been equal to 1. If all "0" classes were predicted correctly (no FNs), the recall would have been equal to 1. Thus, the resulting F1 score would have been equal to 1. The precision of 0.75 and recall of 0.60 indicate that the model is better at minimizing FPs than FNs. To find the equilibrium between precision and recall, F1 score is calculated, which is equal to 0.67. The F1 score is closer to the smaller value among both precision and recall, giving the model a more appropriate score rather than just an arithmetic mean. Accuracy, the percentage of correct predictions (for both classes) made by the model, is equal to 0.70. Accuracy would be equal to 1 if all predictions were correct. As the number of "1" and "0" classes is equal in the dataset (i.e., a balanced dataset), accuracy can be considered an appropriate performance measure in this case. If the number of classes in a dataset is unequal; however, then F1 score should be used as the overall performance measure for the model.

2 *Receiver Operating Characteristic (ROC).* This technique for assessing model performance involves the plot of a curve showing the trade-off between the false positive rate (x-axis) and true positive rate (y-axis) for various cutoff points—for example, for the predicted probability (p) in a logistic regression. The formulas for false positive rate and true positive rate (note that true positive rate is the same as recall) are:

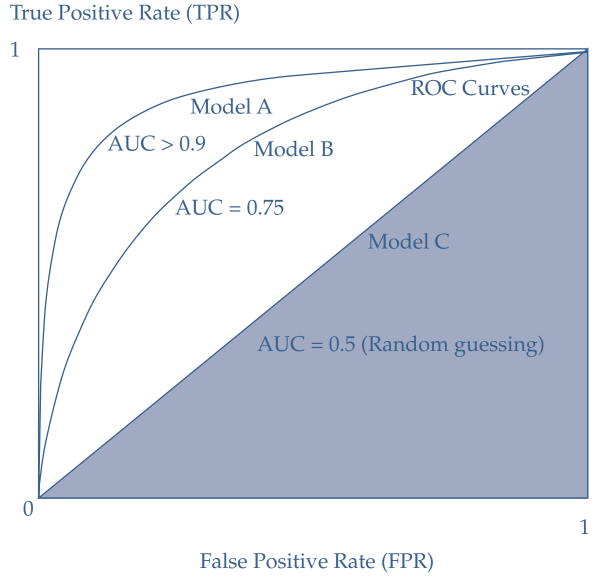
False positive rate (FPR) = FP/(TN + FP) and **(7)**

True positive rate (TPR) = TP/(TP + FN). **(8)**

If p from a logistic regression model for a given observation is greater than the cutoff point (or threshold), then the observation is classified as class = 1. Otherwise, the observation will be classified as class = 0.

The shape of the ROC curve provides insight into the model's performance. A more convex curve indicates better model performance. Area under the curve (AUC) is the metric that measures the area under the ROC curve. An AUC close to 1.0 indicates near perfect prediction, while an AUC of 0.5 signifies random guessing. Exhibit 25 displays three ROC curves and indicates their respective AUC values. It is clear from observing the shapes of the ROC curves and their AUCs that Model A—with the most convex ROC curve with AUC of more than 0.9 (or 90%)—is the best performing among the three models.

Exhibit 25 ROC Curves and AUCs



3 *Root Mean Squared Error (RMSE).* This measure is appropriate for continuous data prediction and is mostly used for regression methods. It is a single metric that captures all the prediction errors in the data (n). The root mean squared error is computed by finding the square root of the mean of the squared differences between the actual values and the model's predicted values (error). A small RMSE indicates potentially better model performance. The formula for RMSE is:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\text{Predicted}_i - \text{Actual}_i)^2}{n}} \quad (9)$$

9 TUNING

d describe objectives, steps, and techniques in model training;

Once the model is evaluated, certain decisions and actions must be taken based on the findings to improve the performance of the model. If the prediction error on the training set is high, the model is underfitting. If the prediction error on the cross-validation (CV) set is significantly higher than on the training set, the model is overfitting. Model fitting has two types of error: bias and variance. Bias error is associated with underfitting, and variance error is associated with overfitting. Bias error is high when a model is overly simplified and does not sufficiently learn from the patterns in the training data. Variance error is high when the model is overly complicated and memorizes the training data so much that it will likely perform poorly on new data. It is not possible to completely eliminate both types of errors. However, both errors can be minimized so the total aggregate error (bias error + variance error) is at a minimum. The bias–variance trade-off is critical to finding an optimum balance where a model neither underfits nor overfits.

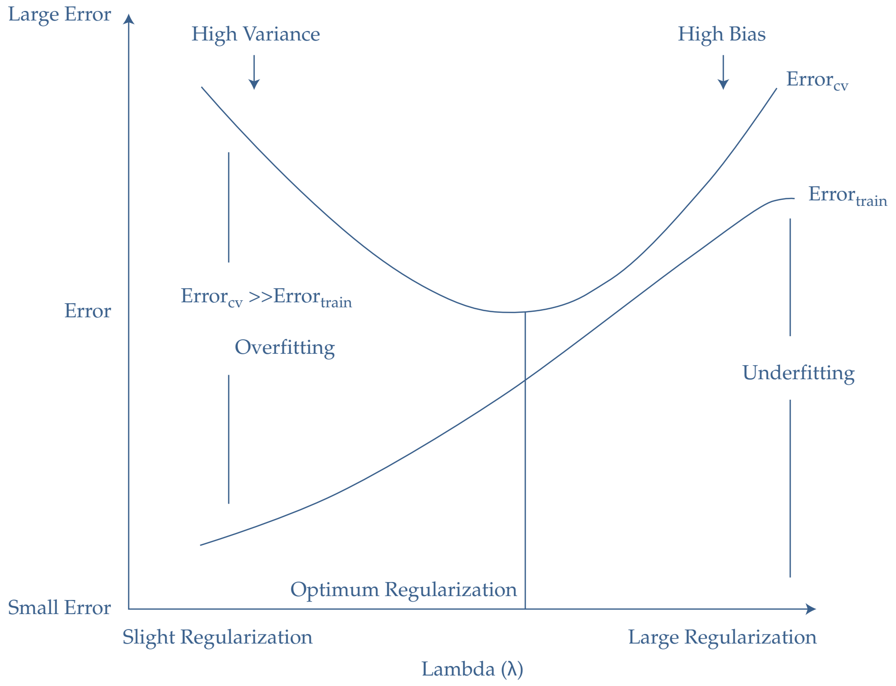
1 *Parameters* are critical for a model and are dependent on the training data. Parameters are learned from the training data as part of the training process by an optimization technique. Examples of parameters include coefficients in regression, weights in NN, and support vectors in SVM.

2 *Hyperparameters* are used for estimating model parameters and are not dependent on the training data. Examples of hyperparameters include the regularization term (λ) in supervised models, activation function and number of hidden layers in NN, number of trees and tree depth in ensemble methods, *k* in k-nearest neighbor classification and k-means clustering, and p-threshold in logistic regression. Hyperparameters are manually set and tuned.

For example, if a researcher is using a logistic regression model to classify sentences from financial statements into positive or negative stock sentiment, the initial cutoff point for the trained model might be a p-threshold of 0.50 (50%). Therefore, any sentence for which the model produces a probability >50% is classified as having positive sentiment. The researcher can create a confusion matrix from the classification results (of running the CV dataset) to determine such model performance metrics as accuracy and F1 score. Next, the researcher can vary the logistic regression's p-threshold—say to 0.55 (55%), 0.60 (60%), or even 0.65 (65%)—and then re-run the CV set, create new confusion matrixes from the new classification results, and compare accuracy and F1 scores. Ultimately, the researcher would select the logistic regression model with a p-threshold value that produces classification results generating the highest accuracy and F1 scores. Note that the process just outlined will be demonstrated in Sections 10–13.

There is no general formula to estimate hyperparameters. Thus, tuning heuristics and such techniques as grid search are used to obtain the optimum values of hyperparameters. **Grid search** is a method of systematically training an ML model by using various combinations of hyperparameter values, cross validating each model, and determining which combination of hyperparameter values ensures the best model performance. The model is trained using different combinations of hyperparameter values until the optimum set of values are found. Optimum values must result in similar performance of the model on training and CV datasets, meaning that the training error and CV error are close. This ensures that the model can be generalized to test data or to new data and thus is less likely to overfit. The plot of training errors for each value of a hyperparameter (i.e., changing model complexity) is called a fitting curve. Fitting curves provide visual insight on the model's performance (for the given hyperparameter and level of model complexity) on the training and CV datasets and are visually helpful to tune hyperparameters. Exhibit 26 shows the bias–variance error trade-off by plotting a generic fitting curve for a regularization hyperparameter (λ).

Exhibit 26 Fitting Curve for Regularization Hyperparameter (λ)



Slight regularization lightly penalizes model complexity, thereby allowing most or all of the features to be included in the model and thus potentially enabling the model to "memorize" the data. Typically with no or slight regularization, the prediction error on the training dataset is small while the prediction error on the CV dataset is significantly larger. This difference in error is variance. High variance error, which typically results from too many features and model complexity, results in model overfitting. When high variance error and low bias error exist, the model performs well on the training dataset but generates many FP and FN errors on the CV dataset; in other words, the model is overfitted and does not generalize to new data well.

Large regularization excessively penalizes model complexity, thereby allowing too few of the features to be included in the model and causing the model to learn less from the data. The model may lack the necessary predictor variables and complexity needed to discern underlying patterns in the data. Typically with large regularization, the prediction errors on the training and CV datasets are both large. Large prediction errors on the training dataset indicate high bias, and high bias error results from model underfitting. When high bias error exists, the model does not perform well on either training or CV datasets because it is typically lacking important predictor variables.

Optimum regularization minimizes both variance and bias errors in a balanced fashion. It penalizes model complexity just enough so that only the most important features are included in the model. This process prevents the model from memorizing the data while enabling the model to learn enough from the data to distinguish important patterns. This results in prediction errors in both training and CV datasets that are similar and also minimal. The range of optimum regularization values can be found heuristically using such techniques as grid search.

If high bias or variance exists after the tuning of hyperparameters, either a larger number of training examples (instances) may be needed or the number of features included in the model may need to be decreased (in the case of high variance) or increased (in the case of high bias). The model then needs to be re-trained and re-tuned using the new training dataset. In the case of a complex model, where a large model is comprised of sub-model(s), ceiling analysis can be performed. **Ceiling analysis** is a systematic process of evaluating different components in the pipeline of model building. It helps to understand what part of the pipeline can potentially improve in performance by further tuning. For example, a stock market prediction model needs historical data from the stock market and perhaps news articles related to the stocks. The sub-model will extract relevant information from the news articles or classify the sentiment of the news articles. The results of the sub-model will feed into the larger model as features. Thus, the performance of the larger model depends on performance of the sub-model(s). Ceiling analysis can help determine which sub-model needs to be tuned to improve the overall accuracy of the larger model.

10 FINANCIAL FORECASTING PROJECT - CLASSIFYING AND PREDICTING SENTIMENT FOR STOCKS, AND TEXT CURATION, PREPARATION AND WRANGLING

e describe preparing, wrangling, and exploring text-based data for financial forecasting;

Robo-readers are automated programs used to analyze large quantities of text, including news articles and social media. In the financial services space, robo-readers are being used by investors to examine how views expressed in text relate to future company performance. One important dimension that robo-readers look to analyze is sentiment polarity—which means how positive, negative, or neutral a particular phrase or statement is regarding a "target." For example, in the statement "XYZ Corporation is doing terrific things with its new product innovation," positive sentiment (i.e., the polarity) is being expressed regarding XYZ Corporation (i.e., the target of the sentiment). Such sentiment can provide invaluable predictive power, both alone and when coupled with structured financial data, for predicting stock price movements for individual firms and for portfolios of companies.

To provide a practical application, we use a financial forecasting project to examine how effectively sentiment—expressed in English news articles on LexisNexis (a searchable database of news articles) related to all companies listed on the NASDAQ OMX Helsinki (Finland)—can be classified. To accomplish this task, we followed the text ML model building steps presented in Sections 2–9 of this reading.

10.1 Text Curation, Preparation, and Wrangling

Text Curation

The text data used in this financial forecasting project are a collection of English language sentences from financial and economic news sources. The text data are acquired from the Financial Phrase Bank located on the website Researchgate.net.[2] The compressed folder contains six text files. The first two files are license and readme files. The other four files contain the text data. The data are presented in a text document format (.txt), which can be opened and viewed using any text editor. Note that this is cross-sectional data (not time series data).

2 https://www.researchgate.net/publication/251231364_FinancialPhraseBank-v10.

A total of 14,780 sentences are in the four files. The sentiment of each sentence has already been labeled with one of three sentiment classes: positive, neutral, or negative. The sentiment classes are provided from an investor's perspective and may be useful for predicting whether a sentence may have a corresponding positive, neutral, or negative influence on the respective company's stock price.

This project uses sentences from two of the text files (Sentences_AllAgree and Sentences_75Agree), labeled as either in the positive or negative sentiment class, for a total of 2,180 sentences. There are 1,457 positive sentiment class sentences and 723 negative sentiment class sentences. A supervised ML model is trained, validated, and tested using these data. The final ML model can be used to predict the sentiment classes of sentences present in similar financial news statements. Exhibit 27 shows a sample of 10 rows of raw text from the Sentences_AllAgree text file. Note the sentiment annotations at the end of each sentence with prefix character "@."

Exhibit 27 Ten Sample Sentences and Sentiment from Raw Text File (Sentences_AllAgree.txt)

```
Profit before taxes amounted to EUR 56.5 mn , down from EUR 232.9 mn a year ago .@negative
Profit before taxes decreased by 9 % to EUR 187.8 mn in the first nine months of 2008 , compared to EUR 207.1 mn a year earlier .@negative
Profit before taxes decreased to EUR 31.6 mn from EUR 50.0 mn the year before .@negative
Profit before taxes was EUR 4.0 mn , down from EUR 4.9 mn .@negative
The company 's profit before taxes fell to EUR 21.1 mn in the third quarter of 2008 , compared to EUR 35.8 mn in the corresponding period in 2007 .@negative
In August-October 2010 , the company 's result before taxes totalled EUR 9.6 mn , up from EUR 0.5 mn in the corresponding period in 2009 .@positive
Finnish Bore that is owned by the Rettig family has grown recently through the acquisition of smaller shipping companies .@positive
The plan is estimated to generate some EUR 5 million ( USD 6.5 m ) in cost savings on an annual basis .@positive
Finnish pharmaceuticals company Orion reports profit before taxes of EUR 70.0 mn in the third quarter of 2010 , up from EUR 54.9 mn in the corresponding period in 2009 .@positive
Finnish Sampo Bank , of Danish Danske Bank group , reports profit before taxes of EUR 152.3 mn in 2010 , up from EUR 32.7 mn in 2009 .@positive
```

Text Preparation (Cleansing)

The raw text data (i.e., sentences) are initially organized into a data table. The data table contains two columns: The first column (sentence) is for the text, and the second column (sentiment) is for the corresponding sentiment class. The separator character, which is "@" in this case, is used to split the data into text and sentiment class columns. A collection of text data in any form, including list, matrix, or data table forms, is called a **corpus**. Exhibit 28 shows a sample of 10 sentences from the data table corpus.

Exhibit 28 Ten Sample Rows of the Data Table (Corpus)

Sentence	Sentiment
Profit before taxes amounted to EUR 56.5 mn , down from EUR 232.9 mn a year ago .	negative
Profit before taxes decreased by 9 % to EUR 187.8 mn in the first nine months of 2008 , compared to EUR 207.1 mn a year earlier .	negative
Profit before taxes decreased to EUR 31.6 mn from EUR 50.0 mn the year before .	negative
Profit before taxes was EUR 4.0 mn , down from EUR 4.9 mn .	negative
The company 's profit before taxes fell to EUR 21.1 mn in the third quarter of 2008 , compared to EUR 35.8 mn in the corresponding period in 2007 .	negative
In August-October 2010 , the company 's result before taxes totalled EUR 9.6 mn , up from EUR 0.5 mn in the corresponding period in 2009 .	positive
Finnish Bore that is owned by the Rettig family has grown recently through the acquisition of smaller shipping companies .	positive
The plan is estimated to generate some EUR 5 million (USD 6.5 m) in cost savings on an annual basis .	positive

Exhibit 28 (Continued)

Sentence	Sentiment
Finnish pharmaceuticals company Orion reports profit before taxes of EUR 70.0 mn in the third quarter of 2010 , up from EUR 54.9 mn in the corresponding period in 2009 .	positive
Finnish Sampo Bank , of Danish Danske Bank group , reports profit before taxes of EUR 152.3 mn in 2010 , up from EUR 32.7 mn in 2009 .	positive

The raw text contains punctuations, numbers, and white spaces that may not be necessary for model training. Text cleansing involves removing, or incorporating appropriate substitutions for, potentially extraneous information present in the text. Operations to remove html tags are unnecessary because none are present in the text

Punctuations: Before stripping out punctuations, percentage and dollar symbols are substituted with word annotations to retain their essence in the financial texts. Such word annotation substitutions convey that percentage and currency-related tokens were involved in the text. As the sentences have already been identified within and extracted from the source text, punctuation helpful for identifying discrete sentences—such as periods, semi-colons, and commas—are removed. Some special characters, such as "+" and "©," are also removed. It is a good practice to implement word annotation substitutions before removing the rest of the punctuations.

Numbers: Numerical values of numbers in the text have no significant utility for sentiment prediction in this project because sentiment primarily depends on the words in a sentence. Here is an example sentence: "*Ragutis, which is based in Lithuania's second-largest city, Kaunas, boosted its sales last year 22.3 percent to 36.4 million litas*." The word "boosted" implies that there was growth in sales, so analysis of this sentiment does not need to rely on interpretation of numerical text data. Sentiment analysis typically does not involve extracting, interpreting, and calculating relevant numbers but instead seeks to understand the context in which the numbers are used. Other commonly occurring numbers are dates and years, which are also not required to predict sentence sentiment. Thus, all numbers present in the text are removed for this financial sentiment project. However, prior to removing numbers, abbreviations representing orders of magnitude, such as million (commonly represented by "m," "mln," or "mn"), billion, or trillion, are replaced with the complete word. Retaining these orders of magnitude-identifying words in the text preserves the original text meaning and can be useful in predicting sentence sentiment.

Whitespaces: White spaces are present in the raw text. Additional white spaces occur after performing the above operations to remove extraneous characters. The white spaces must be removed to keep the text intact. Exhibit 29 shows the sample text after cleansing. The cleansed text is free of punctuations and numbers, with useful substitutions.

Exhibit 29 Ten Sample Rows After Cleansing Process

Sentence	Sentiment
Profit before taxes amounted to EUR million down from EUR million a year ago	negative
Profit before taxes decreased by percentSign to EUR million in the first nine months of compared to EUR million a year earlier	negative
Profit before taxes decreased to EUR million from EUR million the year before	negative
Profit before taxes was EUR million down from EUR million	negative

(continued)

Exhibit 29 (Continued)

Sentence	Sentiment
The companys profit before taxes fell to EUR million in the third quarter of compared to EUR million in the corresponding period in	negative
In August October the companys result before taxes totalled EUR million up from EUR million in the corresponding period in	positive
Finnish Bore that is owned by the Rettig family has grown recently through the acquisition of smaller shipping companies	positive
The plan is estimated to generate some EUR million USD million in cost savings on an annual basis	positive
Finnish pharmaceuticals company Orion reports profit before taxes of EUR million in the third quarter of up from EUR million in the corresponding period in	positive
Finnish Sampo Bank of Danish Danske Bank group reports profit before taxes of EUR million in up from EUR million in	positive

Text Wrangling (Preprocessing)

The cleansed text needs to be normalized using the following normalization procedures:

1. *Lowercasing* of all text to consolidate duplicate words (example, "THE," "The," and "the").
2. *Stop words* are not removed because some stop words (e.g., not, more, very, and few) carry significant meaning in the financial texts that is useful for sentiment prediction. Some stop words, such as articles (a, an, the), may be removed. Nevertheless, to avoid confusion no words are removed at this point. This issue will be revisited during the data exploration stage, which will carefully examine the text using frequency analysis and find custom stop words (common words) for these particular text data.
3. *Stemming*, the converting of inflected forms of a word into its base word (stem), is performed on the text as it is simple to perform and is appropriate for training an ML model for sentiment prediction.

White spaces are stripped after performing these operations. As part of text normalization, different currency abbreviations, such as EUR and USD, can be converted into a single token, such as "currencysign." As we are dealing with financial domain text, the earlier substitution of dollarsign can be replaced with currencysign as well. This step will remove tokens that are different but redundant in nature while maintaining their meaning. Through careful examination of the text and use of domain knowledge, similar substitutions of redundant tokens can be performed. Exhibit 30 shows how the sample text appears after normalization.

Exhibit 30 Ten Sample Rows After Normalization Process

Sentence	Sentiment
profit befor tax amount to currencysign million down from currencysign million a year ago	negative
profit befor tax decreas by percentsign to currencysign million in the first nine month of compar to currencysign million a year earlier	negative
profit before tax decreas to currencysign million from currencysign million the year befor	negative
profit befor tax was currencysign million down from currencysign million	negative
the compani profit befor tax fell to currencysign million in the third quarter of compar to currencysign million in the correspond period in	negative

Exhibit 30 (Continued)

Sentence	Sentiment
in august octob the compani result befor tax total currencysign million up from currencysign million in the correspond period in	positive
finnish bore that is own by the rettig famili has grown recent through the acquisit of smaller shipping company	positive
the plan is estim to generat some currencysign million currencysign million in cost save on an annual basi	positive
finnish pharmaceut compani orion report profit befor tax of currencysign million in the third quarter of up from currencysign million in the correspond period in	positive
finnish sampo bank of danish danske bank group report profit befor tax of currencysign million in up from currencysign million in	positive

The normalized text is tokenized, resulting in 2,673 unique tokens. Altogether, these unique tokens comprise the bag-of-words (BOW) of the text corpus. Exhibit 31 shows a sample of 100 tokens from the BOW. This preliminary unigram BOW can be used to construct a document term matrix (DTM) for ML training.

Exhibit 31 One Hundred Sample Tokens from Preliminary Unigram BOW

"for"	"foundri"	"quarter"	"shop"	"net"	"share"	"to"
"currencysign"	"nokia"	"same"	"plan"	"year"	"sanyo"	"it"
"move"	"nokian"	"tax"	"earn"	"in"	"expect"	"by"
"percentsign"	"director"	"rose"	"dividned"	"total"	"megafon"	"talentum"
"report"	"as"	"chain"	"number"	"consolid"	"accord"	"compar"
"prior"	"last"	"machin"	"componenta"	"afx"	"doubl"	"higher"
"led"	"from"	"announc"	"a"	"with"	"while"	"g"
"handset"	"pre"	"fourth"	"loss"	"analyst"	"increas"	"said"
"board"	"oper"	"propos"	"repres"	"paid"	"finnish"	"base"
"user"	"retail"	"market"	"is"	"late"	"amount"	"estim"
"the"	"divis"	"of"	"helsinki"	"sale"	"close"	
"million"	"after"	"period"	"team"	"earlier"	"manufactur"	
"zero"	"tyre"	"profit"	"beat"	"third"	"dealer"	
"and"	"will"	"correspond"	"per"	"up"	"subscrib"	
"cloth"	"decemb"	"sepp"	"custom"	"reach"	"teliasonera"	

The final DTM for ML model training will be prepared after the data exploration stage. Data exploration may reveal unnecessary tokens or anomalies in the data. Any unnecessary tokens that are not informative must be removed, which will also impact the creation of n-grams. Thus, the final DTM must be made after further analyses and operations, such as exploratory data analysis and feature selection.

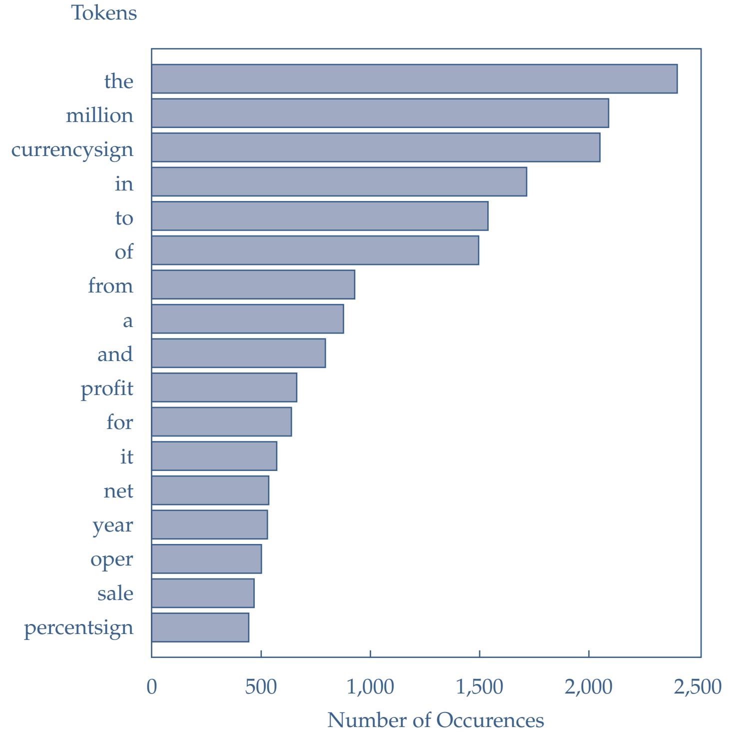
DATA EXPLORATION 11

e describe preparing, wrangling, and exploring text-based data for financial forecasting;

Exploratory Data Analysis

Exploratory data analysis (EDA) performed on text data provides insights on word distribution in the text. Word counts from all the sentences are computed. These word counts can be used to examine outlier tokens—words that are most commonly and least commonly present in the texts. The most frequent word occurrences in all sentences from the dataset are shown in Exhibit 32. These common words will be removed during the feature selection step. Notably, the tokens "million" and "currencysign" occur frequently due to the financial nature of the data.

Exhibit 32 Most Frequently Used Tokens in the Corpus



The most frequent word occurrences in the sentences in the negative sentiment and the positive sentiment classes are shown in Exhibit 33. The most commonly occurring words are similar for both sentiment classes, meaning that they are not useful in discriminating between the two sentiment classes. This finding demonstrates the utility of removing the most commonly used tokens from the BOW.

Exhibit 33 Most Frequently Used Tokens in Two Sentiment Classes of the Corpus

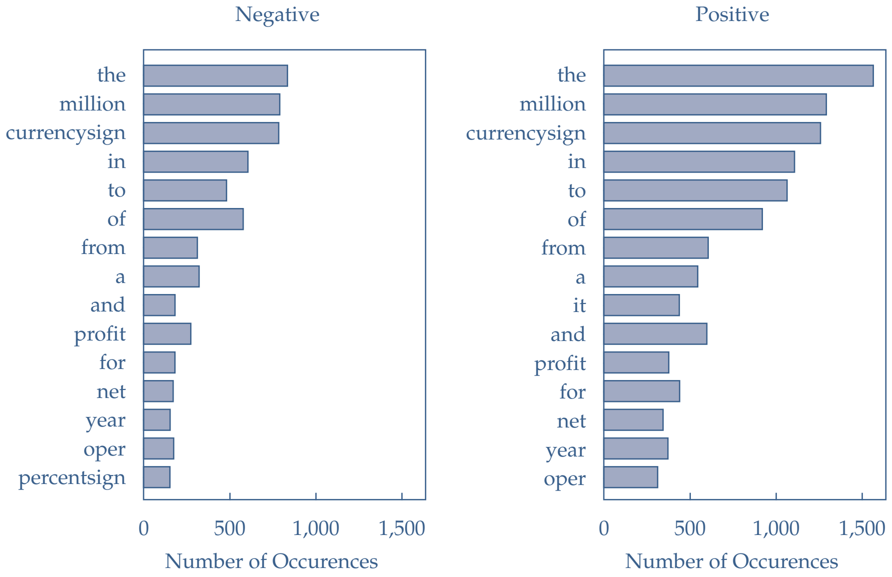
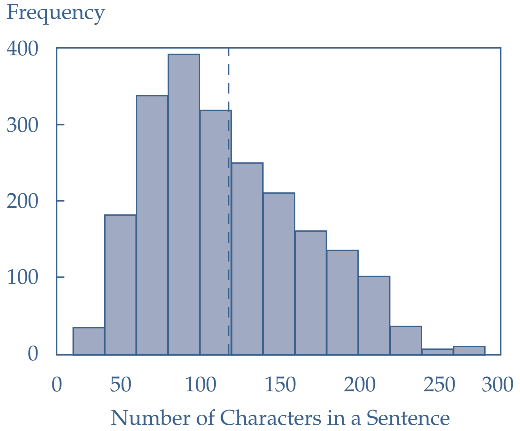


Exhibit 34 shows a histogram of sentence length distribution. **Sentence length** is defined as the number of characters, including spaces, in a sentence. The longest sentence has 273 characters; the shortest sentence has 26 characters; and the average number of characters is about 120 (indicated by the vertical line). Although this distribution does not have any direct impact on model training, this histogram visually demonstrates the range of sentence lengths and helps identify any extremely long or short sentences. This histogram does not appear unusual, so no outlier sentences need to be removed.

Exhibit 34 Histogram of Sentence Lengths with Mean Sentence Length



Word clouds are a convenient method of visualizing the text data because they enable rapid comprehension of a large number of tokens and their corresponding weights. Exhibit 35 shows a word cloud for all the sentences in the corpus. The font sizes of the words are proportionate to the number of occurrences of each word in the corpus. Similarly, Exhibit 36 shows the word cloud divided into two halves: one half representing negative sentiment class sentences (upper half); one half representing positive sentiment class sentences (lower half). Notably, some highly discriminative stems and words, such as “decreas” and “down” in the negative half and “increas” and “rose” in the positive half, are present. The feature selection process will eliminate common words and highlight useful words for better model training.

Exhibit 35 Word Cloud of Entire Corpus

Exhibit 36 Word Cloud Divided by Two Sub-Groups of the Corpus

Feature Selection

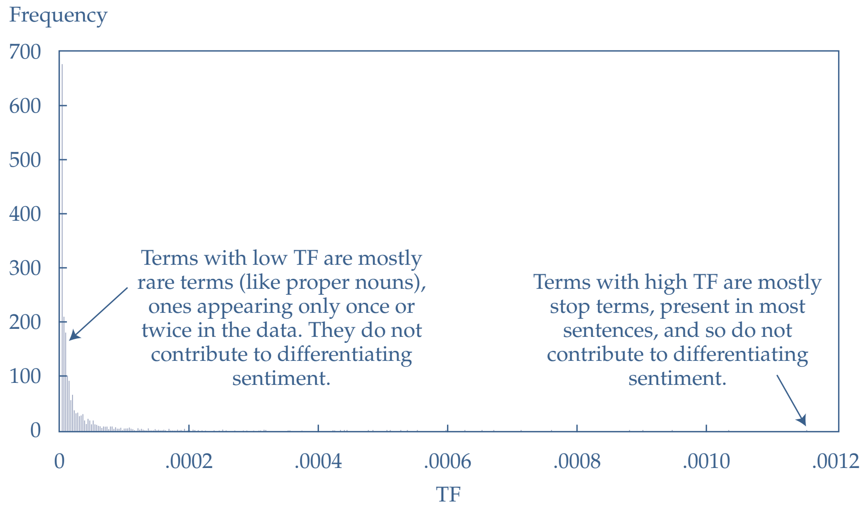
Exploratory data analysis revealed the most frequent tokens in the texts that could potentially add noise to this ML model training process. In addition to common tokens, many rarely occurring tokens, often proper nouns (i.e., names), are not informative for understanding the sentiment of the sentence. Further analyses must be conducted to decide which words to eliminate. Feature selection for text data involves keeping the useful tokens in the BOW that are informative and help to discriminate different classes of texts—those with positive sentiment and those with negative sentiment. At this point, a total of 44,151 non-unique tokens are in the 2,180 sentences.

Frequency analysis on the processed text data helps in filtering unnecessary tokens (or features) by quantifying how important tokens are in a sentence and in the corpus as a whole. Term frequency (TF) at the corpus level—also known as **collection frequency (CF)**—is the number of times a given word appears in the whole corpus (i.e., collection of sentences) divided by the total number of words in the corpus. Term frequency can be calculated and examined to identify outlier words. Exhibit 37 shows the descriptive statistics of term frequency for the words at the collection level. The statistics of TF range between 0 and 1 because TF values are ratios of total occurrences of a particular word to total number of words in the collection. A sample of words with the highest TF and lowest TF values is also shown to gain insight into what kinds of words occur at these extreme frequencies.

Exhibit 37 Summary Statistics of TF for Words at the Collection Level, Sample Words with High and Low TF Values, and Histogram of TF Values

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.265e-05	2.265e-05	4.530e-05	3.741e-04	1.585e-04	5.429e-02

word	TF	word	TF
<chr>	<dbl>	<chr>	<dbl>
the	0.05429096	yet	2.264954e-05
million	0.04722430	yihn	2.264954e-05
currencysign	0.04627302	young	2.264954e-05
in	0.03870807	zahariev	2.264954e-05
to	0.03476705	zone	2.264954e-05
of	0.03377047	zoo	2.264954e-05



Calculating highest and lowest TFs at the collection level is a general strategy to identify noisy terms. The histogram in Exhibit 37 shows a long tail to the right, which represents common terms that must be removed. The high frequency bars on the left show that there are also many rare terms (e.g., ones appearing only once or twice across the data). Such rare terms do not appear enough to be used as meaningful features and are often removed. The words with the highest TF are mostly stop words that are not useful because they are present in most of the sentences and thus do not contribute to differentiating the sentiment embedded in the text. The words with the lowest TF values are mostly proper nouns or sparse terms that are also not important to the meaning of the text. In this example, after careful examination of words with extreme frequencies, the words with high TF values (>99.5th percentile, 14 words) and low TF values (<30th percentile, 714 words) are removed before forming the final document term matrix (DTM). Exhibit 38 shows the 14 words with the highest TF values (>99.5th percentile) that are the custom stop words for this project.

Exhibit 38 Fourteen Custom Stop Words for the Project

"the"	"million"	"currencysign"	"in"	"to"	"of"	"from"
"and"	"profit"	"for"	"it"	"not"	"year"	"a"

To construct a DTM for ML training, different TF measures need to be computed to fill in the cells of the DTM. Exhibit 39 displays part of a TF measures table that is computed for the text data before the removal of custom stop words.

Exhibit 39 Sample Output of Term Frequency (TF) Measures Table

SentenceNo	TotalWordsInSentence	Word	TotalWordCount	WordCountInSentence	SentenceCountWithWord	TF	DF	IDF	TFIDF
<int>	<int>	<chr>	<int>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
624	34	a	873	6	687	0.1764706	0.3151376	1.1547459	0.20377868
701	39	the	2397	6	1453	0.1538462	0.6665138	0.4056945	0.06241454
1826	34	a	873	6	687	0.1764706	0.3151376	1.1547459	0.20377868
1963	39	the	2397	6	1453	0.1538462	0.6665138	0.4056945	0.06241454
128	30	of	1491	5	984	0.1666667	0.4513761	0.7954543	0.13257571
223	37	the	2397	5	1453	0.1351351	0.6665138	0.4056945	0.05482358

The columns of the term frequency measures table are as follows:

1 *SentenceNo*: A unique identification number assigned to each sentence in the order they are present in the original dataset. For example, sentence number 701 is a sentence in row 701 from the data table: "*the airlin estim that the cancel of it flight due to the closur of european airspac and the process of recommenc traffic have caus a the compani a loss of currencysign million includ the cost of strand passeng accommod*."

2 *TotalWordsInSentence*: Count of total number of words present in the sentence. For example, sentence number 701 has a total of 39 words.

3 *Word*: A word token that is present in the corresponding sentence.

4 *TotalWordCount*: Total number of occurrences of the word in the entire corpus or collection. For example, the token "the" occurs 2,397 times in the whole collection of sentences. The following equation can be used to compute TF at the collection level:

$$\text{TF (Collection Level)} = \text{TotalWordCount/Total number of words in collection.} \tag{10}$$

The TF of the word "the" at the collection level is calculated as 2,397/44,151 = 0.05429096. Note that this result was seen previously in Exhibit 37.

5 *WordCountInSentence*: Number of times the token is present in the corresponding sentence. For example, token "the" is present six times in sentence number 701.

6 *SentenceCountWithWord*: Number of sentences in which the word is present. For example, the token "the" is present in 1,453 sentences.

7 *TF (Term Frequency) at Sentence Level*: Number of times a word is present in a sentence divided by the total number of words in that sentence. The following equation can be used to compute TF at the sentence level:

$$\text{TF (Sentence Level)} = \text{WordCountInSentence/TotalWordsInSentence}. \quad (11)$$

For example, TF at the sentence level for the word "the" in sentences number 701 and 223 is calculated as 6/39 = 0.1538462 and 5/37 = 0.1351351, respectively.

8 *DF (Document Frequency)*: Defined as the number of documents (i.e., sentences) that contain a given word divided by the total number of sentences (here, 2,180). Document frequency is important since words frequently occurring across sentences provide no differentiating information in each sentence. The following equation can be used to compute DF:

$$\text{DF} = \text{SentenceCountWithWord/Total number of sentences}. \quad (12)$$

For example, DF of the word "the" is 1,453/2,180 = 0.6665138; so, 66.7% of the sentences contain the word "the." A high DF indicates high word frequency in the text.

9 *IDF (Inverse Document Frequency)*: A relative measure of how unique a term is across the entire corpus. Its meaning is not directly related to the size of the corpus. The following equation can be used to compute IDF:

$$\text{IDF} = \log(1/\text{DF}). \quad (13)$$

For example, IDF of the word "the" is log(1/0.6665138) = 0.4056945. A low IDF indicates high word frequency in the text.

10 *TF–IDF*: To get a complete representation of the value of each word, TF at the *sentence level* is multiplied by the IDF of a word across the entire dataset. Higher TF–IDF values indicate words that appear more frequently within a smaller number of documents. This signifies relatively more unique terms that are important. Conversely, a low TF–IDF value indicates terms that appear in many documents. TF–IDF values can be useful in measuring the key terms across a compilation of documents and can serve as word feature values for training an ML model. The following equation can be used to compute TF–IDF:

$$\text{TF–IDF} = \text{TF} \times \text{IDF}. \quad (14)$$

For example, TF–IDF of the token "of" is calculated as 0.1666667 × 0.7954543 = 0.13257571.

Similarly, Exhibit 40 shows high TF–IDF words for the text data before the removal of custom stop words.

Exhibit 40 Sample Output of High TF–IDF Words

SentenceNo	TotalWordsInSentence	Word	TotalWordCount	WordCountInSentence	SentenceCountWithWord	TF	DF	IDF	TFIDF
<int>	<int>	<chr>	<int>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
28	7	risen	3	1	3	0.1428571	0.0013761468	6.588468	0.9412097
830	7	diminish	2	1	2	0.1428571	0.0009174312	6.993933	0.9991333
1368	9	great	4	1	4	0.1111111	0.0018348624	6.300786	0.7000873
1848	8	injuri	1	1	1	0.1250000	0.0004587156	7.687080	0.9608850
1912	7	cheaper	1	1	1	0.1428571	0.0004587156	7.687080	1.0981543
1952	6	argument	1	1	1	0.1666667	0.0004587156	7.687080	1.2811800

TF or TF–IDF values are placed at the intersection of sentences (rows) and terms (columns) of the document term matrix. For this project, TF values are used for the DTM as the texts are sentences rather than paragraphs or other larger bodies of text. TF–IDF values vary by the *number* of documents in the dataset; therefore, the model performance can vary when applied to a dataset with just a few documents. In addition to removing custom stop words and sparse terms, single character letters are also eliminated because they do not add any value to the sentiment significance.

Feature Engineering

N-grams are used as a feature engineering process in this project. Use of n-grams helps to understand the sentiment of a sentence as a whole. As mentioned previously, the objective of this project is to predict sentiment class (positive and negative) from financial texts. Both unigram and bigrams are implemented, and the BOW is created from them. Bigram tokens are helpful for keeping negations intact in the text, which is vital for sentiment prediction. For example, the tokens "not" and "good" or "no" and "longer" can be formed into single tokens, now bigrams, such as "not_good" and "no_longer." These and similar tokens can be useful during ML model training and can improve model performance. Exhibit 41 shows a sample of 100 words from the BOW containing both unigram and bigram tokens after removal of custom stop words, sparse terms, and single characters. Note that the BOW contains such tokens as increas, loss, loss_prior, oper_rose, tax_loss, and sale_increas. Such tokens are informative about the embedded sentiment in the texts and are useful for training an ML model. The corresponding word frequency measures for the document term matrix are computed based on this new BOW.

Exhibit 41 One-Hundred Sample Tokens from Final BOW of Entire Corpus

“last”	“last_quarter”	“quarter”	“quarter_componenta”	“componenta”
“componenta_sale”	“sale”	“sale_doubl”	“doubl”	“doubl_same”
“same”	“same_period”	“period”	“period_earlier”	“earlier”
“earlier_while”	“while”	“while_move”	“move”	“move_zero”
“zero”	“zero_pre”	“pre”	“pre_tax”	“tax”
“tax_pre”	“tax_loss”	“loss”	“third”	“third_quarter”
“quarter_sale”	“sale_increas”	“increas”	“increas_by”	“by”
“by_percentsign”	“percentsign”	“percentsign_oper”	“oper”	“oper_by”
“oper_rose”	“rose”	“rose_correspond”	“correspond”	“correspond_period”
“period_repres”	“repres”	“repres_percentsign”	“percentsign_sale”	“oper_total”
“total”	“total_up”	“up”	“up_repres”	“finnish”
“finnish_talentum”	“talentum”	“talentum_report”	“report”	“report_oper”
“oper_increas”	“increas_sale”	“sale_total”	“cloth”	“cloth_retail”
“retail”	“retail_chain”	“chain”	“chain_sepp”	“sepp”
“sepp_ls”	“ls”	“ls_sale”	“consolid”	“consolid_sale”
“incres_percentsign”	“percentsign_reach”	“reach”	“reach_while”	“while_oper”
“oper_amount”	“amount”	“amount_compar”	“compar”	“compar_loss”
“loss_prior”	“prior”	“prior_period”	“foundri”	“foundri_divis”
“divis”	“divis_report”	“report_sale”	“percentsign_correspond”	“period_sale”
“sale_machin”	“machin”	“machin_shop”	“shop”	“shop_divis”

EXAMPLE 6

Calculating and Interpreting Term Frequency Measures

Data scientists Jack and Jill are using financial text data to develop sentiment indicators for forecasting future stock price movements. They have assembled a BOW from the corpus of text being examined and have pulled the following abbreviated term frequency measures tables.

Term Frequency Measures Table 1

SentenceNo	TotalWordsInSentence	Word	TotalWordCount	WordCountInSentence	SentenceCountWithWord
<int>	<int>	<chr>	<int>	<int>	<int>
624	34	a	873	6	687
701	39	the	2397	6	1453
1826	34	a	873	6	687
1963	39	the	2397	6	1453
128	30	of	1491	5	984
223	37	the	2397	5	1453

Term Frequency Measures Table 2

SentenceNo	TotalWordsInSentence	Word	TotalWordCount	WordCountInSentence	SentenceCountWithWord
<int>	<int>	<chr>	<int>	<int>	<int>
28	7	risen	3	1	3
830	7	diminish	2	1	2
1368	9	great	4	1	4
1848	8	injuri	1	1	1
1912	7	cheaper	1	1	1
1952	6	argument	1	1	1

1 Determine and interpret term frequency (TF) at the collection level and at the sentence level for the word (i.e., token) "a" in sentence 1,826 in term frequency measures Table 1 and then for the token "great" in sentence 1,368 in term frequency measures Table 2.

2 Determine and interpret TF–IDF (term frequency–inverse document frequency) for the word "a" in sentence 1,826 in term frequency measures Table 1 and then for the token "great" in sentence 1,368 in term frequency measures Table 2.

Solution to 1:

TF at the collection level is calculated using Equation 10:

TF (Collection Level) = TotalWordCount/Total number of words in collection.

For token "a" in sentence 1,826 (Table 1), TF (Collection Level) is 873/44,151 = 0.019773 or 1.977%.For token "great" in sentence 1,368 (Table 2), TF (Collection Level) is 4/44,151 = 0.000091 or 0.009%.TF at the collection level is an indicator of the frequency, in percentage terms, that a token is used throughout the whole collection of texts (here, 44,151). It is useful for identifying outlier words: Tokens with highest TF values are mostly stop words that do not contribute to differentiating the sentiment embedded in the text (such as "a"), and tokens with lowest TF values are mostly proper nouns or sparse terms that are also not important to the meaning of the text. Conversely, tokens with intermediate TF values potentially carry important information useful for differentiating the sentiment embedded in the text.TF at the sentence level is calculated using Equation 11:

TF (Sentence Level) = WordCountInSentence/TotalWordsInSentence.

For token "a" in sentence 1,826, TF (Sentence Level) is 6/34 = 0.176471 or 17.647%.

For token "great" in sentence 1,368, TF (Sentence Level) is 1/9 = 0.111111 or 11.111%.

TF at the sentence level is an indicator of the frequency, in percentage terms, that a token is used in a particular sentence (i.e., instance). Therefore, it is useful for understanding the importance of the specific token in a given sentence.

Solution to 2:

To calculate TF–IDF, besides TF at the sentence level, document frequency (DF) and inverse document frequency (IDF) are also required.

DF is the number of documents (i.e., sentences) that contain a given word divided by the total number of sentences in the corpus (here, 2,180). DF is calculated using Equation 12:

DF = SentenceCountWithWord/Total number of sentences.

For token "a" in sentence 1,826, DF is 687/2,180 = 0.315138 or 31.514%.

For token "great" in sentence 1,368, DF is 4/2,180 = 0.001835 or 0.184%.

Document frequency is important since tokens occurring frequently across sentences (such as "a") provide no differentiating information in each sentence. Tokens occurring less frequently across sentences (such as "great"), however, may provide useful differentiating information.

IDF is a relative measure of how important a term is across the entire corpus (i.e., collection of texts/sentences). IDF is calculated using Equation 13:

$$\text{IDF} = \log(1/\text{DF}).$$

For token "a" in sentence 1,826, IDF is log(1/0.315138) = 1.154746.

For token "great" in sentence 1,368, IDF is log(1/0.001835) = 6.300786.

Using TF and IDF, TF–IDF can now be calculated using Equation 14:

$$\text{TF–IDF} = \text{TF} \times \text{IDF}.$$

For token "a" in sentence 1,826, TF–IDF = 0.176471 × 1.154746 = 0.203779, or 20.378%.

For token "great" in sentence 1,368, TF–IDF = 0.111111 × 6.300786 = 0.700087, or 70.009%.

As TF–IDF combines TF at the *sentence level* with IDF across the entire corpus, it provides a complete representation of the value of each word. A high TF–IDF value indicates the word appears many times within a small number of documents, signifying an important yet unique term within a sentence (such as "great"). A low TF–IDF value indicates tokens that appear in most of the sentences and are not discriminative (such as "a"). TF–IDF values are useful in extracting the key terms in a document for use as features for training an ML model.

MODEL TRAINING 12

e describe preparing, wrangling, and exploring text-based data for financial forecasting;

The sentiment class labels (positive and negative) constitute the target variable (y) for model training. They are relabeled as 1 (for positive) and 0 (for negative) to enable calculating the performance metrics, such as receiver operating characteristic (ROC) curve and area under the curve (AUC) from the trained model results. The master dataset that has been cleansed and preprocessed is partitioned into three separate sets: 1) training set; 2) cross-validation (CV) set; and 3) test set. These are in the ratio of 60:20:20, respectively (following common practice). For splitting, simple random sampling is applied within levels of the target variable to balance the class distributions within the splits. The final DTM is built using the sentences (rows), which are the instances, and resulting tokens (columns), which are the feature variables, from the BOW of the training dataset. The final BOW consists of unigram and bigram tokens from the sentences in the training corpus only. The DTM is then filled in with resultant TF values of the tokens from the training corpus.

Similarly, the DTMs for the CV set and the test set are built using tokens from the final training BOW for tuning, validating, and testing of the model. To be clear, the final BOW from the training corpus is used for building DTMs across all the splits because the model has been trained on that final BOW. Thus, the columns (think, features) of all three DTMs are the same, but the number of rows varies because a different number of sentences are in each split. The DTMs are filled with resultant term frequency values calculated using sentences in the corpuses of the respective splits—sentences from the CV set corpus and sentences from the test set corpus. Exhibit 42 tabulates the summary of dimensions of the data splits and their uses in the model training process. As mentioned, the columns of DTMs for the splits are the same, equal to the number of unique tokens (i.e., features) from the final training corpus BOW, which is 9,188. Note that this number of unique tokens (9,188) differs from that in the master corpus (11,501) based on the sentences that are included in the training corpus after the random sampling.

Exhibit 42 Summary of the Three Data Splits

Corpus	Split %	Number of Sentences	DTM Dimensions	Purpose
Master	100%	2180	2180 × 11501	Used for data exploration
Training	60%	1309	1309 × 9188	Used for ML model training
CV	20%	435	435 × 9188	Used for tuning and validating the trained model
Test	20%	436	436 × 9188	Used for testing the trained, tuned, and validated model

Method Selection

Alternative ML methods, including SVM, decision trees, and logistic regression, were examined because these techniques are all considered potentially suitable for this particular task (i.e., supervised learning), type of data (i.e., text), and size of data (i.e., wider data with many potential variables). The SVM and logistic regression methods appeared to offer better performance than decision trees. For brevity, we discuss logistic regression in the remainder of the chapter. Logistic regression was used to train the model, using the training corpus DTM containing 1,309 sentences. As a reminder, in this project texts are the sentences and the classifications are positive and negative sentiment classes (labeled 1 and 0, respectively). The tokens are feature variables, and the sentiment class is the target variable. Text data typically contain thousands of tokens. These result in sparse DTMs because each column represents a token feature and the values are mostly zeros (i.e., not all the tokens are present in every text). Logistic regression can deal with such sparse training data because the regression coefficients will be close to zero for tokens that are not present in a significant number of sentences. This allows the model to ignore a large number of minimally useful features. Regularization further helps lower the coefficients when the features rarely occur and do not contribute to the model training.

Logistic regression is applied on the final training DTM for model training. As this method uses maximum likelihood estimation, the output of the logistic model is a probability value ranging from 0 to 1. However, because the target variable is binary, coefficients from the logistic regression model are not directly used to predict the value of the target variable. Rather, a mathematical function uses the logistic regression coefficient (β) to calculate probability (p) of sentences having positive sentiment (y = 1).[3] If p for a sentence is 0.90, there is a 90% likelihood that the sentence has positive sentiment. Theoretically, the sentences with p > 0.50 likely have positive sentiment. Because this is not always true in practice, however, it is important to find an ideal threshold value of p. We elaborate on this point in a subsequent example. The threshold

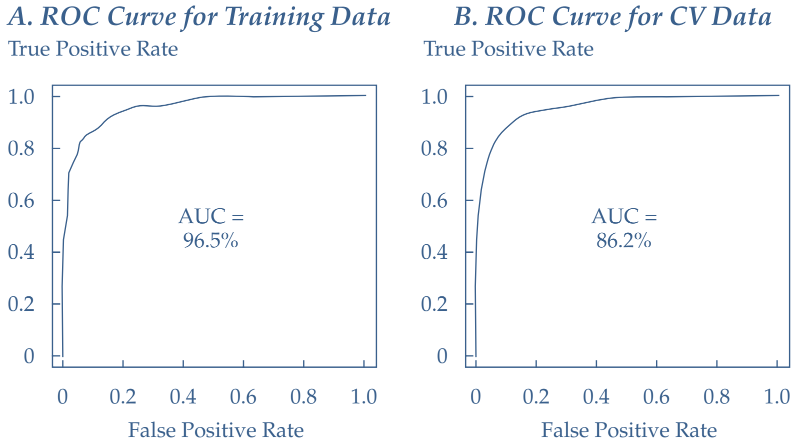
3 This mathematical function is an exponential function of the form: $P(y = 1) = \dfrac{1}{1 + \exp^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}}$ where the βs are the logistic regression coefficients.

value is a cutoff point for p values, and the ideal threshold p value is influenced by the dataset and model training. When the p values (i.e., probability of sentences having positive sentiment) of sentences are above this ideal threshold p value, then the sentences are *highly* likely to have positive sentiment (y = 1). The ideal threshold p value is estimated heuristically using performance metrics and ROC curves, as will be demonstrated shortly.

Performance Evaluation and Tuning

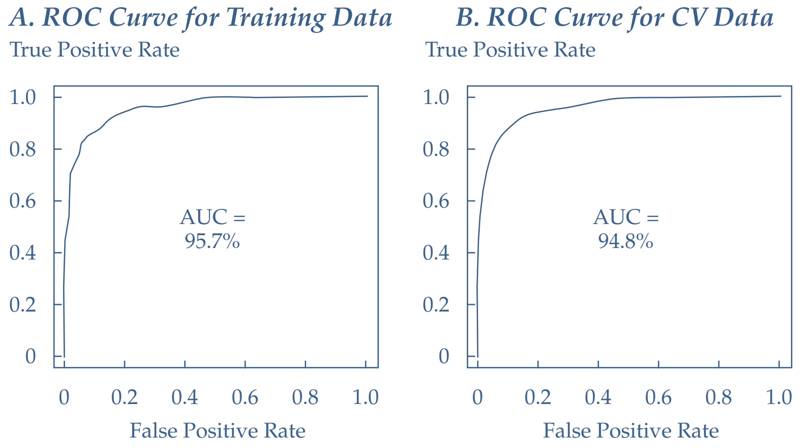
The trained ML model is used to predict the sentiments of the sentences in the training and CV DTMs. Exhibit 43 displays the ROC curves for the training (Panel A) and CV (Panel B) data. Remember that the x-axis is false positive rate, FP/(TN + FP), and the y-axis is true positive rate, TP/(TP + FN). As the model is trained using the training DTM, it clearly performs well on the same training data (so there is no concern about underfitting) but does not perform as well on the CV data. This is apparent as the ROC curves are significantly different between the training and CV datasets. The AUC is 96.5% on training data and 86.2% on CV data. This finding suggests that the model performs comparatively poorly (with a higher rate of error or misclassification) on the CV data when compared to training data. Thus, the implication is that the model is overfitted.

Exhibit 43 ROC Curves of Model Results for Training and CV Data Before Regularization



As the model is overfitted, least absolute shrinkage and selection operator (LASSO) regularization is applied to the logistic regression. LASSO regularization penalizes the coefficients of the logistic regression to prevent overfitting of the model. The penalized regression will select the tokens (features) that have statistically significant (i.e., non-zero) coefficients and that contribute to the model fit; LASSO does this while disregarding the other tokens. Exhibit 44 shows the ROC curves for the new model that uses regularized logistic regression. The ROC curves look similar for model performance on both datasets, with an AUC of 95.7% on the training dataset (Panel A) and 94.8% on the CV dataset (Panel B). These findings suggest that the model performs similarly on both training and CV data and thus indicate a good fitting model (one that is not overfitted).

Exhibit 44 ROC Curves of Model Results for Training and CV Data After Regularization



Regularization along with careful feature selection help to prevent overfitting in logistic regression models. Another model was trained using all token features, including stop words, sparse terms, and single characters, with no regularization. That model showed an AUC of 99.1% when applied on the training dataset and an AUC of 89.4% when applied on the CV dataset, suggesting that the model is overfitting. As the AUC values in all of the models discussed are not far from 100%, these models are clearly not underfitting. In sum, the final ML model for this project uses logistic regression with LASSO regularization.

To further evaluate the model, error analysis is conducted by calculating a confusion matrix using the ML model results from the cross-validation dataset. The threshold p value of 0.5 is used as a cutoff point. When target value $p > 0.5$, the prediction is assumed to be $y = 1$ (meaning, positive sentiment). Otherwise, the prediction is assumed to be $y = 0$ (negative sentiment). A confusion matrix, with performance metrics and overall scores for the model results using the CV data, is shown in Exhibit 45.

Exhibit 45 Confusion Matrix of Model Results for CV Data with Threshold p Value = 0.50

Confusion Matrix for CV Data with Threshold = 0.5

		Actual Training Labels	
		Class "1"	Class "0"
Predicted Results	Class "1"	284 (TP)	38 (FP)
	Class "0"	7 (FN)	106 (TN)

Performance Metrics

TP = 284, FP = 38, FN = 7, TN = 106

$P = 284 / (284+38) = 0.88$

$R = 284 / (284+7) = 0.98$

$\text{F1 Score} = (2 \times 0.88 \times 0.98) / (0.88 + 0.98) = 0.93$

$\text{Accuracy} = (284 + 106) / (284 + 38 + 106 + 7) = 0.90$

The model accuracy is 90% with a theoretically suggested (default) threshold p value of 0.5. The CV data are used to tune the threshold value for best model performance. Various p values from 0.01 to 0.99 are systematically evaluated individually, and confusion matrixes and performance metrics are calculated using each of these p values. Based on these metrics, the p value resulting in the highest model accuracy is selected as the ideal threshold p value. However, there are often trade-offs: Minimizing false positives (FPs) comes at a cost of increasing false negatives (FNs), and vice versa. Prioritizing various performance statistics (e.g., precision versus recall) depends on the context and relative consequences of FP and FN on the project applications. In this project, the values of negative sentiment and positive sentiment sentences are assumed to be equal, thus the impacts of FP and FN are also equal. It is common practice to simulate many model results using different threshold p values and to search for maximized accuracy and F1 statistics that minimize these trade-offs. As noted earlier, accuracy and F1 scores are overall performance measures that give equal weight to FP and FN.

Exhibit 46 shows the overall performance measures (i.e., F1 score and accuracy) for various threshold p values. The threshold p value that results in the highest accuracy and F1 score can now be identified. From the charts in Exhibit 45, the ideal threshold p value appears to be around 0.60. To investigate further, a table of performance measures (i.e., precision, recall, F1 score, and accuracy) is generated for a series of threshold p values ranging from 0.45 to 0.75. The table in Exhibit 47 demonstrates that threshold p values between 0.60 and 0.63 result in the highest accuracy and F1 score for the CV dataset. As a result of this analysis, a final threshold p value of 0.60 is selected.

Exhibit 46 Threshold Values Versus Overall Performance Measures

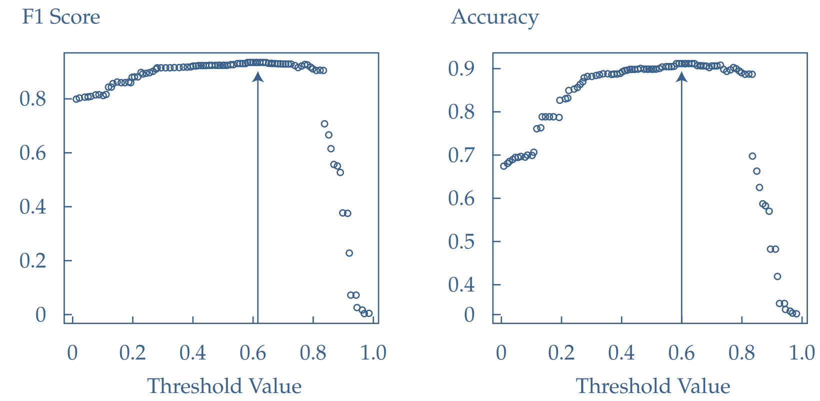


Exhibit 47 Performance Measures of the Model for a Series of Threshold Values

Threshold	Precision	Recall	F1	Accuracy
0.45	0.8750000	0.986254296	0.927302100	0.8965517
0.46	0.8827160	0.982817869	0.930081301	0.9011494
0.47	0.8827160	0.982817869	0.930081301	0.9011494
0.48	0.8819876	0.975945017	0.926590538	0.8965517
0.49	0.8819876	0.975945017	0.926590538	0.8965517
0.50	0.8819876	0.975945017	0.926590538	0.8965517

(continued)

Exhibit 47 (Continued)

Threshold	Precision	Recall	F1	Accuracy
0.51	0.8819876	0.975945017	0.926590538	0.8965517
0.52	0.8819876	0.975945017	0.926590538	0.8965517
0.53	0.8902821	0.975945017	0.931147541	0.9034483
0.54	0.8930818	0.975945017	0.932676519	0.9057471
0.55	0.8930818	0.975945017	0.932676519	0.9057471
0.56	0.8958991	0.975945017	0.934210526	0.9080460
0.57	0.8958991	0.975945017	0.934210526	0.9080460
0.58	0.8958991	9.975945017	0.934210526	0.9080460
0.59	0.9015873	0.975945017	0.937293729	0.9126437
0.60	0.9044586	0.975945017	0.938842975	0.9149425
0.61	0.9044586	0.975945017	0.938842975	0.9149425
0.62	0.9044586	0.975945017	0.938842975	0.9149425
0.63	0.9041534	0.972508591	0.937086093	0.9126437
0.64	0.9041534	0.972508591	0.937086093	0.9126537
0.65	0.9041534	0.972508591	0.937086093	0.9126437
0.66	0.9035370	0.965635739	0.933554817	0.9080460
0.67	0.9035370	0.965635739	0.933554817	0.9080460
0.68	0.9064516	0.965635739	0.935108153	0.9103448
0.69	0.9064516	0.965635739	0.935108153	0.9103448
0.70	0.9061489	0.962199313	0.933333333	0.9080460
0.71	0.9061489	0.962199313	0.933333333	0.9080460
0.72	0.9090909	0.962199313	0.934891486	0.9103448
0.73	0.9090909	0.962199313	0.934891486	0.9103448
0.74	0.9078947	0.948453608	0.927731092	0.9011494
0.75	0.9072848	0.941580756	0.924114671	0.8965517

* The shaded row shows the selected threshold p value (0.60) and the performance metrics for the selected model.

Finally, the confusion matrix using the ideal threshold p value of 0.60 is constructed to observe the performance of the final model. When target value p > 0.60, the prediction is assumed to be $y = 1$ (indicating positive sentiment); otherwise, the prediction is assumed to be $y = 0$ (negative sentiment). The confusion matrix for the CV data is shown in Exhibit 48. It is clear that the model performance metrics have improved in the final model compared to the earliest case when the threshold p value was 0.50. Now, accuracy and F1 score have both increased by one percentage point to 91% and 94%, respectively, while precision has increased by two percentage points to 90%.

Exhibit 48 Confusion Matrix of Model Results for CV Data with Threshold p Value = 0.60

Confusion Matrix for CV Data with Threshold = 0.6

		Actual Training Labels	
		Class "1"	Class "0"
Predicted Results	Class "1"	284 (TP)	30 (FP)
	Class "0"	7 (FN)	114 (TN)

Performance Metrics

TP = 284, FP = 30, FN = 7, TN = 114

P = 284 / (284+30) = 0.90

R = 284 / (284+7) = 0.98

F1 Score = (2 × 0.90 × 0.98) / (0.90 + 0.98) = 0.94

Accuracy = (284 + 114) / (284 + 30 + 114 + 7) = 0.91

RESULTS AND INTERPRETATION

13

e describe preparing, wrangling, and exploring text-based data for financial forecasting;

The final ML model with the appropriate threshold p value has been validated and is now ready for use. The model can be used to predict the sentiment of new sentences from the test data corpus as well as new sentences from similar financial text data sources, such as news wires, earnings call transcripts, and quarterly financial reports. The final model is a collection of penalized regression coefficients for unigram and bigram tokens from the BOW of the training corpus. To use the model to predict the sentiment of new sentences, tokenization and identical cleansing and preprocessing operations must be performed on the new sentences. All the processes performed on the training data must be performed on the new data to which the model will be applied (as was done for the test dataset). The model will use the trained penalized regression coefficients on the term frequency (TF) values of the tokens in the document term matrix (DTM) of the new sentences and will determine the target value (p). The columns of the DTM of the new sentences are the same as those of the training DTM, but the TF values are calculated based on the test corpus. Using the threshold p value of 0.60, the sentiment class for each sentence in the test corpus will be predicted.

The model is now applied on the test data that contains 436 sentences. Note that the test data were not used to train or validate/tune the model and are new to the model. The test data were preprocessed identically to the training and CV data while a part of the master corpus. The model is then applied to the test DTM, and the results are obtained. Exhibit 49 displays 30 sample results from the test corpus. The results table contains cleansed and preprocessed sentences, actual sentiment, target p values from the model, and predicted sentiment. Note that this sample contains three cases of misclassification: the 10th sentence (text), where p = 0.46; the 26th text, where p = 0.77; and the 30th text, where p = 0.71. Therefore, accuracy of this 30-text sample is 27/30 = 90%.

Exhibit 49 Thirty Sample Results of Test Data

Sentence	Sentiment	p	Predicted Sentiment
exclude non recur item pre tax surg percentsign	1	0.81	1
adp news feb finnish retail kesko oyj hel kesbv said today total sale exclud valu ad tax vat stood at januari down percentsign on yea	0	0.12	0
india trade with russia current stand at four billion dollar grow per cent fiscal	1	0.83	1
refin margin was bbl combar bbl prior	1	0.81	1
scania morgan Stanley lift share target on swedish heavi duti truck bus maker scania ab crown euro crown euro	1	0.83	1
deal is like bring save	1	0.83	1
will also strengthen ruukki offshore busi	1	0.83	1
last week finnish metl technolog group announc plan sell more than percent technolog unit further compani strategy goal becom world largest stainless steel maker	1	0.83	1
nest oil board propos dividend full compar with ago	1	0.81	1
pre tax loss total compar loss first quarter	1	0.46	0
pretax total compar loss fourth quarter	1	0.74	1
re use back into pet bottle has also steadili increas rate use strap tape has pick up again after dip pector said previous	1	0.95	1
satama sale would be higher than befor	1	0.83	1
octob finnish wood product technolog supplier raut oyj hel rutav said today swung first nine month versus loss same period earlier	1	0.79	1
ebit total compar loss correspond period	1	0.74	1
finnish consum packag manufactur huhtamaki oyj said swung euro first nine month loss euro same period	1	0.77	1
finnish dental care group oral hammaslaakarit oyj post total euro first nine month versus loss euro same period	1	0.79	1
finnish silicon water manufactur okmet oyj said swung euro first nine month loss euro earlier	1	0.77	1
adp news feb finnish print circuit board pcb maker aspocomp group oyj hel acg said today swung versus loss	1	0.79	1
mn pretax third quarter	1	0.83	1
oper total compar correspond period	1	0.81	1
raut post euro third quarter compar loss euro correspond period	1	0.74	1
russian export duti will active harvest finland sale russia will increas also	1	0.91	1
compani expect sale signific increas	1	0.91	1
compani amount ee which was percentsign more than	1	0.81	1
third quarter fiscal efor swung loss versus correspond period fiscal	0	0.77	1
acando ab acanb ss fell percent kronor lowest close sinc dec	0	0.20	0
compani oper loss total compar	0	0.27	0
last paseng flew airlin down percent	0	0.12	0
loss after financi item total compar correspond period	0	0.71	1

Exhibit 50 shows the confusion matrix for the test data. Accuracy and F1 score are 90% and 93%, respectively, while precision and recall are 89% and 98%, respectively. Therefore, it is apparent that the model performs similarly on the training, CV, and test datasets. These findings suggest that the model is robust and is not overfitting. They also suggest that the model should generalize well out-of-sample and can thus be used to predict the sentiment classes for new sentences from similar financial text data sources. Of course, these new text data must first be subjected to identical tokenization, cleansing, and preprocessing as done for the training dataset.

Exhibit 50 Confusion Matrix of Model Results for Test Data with Threshold p Value = 0.60

Confusion Matrix for Test Data

		Actual Training Labels	
		Class "1"	Class "0"
Predicted Results	Class "1"	284 (TP)	35 (FP)
	Class "0"	7 (FN)	110 (TN)

Performance Metrics

TP = 284, FP = 35, FN = 7, TN = 110

$P = 284 / (284+35) = 0.89$

$R = 284 / (284+7) = 0.98$

F1 Score = $(2 \times 0.89 \times 0.98) / (0.89 + 0.98) = 0.93$

Accuracy = $(284 + 110) / (284 + 35 + 110 + 7) = 0.90$

To recap, this project involves converting unstructured data (i.e., text data from financial data sources) into structured data (i.e., tokens, sentences, and term frequency values) in a document term matrix that is used as input for training, validating, and testing machine learning-based models (here, logistic regression) for predicting classification (here, sentiment classes). Similar models can be built and used in different contexts to understand the sentiment embedded in larger texts. The derived sentiment classification can be useful as a visualization tool to provide insight about the text without reading large documents. These sentiment classifications can also be used as structured input data for larger ML models that have a specific purpose, such as to predict future stock price movements.

EXAMPLE 7

Comparing Performance Metrics for Confusion Matrixes with Different Threshold p Values

In the previous analysis using the cross-validation dataset, performance measures for the sentiment classification ML model were calculated for a wide range (from 0.45 to 0.75) of threshold p values. The threshold value of 0.60 was determined to be the p value that maximizes model accuracy and F1 score; the confusion matrix for this model is shown in Exhibit 48. Use the following confusion matrixes with threshold p values of 0.75 and 0.45, A and B, respectively, to answer the following questions.

Confusion Matrix A

Confusion Matrix for CV, Threshold = 0.75

N = 436		Actual Training Labels	
		Class "1"	Class "0"
Predicted Results	Class "1"	281	28
	Class "0"	17	110

Performance Metrics

TP = 281, FP = 28, FN = 17, TN = 110
Precision = TP/(TP + FP) = 0.91
Recall = TP/(TP + FN) = 0.94
F1 Score = HMean: Prec. & Recall = 0.93
Accuracy = (TP + TN)/N = 0.90

Confusion Matrix B

Confusion Matrix for CV, Threshold = 0.45

N = 436		Actual Training Labels	
		Class "1"	Class "0"
Predicted Results	Class "1"	281	41
	Class "0"	4	110

Performance Metrics

TP = 281, FP = 41, FN = 4, TN = 110
Precision = TP/(TP + FP) = 0.87
Recall = TP/(TP + FN) = 0.99
F1 Score = HMean: Prec. & Recall = 0.93
Accuracy = (TP + TN)/N = 0.90

1 Compare the performance metrics of confusion matrix A (using a threshold p value of 0.75) with the confusion matrix in Exhibit 48 (using a threshold p value of 0.60).

2 Compare the performance metrics of confusion matrix B (using a threshold p value of 0.45) with the confusion matrix in Exhibit 48 (using a threshold p value of 0.60).

3 Contrast the performance metrics of confusion matrixes A and B, and explain the trade-offs implied between them.

Solution to 1:

Since confusion matrix A has fewer true positives (TPs) and fewer true negatives (TNs) than the confusion matrix in Exhibit 48 (281 vs. 284 and 110 vs. 114, respectively), confusion matrix A has lower accuracy and a lower F1 score compared to the one in Exhibit 48 (0.90 vs. 0.91 and 0.93 vs. 0.94, respectively). Also, although confusion matrix A has slightly better precision, 0.91 vs. 0.90, due to a few less false positives (FPs), it has significantly lower recall, 0.94 vs. 0.98, due to having many more false negatives (FNs), 17 vs. 7, than the confusion matrix in Exhibit 48. On balance, the ML model using the threshold p value of 0.60 is the superior model for this sentiment classification problem.

Solution to 2:

Confusion matrix B has the same number of TPs (281) and TNs (110) as confusion matrix A. Therefore, confusion matrix B also has lower accuracy (0.90) and a lower F1 score (0.93) compared to the one in Exhibit 48. Although confusion matrix B has slightly better recall, 0.99 vs. 0.98, due to fewer FNs, it has somewhat lower precision, 0.87 vs. 0.90, due to having many more FPs, 41 vs. 30, than the confusion matrix in Exhibit 48. Again, it is apparent that the ML model using the threshold p value of 0.60 is the better model in this sentiment classification context.

Solution to 3:

The main differences in performance metrics between confusion matrixes A and B are in precision and recall. Confusion matrix A has higher precision, at 0.91 vs. 0.87, but confusion matrix B has higher recall, at 0.99 vs. 0.94. These differences highlight the trade-off between FP (Type I error) and FN (Type II error). Precision is useful when the cost of FP is high, such as when an expensive product that is fine mistakenly fails quality inspection and is scrapped; in this case, FP should be minimized. Recall is useful when the cost of FN is high, such as when an expensive product is defective but mistakenly passes quality inspection and is sent to the customer; in this case, FN should be minimized. In the context of sentiment classification, FP might result in buying a stock for which sentiment is incorrectly classified as positive when it is actually negative.

Conversely, FN might result in avoiding (or even shorting) a stock for which the sentiment is incorrectly classified as negative when it is actually positive. The model behind the confusion matrix in Exhibit 48 strikes a balance in the trade-off between precision and recall.

SUMMARY

In this reading, we have discussed the major steps in big data projects involving the development of machine learning (ML) models—namely, those combining textual big data with structured inputs.

- Big data—defined as data with volume, velocity, variety, and potentially lower veracity—has tremendous potential for various fintech applications, including several related to investment management.
- The main steps for traditional ML model building are conceptualization of the problem, data collection, data preparation and wrangling, data exploration, and model training.
- For textual ML model building, the first four steps differ somewhat from those used in the traditional model: Text problem formulation, text curation, text preparation and wrangling, and text exploration are typically necessary.
- For structured data, data preparation and wrangling entail data cleansing and data preprocessing. Data cleansing typically involves resolving incompleteness errors, invalidity errors, inaccuracy errors, inconsistency errors, non-uniformity errors, and duplication errors.
- Preprocessing for structured data typically involves performing the following transformations: extraction, aggregation, filtration, selection, and conversion.
- Preparation and wrangling text (unstructured) data involves a set of text-specific cleansing and preprocessing tasks. Text cleansing typically involves removing the following: html tags, punctuations, most numbers, and white spaces.
- Text preprocessing requires performing normalization that involves the following: lowercasing, removing stop words, stemming, lemmatization, creating bag-of-words (BOW) and n-grams, and organizing the BOW and n-grams into a document term matrix (DTM).
- Data exploration encompasses exploratory data analysis, feature selection, and feature engineering. Whereas histograms, box plots, and scatterplots are common techniques for exploring structured data, word clouds are an effective way to gain a high-level picture of the composition of textual content. These visualization tools help share knowledge among the team (business subject matter experts, quants, technologists, etc.) to help derive optimal solutions.
- Feature selection methods used for text data include term frequency, document frequency, chi-square test, and a mutual information measure. Feature engineering for text data includes converting numbers into tokens, creating n-grams, and using name entity recognition and parts of speech to engineer new feature variables.
- The model training steps (method selection, performance evaluation, and model tuning) often do not differ much for structured versus unstructured data projects.

- Model selection is governed by the following factors: whether the data project involves labeled data (supervised learning) or unlabeled data (unsupervised learning); the type of data (numerical, continuous, or categorical; text data; image data; speech data; etc.); and the size of the dataset.
- Model performance evaluation involves error analysis using confusion matrixes, determining receiver operating characteristics, and calculating root mean square error.
- To carry out an error analysis for each model, a confusion matrix is created; true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) are determined. Then, the following performance metrics are calculated: accuracy, F1 score, precision, and recall. The higher the accuracy and F1 score, the better the model performance.
- To carry out receiver operating characteristic (ROC) analysis, ROC curves and area under the curve (AUC) of various models are calculated and compared. The more convex the ROC curve and the higher the AUC, the better the model performance.
- Model tuning involves managing the trade-off between model bias error, associated with underfitting, and model variance error, associated with overfitting. A fitting curve of in-sample (training sample) error and out-of-sample (cross-validation sample) error on the y-axis versus model complexity on the x-axis is useful for managing the bias vs. variance error trade-off.
- In a real-world big data project involving text data analysis for classifying and predicting sentiment of financial text for particular stocks, the text data are transformed into structured data for populating the DTM, which is then used as the input for the ML algorithm.
- To derive term frequency (TF) at the sentence level and TF–IDF, both of which can be inputs to the DTM, the following frequency measures should be used to create a term frequency measures table: TotalWordsInSentence; TotalWordCount; TermFrequency (Collection Level); WordCountInSentence; SentenceCountWithWord; Document Frequency; and Inverse Document Frequency.

PRACTICE PROBLEMS

The following information relates to Questions 1–15

Aaliyah Schultz is a fixed-income portfolio manager at Aries Investments. Schultz supervises Ameris Steele, a junior analyst.

A few years ago, Schultz developed a proprietary machine learning (ML) model that aims to predict downgrades of publicly-traded firms by bond rating agencies. The model currently relies only on structured financial data collected from different sources. Schultz thinks the model's predictive power may be improved by incorporating sentiment data derived from textual analysis of news articles and Twitter content relating to the subject companies.

Schultz and Steele meet to discuss plans for incorporating the sentiment data into the model. They discuss the differences in the steps between building ML models that use traditional structured data and building ML models that use textual big data. Steele tells Schultz:

Statement 1 The second step in building text-based ML models is text preparation and wrangling, whereas the second step in building ML models using structured data is data collection.

Statement 2 The fourth step in building both types of models encompasses data/text exploration.

Steele expresses concern about using Twitter content in the model, noting that research suggests that as much as 10%–15% of social media content is from fake accounts. Schultz tells Steele that she understands her concern but thinks the potential for model improvement outweighs the concern.

Steele begins building a model that combines the structured financial data and the sentiment data. She starts with cleansing and wrangling the raw structured financial data. Exhibit 1 presents a small sample of the raw dataset before cleansing: Each row represents data for a particular firm.

Exhibit 1 Sample of Raw Structured Data Before Cleansing

ID	Ticker	IPO Date	Industry (NAICS)	EBIT	Interest Expense	Total Debt
1	ABC	4/6/17	44	9.4	0.6	10.1
2	BCD	November 15, 2004	52	5.5	0.4	6.2
3	HIJ	26-Jun-74	54	8.9	1.2	15.8
4	KLM	14-Mar-15	72	5.7	1.5	0.0

After cleansing the data, Steele then preprocesses the dataset. She creates two new variables: an "Age" variable based on the firm's IPO date and an "Interest Coverage Ratio" variable equal to EBIT divided by interest expense. She also deletes the "IPO Date" variable from the dataset. After applying these transformations, Steele scales

© 2020 CFA Institute. All rights reserved.

the financial data using normalization. She notes that over the full sample dataset, the "Interest Expense" variable ranges from a minimum of 0.2 and a maximum of 12.2, with a mean of 1.1 and a standard deviation of 0.4.

Steele and Schultz then discuss how to preprocess the raw text data. Steele tells Schultz that the process can be completed in the following three steps:

Step 1 Cleanse the raw text data.

Step 2 Split the cleansed data into a collection of words for them to be normalized.

Step 3 Normalize the collection of words from Step 2 and create a distinct set of tokens from the normalized words.

With respect to Step 1, Steele tells Schultz:

> "I believe I should remove all html tags, punctuations, numbers, and extra white spaces from the data before normalizing them."

After properly cleansing the raw text data, Steele completes Steps 2 and 3. She then performs exploratory data analysis. To assist in feature selection, she wants to create a visualization that shows the most informative words in the dataset based on their term frequency (TF) values. After creating and analyzing the visualization, Steele is concerned that some tokens are likely to be noise features for ML model training; therefore, she wants to remove them.

Steele and Schultz discuss the importance of feature selection and feature engineering in ML model training. Steele tells Schultz:

> "Appropriate feature selection is a key factor in minimizing model overfitting, whereas feature engineering tends to prevent model underfitting."

Once satisfied with the final set of features, Steele selects and runs a model on the training set that classifies the text as having positive sentiment (Class "1" or negative sentiment (Class "0"). She then evaluates its performance using error analysis. The resulting confusion matrix is presented in Exhibit 2.

Exhibit 2 Confusion Matrix

		Actual Training Results	
		Class "1"	Class "0"
Predicted Results	Class "1"	TP = 182	FP = 52
	Class "0"	FN = 31	TN = 96

1 Which of Steele's statements relating to the steps in building structured data-based and text-based ML models is correct?

A Only Statement 1 is correct.

B Only Statement 2 is correct.

C Statement 1 and Statement 2 are correct.

2 Steele's concern about using Twitter data in the model *best* relates to:

A volume.

B velocity.

C veracity.

3 What type of error appears to be present in the IPO Date column of Exhibit 1?

A invalidity error.

B inconsistency error.

C non-uniformity error.

4 What type of error is most likely present in the last row of data (ID #4) in Exhibit 1?

A Inconsistency error

B Incompleteness error

C Non-uniformity error

5 During the preprocessing of the data in Exhibit 1, what type of data transformation did Steele perform during the data preprocessing step?

A Extraction

B Conversion

C Aggregation

6 Based on Exhibit 1, for the firm with ID #3, Steele should compute the scaled value for the "Interest Expense" variable as:

A 0.008.

B 0.083.

C 0.250.

7 Is Steele's statement regarding Step 1 of the preprocessing of raw text data correct?

A Yes.

B No, because her suggested treatment of punctuation is incorrect.

C No, because her suggested treatment of extra white spaces is incorrect.

8 Steele's Step 2 can be *best* described as:

A tokenization.

B lemmatization.

C standardization.

9 The output created in Steele's Step 3 can be *best* described as a:

A bag-of-words.

B set of n-grams.

C document term matrix.

10 Given her objective, the visualization that Steele should create in the exploratory data analysis step is a:

A scatter plot.

B word cloud.

C document term matrix.

11 To address her concern in her exploratory data analysis, Steele should focus on those tokens that have:

A low chi-square statistics.

B low mutual information (ML) values.

C very low and very high term frequency (TF) values.

12 Is Steele's statement regarding the relationship between feature selection/feature engineering and model fit correct?

A Yes.

B No, because she is incorrect with respect to feature selection.

C No, because she is incorrect with respect to feature engineering.

13 Based on Exhibit 2, the model's precision metric is *closest* to:

A 78%.

B 81%.

C 85%.

14 Based on Exhibit 2, the model's F1 score is *closest* to:

A 77%.

B 81%.

C 85%.

15 Based on Exhibit 2, the model's accuracy metric is *closest* to:

A 77%.

B 81%.

C 85%.

The following information relates to Questions 16–22

Iesha Azarov is a senior analyst at Ganymede Moon Partners (Ganymede), where he works with junior analyst Pàola Bector. Azarov would like to incorporate machine learning (ML) models into the company's analytical process. Azarov asks Bector to develop ML models for two unstructured stock sentiment datasets, Dataset ABC and Dataset XYZ. Both datasets have been cleaned and preprocessed in preparation for text exploration and model training.

Following an exploratory data analysis that revealed Dataset ABC's most frequent tokens, Bector conducts a collection frequency analysis. Bector then computes TF–IDF (term frequency–inverse document frequency) for several words in the collection and tells Azarov the following:

Statement 1 IDF is equal to the inverse of the document frequency measure.

Statement 2 TF at the collection level is multiplied by IDF to calculate TF–IDF.

Statement 3 TF–IDF values vary by the number of documents in the dataset, and therefore, model performance can vary when applied to a dataset with just a few documents.

Bector notes that Dataset ABC is characterized by the absence of ground truth.

Bector turns his attention to Dataset XYZ, containing 84,000 tokens and 10,000 sentences. Bector chooses an appropriate feature selection method to identify and remove unnecessary tokens from the dataset and then focuses on model training. For performance evaluation purposes, Dataset XYZ is split into a training set, cross-validation (CV) set, and test set. Each of the sentences has already been labeled as either a positive sentiment (Class "1") or a negative sentiment (Class "0") sentence. There is an unequal class distribution between the positive sentiment and negative sentiment sentences in Dataset XYZ. Simple random sampling is applied within levels of the sentiment class labels to balance the class distributions within the splits. Bector's view is that the false positive and false negative evaluation metrics should be given equal weight. Select performance data from the cross-validation set confusion matrices is presented in Exhibit 1:

Exhibit 1 Performance Metrics for Dataset XYZ

Confusion Matrix	CV Data (threshold *p*-value)	Performance Metrics			
		Precision	Recall	F1 Score	Accuracy
A	0.50	0.95	0.87	0.91	0.91
B	0.35	0.93	0.90	0.91	0.92
C	0.65	0.86	0.97	0.92	0.91

Azarov and Bector evaluate the Dataset XYZ performance metrics for Confusion Matrices A, B, and C in Exhibit 1. Azarov says, "For Ganymede's purposes, we should be most concerned with the cost of Type I errors."

Azarov requests that Bector apply the ML model to the test dataset for Dataset XYZ, assuming a threshold *p*-value of 0.65. Exhibit 2 contains a sample of results from the test dataset corpus.

Exhibit 2 10 Sample Results of Test Data for Dataset XYZ

Sentence #	Actual Sentiment	Target *p*-Value
1	1	0.75
2	0	0.45
3	1	0.64
4	1	0.81
5	0	0.43
6	1	0.78
7	0	0.59
8	1	0.60
9	0	0.67
10	0	0.54

Bector makes the following remarks regarding model training:

Remark 1: Method selection is governed by such factors as the type of data and the size of data.

Remark 2: In the performance evaluation stage, model fitting errors, such as bias error and variance error, are used to measure goodness of fit.

16 Based on the text exploration method used for Dataset ABC, tokens that potentially carry important information useful for differentiating the sentiment embedded in the text are *most likely* to have values that are:

A low.

B intermediate.

C high.

17 Which of Bector's statements regarding TF, IDF, and TF–IDF is correct?

A Statement 1

B Statement 2

C Statement 3

18 What percentage of Dataset ABC should be allocated to a training subset?

A 0%

B 20%

C 60%

19 Based only on Dataset XYZ's composition and Bector's view regarding false positive and false negative evaluation metrics, which performance measure is *most appropriate*?

A Recall

B F1 score

C Precision

20 Based on Exhibit 1, which confusion matrix demonstrates the *most* favorable value of the performance metric that *best* addresses Azarov's concern?

A Confusion Matrix A

B Confusion Matrix B

C Confusion Matrix C

21 Based on Exhibit 2, the accuracy metric for Dataset XYZ's test set sample is *closest to*:

A 0.67.

B 0.70.

C 0.75.

22 Which of Bector's remarks related to model training is correct?

A Only Remark 1

B Only Remark 2

C Both Remark 1 and Remark 2

The following information relates to Questions 23–31

Bernadette Rivera is a portfolio manager at Voxkor, a private equity company that provides financing to early-stage start-up businesses. Rivera is working with a data analyst, Tim Achler, on a text-based machine-learning (ML) model to enhance Voxkor's predictive ability to identify successful start-ups.

Voxkor currently uses ML models based only on traditional, structured financial data but would like to develop a new ML model that analyzes textual big data gathered from the internet. The model will classify text information into positive or negative sentiment classes for each respective start-up. Rivera wants to confirm her understanding of any differences in the ML model building steps between data analysis projects that use traditional structured data and projects that involve unstructured, text-based data. Rivera makes the following statements:

Statement 1 Some of the methods used in the exploration step are different for structured and unstructured data, but for both types of data, the step involves feature selection and feature engineering.

Statement 2 A major difference when developing a text-based ML model is the curation step, which involves cleansing, preprocessing, and converting the data into a structured format usable for model training.

Achler uses a web spidering program to obtain the data for the text-based model. The program extracts raw content from social media webpages, which contains English language sentences and special characters. After curating the text, Achler removes unnecessary elements from the raw text using regular expression software and completes additional text cleansing and preprocessing tasks.

Next, Achler and Rivera discuss remaining text wrangling tasks—specifically, which tokens to include in the document term matrix (DTM). Achler divides unique tokens into three groups; a sample of each group is shown in Exhibit 1.

Exhibit 1 Summary of Sample Tokens

Token Group 1	Token Group 2	Token Group 3
"not_increas_market"	"not_increased_market"	"not," "increased," "market"
"currencysign"	"currencysign"	"EUR"
"sale_decreas"	"sale_decreased"	"Sales," "decreased"

The dataset is now ready for the text exploration step. At this point in the process, Rivera wants to better comprehend the collection of unique words. Achler recommends an exploratory data analysis technique that visualizes words by varying their font size proportionately to the number of occurrences of each word in the corpus.

As an additional part of the text exploration step, Achler conducts a term frequency analysis to identify outliers. Achler summarizes the analysis in Exhibit 2.

Exhibit 2 Words with Highest and Lowest Frequency Value

Group 1		Group 2	
Word	Frequency	Word	Frequency
the	0.04935	naval	1.0123e-05
and	0.04661	stereotype	1.5185e-05
to	0.04179	till	1.5185e-05
that	0.03577	ribbon	2.0247e-05
in	0.03368	deposit	2.5308e-05

Note: "e-05" represents 10^{-5}.

Achler has the data ready for the model training process. Rivera asks Achler to include start-up failure rates as a feature. Achler notices that the number of start-ups that fail (majority class) is significantly larger than the number of the start-ups that are successful (minority class). Achler is concerned that because of class imbalance, the model will not be able to discriminate between start-ups that fail and start-ups that are successful.

Achler splits the DTM into training, cross-validation, and test datasets. Achler uses a supervised learning approach to train the logistic regression model in predicting sentiment. Applying the receiver operating characteristics (ROC) technique and area under the curve (AUC) metrics, Achler evaluates model performance on both the training and the cross-validation datasets. The trained model performance for three different logistic regressions' threshold *p*-values is presented in Exhibit 3.

Exhibit 3 AUC for Different Threshold p-values

Threshold *p*-Value	Training Set	Cross-Validation Set
$p = 0.57$	56.7%	57.3%
$p = 0.79$	91.3%	89.7%
$p = 0.84$	98.4%	87.1%

Rivera suggests adjusting the model's hyperparameters to improve performance. Achler runs a grid search that compares the difference between the prediction error on both the training and the cross-validation datasets for various combinations of hyperparameter values. For the current values of hyperparameters, Achler observes that the prediction error on the training dataset is small, whereas the prediction error on the cross-validation dataset is significantly larger.

23 Which of Rivera's statements about differences in ML model building steps is correct?

- **A** Only Statement 1
- **B** Only Statement 2
- **C** Both Statement 1 and Statement 2

24 Based on the source of the data, as part of the data cleansing and wrangling process, Achler *most likely* needs to remove:

- **A** html tags and perform scaling.
- **B** numbers and perform lemmatization.
- **C** white spaces and perform winsorization.

25 Based on Exhibit 1, which token group has *most likely* undergone the text preparation and wrangling process?

- **A** Token Group 1
- **B** Token Group 2
- **C** Token Group 3

26 The visual text representation technique that Achler recommends to Rivera is a:

- **A** word cloud.
- **B** bag of words.
- **C** collection frequency.

27 Based on Exhibit 2, Achler should exclude from further analysis words in:

- **A** only Group 1.
- **B** only Group 2.
- **C** both Group 1 and Group 2.

28 Achler's model training concern related to the model's ability to discriminate could be addressed by randomly:

- **A** oversampling the failed start-up data.
- **B** oversampling the successful start-up data.
- **C** undersampling the successful start-up data.

29 Based on Exhibit 3, which threshold *p*-value indicates the *best* fitting model?

A 0.57

B 0.79

C 0.84

30 Based on Exhibit 3, if Achler wants to improve model performance at the threshold *p*-value of 0.84, he should:

A tune the model to lower the AUC.

B adjust model parameters to decrease ROC convexity.

C apply LASSO regularization to the logistic regression.

31 Based on Achler's grid search analysis, the current model can be characterized as:

A underfitted.

B having low variance.

C exhibiting slight regularization.

SOLUTIONS

1 B is correct. The five steps in building structured data-based ML models are: 1) conceptualization of the modeling task, 2) data collection, 3) data preparation and wrangling, 4) data exploration, and 5) model training. The five steps in building text-based ML models are: 1) text problem formulation, 2) data (text) curation, 3) text preparation and wrangling, 4) text exploration, and 5) model training. Statement 1 is incorrect: Text preparation and wrangling is the third step in building text ML models and occurs after the second data (text) curation step. Statement 2 is correct: The fourth step in building both types of models encompasses data/text exploration.

2 C is correct. Veracity relates to the credibility and reliability of different data sources. Steele is concerned about the credibility and reliability of Twitter content, noting that research suggests that as much as 10%–15% of social media content is from fake accounts.

3 C is correct. A non-uniformity error occurs when the data are not presented in an identical format. The data in the "IPO Date" column represent the IPO date of each firm. While all rows are populated with valid dates in the IPO Date column, the dates are presented in different formats (e.g., mm/dd/yyyy, dd/mm/yyyy).

4 A is correct. There appears to be an inconsistency error in the last row (ID #4). An inconsistency error occurs when a data point conflicts with corresponding data points or reality. In the last row, the interest expense data item has a value of 1.5, and the total debt item has a value of 0.0. This appears to be an error: Firms that have interest expense are likely to have debt in their capital structure, so either the interest expense is incorrect or the total debt value is incorrect. Steele should investigate this issue by using alternative data sources to confirm the correct values for these variables.

5 A is correct. During the data preprocessing step, Steele created a new "Age" variable based on the firm's IPO date and then deleted the "IPO Date" variable from the dataset. She also created a new "Interest Coverage Ratio" variable equal to EBIT divided by interest expense. Extraction refers to a data transformation where a new variable is extracted from a current variable for ease of analyzing and using for training an ML model, such as creating an age variable from a date variable or a ratio variable. Steele also performed a selection transformation by deleting the IPO Date variable, which refers to deleting the data columns that are not needed for the project.

6 B is correct. Steele uses normalization to scale the financial data. Normalization is the process of rescaling numeric variables in the range of [0, 1]. To normalize variable X, the minimum value (X_{min}) is subtracted from each observation (X_i), and then this value is divided by the difference between the maximum and minimum values of X ($X_{max} - X_{min}$):

$$X_{i\ (normalized)} = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

The firm with ID #3 has an interest expense of 1.2. So, its normalized value is calculated as:

$$X_{i\ (normalized)} = \frac{1.2 - 0.2}{12.2 - 0.2} = 0.083$$

7 B is correct. Although most punctuations are not necessary for text analysis and should be removed, some punctuations (e.g., percentage signs, currency symbols, and question marks) may be useful for ML model training. Such punctuations should be substituted with annotations (e.g., /percentSign/, /dollarSign/, and /questionMark/) to preserve their grammatical meaning in the text. Such annotations preserve the semantic meaning of important characters in the text for further text processing and analysis stages.

8 A is correct. Tokenization is the process of splitting a given text into separate tokens. This step takes place after cleansing the raw text data (removing html tags, numbers, extra white spaces, etc.). The tokens are then normalized to create the bag-of-words (BOW).

9 A is correct. After the cleansed text is normalized, a bag-of-words is created. A bag-of-words (BOW) is a collection of a distinct set of tokens from all the texts in a sample dataset.

10 B is correct. Steele wants to create a visualization for Schultz that shows the most informative words in the dataset based on their term frequency (TF, the ratio of the number of times a given token occurs in the dataset to the total number of tokens in the dataset) values. A word cloud is a common visualization when working with text data as it can be made to visualize the most informative words and their TF values. The most commonly occurring words in the dataset can be shown by varying font size, and color is used to add more dimensions, such as frequency and length of words.

11 C is correct. Frequency measures can be used for vocabulary pruning to remove noise features by filtering the tokens with very high and low TF values across all the texts. Noise features are both the most frequent and most sparse (or rare) tokens in the dataset. On one end, noise features can be stop words that are typically present frequently in all the texts across the dataset. On the other end, noise features can be sparse terms that are present in only a few text files. Text classification involves dividing text documents into assigned classes. The frequent tokens strain the ML model to choose a decision boundary among the texts as the terms are present across all the texts (an example of underfitting). The rare tokens mislead the ML model into classifying texts containing the rare terms into a specific class (an example of overfitting). Thus, identifying and removing noise features are critical steps for text classification applications.

12 A is correct. A dataset with a small number of features may not carry all the characteristics that explain relationships between the target variable and the features. Conversely, a large number of features can complicate the model and potentially distort patterns in the data due to low degrees of freedom, causing overfitting. Therefore, appropriate feature selection is a key factor in minimizing such model overfitting. Feature engineering tends to prevent underfitting in the training of the model. New features, when engineered properly, can elevate the underlying data points that better explain the interactions of features. Thus, feature engineering can be critical to overcome underfitting.

13 A is correct. Precision, the ratio of correctly predicted positive classes (true positives) to all predicted positive classes, is calculated as:

$$\text{Precision (P)} = \text{TP}/(\text{TP} + \text{FP}) = 182/(182 + 52) = 0.7778\ (78\%).$$

14 B is correct. The model's F1 score, which is the harmonic mean of precision and recall, is calculated as:

$$\text{F1 score} = (2 \times \text{P} \times \text{R})/(\text{P} + \text{R}).$$

$$\text{F1 score} = (2 \times 0.7778 \times 0.8545)/(0.7778 + 0.8545) = 0.8143\ (81\%).$$

15 A is correct. The model's accuracy, which is the percentage of correctly predicted classes out of total predictions, is calculated as:

Accuracy = (TP + TN)/(TP + FP + TN + FN).

Accuracy = (182 + 96)/(182 + 52 + 96 + 31) = 0.7701 (77%).

16 B is correct. When analyzing term frequency at the corpus level, also known as collection frequency, tokens with intermediate term frequency (TF) values potentially carry important information useful for differentiating the sentiment embedded in the text. Tokens with the highest TF values are mostly stop words that do not contribute to differentiating the sentiment embedded in the text, and tokens with the lowest TF values are mostly proper nouns or sparse terms that are also not important to the meaning of the text.

A is incorrect because tokens with the lowest TF values are mostly proper nouns or sparse terms (noisy terms) that are not important to the meaning of the text.

C is incorrect because tokens with the highest TF values are mostly stop words (noisy terms) that do not contribute to differentiating the sentiment embedded in the text.

17 C is correct. Statement 3 is correct. TF–IDF values vary by the number of documents in the dataset, and therefore, the model performance can vary when applied to a dataset with just a few documents.

Statement 1 is incorrect because IDF is calculated as the log of the inverse, or reciprocal, of the document frequency measure. Statement 2 is incorrect because TF at the sentence (not collection) level is multiplied by IDF to calculate TF–IDF.

A is incorrect because Statement 1 is incorrect. IDF is calculated as the log of the inverse, or reciprocal, of the document frequency (DF) measure.

B is incorrect because Statement 2 is incorrect. TF at the sentence (not collection) level is multiplied by IDF to calculate TF–IDF.

18 A is correct; 0% of the master dataset of Dataset ABC should be allocated to a training subset. Dataset ABC is characterized by the absence of ground truth (i.e., no known outcome or target variable) and is therefore an unsupervised ML model. For unsupervised learning models, no splitting of the master dataset is needed, because of the absence of labeled training data. Supervised ML datasets (with labeled training data) contain ground truth, the known outcome (target variable) of each observation in the dataset.

B is incorrect because 20% is the commonly recommended split for the cross-validation set and test set in supervised training ML datasets. Dataset ABC is an unsupervised ML dataset, for which no splitting (0%) of the master dataset is needed, because of the absence of labeled training data. In supervised ML models (which contain labeled training data), the master dataset is split into three subsets (a training set, cross-validation set, and test set), which are used for model training and testing purposes.

C is incorrect because 60% is the commonly recommended split for the training set in supervised training ML datasets. Dataset ABC is an unsupervised ML dataset, for which no splitting (0%) of the master dataset is needed, because of the absence of labeled training data. In supervised ML models (which contain labeled training data), the master dataset is split into three subsets (a training set, cross-validation set, and test set), which are used for model training and testing purposes.

19 B is correct. F1 score is the most appropriate performance measure for Dataset XYZ. Bector gives equal weight to false positives and false negatives. Accuracy and F1 score are overall performance measures that give equal weight to false positives and false negatives. Accuracy is considered an appropriate performance measure for balanced datasets, where the number of "1" and "0" classes are equal. F1 score is considered more appropriate than accuracy when there is unequal class distribution in the dataset and it is necessary to measure the equilibrium of precision and recall. Since Dataset XYZ contains an unequal class distribution between positive and negative sentiment sentences, F1 score is the most appropriate performance measure.

Precision is the ratio of correctly predicted positive classes to all predicted positive classes and is useful in situations where the cost of false positives or Type I errors is high. Recall is the ratio of correctly predicted positive classes to all actual positive classes and is useful in situations where the cost of false negatives or Type II errors is high.

A is incorrect because Bector gives equal weight to false positives and false negatives. Accuracy and F1 score are overall performance measures that give equal weight to false positives and false negatives. Recall is the ratio of correctly predicted positive classes to all actual positive classes and is useful in situations where the cost of false negatives or Type II errors is high.

C is incorrect because Bector gives equal weight to false positive and false negatives. Accuracy and F1 score are overall performance measures that give equal weight to false positives and false negatives. Precision is the ratio of correctly predicted positive classes to all predicted positive classes and is useful in situations where the cost of false positives or Type-I error is high.

20 A is correct. Precision is the ratio of correctly predicted positive classes to all predicted positive classes and is useful in situations where the cost of false positives or Type I errors is high. Confusion Matrix A has the highest precision and therefore demonstrates the most favorable value of the performance metric that best addresses Azarov's concern about the cost of Type I errors. Confusion Matrix A has a precision score of 0.95, which is higher than the precision scores of Confusion Matrix B (0.93) and Confusion Matrix C (0.86).

B is incorrect because precision, not accuracy, is the performance measure that best addresses Azarov's concern about the cost of Type I errors. Confusion Matrix B demonstrates the most favorable value for the accuracy score (0.92), which is higher than the accuracy scores of Confusion Matrix A (0.91) and Confusion Matrix C (0.91). Accuracy is a performance measure that gives equal weight to false positives and false negatives and is considered an appropriate performance measure when the class distribution in the dataset is equal (a balanced dataset). However, Azarov is most concerned with the cost of false positives, or Type I errors, and not with finding the equilibrium between precision and recall. Furthermore, Dataset XYZ has an unequal (unbalanced) class distribution between positive sentiment and negative sentiment sentences.

C is incorrect because precision, not recall or F1 score, is the performance measure that best addresses Azarov's concern about the cost of Type I errors. Confusion Matrix C demonstrates the most favorable value for the recall score (0.97), which is higher than the recall scores of Confusion Matrix A (0.87) and Confusion Matrix B (0.90). Recall is the ratio of correctly predicted positive classes to all actual positive classes and is useful in situations where the cost of false negatives, or Type II errors, is high. However, Azarov is most concerned with the cost of Type I errors, not Type II errors.

F1 score is more appropriate (than accuracy) when there is unequal class distribution in the dataset and it is necessary to measure the equilibrium of precision and recall. Confusion Matrix C demonstrates the most favorable value for the F1 score (0.92), which is higher than the F1 scores of Confusion Matrix A (0.91) and Confusion Matrix B (0.91). Although Dataset XYZ has an unequal class distribution between positive sentiment and negative sentiment sentences, Azarov is most concerned with the cost of false positives, or Type I errors, and not with finding the equilibrium between precision and recall.

21 B is correct. Accuracy is the percentage of correctly predicted classes out of total predictions and is calculated as (TP + TN)/(TP + FP + TN + FN).

In order to obtain the values for true positive (TP), true negative (TN), false positive (FP), and false negative (FN), predicted sentiment for the positive (Class “1”) and the negative (Class “0”) classes are determined based on whether each individual target *p*-value is greater than or less than the threshold *p*-value of 0.65. If an individual target *p*-value is greater than the threshold *p*-value of 0.65, the predicted sentiment for that instance is positive (Class “1”). If an individual target *p*-value is less than the threshold *p*-value of 0.65, the predicted sentiment for that instance is negative (Class “0”). Actual sentiment and predicted sentiment are then classified as follows:

Actual Sentiment	Predicted Sentiment	Classification
1	1	TP
0	1	FP
1	0	FN
0	0	TN

Exhibit 2, with added “Predicted Sentiment” and “Classification” columns, is presented below:

Exhibit 2 10 Sample Results of Test Data for Dataset XYZ

Sentence #	Actual Sentiment	Target *p*-Value	Predicted Sentiment	Classification
1	1	0.75	1	TP
2	0	0.45	0	TN
3	1	0.64	0	FN
4	1	0.81	1	TP
5	0	0.43	0	TN
6	1	0.78	1	TP
7	0	0.59	0	TN
8	1	0.60	0	FN
9	0	0.67	1	FP
10	0	0.54	0	TN

Based on the classification data obtained from Exhibit 2, a confusion matrix can be generated:

Confusion Matrix for Dataset XYZ Sample Test Data with Threshold *p*-Value = 0.65

	Actual Training Labels	
Predicted Results	Class “1”	Class “0”
Class “1”	TP = 3	FP = 1
Class “0”	FN = 2	TN = 4

Using the data in the confusion matrix above, the accuracy metric is computed as follows:

Accuracy = (TP + TN)/(TP + FP + TN + FN).

Accuracy = (3 + 4)/(3 + 1 + 4 + 2) = 0.70.

A is incorrect because 0.67 is the F1 score, not accuracy metric, for the sample of the test set for Dataset XYZ, based on Exhibit 2. To calculate the F1 score, the precision (P) and the recall (R) ratios must first be calculated. Precision and recall for the sample of the test set for Dataset XYZ, based on Exhibit 2, are calculated as follows:

Precision (P) = TP/(TP + FP) = 3/(3 + 1) = 0.75.

Recall (R) = TP/(TP + FN) = 3/(3 + 2) = 0.60.

The F1 score is calculated as follows:

F1 score = (2 × P × R)/(P + R) = (2 × 0.75 × 0.60)/(0.75 + 0.60)

= 0.667, or 0.67.

C is incorrect because 0.75 is the precision ratio, not the accuracy metric, for the sample of the test set for Dataset XYZ, based on Exhibit 2. The precision score is calculated as follows:

Precision (P) = TP/(TP + FP) = 3/(3 + 1) = 0.75.

22 A is correct. Only Remark 1 is correct. Method selection is the first task of ML model training and is governed by the following factors: (1) supervised or unsupervised learning, (2) the type of data, and (3) the size of data. The second and third tasks of model training, respectively, are performance evaluation and tuning.

Remark 2 is incorrect because model fitting errors (bias error and variance error) are used in tuning, not performance evaluation. The techniques used in performance evaluation, which measure the goodness of fit for validation of the model, include (1) error analysis, (2) receiver operating characteristic (ROC) plots, and (3) root mean squared error (RMSE) calculations.

B and C are incorrect because Remark 2 is incorrect. Model fitting errors (bias error and variance error) are used in tuning, not performance evaluation. The techniques used in performance evaluation, which measure the goodness of fit for validation of the model, include (1) error analysis, (2) receiver operating characteristic plots, and (3) root mean squared error calculations.

23 A is correct. Statement 1 is correct because some of the methods used in the fourth step of ML model building (data/text exploration) are different for structured and unstructured data, and for both structured and unstructured data, the exploration step encompasses feature selection and feature engineering. Statement 2 is incorrect because Rivera described the text preparation and wrangling step, not the text curation step. The data (text) curation step involves gathering relevant external text data via web services or programs that extract raw content from a source.

B and C are incorrect because Statement 2 is incorrect. Rivera described the text preparation and wrangling step, not the text curation step. The data (text) curation step involves gathering relevant external text data via web services or programs that extract raw content from a source.

24 B is correct. Achler uses a web spidering program that extracts unstructured raw content from social media webpages. Raw text data are a sequence of characters and contain other non-useful elements including html tags, punctuation, and white spaces (including tabs, line breaks, and new lines). Removing numbers is one of the basic operations in the text cleansing/preparation process for unstructured data. When numbers (or digits) are present in the text, they should be removed or substituted with the annotation "/number/." Lemmatization, which takes places during the text wrangling/preprocessing process for unstructured data, is the process of converting inflected forms of a word into its morphological root (known as lemma). Lemmatization reduces the repetition of words occurring in various forms while maintaining the semantic structure of the text data, thereby aiding in training less complex ML models.

A is incorrect because although html tag removal is part of text cleansing/preparation for unstructured data, scaling is a data wrangling/preprocessing process applied to structured data. Scaling adjusts the range of a feature by shifting and changing the scale of data; it is performed on numeric variables, not on text data.

C is incorrect because although raw text contains white spaces (including tabs, line breaks, and new lines) that need to be removed as part of the data cleansing/preparation process for unstructured data, winsorization is a data wrangling/preprocessing task performed on values of data points, not on text data. Winsorization is used for structured numerical data and replaces extreme values and outliers with the maximum (for large-value outliers) and minimum (for small-value outliers) values of data points that are not outliers.

25 A is correct. Data preparation and wrangling involve cleansing and organizing raw data into a consolidated format. Token Group 1 includes n-grams ("not_increas_market," "sale_decreas") and the words that have been converted from their inflected forms into their base word ("increas," "decreas"), and the currency symbol has been replaced with a "currencysign" token. N-gram tokens are helpful for keeping negations intact in the text, which is vital for sentiment prediction. The process of converting inflected forms of a word into its base word is called stemming and helps decrease data sparseness, thereby aiding in training less complex ML models.

B is incorrect because Token Group 2 includes inflected forms of words ("increased," "decreased") before conversion into their base words (known as stems). Stemming (along with lemmatization) decreases data sparseness by aggregating many sparsely occurring words in relatively less sparse stems or lemmas, thereby aiding in training less complex ML models.

C is incorrect because Token Group 3 includes inflected forms of words ("increased," "decreased") before conversion into their base words (known as stems). In addition, the "EUR" currency symbol has not been replaced with the "currencysign" token and the word "Sales" has not been lowercased.

26 A is correct. Achler recommends creating a word cloud, which is a common text visualization technique at the data exploration phase in ML model building. The most commonly occurring words in the dataset can be visualized by varying font size, and color is used to add more dimensions, such as frequency and length of words.

B is incorrect because Achler recommends creating a word cloud and not a bag of words (BOW). A BOW is a collection of a distinct set of tokens from all the texts in a sample dataset. A BOW representation is a basic procedure used primarily to analyze text during Step 3 (text wrangling/preprocessing), although it may also be used in Step 4 during the feature engineering process. In contrast to a word cloud, which visually varies font size and color, BOW is simply a set of words (typically displayed in table).

C is incorrect because Achler recommends creating a word cloud and not a collection frequency. Collection frequency (or term frequency) is the ratio of the number of times a given token occurs in all the texts in the dataset to the total number of tokens in the dataset. Collection frequency can be calculated and examined to identify outlier words, but it is not a visual text representation tool.

27 C is correct. Achler should remove words that are in both Group 1 and Group 2. Term frequency values range between 0 and 1. Group 1 consists of the highest frequency values (e.g., "the" = 0.04935), and Group 2 consists of the lowest frequency values (e.g., "naval" = 1.0123e–05). Frequency analysis on the processed text data helps in filtering unnecessary tokens (or features) by quantifying how important tokens are in a sentence and in the corpus as a whole. The most frequent tokens (Group 1) strain the machine-learning model to choose a decision boundary among the texts as the terms are present across all the texts, which leads to model underfitting. The least frequent tokens (Group 2) mislead the machine-learning model into classifying texts containing the rare terms into a specific class, which leads to model overfitting. Identifying and removing noise features is critical for text classification applications.

A is incorrect because words in both Group 1 and Group 2 should be removed. The words with high term frequency value are mostly stop words, present in most sentences. Stop words do not carry a semantic meaning for the purpose of text analyses and ML training, so they do not contribute to differentiating sentiment.

B is incorrect because words in both Group 1 and Group 2 should be removed. Terms with low term frequency value are mostly rare terms, ones appearing only once or twice in the data. They do not contribute to differentiating sentiment.

28 B is correct. Achler is concerned about class imbalance, which can be resolved by balancing the training data. The majority class (the failed start-up data) can be randomly undersampled, or the minority class (the successful start-up data) can be randomly oversampled.

29 B is correct. The higher the AUC, the better the model performance. For the threshold p-value of 0.79, the AUC is 91.3% on the training dataset and 89.7% on the cross-validation dataset, and the ROC curves are similar for model performance on both datasets. These findings suggest that the model performs similarly on both training and CV data and thus indicate a good fitting model.

A is incorrect because for the threshold p-value of 0.57, the AUC is 56.7% on the training dataset and 57.3% on the cross-validation dataset. The AUC close to 50% signifies random guessing on both the training dataset and the cross-validation dataset. The implication is that for the threshold p-value of 0.57, the model is randomly guessing and is not performing well.

C is incorrect because for the threshold p-value of 0.84, there is a substantial difference between the AUC on the training dataset (98.4%) and the AUC on the cross-validation dataset (87.1%). This suggests that the model performs comparatively poorly (with a higher rate of error or misclassification) on the cross-validation dataset when compared with training data. Thus, the implication is that the model is overfitted.

30 C is correct. At the threshold p-value of 0.84, the AUC is 98.4% for the training dataset and 87.1% for the cross-validation dataset, which suggests that the model is currently overfitted. Least absolute shrinkage and selection operator (LASSO) regularization can be applied to the logistic regression to prevent overfitting of logistic regression models.

A is incorrect because the higher the AUC, the better the model performance.

B is incorrect because the more convex the ROC curve and the higher the AUC, the better the model performance. Adjusting model parameters with the aim of achieving lower ROC convexity would result in worse model performance on the cross-validation dataset.

31 C is correct. Slight regularization occurs when the prediction error on the training dataset is small, while the prediction error on the cross-validation data set is significantly larger. This difference in error is variance. High variance error, which typically is due to too many features and model complexity, results in model overfitting.

A is incorrect. The current model has high variance which results in model overfitting, not underfitting.

B is incorrect. The difference between the prediction error on the training dataset and the prediction error on the cross-validation dataset is high, which means that the current model has high variance, not low.

Economics

STUDY SESSION

Study Session 3	Economics

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to explain and demonstrate the use of economic concepts and methods in the determination and forecasting of currency exchange rates, the analysis of economic growth, and the analysis of business and financial market regulation.

A country's exchange rates, level of economic activity, and regulatory environment have significant implications for companies operating within its borders. Although predicting exchange rates is extremely difficult, exchange rate equilibrium relationships provide valuable insights for understanding the currency risks inherent in overseas operations and international investments.

© 2021 CFA Institute. All rights reserved.

ECONOMICS
STUDY SESSION 3

Economics

This study session begins with fundamental foreign exchange concept and theories of exchange rate determination. As a means to understanding exchange rate risk exposures, discussion centers on theoretical long-term equilibrium values. International parity conditions and the carry trade are described. Both direct (capital controls, foreign exchange intervention) and indirect (monetary, fiscal policy) exchange rate influencers are considered. A discussion of long-term growth and its relationship to investment returns follows. The three theories of growth (classical, neoclassical, endogenous) are presented. The session concludes with an overview of regulation, including the types of regulation, roles played by regulation, and considerations to use when evaluating the effects of regulation on an industry.

READING ASSIGNMENTS

Reading 6	Currency Exchange Rates: Understanding Equilibrium Value by Michael R. Rosenberg and William A. Barker, PhD, CFA
Reading 7	Economic Growth by Paul R. Kutasovic, PhD, CFA
Reading 8	Economics of Regulation by Chester S. Spatt, PhD

© 2021 CFA Institute. All rights reserved.

READING

6

Currency Exchange Rates: Understanding Equilibrium Value

by Michael R. Rosenberg and William A. Barker, PhD, CFA

Michael R. Rosenberg (USA). William A. Barker, PhD, CFA (Canada).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** calculate and interpret the bid–offer spread on a spot or forward currency quotation and describe the factors that affect the bid–offer spread;
☐	**b.** identify a triangular arbitrage opportunity and calculate its profit, given the bid–offer quotations for three currencies;
☐	**c.** explain spot and forward rates and calculate the forward premium/discount for a given currency;
☐	**d.** calculate the mark-to-market value of a forward contract;
☐	**e.** explain international parity conditions (covered and uncovered interest rate parity, forward rate parity, purchasing power parity, and the international Fisher effect);
☐	**f.** describe relations among the international parity conditions;
☐	**g.** evaluate the use of the current spot rate, the forward rate, purchasing power parity, and uncovered interest parity to forecast future spot exchange rates;
☐	**h.** explain approaches to assessing the long-run fair value of an exchange rate;
☐	**i.** describe the carry trade and its relation to uncovered interest rate parity and calculate the profit from a carry trade;
☐	**j.** explain how flows in the balance of payment accounts affect currency exchange rates;
☐	**k.** explain the potential effects of monetary and fiscal policy on exchange rates;
☐	**l.** describe objectives of central bank or government intervention and capital controls and describe the effectiveness of intervention and capital controls;
☐	**m.** describe warning signs of a currency crisis.

© 2020 CFA Institute. All rights reserved.

1 INTRODUCTION

Exchange rates are well known to follow a random walk, whereby fluctuations from one day to the next are unpredictable. The business of currency forecasting can be a humbling experience. Alan Greenspan, former chairman of the US Federal Reserve Board, famously noted that "having endeavored to forecast exchange rates for more than half a century, I have understandably developed significant humility about my ability in this area."

Hence, our discussion is not about predicting exchange rates but about the tools the reader can use to better understand long-run equilibrium value. This outlook helps guide the market participant's decisions with respect to risk exposures, as well as whether currency hedges should be implemented and, if so, how they should be managed. After discussing the basics of exchange rate transactions, we present the main theories for currency determination—starting with the international parity conditions—and then describe other important influences, such as current account balances, capital flows, and monetary and fiscal policy.

Although these fundamentals-based models usually perform poorly in predicting future exchange rates in the short run, they are crucial for understanding long-term currency value. Thus, we proceed as follows:

- We review the basic concepts of the foreign exchange market covered in the CFA Program Level I curriculum and expand this previous coverage to incorporate more material on bid–offer spreads.
- We then begin to examine determinants of exchange rates, starting with longer-term interrelationships among exchange rates, interest rates, and inflation rates embodied in the international parity conditions. These parity conditions form the key building blocks for many long-run exchange rate models.
- We also examine the foreign exchange (FX) carry trade, a trading strategy that exploits deviations from uncovered interest rate parity and discuss the relationship between a country's exchange rate and its balance of payments.
- We then examine how monetary and fiscal policies can *indirectly* affect exchange rates by influencing the various factors described in our exchange rate model.
- The subsequent section focuses on *direct* public sector actions in foreign exchange markets, both through capital controls and by foreign exchange market intervention (buying and selling currencies for policy purposes).
- The last section examines historical episodes of currency crisis and some leading indicators that may signal the increased likelihood of a crisis.
- We conclude with a summary.

2 FOREIGN EXCHANGE MARKET CONCEPTS

a calculate and interpret the bid–offer spread on a spot or forward currency quotation and describe the factors that affect the bid–offer spread;

We begin with a brief review of some of the basic conventions of the FX market that were covered in the CFA Program Level I curriculum. In this section, we cover (1) the basics of exchange rate notation and pricing, (2) arbitrage pricing constraints on spot rate foreign exchange quotes, and (3) forward rates and covered interest rate parity.

An exchange rate is the price of the *base* currency expressed in terms of the *price* currency. For example, a USD/EUR rate of 1.1650 means the euro, the base currency, costs 1.1650 US dollars (an appendix defines the three-letter currency codes). The exact notation used to represent exchange rates can vary widely between sources, and occasionally the same exchange rate notation will be used by different sources to mean completely different things. *The reader should be aware that the notation used here may not be the same as that encountered elsewhere.* To avoid confusion, we will identify exchange rates using the convention of "P/B," referring to the price of the base currency, "B," expressed in terms of the price currency, "P."

Notation Conventions

Notation is generally not standardized in global foreign exchange markets, and there are several common ways of expressing the same currency pair (e.g., JPY/USD, USD:JPY, $/¥). What is common in FX markets, however, is the concept of a "base" and a "price" currency when setting exchange rates. We will sometimes switch to discussing a "domestic" and a "foreign" currency, quoted as foreign/domestic (f/d). This is only an illustrative device for more easily explaining various theoretical concepts. The candidate should be aware that currency pairs are not described in terms of "foreign" and "domestic" currencies in professional FX markets. This is because what is the "foreign" and what is the "domestic" currency depend on where one is located, which can lead to confusion. For instance, what is "foreign" and what is "domestic" for a Middle Eastern investor trading CHF against GBP with the New York branch of a European bank, with the trade ultimately booked at the bank's headquarters in Paris?

The spot exchange rate is usually used for settlement on the second business day after the trade date, referred to as $T + 2$ settlement (the exception being CAD/USD, for which standard spot settlement is $T + 1$). In foreign exchange markets—as in other financial markets—market participants are presented with a two-sided price in the form of a bid price and an offer price (also called an ask price) quoted by potential counterparties. The bid price is the price, defined in terms of the price currency, at which the counterparty is willing to buy one unit of the base currency. Similarly, the offer price is the price, in terms of the price currency, at which that counterparty is willing to sell one unit of the base currency. For example, given a price request from a client, a dealer might quote a two-sided price on the spot USD/EUR exchange rate of 1.1648/1.1652. This means that the dealer is willing to pay USD 1.1648 to buy one euro and that the dealer is willing to sell one euro for USD 1.1652.

There are two points to bear in mind about bid–offer quotes:

1 *The offer price is always higher than the bid price.* The bid–offer spread—the difference between the offer price and the bid price—is the compensation that counterparties seek for providing foreign exchange to other market participants.

2 *The party in the transaction who requests a two-sided price quote has the option (but not the obligation) to deal at either the bid (to sell the base currency) or the offer (to buy the base currency) quoted by the dealer.* If the party chooses to trade at the quoted prices, the party is said to have either "*hit the bid*" or "*paid the offer*." If the base currency is being sold, the party is said to have hit the bid. If the base currency is being bought, the party is said to have paid the offer.

We will distinguish here between the bid–offer pricing *a client receives from a dealer* and the pricing *a dealer receives from the interbank market*. Dealers buy and sell foreign exchange among themselves in what is called the interbank market. This global network for exchanging currencies among professional market participants allows dealers to adjust their inventories and risk positions, distribute foreign currencies to end users who need them, and transfer foreign exchange rate risk to market participants who are willing to bear it. The interbank market is typically for dealing sizes of at least 1 million units of the base currency. Of course, the dealing amount can be larger than 1 million units; indeed, interbank market trades generally are measured in terms of multiples of a million units of the base currency. Please note that many non-bank entities can now access the interbank market. They include institutional asset managers and hedge funds.

The bid–offer spread a dealer provides to most clients typically is slightly wider than the bid–offer spread observed in the interbank market. Most currencies, except for the yen, are quoted to four decimal places. The fourth decimal place (0.0001) is referred to as a "pip." The yen is typically quoted to just two decimal places; in yen quotes, the second decimal place (0.01) is referred to as a pip.

For example, if the quote in the interbank USD/EUR spot market is 1.1649/1.1651 (two pips wide), the dealer might quote a client a bid–offer of 1.1648/1.1652 (four pips wide) for a spot USD/EUR transaction. When the dealer buys (sells) the base currency from (to) a client, the dealer is typically expecting to quickly turn around and sell (buy) the base currency in the interbank market. This offsetting transaction allows the dealer to divest the risk exposure assumed by providing a two-sided price to the client and to hopefully make a profit. Continuing our example, suppose the dealer's client hits the dealer's bid and sells EUR to the dealer for USD 1.1648. The dealer is now long EUR (and short USD) and wants to cover this position in the interbank market. To do this, the dealer sells the EUR in the interbank market by hitting the interbank bid. As a result, the dealer *bought* EUR from the client at USD 1.1648 and then *sold* the EUR in the interbank for USD 1.1649. This gives the dealer a profit of USD 0.0001 (one pip) for every EUR transacted. This one pip translates into a profit of USD 100 per EUR million bought from the client. If, instead of hitting his bid, the client paid the offer (1.1652), then the dealer could pay the offer in the interbank market (1.1651), earning a profit of one pip.

The size of the bid–offer spread quoted to dealers' clients in the FX market can vary widely across exchange rates and is not constant over time, even for a single exchange rate. The size of this spread depends primarily on three factors:

- the bid–offer spread in the interbank foreign exchange market for the two currencies involved,
- the size of the transaction, and
- the relationship between the dealer and the client.

We examine each factor in turn.

The size of the bid–offer spread quoted in the interbank market depends on the liquidity in this market. Liquidity is influenced by several factors:

1 *The currency pair involved.* Market participation is greater for some currency pairs than for others. Liquidity in the major currency pairs—for example, USD/EUR, JPY/USD, and USD/GBP—can be quite high. These markets are almost always deep, with multiple bids and offers from market participants around the world. In other currency pairs, particularly some of the more obscure currency cross rates (e.g., MXN/CHF), market participation is much thinner and consequently the bid–offer spread in the interbank market will be wider.

2 *The time of day.* The interbank FX markets are most liquid when the major FX trading centers are open. Business hours in London and New York—the two largest FX trading centers—overlap from approximately 8:00 a.m. to 11:00 a.m. New York time. The interbank FX market for most currency pairs is typically most liquid during these hours. After London closes, liquidity is thinner through the New York afternoon. The Asian session starts when dealers in

Tokyo, Singapore, and Hong Kong SAR open for business, typically by 7:00 p.m. New York time. For most currency pairs, however, the Asian session is not as liquid as the London and New York sessions. Although FX markets are open 24 hours a day on business days, between the time New York closes and the time Asia opens, liquidity in interbank markets can be very thin because Sydney, Australia, tends to be the only active trading center during these hours. For reference, the chart below shows a 24-hour period from midnight (00:00) to midnight (24:00) London time, corresponding standard times in Tokyo and New York, and, shaded in grey, the *approximate* hours of the most liquid trading periods in each market.

Standard Time and Approximate FX Trading Hours in Major Markets: Midnight to Midnight (London Time)

Tokyo	09:00	13:00	17:00	21:00	01:00 Day+1	05:00 Day+1	09:00 Day+1
London	00:00	04:00	08:00	12:00	16:00	20:00	24:00
New York	19:00 Day−1	23:00 Day−1	03:00	07:00	11:00	15:00	19:00

3 *Market volatility.* As in any financial market, when major market participants have greater uncertainty about the factors influencing market pricing, they will attempt to reduce their risk exposures and/or charge a higher price for taking on risk. In the FX market, this response implies wider bid–offer spreads in both the interbank and broader markets. Geopolitical events (e.g., war, civil strife), market crashes, and major data releases (e.g., US non-farm payrolls) are among the factors that influence spreads and liquidity.

The size of the transaction can also affect the bid–offer spread shown by a dealer to clients. Typically, the larger the transaction, the further away from the current spot exchange rate the dealing price will be. Hence, a client who asks a dealer for a two-sided spot CAD/USD price on, for example, USD 50 million will be shown a wider bid–offer spread than a client who asks for a price on USD 1 million. The wider spread reflects the greater difficulty the dealer faces in offsetting the foreign exchange risk of the position in the interbank FX market. Smaller dealing sizes can also affect the bid–offer quote shown to clients. "Retail" quotes are typically for dealing sizes smaller than 1 million units of the base currency and can range all the way down to foreign exchange transactions conducted by individuals. The bid–offer spreads for these retail transactions can be very large compared with those in the interbank market.

The relationship between the dealer and the client can also affect the size of the bid–offer spread shown by the dealer. For many clients, the spot foreign exchange business is only one business service among many that a dealer provides to that client. For example, the dealer firm might also transact in bond and/or equity securities with the same client. In a competitive business environment, in order to win the client's business for these other services, the dealer might provide a tighter (i.e., smaller) bid–offer spot exchange rate quote. The dealer might also give tighter bid–offer quotes in order to win repeat FX business. A client's credit risk can also be a factor. A client with a poor credit profile may be quoted a wider bid–offer spread than one with good credit. Given the short settlement cycle for spot FX transactions (typically two business days), however, credit risk is not the most important factor in determining the client's bid–offer spread on spot exchange rates.

3 ARBITRAGE CONSTRAINTS ON SPOT EXCHANGE RATE QUOTES

b identify a triangular arbitrage opportunity and calculate its profit, given the bid–offer quotations for three currencies;

The bid–offer quotes a dealer shows in the interbank FX market must respect two arbitrage constraints; otherwise the dealer creates riskless arbitrage opportunities for other interbank market participants. We will confine our attention to the interbank FX market because arbitrage presumes the ability to deal simultaneously with different market participants and in different markets, the ability to access "wholesale" bid–offer quotes, and the market sophistication to spot arbitrage opportunities.

First, the bid shown by a dealer in the interbank market cannot be higher than the current interbank offer, and the offer shown by a dealer cannot be lower than the current interbank bid. If the bid–offer quotes shown by a dealer are inconsistent with the then-current interbank market quotes, other market participants will buy from the cheaper source and sell to the more expensive source. This arbitrage will eventually bring the two prices back into line. For example, suppose that the current spot USD/EUR price in the interbank market is 1.1649/1.1651. If a dealer showed a misaligned price quote of 1.1652/1.1654, then other market participants would pay the offer in the interbank market, *buying* EUR at a price of USD 1.1651, and then *sell* the EUR to the dealer by hitting the dealer's bid at USD 1.1652—thereby making a riskless profit of one pip on the trade. This arbitrage would continue as long as the dealer's bid–offer quote violated the arbitrage constraint.

Second, the cross-rate bids (offers) posted by a dealer must be lower (higher) than the implied cross-rate offers (bids) available in the interbank market. A currency dealer located in a given country typically provides exchange rate quotations between that country's currency and various foreign currencies. If a particular currency pair is not explicitly quoted, it can be inferred from the quotes for each currency in terms of the exchange rate with a third nation's currency. For example, given exchange rate quotes for the currency pairs A/B and C/B, we can back out the implied cross rate of A/C. This implied A/C cross rate must be consistent with the A/B and C/B rates. This again reflects the basic principle of arbitrage: If identical financial products are priced differently, then market participants will buy the cheaper one and sell the more expensive one until the price difference is eliminated. In the context of FX cross rates, there are two ways to trade currency A against currency C: (1) using the cross rate A/C or (2) using the A/B and C/B rates. Because, in the end, both methods involve selling (buying) currency C in order to buy (sell) currency A, the exchange rates for these two approaches must be consistent. If the exchange rates are not consistent, the arbitrageur will buy currency C from a dealer if it is undervalued (relative to the cross rate) and sell currency A. If currency C is overvalued by a dealer (relative to the cross rate), it will be sold and currency A will be bought.

To illustrate this **triangular arbitrage** among three currencies, suppose that the interbank market bid–offer in USD/EUR is 1.1649/1.1651 and that the bid–offer in JPY/USD is 105.39/105.41. We need to use these two interbank bid–offer quotes to calculate the market-implied bid–offer quote on the JPY/EUR cross rate.

Begin by considering the transactions required to *sell* JPY and *buy* EUR, going through the JPY/USD and USD/EUR currency pairs. We can view this process intuitively as follows:

$$\begin{matrix}\text{Sell JPY} \\ \text{Buy EUR}\end{matrix} = \begin{matrix}\text{Sell JPY} \\ \text{Buy USD}\end{matrix} \text{ then } \begin{matrix}\text{Sell USD} \\ \text{Buy EUR}\end{matrix}$$

Note that "Buy USD" and "Sell USD" in the expressions on the right-hand side of the equal sign will cancel out to give the JPY/EUR cross rate. In equation form, we can represent this relationship as follows:

$$\left(\frac{\text{JPY}}{\text{EUR}}\right) = \left(\frac{\text{JPY}}{\cancel{\text{USD}}}\right)\left(\frac{\cancel{\text{USD}}}{\text{EUR}}\right).$$

Now, let's incorporate the bid–offer rates in order to do the JPY/EUR calculation. A rule of thumb is that when we speak of a bid or offer exchange rate, we are referring to the bid or offer for the currency in the denominator (the base currency).

- **i.** The left-hand side of the above equal sign is "Sell JPY, Buy EUR." In the JPY/EUR price quote, the EUR is in the denominator (it is the base currency). Because we want to buy the currency in the denominator, we need an exchange rate that is an offer rate. Thus, we will be calculating the *offer* rate for JPY/EUR.
- **ii.** The first term on the right-hand side of the equal sign is "Sell JPY, Buy USD." Because we want to buy the currency in the denominator of the quote, we need an exchange rate that is an offer rate. Thus, we need the *offer* rate for JPY/USD.
- **iii.** The second term on the right-hand side of the equal sign is "Sell USD, Buy EUR." Because we want to buy the currency in the denominator of the quote, we need an exchange rate that is an offer rate. Thus, we need the *offer* rate for USD/EUR.

Combining all of this conceptually and putting in the relevant offer rates leads to a JPY/EUR offer rate of

$$\left(\frac{\text{JPY}}{\text{EUR}}\right)_{offer} = \left(\frac{\text{JPY}}{\boxed{\text{USD}}}\right)_{offer}\left(\frac{\boxed{\text{USD}}}{\text{EUR}}\right)_{offer} = 105.41 \times 1.1651 = 122.81.$$

Perhaps not surprisingly, calculating the implied JPY/EUR *bid* rate uses the same process as above but with "Buy JPY, Sell EUR" for the left-hand side of the equation, which leads to

$$\left(\frac{\text{JPY}}{\text{EUR}}\right)_{bid} = \left(\frac{\text{JPY}}{\boxed{\text{USD}}}\right)_{bid}\left(\frac{\boxed{\text{USD}}}{\text{EUR}}\right)_{bid} = 105.39 \times 1.1649 = 122.77.$$

As one would expect, the implied cross-rate bid (122.77) is less than the offer (122.81).

This simple formula seems relatively straightforward: To get the implied *bid* cross rate, simply multiply the *bid* rates for the other two currencies. However, depending on the quotes provided, it may be necessary to *invert* one of the quotes in order to complete the calculation.

This is best illustrated with an example. Consider the case of calculating the implied GBP/EUR cross rate if you are given USD/GBP and USD/EUR quotes. Simply using the provided quotes will not generate the desired GBP/EUR cross rate:

$$\frac{\text{GBP}}{\text{EUR}} \neq \left(\frac{\text{USD}}{\text{GBP}}\right)\left(\frac{\text{USD}}{\text{EUR}}\right).$$

Instead, because the USD is in the numerator in both currency pairs, we will have to invert one of the pairs to derive the GBP/EUR cross rate.

The following equation represents the cross-rate relationship we are trying to derive:

$$\frac{\text{GBP}}{\text{EUR}} = \left(\frac{\text{GBP}}{\text{USD}}\right)\left(\frac{\text{USD}}{\text{EUR}}\right).$$

But we don't have the GBP/USD quote. We can, however, invert the USD/GBP quote and use that in our calculation. Let's assume the bid–offer quote provided is for USD/GBP and is 1.2302/1.2304. With this quote, if we want to *buy* GBP (the currency in the denominator), we will buy GBP at the offer and the relevant quote is 1.2304. We can invert this quote to arrive at the needed GBP/USD quote: 1 ÷ 1.2304 = 0.81274. Note that, in this example, when we buy the GBP, we are also selling the USD. When we invert the provided USD/GBP offer quote, we obtain 0.81274 GBP/USD. This is the price at which we sell the USD—that is, the GBP/USD *bid*. It may help here to remember our rule of thumb from above: When we speak of a bid or offer exchange rate, we are referring to the bid or offer for the currency in the denominator (the base currency).

Similarly, to get a GBP/USD *offer*, we use the inverse of the USD/GBP *bid* of 1.2302: 1 ÷ 1.2302 = 0.81288. (Note that we extended the calculated GBP/USD 0.81274/0.81288 quotes to five decimal places to avoid truncation errors in subsequent calculations.)

We can now finish the calculation of the bid and offer cross rates for GBP/EUR. Using the previously provided 1.1649/1.1651 as the bid–offer in USD/EUR, we calculate the GBP/EUR *bid* rate as follows:

$$\left(\frac{\text{GBP}}{\text{EUR}}\right)_{bid} = \left(\frac{\text{GBP}}{\boxed{\text{USD}}}\right)_{bid}\left(\frac{\boxed{\text{USD}}}{\text{EUR}}\right)_{bid} = 0.81274 \times 1.1649 = 0.9468.$$

Similarly, the implied GBP/EUR *offer* rate is

$$\left(\frac{\text{GBP}}{\text{EUR}}\right)_{offer} = \left(\frac{\text{GBP}}{\boxed{\text{USD}}}\right)_{offer}\left(\frac{\boxed{\text{USD}}}{\text{EUR}}\right)_{offer} = 0.81288 \times 1.1651 = 0.9471.$$

Note that the implied *bid* rate is less than the implied *offer* rate, as it must be to prevent arbitrage.

We conclude this section on arbitrage constraints with some simple observations:

- The arbitrage constraint on implied cross rates is similar to that for spot rates (posted bid rates cannot be higher than the market's offer; posted offer rates cannot be lower than the market's bid). The only difference is that this second arbitrage constraint is applied *across* currency pairs instead of involving a *single* currency pair.
- In reality, any violations of these arbitrage constraints will quickly disappear. Both human traders and automatic trading algorithms are constantly on alert for any pricing inefficiencies and will arbitrage them away almost instantly. If Dealer 1 is buying a currency at a price higher than the price at which Dealer 2 is selling it, the arbitrageur will buy the currency from Dealer 2 and resell it to Dealer 1. As a result of buying and selling pressures, Dealer 2 will raise his offer prices and Dealer 1 will reduce her bid prices to the point where arbitrage profits are no longer available.
- Market participants do not need to calculate cross rates *manually* because electronic dealing machines (which are essentially just specialized computers) will automatically calculate cross bid–offer rates given any two underlying bid–offer rates.

EXAMPLE 1

Bid–Offer Rates

The following are spot rate quotes in the interbank market:

USD/EUR	1.1649/1.1651
JPY/USD	105.39/105.41
CAD/USD	1.3199/1.3201
SEK/USD	9.6300/9.6302

1 What is the bid–offer on the SEK/EUR cross rate implied by the interbank market?

A 0.1209/0.1211

B 8.2656/8.2668

C 11.2180/11.2201

2 What is the bid–offer on the JPY/CAD cross rate implied by the interbank market?

A 78.13/78.17

B 79.85/79.85

C 79.84/79.86

3 If a dealer quoted a bid–offer rate of 79.81/79.83 in JPY/CAD, then a triangular arbitrage would involve buying:

A CAD in the interbank market and selling it to the dealer, for a profit of JPY 0.01 per CAD.

B JPY from the dealer and selling it in the interbank market, for a profit of CAD 0.01 per JPY.

C CAD from the dealer and selling it in the interbank market, for a profit of JPY 0.01 per CAD.

4 If a dealer quoted a bid–offer of 79.82/79.87 in JPY/CAD, then you could:

A not make any arbitrage profits.

B make arbitrage profits buying JPY from the dealer and selling it in the interbank market.

C make arbitrage profits buying CAD from the dealer and selling it in the interbank market.

5 A market participant is considering the following transactions:

Transaction 1	Buy CAD 100 million against the USD at 15:30 London time.
Transaction 2	Sell CAD 100 million against the KRW at 21:30 London time.
Transaction 3	Sell CAD 10 million against the USD at 15:30 London time.

Given the proposed transactions, what is the *most likely* ranking of the bid–offer spreads, from tightest to widest, under normal market conditions?

A Transactions 1, 2, 3

B Transactions 2, 1, 3

C Transactions 3, 1, 2

Solution to 1:

C is correct. Using the provided quotes and setting up the equations so that the cancellation of terms results in the SEK/EUR quote,

$$\frac{\text{SEK}}{\text{EUR}} = \frac{\text{SEK}}{\text{USD}} \times \frac{\text{USD}}{\text{EUR}}.$$

Hence, to calculate the SEK/EUR bid (offer) rate, we multiply the SEK/USD and USD/EUR bid (offer) rates to get the following:

Bid: 11.2180 = 9.6300 × 1.1649.

Offer: 11.2201 = 9.6302 × 1.1651.

Solution to 2:

C is correct. Using the intuitive equation-based approach,

$$\frac{\text{JPY}}{\text{CAD}} = \frac{\text{JPY}}{\text{USD}} \times \left(\frac{\text{CAD}}{\text{USD}}\right)^{-1} = \frac{\text{JPY}}{\cancel{\text{USD}}} \times \frac{\cancel{\text{USD}}}{\text{CAD}}.$$

This equation shows that we have to invert the CAD/USD quotes to get the USD/CAD bid–offer rates of 0.75752/0.75763. That is, given the CAD/USD quotes of 1.3199/1.3201, we take the inverse of each and interchange bid and offer, so that the USD/CAD quotes are (1/1.3201)/(1/1.3199), or 0.75752/0.75763. Multiplying the JPY/USD and USD/CAD bid–offer rates then leads to the following:

Bid: 79.84 = 105.39 × 0.75752.

Offer: 79.86 = 105.41 × 0.75763.

Solution to 3:

C is correct. The implied interbank cross rate for JPY/CAD is 79.84/79.86 (the answer to Question 2). Hence, the dealer is offering to sell the CAD (the base currency in the quote) too cheaply, at an offer rate that is below the interbank bid rate (79.83 versus 79.84, respectively). Triangular arbitrage would involve buying CAD from the dealer (paying the dealer's offer) and selling CAD in the interbank market (hitting the interbank bid), for a profit of JPY 0.01 (79.84 – 79.83) per CAD transacted.

Solution to 4:

A is correct. The arbitrage relationship is not violated: The dealer's bid (offer) is not above (below) the interbank market's offer (bid). The implied interbank cross rate for JPY/CAD is 79.84/79.86 (the solution to Question 2).

Solution to 5:

C is correct. The CAD/USD currency pair is most liquid when New York and London are both in their most liquid trading periods at the same time (approximately 8:00 a.m. to 11:00 a.m. New York time, or about 13:00 to 16:00 London time). Transaction 3 is for a smaller amount than Transaction 1. Transaction 2 is for a less liquid currency pair (KRW/CAD is traded less than CAD/USD) and occurs outside of normal dealing hours in all three major centers (London, North America, and Asia); the transaction is also for a large amount.

4 FORWARD MARKETS

c explain spot and forward rates and calculate the forward premium/discount for a given currency;

Outright forward contracts (often referred to simply as forwards) are agreements to exchange one currency for another on a future date at an exchange rate agreed upon today. Any exchange rate transaction that has a settlement date longer than $T + 2$ is a forward contract.

Forward exchange rates must satisfy an arbitrage relationship that equates the investment return on two alternative but equivalent investments. To simplify the explanation of this arbitrage relationship and to focus on the intuition behind forward rate calculations, we will ignore the bid–offer spread on exchange rates and money market instruments. In addition, we will alter our exchange rate notation from price/base currency (P/B) to "foreign/domestic currency" (f/d), making the assumption that the domestic currency for an investor is the base currency in the exchange rate quotation. Using this (f/d) notation will make it easier to illustrate the choice an investor faces between domestic and foreign investments, as well as the arbitrage relationships that equate the returns on these investments when their risk characteristics are equivalent.

Consider an investor with one unit of domestic currency to invest for one year. The investor faces two alternatives:

A One alternative is to invest cash for one year at the domestic risk-free rate (i_d). At the end of the year, the investment would be worth $(1 + i_d)$.

B The other alternative is to convert the domestic currency to foreign currency at the spot rate of $S_{f/d}$ and invest for one year at the foreign risk-free rate (i_f). At the end of the period, the investor would have $S_{f/d}(1 + i_f)$ units of foreign currency. These funds then must be converted back to the investor's domestic currency. If the exchange rate to be used for this end-of-year conversion is set at the start of the period using a one-year forward contract, then the investor will have eliminated the foreign exchange risk associated with converting at an unknown future spot rate. If we let $F_{f/d}$ denote the forward rate, the investor would obtain $(1/F_{f/d})$ units of the domestic currency for each unit of foreign currency sold forward. Hence, in domestic currency, at the end of the year, the investment would be worth $S_{f/d}(1 + i_f)(1/F_{f/d})$.

The two investment alternatives above (A and B) are risk free and therefore must offer the same return. If they did not offer the same return, investors could earn a riskless arbitrage profit by borrowing in one currency, lending in the other, and using the spot and forward exchange markets to eliminate currency risk. Equating the returns on these two investment alternatives—that is, putting investments A and B on opposite sides of the equal sign—leads to the following relationship:

$$\left(1 + i_d\right) = S_{f/d}\left(1 + i_f\right)\left(\frac{1}{F_{f/d}}\right).$$

To see the intuition behind forward rate calculations, note that the right-hand side of the expression (for investment B) also shows the chronological order of this investment: Convert from domestic to foreign currency at the spot rate ($S_{f/d}$); invest this foreign currency amount at the foreign risk-free interest rate $(1 + i_f)$; and then at maturity, convert the foreign currency investment proceeds back into the domestic currency using the forward rate $(1/F_{f/d})$.

For simplicity, we assumed a one-year horizon in the preceding example. However, the argument holds for any investment horizon. The risk-free assets used in this arbitrage relationship are typically bank deposits quoted using the reference rate (typically the historically most commonly used London Interbank Offered Rate, or Libor) for each currency involved. The day count convention for almost all Libor deposits is Actual/360. This notation means that interest is calculated as if there were 360 days in a year. However, the actual number of days the funds are on deposit is used to calculate the interest payable (the main exception to the Actual/360 day count convention is the GBP, for which the convention is Actual/365). For the purposes of our discussion, we will use Actual/360 consistently in order to avoid complication. Incorporating this day count convention into our arbitrage formula leads to

$$\left(1+i_d\left[\frac{\text{Actual}}{360}\right]\right)=S_{f/d}\left(1+i_f\left[\frac{\text{Actual}}{360}\right]\right)\left(\frac{1}{F_{f/d}}\right).$$

This equation can be rearranged to isolate the forward rate:

$$F_{f/d}=S_{f/d}\left(\frac{1+i_f\left[\frac{\text{Actual}}{360}\right]}{1+i_d\left[\frac{\text{Actual}}{360}\right]}\right). \quad (1)$$

Equation 1 describes **covered interest rate parity**. Our previous work shows that covered interest rate parity is based on an arbitrage relationship among risk-free interest rates and spot and forward exchange rates. Because of this arbitrage relationship between investment alternatives, Equation 1 can also be described as saying that the covered (i.e., currency-hedged) interest rate differential between the two markets is zero.

The covered interest rate parity equation can also be rearranged to give an expression for the forward premium or discount:

$$F_{f/d}-S_{f/d}=S_{f/d}\left(\frac{\left[\frac{\text{Actual}}{360}\right]}{1+i_d\left[\frac{\text{Actual}}{360}\right]}\right)\left(i_f-i_d\right).$$

The domestic currency will trade at a forward premium ($F_{f/d} > S_{f/d}$) if, and only if, the foreign risk-free interest rate exceeds the domestic risk-free interest rate ($i_f > i_d$). Equivalently, in this case, the foreign currency will trade at a lower rate in the forward contract (relative to the spot rate), and we would say that the foreign currency trades at a forward discount. In other words, if it is possible to earn more interest in the foreign market than in the domestic market, then the forward discount for the foreign currency will offset the higher foreign interest rate. Otherwise, covered interest rate parity would not hold and arbitrage opportunities would exist.

When the foreign currency is at a higher rate in the forward contract, relative to the spot rate, we say that the foreign currency trades at a forward premium. In the case of a forward premium for the foreign currency, the foreign risk-free interest rate will be less than the domestic risk-free interest rate. Additionally, as can be seen in the equation above, the premium or discount is proportional to the spot exchange rate ($S_{f/d}$), proportional to the interest rate differential ($i_f - i_d$) between the markets, and approximately proportional to the time to maturity (Actual/360).

Although we have illustrated the covered interest rate parity equation (Equation 1) in terms of foreign and domestic currencies (using the notation f/d), this equation can also be expressed in our more standard exchange rate quoting convention of price and base currencies (P/B):

$$F_{P/B}=S_{P/B}\left(\frac{1+i_P\left[\frac{\text{Actual}}{360}\right]}{1+i_B\left[\frac{\text{Actual}}{360}\right]}\right).$$

When dealing in professional FX markets, it may be more useful to think of the covered interest rate parity equation and the calculation of forward rates in this P/B notation rather than in foreign/domestic (*f*/*d*) notation. Domestic and foreign are relative concepts that depend on where one is located, and because of the potential for confusion, these terms are not used for currency quotes in professional FX markets.

EXAMPLE 2

Calculating the Forward Premium (Discount)

The following table shows the mid-market rate (i.e., the average of the bid and offer) for the current CAD/AUD spot exchange rate as well as for AUD and CAD 270-day Libor (annualized):

Spot (CAD/AUD)	0.9000
270-day Libor (AUD)	1.47%
270-day Libor (CAD)	0.41%

The forward premium (discount) for a 270-day forward contract for CAD/AUD would be *closest* to:

A −0.0094.

B −0.0071.

C +0.0071.

Solution:

B is correct. The equation to calculate the forward premium (discount) is as follows:

$$F_{P/B} - S_{P/B} = S_{P/B}\left(\frac{\left[\frac{\text{Actual}}{360}\right]}{1 + i_B\left[\frac{\text{Actual}}{360}\right]}\right)(i_P - i_B).$$

Because AUD is the base currency in the CAD/AUD quote, putting in the information from the table gives us

$$F_{P/B} - S_{P/B} = 0.9000\left(\frac{\left[\frac{270}{360}\right]}{1 + 0.0147\left[\frac{270}{360}\right]}\right)(0.0041 - 0.0147) = -0.0071.$$

In professional FX markets, forward exchange rates are typically quoted in terms of points—the difference between the forward exchange rate quote and the spot exchange rate quote, scaled so that the points can be directly related to the last decimal place in the spot quote. Thus, the forward rate quote is typically shown as the bid–offer on the spot rate and the number of forward points at each maturity, as shown in Exhibit 1 ("Maturity" is defined in terms of the time between spot settlement—usually T + 2—and the settlement of the forward contract).

Exhibit 1 Sample Spot and Forward Quotes (Bid–Offer)

Maturity	Spot Rate
Spot (USD/EUR)	1.1649/1.1651
	Forward Points
1 month	–5.6/–5.1
3 months	–15.9/–15.3
6 months	–37.0/–36.3
12 months	–94.3/–91.8

Note the following:

- As always, the offer in the bid–offer quote is larger than the bid. In this example, the forward points are negative (i.e., the forward rate for the EUR is at a discount to the spot rate) but the bid is a smaller number (–5.6 versus –5.1 at the one-month maturity).
- The absolute number of forward points is a function of the term of the forward contract: A longer contract term results in a larger number of points.
- Because this is an OTC market, a client is not restricted to dealing *only* at the dates/maturities shown. Dealers typically quote standard forward dates, but forward deals can be arranged for any forward date the client requires. The forward points for these non-standard (referred to as "broken") forward dates will typically be interpolated on the basis of the points shown for the standard settlement dates.
- The quoted points are already scaled to each maturity—they are not annualized—so there is no need to adjust them.

To convert any of these quoted forward points into a forward rate, divide the number of points by 10,000 (to scale it down to the same four decimal places in the USD/EUR spot quote) and then add the result to the spot exchange rate quote (because the JPY/USD exchange rate is quoted to only two decimal places, forward points for the dollar–yen currency pair are divided by 100). Be careful, however, about which side of the market (bid or offer) is being quoted. For example, suppose a market participant is *selling* the EUR forward against the USD and is given a USD/EUR quote. The EUR is the base currency; thus, the market participant must use the *bid* rates (i.e., hit the bid). Using the data in Exhibit 1, the three-month forward *bid* rate in this case would be based on the spot bid and the forward points bid and hence would be

$$1.1649 + (-15.9/10{,}000) = 1.16331.$$

The market participant would be selling EUR three months forward at a price of USD 1.16331 per EUR.

5 THE MARK-TO-MARKET VALUE OF A FORWARD CONTRACT

d calculate the mark-to-market value of a forward contract;

Next, we consider the mark-to-market value of forward contracts. As with other financial instruments, the mark-to-market value of forward contracts reflects the profit (or loss) that would be realized from closing out the position at current market prices. To close out a forward position, we must offset it with an equal and opposite forward position using the spot exchange rate and forward points available in the market when the offsetting position is created. When a forward contract is initiated, the mark-to-market value of the contract is zero, and no cash changes hands. From that moment onward, however, the mark-to-market value of the forward contract will change as the spot exchange rate changes and as interest rates change in either of the two currencies.

Let's look at an example. Suppose that a market participant bought GBP 10 million for delivery against the AUD in six months at an "all-in" forward rate of 1.8100 AUD/GBP. (The all-in forward rate is simply the sum of the spot rate and the scaled forward points.) Three months later, the market participant wants to close out this forward contract. This would require selling GBP 10 million three months forward using the AUD/GBP spot exchange rate and forward points in effect at that time. Before looking at this exchange rate, note that the offsetting forward contract is defined in terms of the original position taken. The original position in this example was "long GBP 10 million," so the offsetting contract is "short GBP 10 million." However, there is ambiguity here: To be long GBP 10 million at 1.8100 AUD/GBP is equivalent to being short AUD 18,100,000 (10,000,000 × 1.8100) at the same forward rate. To avoid this ambiguity, for the purposes of this discussion, we will state what the relevant forward position is for mark-to-market purposes. The net gain or loss from the transaction will be reflected in the alternate currency.

Assume the bid–offer quotes for spot and forward points three months prior to the settlement date are as follows:

Spot rate (AUD/GBP)	1.8210/1.8215
Three-month points	130/140

To sell GBP (the base currency in the AUD/GBP quote), we will be calculating the *bid* side of the market. Hence, the appropriate all-in three-month forward rate to use is

$$1.8210 + 130/10{,}000 = 1.8340.$$

This means that the market participant originally bought GBP 10 million at an AUD/GBP rate of 1.8100 and subsequently sold that amount at a rate of 1.8340. These GBP amounts will net to zero at the settlement date (GBP 10 million both bought and sold), but the AUD amounts will not, because the forward rate has changed. The AUD cash flow at the settlement date will be

$$(1.8340 - 1.8100) \times 10{,}000{,}000 = +\text{AUD } 240{,}000.$$

This is a cash *inflow* because the market participant was long the GBP with the original forward position and the GBP subsequently appreciated (the AUD/GBP rate increased).

This cash flow will be paid at the settlement day, which is still three months away. To calculate the mark-to-market value of the dealer's position, we must discount this cash flow to the present. The present value of this amount is found by discounting the settlement day cash flow by the three-month discount rate. Because this amount is in AUD, we use the three-month AUD discount rate. Let's use Libor and suppose that three-month AUD Libor is 2.40% (annualized). The present value of this future AUD cash flow is then

$$\frac{\text{AUD } 240{,}000}{1 + 0.024\left(\frac{90}{360}\right)} = \text{AUD } 238{,}569.$$

This result is the mark-to-market value of the original long GBP 10 million six-month forward when it is closed out three months prior to settlement.

To summarize, the process for marking to market a forward position is relatively straightforward:

1 Create an offsetting forward position that is equal to the original forward position. (In the example above, the market participant was long GBP 10 million forward, so the offsetting forward contract would be to sell GBP 10 million.)
2 Determine the appropriate all-in forward rate for this new, offsetting forward position. If the base currency of the exchange rate quote is being sold (bought), then use the bid (offer) side of the market.
3 Calculate the cash flow at the settlement day. This amount will be based on the original contract size times the difference between the original forward rate and that calculated in Step 2. If the currency the market participant was originally long (short) subsequently appreciated (depreciated), then there will be a cash *inflow* (*outflow*). (In the above example, the market participant was long the GBP, which subsequently appreciated, leading to a cash inflow at the settlement day.)
4 Calculate the present value of this cash flow at the future settlement date. The currency of the cash flow and the discount rate must match. (In the example above, the cash flow at the settlement date was in AUD, so an AUD Libor was used to calculate the present value.)

The factors that affect the bid–offer spread for forward points are the same as those we discussed for spot bid–offer rates: the interbank market liquidity of the underlying currency pair, the size of the transaction, and the relationship between the client and the dealer. For forward bid–offer spreads, we can also add a fourth factor: the term of the forward contract. Generally, the longer the term of the forward contract, the wider the bid–offer spread. This relationship holds because as the term of the contract increases,

- liquidity in the forward market tends to decline,
- the exposure to counterparty credit risk increases, and
- the interest rate risk of the contract increases (forward rates are based on interest rate differentials, and a longer duration means greater price sensitivity to movements in interest rates).

EXAMPLE 3

Forward Rates and the Mark-to-Market Value of Forward Positions

A dealer is contemplating trade opportunities in the CHF/GBP currency pair. The following are the current spot rates and forward points being quoted for the CHF/GBP currency pair:

Spot rate (CHF/GBP)	1.2939/1.2941
One month	–8.3/–7.9
Two months	–17.4/–16.8
Three months	–25.4/–24.6
Four months	–35.4/–34.2

Five months	–45.9/–44.1
Six months	–56.5/–54.0

1 The current all-in bid rate for delivery of GBP against the CHF in three months is *closest* to:

A 1.29136.

B 1.29150.

C 1.29164.

2 The all-in rate that the dealer will be quoted today by another dealer to sell the CHF six months forward against the GBP is *closest* to:

A 1.28825.

B 1.28835.

C 1.28870.

Some time ago, Laurier Bay Capital, an investment fund based in Los Angeles, hedged a long exposure to the New Zealand dollar by selling NZD 10 million forward against the USD; the all-in forward price was 0.7900 (USD/NZD). Three months prior to the settlement date, Laurier Bay wants to mark this forward position to market. The bid–offer for the USD/NZD spot rate, the three-month forward points, and the three-month Libors (annualized) are as follows:

Spot rate (USD/NZD)	0.7825/0.7830
Three-month points	–12.1/–10.0
Three-month Libor (NZD)	3.31%
Three-month Libor (USD)	0.31%

3 The mark-to-market value for Laurier Bay's forward position is *closest* to:

A –USD 87,100.

B +USD 77,437.

C +USD 79,938.

Now, suppose that instead of having a long exposure to the NZD, Laurier Bay Capital had a long forward exposure to the USD, which it hedged by selling USD 10 million forward against the NZD at an all-in forward rate of 0.7900 (USD/NZD). Three months prior to settlement date, it wants to close out this short USD forward position.

4 Using the above table, the mark-to-market value for Laurier Bay's short USD forward position is *closest* to:

A –NZD 141,117.

B –NZD 139,959.

C –NZD 87,100.

Solution to 1:

A is correct. The current all-in three-month bid rate for GBP (the base currency) is equal to 1.2939 + (–25.4/10,000) = 1.29136.

Solution to 2:

C is correct. The dealer will sell CHF against the GBP, which is equivalent to buying GBP (the base currency) against the CHF. Hence, the *offer* side of the market will be used for forward points. The all-in forward price will be 1.2941 + (–54.0/10,000) = 1.28870.

Solution to 3:

C is correct. Laurier Bay sold NZD 10 million forward to the settlement date at an all-in forward rate of 0.7900 (USD/NZD). To mark this position to market, the fund would need an offsetting forward transaction involving buying NZD 10 million three months forward to the settlement date. The NZD amounts on the settlement date net to zero. For the offsetting forward contract, because the NZD is the base currency in the USD/NZD quote, buying NZD forward means paying the offer for both the spot rate and the forward points. This scenario leads to an all-in three-month forward rate of 0.7830 – 0.0010 = 0.7820. On the settlement day, Laurier Bay will receive USD 7,900,000 (NZD 10,000,000 × 0.7900 USD/NZD) from the original forward contract and pay out USD 7,820,000 (NZD 10,000,000 × 0.7820 USD/NZD) based on the offsetting forward contract. The result is a net cash flow on the settlement day of 10,000,000 × (0.7900 – 0.7820) = +USD 80,000.

This is a cash inflow because Laurier Bay sold the NZD forward and the NZD depreciated against the USD. This USD cash inflow will occur in three months. To calculate the mark-to-market value of the original forward position, we need to calculate the present value of this USD cash inflow using the three-month USD discount rate (we use USD Libor for this purpose):

$$\frac{\text{USD } 80{,}000}{1 + 0.0031\left(\frac{90}{360}\right)} = +\text{USD } 79{,}938.$$

Solution to 4:

B is correct. Laurier Bay initially sold USD 10 million forward, and it will have to buy USD 10 million forward to the same settlement date (i.e., in three months' time) in order to close out the initial position. Buying USD using the USD/NZD currency pair is the same as selling the NZD. Because the NZD is the base currency in the USD/NZD quote, selling the NZD means calculating the *bid* rate:

$$0.7825 + (-12.1/10{,}000) = 0.78129.$$

At settlement, the USD amounts will net to zero (USD 10 million both bought and sold). The NZD amounts will not net to zero, however, because the all-in forward rate changed between the time Laurier Bay initiated the original position and the time it closed out this position. At initiation, Laurier Bay contracted to sell USD 10 million and receive NZD 12,658,228 (i.e., 10,000,000/0.7900) on the settlement date. To close out the original forward contract, Laurier Bay entered into an offsetting forward contract to receive USD 10 million and pay out NZD 12,799,345 (i.e., 10,000,000/0.78129) at settlement. The difference between the NZD amounts that Laurier Bay will receive and pay out on the settlement date equals

$$\text{NZD } 12{,}658{,}228 - \text{NZD } 12{,}799{,}345 = -\text{NZD } 141{,}117.$$

This is a cash *outflow* for Laurier Bay because the fund was *short* the USD in the original forward position and the USD subsequently *appreciated* (i.e., the NZD subsequently depreciated, because the all-in forward rate in USD/NZD dropped from 0.7900 to 0.78129). This NZD cash outflow occurs in three months' time, and we must calculate its present value using the three-month NZD Libor:

$$\frac{-\text{NZD } 141{,}117}{1 + 0.0331\left(\frac{90}{360}\right)} = -\text{NZD } 139{,}959.$$

A LONG-TERM FRAMEWORK FOR EXCHANGE RATES: INTERNATIONAL PARITY CONDITIONS

6

e explain international parity conditions (covered and uncovered interest rate parity, forward rate parity, purchasing power parity, and the international Fisher effect);

Having reviewed the basic tools of the FX market, we now turn our focus to how they are used in practice. At the heart of the trading decision in FX markets lies a view on equilibrium market prices. An understanding of equilibrium pricing will assist the investor in framing decisions regarding risk exposures and how they should be managed.

In this and the following sections, we lay out a framework for developing a view on equilibrium exchange rates. We begin by examining international parity conditions, which describe the inter-relationships that jointly determine *long-run* movements in exchange rates, interest rates, and inflation. These parity conditions are the basic building blocks for describing long-term equilibrium levels for exchange rates. In subsequent sections, we expand beyond the parity conditions by discussing other factors that influence a currency's value.

Always keep in mind that exchange rate movements reflect complex interactions among multiple forces. In trying to untangle this complex web of interactions, we must clearly delineate the following concepts:

1 Long run versus short run: Many of the factors that determine exchange rate movements exert subtle but persistent influences over long periods of time. Although a poor guide for short-term prediction, longer-term equilibrium values act as an anchor for exchange rate movements.

2 Expected versus unexpected changes: In reasonably efficient markets, prices will adjust to reflect market participants' expectations of future developments. When a key factor—say, inflation—is trending gradually in a particular direction, market pricing will eventually come to reflect expectations that this trend will continue. In contrast, large, unexpected movements in a variable (for example, a central bank intervening in the foreign exchange market) can lead to immediate, discrete price adjustments. This concept is closely related to risk. For example, a moderate but steady rate of inflation will not have the same effect on market participants as an inflation rate that is very unpredictable. The latter clearly describes a riskier financial environment. Market pricing will reflect risk premiums—that is, the compensation that traders and investors demand for being exposed to unpredictable outcomes. Whereas expectations of long-run equilibrium values tend to evolve slowly, risk premiums—which are closely related to confidence and reputation—can change quickly in response to unexpected developments.

3 Relative movements: An exchange rate represents the relative price of one currency in terms of another. Hence, for exchange rate determination, the level or variability of key factors in any particular country is typically much less important than the *differences* in these factors across countries. For example, knowing that inflation is increasing in Country A may not give much insight into the direction of the A/B exchange rate without also knowing what is happening with the inflation rate in Country B.

As a final word of caution—and this cannot be emphasized enough—*there is no simple formula, model, or approach that will allow market participants to precisely forecast exchange rates.* Models that work well in one period may fail in others. Models that work for one set of exchange rates may fail to work for others.

Nonetheless, market participants must have a market view to guide their decisions, even if this view requires significant revision as new information becomes available. The following sections provide a framework for understanding FX markets, a guide for thinking through the complex forces driving exchange rates. As with all theory, however, it does not eliminate the need for insightful analysis of actual economic and market conditions.

6.1 International Parity Conditions

International parity conditions form the building blocks of most models of exchange rate determination. The key international parity conditions are as follows:

1 covered interest rate parity,

2 uncovered interest rate parity,

3 forward rate parity,

4 purchasing power parity, and

5 the international Fisher effect.

Parity conditions show how expected inflation differentials, interest rate differentials, forward exchange rates, current spot exchange rates, and expected future spot exchange rates would be linked in an ideal world. These conditions typically make simplifying assumptions, such as zero transaction costs, perfect information that is available to all market participants, risk neutrality, and freely adjustable market prices.

Although empirical studies have found that the parity conditions rarely hold in the short term, they do help form a broadly based, long-term view of exchange rates and accompanying risk exposures. The exception to the rule that parity conditions do not hold in the short term is covered interest rate parity, which is the only parity condition that is enforced by arbitrage. We examine this parity condition first.

7 COVERED INTEREST RATE PARITY, UNCOVERED INTEREST RATE PARITY, & FORWARD RATE PARITY

e explain international parity conditions (covered and uncovered interest rate parity, forward rate parity, purchasing power parity, and the international Fisher effect);

f describe relations among the international parity conditions;

g evaluate the use of the current spot rate, the forward rate, purchasing power parity, and uncovered interest parity to forecast future spot exchange rates;

We have already discussed covered interest rate parity in our examination of forward exchange rates. Under this parity condition, *an investment in a foreign money market instrument that is completely hedged against exchange rate risk should yield exactly the same return as an otherwise identical domestic money market investment.* Given the spot exchange rate and the domestic and foreign yields, the forward exchange rate must equal the rate that gives these two alternative investment strategies—invest either in a domestic money market instrument or in a fully currency-hedged foreign money market instrument—exactly the same holding period return. If one strategy gave a higher holding period return than the other, then an investor could short-sell the lower-yielding approach and invest the proceeds in the higher-yielding approach, earning riskless arbitrage profits in the process. In real-world financial markets, such a disparity will be quickly arbitraged away so that no further arbitrage profits are available. Covered interest rate parity is thus said to be a no-arbitrage condition.

For covered interest rate parity to hold exactly, it must be assumed that there are zero transaction costs and that the underlying domestic and foreign money market instruments being compared are identical in terms of liquidity, maturity, and default risk. Where capital is permitted to flow freely, spot and forward exchange markets are liquid, and financial market conditions are relatively stress free, covered interest rate differentials are generally found to be close to zero and covered interest rate parity holds.

7.1 Uncovered Interest Rate Parity

According to the **uncovered interest rate parity** condition, the *expected* return on an uncovered (i.e., unhedged) foreign currency investment should equal the return on a comparable domestic currency investment. Uncovered interest rate parity states that *the change in spot rate over the investment horizon should, on average, equal the differential in interest rates between the two countries. That is, the expected appreciation/depreciation of the exchange rate will just offset the yield differential.*

To explain the intuition behind this concept, let's switch, as we did with the examples for covered interest rate parity, from the standard price/base currency notation (P/B) to foreign/domestic currency notation (*f*/*d*) in order to emphasize the choice between foreign and domestic investments. As before, we also will assume that for the investor, the base currency is the domestic currency. (In *covered* interest rate parity, we assumed the investor transacted at a forward rate that was locked in at strategy initiation. In *uncovered* interest rate parity, the investor is assumed to transact at a future spot rate that is unknown at the time the strategy is initiated and the investor's currency position in the future is not hedged—that is, uncovered.)

For our example, assume that this investor has a choice between a one-year domestic money market instrument and an unhedged one-year foreign-currency-denominated money market investment. Under the assumption of uncovered interest rate parity, the investor will compare the *known* return on the domestic investment with the *expected* all-in return on the unhedged foreign-currency-denominated investment (which includes the foreign yield as well as any movements in the exchange rate, in $S_{f/d}$ terms). The choice between these two investments will depend on which market offers the higher expected return on an unhedged basis.

For example, assume that the return on the one-year foreign money market instrument is 10% while the return on the one-year domestic money market instrument is 4%. From the investor's perspective, the 4% expected return on the one-year domestic investment in domestic currency terms is known with complete certainty. This is not the case for the uncovered investment in the foreign currency money market instrument. In domestic currency terms, the investment return on an uncovered (or unhedged) foreign-currency-denominated investment is equal to

$$\left(1 + i_f\right)\left(1 - \%\Delta S_{f/d}\right) - 1.$$

Intuitively, the formula says that the investor's return on a foreign investment is a function of both the foreign interest rate and the change in the spot rate, whereby a depreciation in the foreign currency reduces the investor's return. The percentage change in $S_{f/d}$ enters with a minus sign because an *increase* in $S_{f/d}$ means the foreign currency *declines* in value, thereby reducing the all-in return from the domestic currency perspective of our investor. This all-in return depends on *future* movements in the $S_{f/d}$ rate, which cannot be known until the end of the period. This return can be approximated using the following equation:

$$\cong i_f - \%\Delta S_{f/d}.$$

Note that this approximate formula holds because the product ($i \times \%\Delta S$) is small compared with the interest rate (i) and the percentage change in the exchange rate ($\%\Delta S$). For simplicity of exposition, we will use the $\cong$ symbol when we introduce an approximation but will subsequently treat the relationship as an equality (=) unless the distinction is important for the issue being discussed.

Using the previous example, consider three cases:

1 The $S_{f/d}$ rate is expected to remain unchanged.
2 The domestic currency is expected to appreciate by 10%.
3 The domestic currency is expected to appreciate by 6%.

In the first case, the investor would prefer the foreign-currency-denominated money market investment because it offers a 10% (= 10% − 0%) expected return, while the comparable domestic investment offers only 4%. In the second case, the investor would prefer the domestic investment because the expected return on the foreign-currency-denominated investment is 0% (= 10% − 10%). In the third case, uncovered interest rate parity holds because both investments offer a 4% (for the foreign investment, 10% − 6%) expected return. In this case, the risk-neutral investor is assumed to be indifferent between the alternatives.

Note that in the third case, in which uncovered interest rate parity holds, while the *expected* return over the one-year investment horizon is the same for both instruments, that expected return is *just a point on the distribution* of possible total return outcomes. The all-in return on the foreign money market instrument is uncertain because the *future* $S_{f/d}$ rate is uncertain. Hence, when we say that the investor would be indifferent between owning domestic and foreign investments because they both offer the same *expected* return (4%), we are assuming that the investor is *risk neutral* (risk-neutral investors base their decisions solely on the expected return and are indifferent to the investments' risk). Thus, uncovered interest rate parity assumes that there are enough risk-neutral investors to force equality of expected returns.

Using our example's foreign/domestic (*f*/*d*) notation, uncovered interest rate parity says the expected change in the spot exchange rate over the investment horizon should be reflected in the interest rate differential:

$$\%\Delta S^e_{f/d} = i_f - i_d, \tag{2}$$

where ΔS^e indicates the change in the spot rate expected for *future* periods. Note that Equation 2 cannot hold simultaneously for $S_{f/d}$ and $S_{d/f}$ (= $1/S_{f/d}$) because their percentage changes are not of exactly equal magnitude. This reflects our earlier approximation. Using the exact return on the unhedged foreign instrument would alleviate this issue but would produce a less intuitive equation.

In our example, if the yield spread between the foreign and domestic investments is 6% ($i_f - i_d$ = 6%), then this spread implicitly reflects the expectation that the domestic currency will strengthen versus the foreign currency by 6%.

Uncovered interest rate parity assumes that the country with the *higher* interest rate or money market yield will see its currency *depreciate*. The depreciation of the currency offsets the initial higher yield so that the (expected) all-in return on the two investment choices is the same. Hence, if the uncovered interest rate parity condition held consistently in the real world, it would rule out the possibility of earning excess returns from going long a high-yield currency and going short a low-yield currency: The depreciation of the high-yield currency would exactly offset the yield advantage that the high-yield currency offers. Taking this scenario to its logical conclusion, if uncovered interest rate parity held at all times, investors would have no incentive to shift capital from one currency to another because expected returns on otherwise identical money market investments would be equal across markets and risk-neutral investors would be indifferent among them.

Most studies have found that over short- and medium-term periods, the rate of depreciation of the high-yield currency is less than what would be implied by uncovered interest rate parity. In many cases, high-yield currencies have been found to *strengthen*, not weaken. There is, however, evidence that uncovered interest rate parity works better over very long-term horizons.

Such findings have significant implications for foreign exchange investment strategies. If high-yield currencies do not depreciate in line with the path predicted by the uncovered interest rate parity condition, then high-yield currencies should exhibit a tendency to outperform low-yield currencies over time. If so, investors could adopt strategies that overweight high-yield currencies at the expense of low-yield currencies and generate attractive returns in the process. Such approaches are known as FX carry trade strategies. We will discuss them in greater detail later.

7.2 Forward Rate Parity

Forward rate parity states that the forward exchange rate will be an unbiased predictor of the future spot exchange rate. It does not state that the forward rate will be a perfect forecast, just an unbiased one; the forward rate may overestimate or underestimate the future spot rate from time to time, but on average, it will equal the future spot rate. Forward rate parity builds upon two other parity conditions, covered interest rate parity and uncovered interest rate parity.

The covered interest rate parity condition describes the relationship among the spot exchange rate, the forward exchange rate, and interest rates. Let's keep using the foreign/domestic exchange rate notation (f/d) to simplify the explanation. The arbitrage condition that underlies covered interest rate parity (illustrated earlier) can be rearranged to give an expression for the forward premium or discount:

$$F_{f/d} - S_{f/d} = S_{f/d}\left(\frac{\left[\frac{\text{Actual}}{360}\right]}{1 + i_d\left[\frac{\text{Actual}}{360}\right]}\right)\left(i_f - i_d\right).$$

The domestic currency will trade at a forward premium ($F_{f/d} > S_{f/d}$) if, and only if, the foreign risk-free interest rate exceeds the domestic risk-free interest rate ($i_f > i_d$).

For the sake of simplicity, we assume that the investment horizon is one year, so that

$$F_{f/d} - S_{f/d} = S_{f/d}\left(\frac{i_f - i_d}{1 + i_d}\right).$$

Because the $1 + i_d$ denominator will be close to 1, we can approximate the above equation as follows:

$$F_{f/d} - S_{f/d} \cong S_{f/d}\left(i_f - i_d\right).$$

This covered interest rate parity equation can be rearranged to show the forward discount or premium as a percentage of the spot rate:

$$\frac{F_{f/d} - S_{f/d}}{S_{f/d}} \cong i_f - i_d.$$

We have also shown that if uncovered interest rate parity holds, then the expected change in the spot rate is equal to the interest rate differential:

$$\%\Delta S^e_{f/d} = i_f - i_d.$$

We can link the covered interest rate parity and uncovered interest rate parity equations as follows:

$$\frac{F_{f/d} - S_{f/d}}{S_{f/d}} = \%\Delta S^e_{f/d} = i_f - i_d.$$

Thus, the forward premium (discount) on a currency, expressed in percentage terms, equals the expected percentage appreciation (depreciation) of the domestic currency (assuming that the uncovered interest rate parity condition holds).

In theory, then, *the forward exchange rate will be an unbiased forecast of the future spot exchange rate if both covered and uncovered interest rate parity hold:*

$$F_{f/d} = S^e_{f/d}.$$

This condition is often referred to as **forward rate parity**.

We know covered interest rate parity must hold because it is enforced by arbitrage. *The question of whether forward rate parity holds is thus dependent upon whether uncovered interest rate parity holds.*

How might uncovered interest rate parity be enforced? It is not enforced by arbitrage because there is no combination of trades that will lock in a (riskless) profit. It could, however, hold if speculators willing to take risk enter the market. If the forward rate is above (below) speculators' expectations of the future spot rate, then risk-neutral speculators will buy the domestic currency in the spot (forward) market and simultaneously sell it in the forward (spot) market. These transactions would push the forward premium into alignment with the consensus expectation of the future spot rate. If the speculators' expectations are correct, they will make a profit.

Note, however, that spot exchange rates are volatile and determined by a complex web of influences: Interest rate differentials are only one among many factors. So, speculators can also lose. Because speculators are rarely, if ever, truly risk neutral and without an arbitrage relationship to enforce it, uncovered interest rate parity is often violated. *As a result, we can conclude that forward exchange rates are typically poor predictors of future spot exchange rates in the short run.* Over the longer term, uncovered interest rate parity and forward rate parity have more empirical support.

EXAMPLE 4

Covered and Uncovered Interest Rate Parity: Predictors of Future Spot Rates

An Australia-based fixed-income asset manager is deciding how to allocate money between Australia and Japan. Note that the base currency in the exchange rate quote (AUD) is the domestic currency for the asset manager.

JPY/AUD spot rate (mid-market)	71.78
One-year forward points (mid-market)	–139.4
One-year Australian deposit rate	3.00%
One-year Japanese deposit rate	1.00%

1 Based on uncovered interest rate parity, over the next year, the expected change in the JPY/AUD rate is *closest* to a(n):

A decrease of 6%.

B decrease of 2%.

C increase of 2%.

2 The *best* explanation of why this prediction may not be very accurate is that:

A covered interest rate parity does hold in this case.

B the forward points indicate that a riskless arbitrage opportunity exists.

C there is no arbitrage condition that forces uncovered interest rate parity to hold.

3 Using the forward points to forecast the future JPY/AUD spot rate one year ahead assumes that:

A investors are risk neutral.

B spot rates follow a random walk.

C it is not necessary for uncovered interest rate parity to hold.

4 Forecasting that the JPY/AUD spot rate one year from now will equal 71.78 assumes that:

A investors are risk neutral.

B spot rates follow a random walk.

C it is necessary for uncovered interest rate parity to hold.

5 If the asset manager completely hedged the currency risk associated with a one-year Japanese deposit using a forward rate contract, the one-year all-in holding return, in AUD, would be *closest* to:

A 0%.

B 1%.

C 3%.

The fixed-income manager collects the following information and uses it, along with the international parity conditions, to estimate investment returns and future exchange rate movements.

Today's One-Year Libor		Currency Pair	Spot Rate Today
JPY	0.10%	JPY/USD	105.40
USD	0.10%	USD/GBP	1.2303
GBP	3.00%	JPY/GBP	129.67

6 If covered interest rate parity holds, the all-in one-year investment return to a Japanese investor whose currency exposure to the GBP is fully hedged is *closest* to:

A 0.10%.

B 0.17%.

C 3.00%.

7 If uncovered interest rate parity holds, today's expected value for the JPY/GBP currency pair one year from now would be *closest* to:

A 126.02.

B 129.67.

C 130.05.

8 If uncovered interest rate parity holds, between today and one year from now, the expected movement in the JPY/USD currency pair is *closest* to:

A −1.60%.

B +0.00%.

C +1.63%.

Solution to 1:

B is correct. The expected depreciation of the Australian dollar (decline in the JPY/AUD rate) is equal to the interest rate differential between Australia and Japan (3% − 1%).

Solution to 2:

C is correct. There is no arbitrage condition that forces uncovered interest rate parity to hold. In contrast, arbitrage virtually always ensures that covered interest rate parity holds. This is the case for our table, where the −139 point discount is calculated from the covered interest rate parity equation.

Solution to 3:

A is correct. Using forward rates (i.e., adding the forward points to the spot rate) to forecast future spot rates assumes that uncovered interest rate parity and forward rate parity hold. Uncovered interest rate parity assumes that investors are risk neutral. If these conditions hold, then movements in the spot exchange rate, although they *approximate* a random walk, will not actually be a random walk because current interest spreads will determine expected exchange rate movements.

Solution to 4:

B is correct. Assuming that the current spot exchange rate is the best predictor of future spot rates assumes that exchange rate movements follow a random walk. If uncovered interest rate parity holds, the current exchange rate will not be the best predictor unless the interest rate differential happens to be zero. Risk neutrality is needed to enforce uncovered interest rate parity, but it will not make the current spot exchange rate the best predictor of future spot rates.

Solution to 5:

C is correct. A fully hedged JPY investment would provide the same return as the AUD investment: 3%. This represents covered interest rate parity, an arbitrage condition.

Solution to 6:

A is correct. If covered interest rate parity holds (and it very likely does, because this is a pure arbitrage relationship), then the all-in investment return to a Japanese investor in a one-year, fully hedged GBP Libor position would be identical to a one-year JPY Libor position: 0.10%. No calculations are necessary.

Solution to 7:

A is correct. If uncovered interest rate parity holds, then forward rate parity will hold and the expected spot rate one year forward is equal to the one-year forward exchange rate. This forward rate is calculated in the usual manner, given the spot exchange rates and Libors:

$$S^e = F = 129.67(1.001/1.03) = 126.02.$$

Solution to 8:

B is correct. Given uncovered interest rate parity, the expected change in a spot exchange rate is equal to the interest rate differential. At the one-year term, there is no difference between USD Libor and JPY Libor.

PURCHASING POWER PARITY

8

e explain international parity conditions (covered and uncovered interest rate parity, forward rate parity, purchasing power parity, and the international Fisher effect);

f describe relations among the international parity conditions;

g evaluate the use of the current spot rate, the forward rate, purchasing power parity, and uncovered interest parity to forecast future spot exchange rates;

So far, we have looked at the relationship between exchange rates and interest rate differentials. Now, we turn to examining the relationship between exchange rates and inflation differentials. The basis for this relationship is known as **purchasing power parity (PPP)**.

Various versions of PPP exist. The foundation for all of the versions is the law of one price. According to the **law of one price**, identical goods should trade at the same price across countries when valued in terms of a common currency. To simplify the explanation, as we did with our examples for covered and uncovered interest rate parity, let's continue to use the foreign/domestic currency quote convention (f/d) and the case where the base currency in the P/B notation is the domestic currency for the investor in the f/d notation.

The law of one price asserts that the foreign price of good x, P_f^x, should equal the exchange rate–adjusted price of the identical good in the domestic country, P_d^x:

$$P_f^x = S_{f/d} \times P_d^x.$$

For example, for a euro-based consumer, if the price of good x in the euro area is EUR 100 and the nominal exchange rate stands at 1.15 USD/EUR, then the price of good x in the United States should equal USD 115.

The **absolute version of PPP** simply extends the law of one price to the broad range of goods and services that are consumed in different countries. Expanding our example above to include all goods and services, not just good x, the broad price level of the foreign country (P_f) should equal the currency-adjusted broad price level in the domestic country (P_d):

$$P_f = (S_{f/d})(P_d).$$

This equation implicitly assumes that all domestic and foreign goods are tradable and that the domestic and foreign price indexes include the same bundle of goods and services with the same exact weights in each country. Rearranging this equation and solving for the nominal exchange rate ($S_{f/d}$), the absolute version of PPP states that the nominal exchange rate will be determined by the ratio of the foreign and domestic broad price indexes:

$$S_{f/d} = P_f/P_d.$$

The absolute version of PPP asserts that the equilibrium exchange rate between two countries is determined entirely by the ratio of their national price levels. However, it is highly unlikely that this relationship actually holds in the real world. The absolute version of PPP assumes that goods arbitrage will equate the prices of all goods and service across countries, but if transaction costs are significant and/or not all goods and services are tradable, then goods arbitrage will be incomplete. Hence, sizable and persistent departures from absolute PPP are likely.

However, if it is assumed that transaction costs and other trade impediments are constant over time, it might be possible to show that *changes* in exchange rates and *changes* in national price levels are related, even if the relationship between exchange rate *levels* and national price *levels* does not hold. According to the **relative version of PPP**, the percentage change in the spot exchange rate ($\%\Delta S_{f/d}$) will be completely determined by the difference between the foreign and domestic inflation rates ($\pi_f - \pi_d$):

$$\%\Delta S_{f/d} \cong \pi_f - \pi_d. \tag{3}$$

Intuitively, the relative version of PPP implies that the exchange rate changes to offset changes in competitiveness arising from inflation differentials. For example, if the foreign inflation rate is assumed to be 9% while the domestic inflation rate is assumed to be 5%, then the $S_{f/d}$ exchange rate must rise by 4% ($\%\Delta S_{f/d}$ = 9% − 5% = 4%) in order to maintain the relative competitiveness of the two regions: The currency of the high-inflation country should depreciate relative to the currency of the low-inflation country. If the $S_{f/d}$ exchange rate remained unchanged, the higher foreign inflation rate would erode the competitiveness of foreign companies relative to domestic companies.

Conversion from Absolute Levels to a Rate of Change

We will occasionally need to convert from a relationship expressed in levels of the relevant variables to a relationship among rates of change. If $X = (Y \times Z)$, then

$$(1 + \%\Delta X) = (1 + \%\Delta Y)(1 + \%\Delta Z)$$

and

$$\%\Delta X \approx \%\Delta Y + \%\Delta Z$$

because ($\%\Delta Y \times \%\Delta Z$) is "small." Similarly, it can be shown that if $X = (Y/Z)$, then

$$(1 + \%\Delta X) = (1 + \%\Delta Y)/(1 + \%\Delta Z)$$

and

$$\%\Delta X \approx \%\Delta Y - \%\Delta Z.$$

Applying this conversion to the equation for absolute PPP gives Equation 3.

Whereas the relative version of PPP focuses on *actual* changes in exchange rates being driven by *actual* differences in national inflation rates, the ***ex ante* version of PPP** asserts that the *expected* changes in the spot exchange rate are entirely driven by *expected* differences in national inflation rates. *Ex ante* PPP tells us that countries that are expected to run *persistently* high inflation rates should expect to see their currencies depreciate over time, while countries that are expected to run relatively low inflation rates on a sustainable basis should expect to see their currencies appreciate over time. *Ex ante* PPP can be expressed as

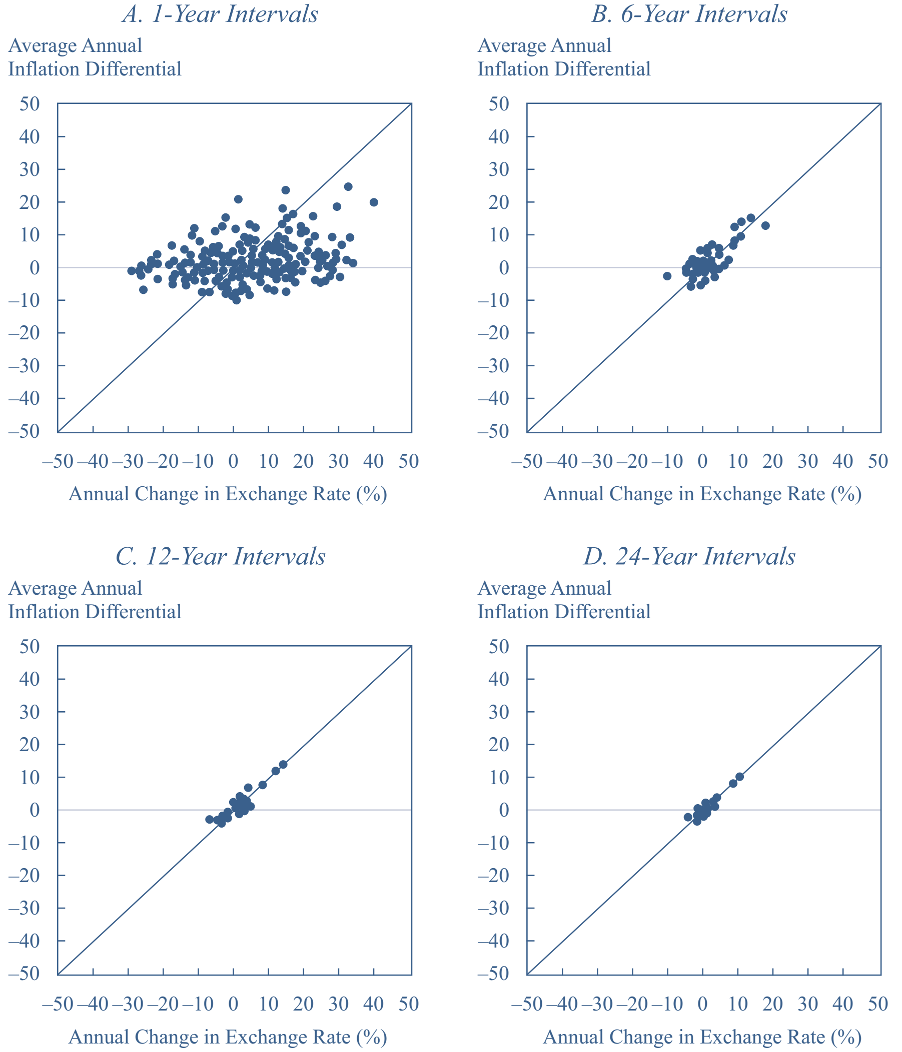
$$\%\Delta S^e_{f/d} = \pi^e_f - \pi^e_d, \tag{4}$$

where it is understood that the use of expectations (the superscript *e*) indicates that we are now focused on *future* periods. That is, $\%\Delta S^e_{f/d}$ represents the expected percentage change in the spot exchange rate, while π^e_d and π^e_f represent the expected domestic and foreign inflation rates over the same period.

Studies have found that while *over shorter horizons nominal exchange rate movements may appear random, over longer time horizons nominal exchange rates tend to gravitate toward their long-run PPP equilibrium values.*

Exhibit 2 illustrates the success, or lack thereof, of the relative version of PPP at different time horizons: 1 year, 6 years, 12 years, and 24 years. Each chart plots the inflation differential (vertical axis) against the percentage change in the exchange rate (horizontal axis). If PPP holds, the points should fall along the upward-sloping diagonal line. As indicated, there is no clear relationship between changes in exchange rates and inflation differentials over a one-year time horizon. If the horizon is lengthened to six years and beyond, however, a strong positive relationship becomes apparent. Hence, *PPP appears to be a valid framework for assessing long-run fair value in the FX markets, even though the path to PPP equilibrium is excruciatingly slow*. On average, it takes roughly three to five years to narrow a given PPP deviation by roughly 50%.

Exhibit 2 Effect of Relative Inflation Rates on Exchange Rates over Different Time Horizons



Source: Isard, Faruqee, Kincaid, and Fetherston (2001).

9 THE FISHER EFFECT, REAL INTEREST RATE PARITY AND TYING THE INTERNATIONAL PARITY CONDITIONS TOGETHER

e explain international parity conditions (covered and uncovered interest rate parity, forward rate parity, purchasing power parity, and the international Fisher effect);

f describe relations among the international parity conditions;

g evaluate the use of the current spot rate, the forward rate, purchasing power parity, and uncovered interest parity to forecast future spot exchange rates;

h explain approaches to assessing the long-run fair value of an exchange rate;

So far, we have examined the relationships between exchange rates and interest rate differentials and between exchange rates and inflation differentials. Now, we will begin to bring these concepts together by examining how exchange rates, interest rates, and inflation rates interact.

According to what economists call the Fisher effect, one can break down the nominal interest rate (i) in a given country into two parts: (1) the real interest rate (r) in that particular country and (2) the expected inflation rate (π^e) in that country:

$$i = r + \pi^e.$$

To relate this concept to exchange rates, we can write the Fisher equation for both the domestic country and a foreign country. If the Fisher effect holds, the nominal interest rates in both countries will equal the sum of their respective real interest rates and expected inflation rates:

$$i_d = r_d + \pi_d^e.$$

$$i_f = r_f + \pi_f^e.$$

Let's take a closer look at the macroeconomic forces that drive the trend in nominal yield spreads. Subtracting the top equation from the bottom equation shows that the nominal yield spread between the foreign and domestic countries ($i_f - i_d$) equals the sum of two parts: (1) the foreign–domestic real yield spread ($r_f - r_d$) and (2) the foreign–domestic expected inflation differential ($\pi_f^e - \pi_d^e$):

$$i_f - i_d = \left(r_f - r_d\right) + \left(\pi_f^e - \pi_d^e\right).$$

We can rearrange this equation to solve for the *real* interest rate differential instead of the *nominal* interest rate differential:

$$\left(r_f - r_d\right) = \left(i_f - i_d\right) - \left(\pi_f^e - \pi_d^e\right).$$

To tie this material to our previous work on exchange rates, recall our expression for uncovered interest rate parity:

$$\%\Delta S_{f/d}^e = i_f - i_d.$$

The nominal interest rate spread ($i_f - i_d$) equals the expected change in the exchange rate ($\%\Delta S_{f/d}^e$).

Recall also the expression for *ex ante* PPP:

$$\%\Delta S_{f/d}^e = \pi_f^e - \pi_d^e.$$

The difference in expected inflation rates equals the expected change in the exchange rate. Combining these two expressions, we derive the following:

$$i_f - i_d = \pi_f^e - \pi_d^e.$$

The nominal interest rate spread is equal to the difference in expected inflation rates. We can therefore conclude that if uncovered interest rate parity and *ex ante* PPP hold,

$$(r_f - r_d) = 0.$$

The real yield spread between the domestic and foreign countries ($r_f - r_d$) will be zero, and the level of real interest rates in the domestic country will be identical to the level of real interest rates in the foreign country.

The proposition that real interest rates will converge to the same level across different markets is known as the **real interest rate parity** condition.

Finally, if real interest rates are equal across markets, then it also follows that the foreign–domestic nominal yield spread is determined solely by the foreign–domestic expected inflation differential:

$$i_f - i_d = \pi_f^e - \pi_d^e.$$

This is known as the **international Fisher effect**. The reader should be aware that some authors refer to uncovered interest rate parity as the "international Fisher effect." We reserve this term for the relationship between nominal interest rate differentials and expected inflation differentials because the original (domestic) Fisher effect is a relationship between interest rates and expected inflation.

The international Fisher effect and, by extension, real interest rate parity assume that currency risk is the same throughout the world. However, not all currencies carry the same risk. For example, an emerging country may have a high level of indebtedness, which could result in an elevated level of currency risk (i.e., likelihood of currency depreciation). In this case, because the emerging market currency has higher risk, subtracting the expected inflation rate from the nominal interest rate will result in a calculated real interest rate that is higher than in other countries. Economists typically separate the nominal interest rate into the real interest rate, an inflation premium, and a risk premium. The emerging country's investors will require a risk premium for holding the currency, which will be reflected in nominal and real interest rates that are higher than would be expected under the international Fisher effect and real interest rate parity conditions.

EXAMPLE 5

PPP and the International Fisher Effect

An Australia-based fixed-income investment manager is deciding how to allocate her portfolio between Australia and Japan. (As before, the AUD is the domestic currency.) Australia's one-year deposit rate is 3%, considerably higher than Japan's 1% rate, but the Australian dollar is estimated to be roughly 10% overvalued relative to the Japanese yen based on purchasing power parity. Before making her asset allocation, the investment manager considers the implications of interest rate differentials and PPP imbalances.

1 All else equal, which of the following events would restore the Australian dollar to its PPP value?

A The Japanese inflation rate increases by 2%.

B The Australian inflation rate decreases by 10%.

C The JPY/AUD exchange rate declines by 10%.

2 If real interest rates in Japan and Australia were equal, then under the international Fisher effect, the inflation rate differential between Japan and Australia would be *closest* to:

A 0%.

B 2%.

C 10%.

3 According to the theory and empirical evidence of purchasing power parity, which of the following would *not* be true if PPP holds in the long run?

A An exchange rate's equilibrium path should be determined by the long-term trend in domestic price levels relative to foreign price levels.

B Deviations from PPP might occur over short- and medium-term periods, but fundamental forces should eventually work to push exchange rates toward their long-term PPP path.

C High-inflation countries should tend to see their currencies appreciate over time.

4 Which of the following would *best* explain the failure of the absolute version of PPP to hold?

A Inflation rates vary across countries.

B Real interest rates are converging across countries.

C Trade barriers exist, and different product mixes are consumed across countries.

Solution to 1:

C is correct. If the Australian dollar is overvalued by 10% on a PPP basis, with all else held equal, a depreciation of the JPY/AUD rate by 10% would move the Australian dollar back to equilibrium.

Solution to 2:

B is correct. If the real interest rates were equal, then the difference in nominal yields would be explained by the difference in inflation rates (3% – 1%).

Solution to 3:

C is correct. According to PPP, high-inflation countries should see their currencies depreciate (at least, over the longer term) in order to re-equilibrate real purchasing power between countries.

Solution to 4:

C is correct. The absolute version of PPP assumes that all goods and services are tradable and that the domestic and foreign price indexes include the same bundle of goods and services with the same exact weights in each country.

9.1 International Parity Conditions: Tying All the Pieces Together

As noted above, the various parity relationships usually do not hold over short time horizons. However, studies show that over longer time periods, there is a discernible interaction among nominal interest rates, exchange rates, and inflation rates across countries, such that the international parity conditions serve as an anchor for longer-term exchange rate movements. We now summarize the key international parity conditions and describe how they are all linked.

1 According to covered interest rate parity, arbitrage ensures that nominal interest rate spreads equal the percentage forward premium (or discount).

2 According to uncovered interest rate parity, the expected percentage change of the spot exchange rate should, on average, be reflected in the nominal interest rate spread.

3 If both covered and uncovered interest rate parity hold—that is, the nominal yield spread equals both the forward premium (or discount) and the expected percentage change in the spot exchange rate—then the forward exchange rate will be an unbiased predictor of the future spot exchange rate.

4 According to the *ex ante* PPP approach to exchange rate determination, the expected change in the spot exchange rate should equal the expected difference between domestic and foreign inflation rates.

5 Assuming the Fisher effect holds in all markets—that is, the nominal interest rate in each market equals the real interest rate plus the expected inflation rate—and also assuming that real interest rates are broadly the same across all markets (real interest rate parity), then the nominal yield spread between domestic and foreign markets will equal the domestic–foreign expected inflation differential, which is the international Fisher effect.

6 If *ex ante* PPP and the international Fisher effect hold, then expected inflation differentials should equal both the expected change in the exchange rate and the nominal interest rate differential. This relationship implies that the expected change in the exchange rate equals the nominal interest rate differential, which is uncovered interest rate parity.

In sum, if all the key international parity conditions held at all times, then the expected percentage change in the *spot* exchange rate would equal

- the forward premium or discount (expressed in percentage terms),
- the nominal yield spread between countries, and
- the difference between expected national inflation rates.

In other words, *if all these parity conditions held, it would be impossible for a global investor to earn consistent profits on currency movements.* If forward exchange rates accurately predicted the future path of spot exchange rates, there would be no way to make money in forward exchange speculation. If high-yield currencies fell in value versus low-yield currencies exactly in line with the path implied by nominal interest rate spreads, all markets would offer the same currency-adjusted total returns over time. Investors would have no incentive to shift funds from one market to another based solely on currency considerations.

EXAMPLE 6

The Relationships among the International Parity Conditions

1 Which of the following is a no-arbitrage condition?

  A Real interest rate parity

  B Covered interest rate parity

  C Uncovered interest rate parity

2 Forward rates are unbiased predictors of future spot rates if two parity conditions hold. Which of the following is *not* one of these conditions?

  A Real interest rate parity

  B Covered interest rate parity

  C Uncovered interest rate parity

3 The international Fisher effect requires all but which of the following to hold?

  A *Ex ante* PPP

  B Absolute PPP

  C Real interest rate parity

4 The forward premium/discount is determined by nominal interest rate differentials because of:
   A the Fisher effect.
   B covered interest parity.
   C real interest rate parity.

5 If all of the key international parity conditions held at all times, then the expected percentage change in the spot exchange rate would equal all *except* which of the following?
   A The real yield spread
   B The nominal yield spread
   C The expected inflation spread

Solution to 1:

B is correct. Covered interest rate parity is enforced by equating the investment return on two riskless investments (domestic and currency-hedged foreign).

Solution to 2:

A is correct. Both covered and uncovered interest rate parity must hold for the forward rate to be an unbiased predictor of the future spot rate. Real interest rate parity is not required.

Solution to 3:

B is correct. The international Fisher effect is based on real interest rate parity and *ex ante* PPP (not absolute PPP).

Solution to 4:

B is correct. The forward premium/discount is determined by covered interest rate arbitrage.

Solution to 5:

A is correct. If all the international parity conditions held, the real yield spread would equal zero, regardless of expected changes in the spot exchange rate.

THE CARRY TRADE 10

i. describe the carry trade and its relation to uncovered interest rate parity and calculate the profit from a carry trade;

According to uncovered interest rate parity, high-yield currencies are expected to depreciate in value, while low-yield currencies are expected to appreciate in value. If uncovered interest rate parity held at all times, investors would not be able to profit from a strategy that undertook long positions in high-yield currencies and short positions in low-yield currencies. The change in spot rates over the tenor of the forward contracts would cancel out the interest rate differentials locked in at the inception of the position.

Uncovered interest rate parity is one of the most widely tested propositions in international finance. The evidence suggests that uncovered interest rate parity does *not* hold over short and medium time periods. Studies have generally found that *high-yield currencies, on average, have not depreciated and low-yield currencies have not appreciated to the levels predicted by interest rate differentials.*

These findings underscore the potential profitability of a trading strategy known as the **FX carry trade**, which involves taking long positions in high-yield currencies and short positions in low-yield currencies. The latter are often referred to as "funding currencies." As a simplified example of the carry trade, assume a trader can borrow Canadian dollars at 1% and earn 9% on an investment in Brazilian reals for one year. To execute the trade to earn 8% from the interest rate differential, the trader will do the following:

1 Borrow Canadian dollars at $t = 0$.
2 Sell the dollars and buy Brazilian reals at the spot rate at $t = 0$.
3 Invest in a real-denominated investment at $t = 0$.
4 Liquidate the Brazilian investment at $t = 1$.
5 Sell the reals and buy dollars at the spot rate at $t = 1$.
6 Pay back the dollar loan.

If the real appreciates, the trader's profits will be greater than 8% because the stronger real will buy more dollars in one year. If the real depreciates, the trader's profits will be less than 8% because the weaker real will buy fewer dollars in the future. If the real falls in value by more than 8%, the trader will experience losses. The carry trader's return consists of the intermarket yield spread, the currency appreciation/depreciation, and the foreign investment appreciation/depreciation. Typically, a carry trade is executed using an investment in highly rated government debt so as to mitigate credit risk. In this simplified example, we use an additive approach to determine the trader's returns (i.e., we ignore the currency gain or loss on the 8% interest rate differential).

Historical evidence shows that carry trade strategies have generated positive returns over extended periods (see for example Dimson, Marsh, McGinnie, Staunton, and Wilmot 2012). One argument for the persistence of the carry trade is that the yields in higher interest rate countries reflect a risk premium due to a more unstable economy, while low-yield currencies represent less risky markets. Although small increases in financial market and/or FX volatility are unlikely to materially affect carry strategy profits, elevated levels of volatility and/or perceived risk in the financial markets can quickly turn these profits into substantial losses. That is, during turbulent periods, the returns on long high-yield currency positions will tend to decline dramatically, while the losses on short low-yield currency positions will tend to rise dramatically.

To understand why, we need to understand the nature of the risk and reward in the carry trade. The reward is the gradual accrual of the interest rate differential—income that is unrelated to exchange rate volatility. The risk arises from the potential for sudden adverse exchange rate movements that result in instantaneous capital losses. During periods of low turbulence, investors may feel relatively confident that exchange rate movements will not jeopardize the gradual accrual of the interest rate differential. Because low-volatility regimes have tended to be the norm and often last for extended periods, investors can become complacent, taking on larger carry trade positions in a search for yield but increasing their risk exposures. When volatility in the currency markets spikes, however, the risk of an adverse exchange rate movement rises sharply relative to the gradual flow of income. As the trade moves toward unprofitability, investors may rush to unwind the carry trade, selling high-yielding currencies and re-purchasing low-yielding currencies. These carry trades are often large-scale trades initiated by trading firms and other opportunistic investors, such as hedge funds. Traders often have stop-loss orders in place that are triggered when price declines reach a certain level. When they all attempt to unwind the trades at once, the selling pressure adds to the losses on the long position currency and the buying pressure on the short position currency drives that currency higher, exacerbating the loss. The "flight to quality" during turbulent times and the leverage inherent in the carry trade further compound the losses. The upshot is that *during periods of low volatility, carry trades tend to generate positive returns, but they are prone to significant crash risk in turbulent times.*

The tendency for carry trades to experience periodic crashes results in a non-normal distribution of returns for both developed and emerging market (EM) carry trades. Relative to a normal distribution, the distributions tend to be more peaked, with fatter tails and negative skewness. The more peaked distribution around the mean implies that carry trades have typically generated a larger number of trades with small gains/losses than would occur with the normal distribution. Although carry trades have generated positive returns on average in the past, the negative skew and fat tails indicate that carry trades have tended to have more frequent and larger losses than would have been experienced had the return distribution been normal.

EXAMPLE 7

Carry Trade Strategies

A currency fund manager is considering allocating a portion of her FX portfolio to carry trade strategies. The fund's investment committee asks the manager a number of questions about why she has chosen to become involved in FX carry trades and how she will manage the risk of potentially large downside moves associated with the unwinding of carry trades. Which of the following would be her *best* responses to the investment committee's questions?

1 Carry trades can be profitable when:

A covered interest rate parity does not hold.

B uncovered interest rate parity does not hold.

C the international Fisher effect does not hold.

2 Over time, the return distribution of the fund's FX carry trades is *most* likely to resemble a:

A normal distribution with fat tails.

B distribution with fat tails and a negative skew.

C distribution with thin tails and a positive skew.

3 The volatility of the fund's returns relative to its equity base is *best* explained by:

A leverage.

B low deposit rates in the funding currency.

C the yield spread between the high- and low-yielding currencies.

4 A Tokyo-based asset manager enters into a carry trade position based on borrowing in yen and investing in one-year Australian Libor.

Today's One-Year Libor		Currency Pair	Spot Rate Today	Spot Rate One Year Later
JPY	0.10%	JPY/USD	105.40	104.60
AUD	1.70%	USD/AUD	0.6810	0.6850

After one year, the all-in return to this trade, measured in JPY terms, would be *closest* to:

A +0.03%.

B +1.53%.

C +1.63%.

Solution to 1:

B is correct. The carry trade is based on the supposition that uncovered interest rate parity does not hold.

Solution to 2:

B is correct. The "crash risk" of carry trades implies a fat-tailed distribution skewed toward a higher probability of large losses (compared with a normal distribution).

Solution to 3:

A is correct. Carry trades are leveraged trades (borrow in the funding currency, invest in the high-yield currency), and leverage increases the volatility in the investor's return on equity.

Solution to 4:

B is correct. To calculate the all-in return for a Japanese investor in a one-year AUD Libor deposit, we must first calculate the current and one-year-later JPY/AUD cross rates. Because USD 1.0000 buys JPY 105.40 today and AUD 1.0000 buys USD 0.6810 today, today's JPY/AUD cross rate is the product of these two numbers: 105.40 × 0.6810 = 71.78 (rounded to two decimal places). Similarly, one year later, the observed cross rate is 104.60 × 0.6850 = 71.65 (rounded to two decimal places).

Accordingly, measured in yen, the investment return for the unhedged Australian Libor deposit is

$$(1/71.78)(1 + 1.70\%)71.65 - 1 = 0.0152.$$

Against this 1.52% *gross* return, however, the manager must charge the borrowing costs to fund the carry trade investment (one-year JPY Libor was 0.10%). Hence, the *net* return on the carry trade is 1.52% − 0.10% = 1.42%.

We can also calculate the profit using a transactional approach. Assuming an initial position of, for example, 100 yen (JPY 100), the investor will obtain JPY 100 × 1/JPY 71.78 = AUD 1.3931. After one year, the investment will be worth AUD 1.3931 × 1.017 = AUD 1.4168. Converting back to yen in one year results in AUD 1.4168 × JPY 71.65/AUD = JPY 101.51. Paying off the yen loan results in a profit of JPY 101.51 − (JPY 100 × 1.001) = JPY 1.41. This is equivalent to the 1.42% profit calculated previously (slight difference arising due to rounding).

11 THE IMPACT OF BALANCE OF PAYMENTS FLOWS: CURRENT ACCOUNT IMBALANCES AND THE DETERMINATION OF EXCHANGE RATES

j explain how flows in the balance of payment accounts affect currency exchange rates;

As noted earlier, the parity conditions may be appropriate for assessing fair value for currencies over long horizons, but they are of little use as a real-time gauge of value. There have been many attempts to find a better framework for determining a currency's short- or long-run equilibrium value. In this section, we examine the influence of trade and capital flows.

A country's balance of payments consists of its current account as well as its capital and financial account. The official balance of payments accounts make a distinction between the "capital account" and the "financial account" based on the nature of the assets involved. For simplicity, we will use the term "capital account" here to reflect all investment/financing flows. Loosely speaking, the current account reflects flows in the real economy, which refers to that part of the economy engaged in the actual production of goods and services (as opposed to the financial sector). The capital account reflects financial flows. Decisions about trade flows (the current account) and investment/financing flows (the capital account) are typically made by different entities with different perspectives and motivations. Their decisions are brought into alignment by changes in market prices and/or quantities. One of the key prices—perhaps *the* key price—in this process is the exchange rate.

Countries that import more than they export will have a negative current account balance and are said to have current account deficits. Those with more exports than imports will have a current account surplus. A country's current account balance must be matched by an equal and opposite balance in the capital account. Thus, countries with current account deficits must attract funds from abroad in order to pay for the imports (i.e., they must have a capital account surplus).

When discussing the effect of the balance of payments components on a country's exchange rate, one must distinguish between short- and intermediate-term influences on the one hand and longer-term influences on the other. Over the long term, countries that run persistent current account deficits (net borrowers) often see their currencies depreciate because they finance their acquisition of imports through the continued use of debt. Similarly, countries that run persistent current account surpluses (net lenders) often see their currencies appreciate over time.

However, investment/financing decisions are usually the dominant factor in determining exchange rate movements, at least in the short to intermediate term. There are four main reasons for this:

- Prices of real goods and services tend to adjust much more slowly than exchange rates and other asset prices.
- Production of real goods and services takes time, and demand decisions are subject to substantial inertia. In contrast, liquid financial markets allow virtually instantaneous redirection of financial flows.
- Current spending/production decisions reflect only purchases/sales of current production, while investment/financing decisions reflect not only the financing of current expenditures but also the reallocation of existing portfolios.
- *Expected* exchange rate movements can induce very large short-term capital flows. This tends to make the *actual* exchange rate very sensitive to the currency views held by owners/managers of liquid assets.

In this section, we first examine the impact of current account imbalances on exchange rates. Then, we take a closer look at capital flows.

11.1 Current Account Imbalances and the Determination of Exchange Rates

Current account trends influence the path of exchange rates over time through several mechanisms:

- The flow supply/demand channel
- The portfolio balance channel
- The debt sustainability channel

We briefly discuss each of these mechanisms next.

11.1.1 *The Flow Supply/Demand Channel*

The flow supply/demand channel is based on a fairly simple model that focuses on the fact that purchases and sales of internationally traded goods and services require the exchange of domestic and foreign currencies in order to arrange payment for those goods and services. For example, if a country sold more goods and services than it purchased (i.e., the country was running a current account surplus), then the demand for its currency should rise, and vice versa. Such shifts in currency demand should exert upward pressure on the value of the surplus nation's currency and downward pressure on the value of the deficit nation's currency.

Hence, countries with persistent current account surpluses should see their currencies appreciate over time, and countries with persistent current account deficits should see their currencies depreciate over time. A logical question, then, would be whether such trends can go on indefinitely. At some point, domestic currency strength should contribute to deterioration in the trade competitiveness of the surplus nation, while domestic currency weakness should contribute to an improvement in the trade competitiveness of the deficit nation. Thus, the exchange rate responses to these surpluses and deficits should eventually help eliminate—in the medium to long run—the source of the initial imbalances.

The amount by which exchange rates must adjust to restore current accounts to balanced positions depends on a number of factors:

- The initial gap between imports and exports
- The response of import and export prices to changes in the exchange rate
- The response of import and export demand to changes in import and export prices

If a country imports significantly more than it exports, export growth would need to far outstrip import growth in percentage terms in order to narrow the current account deficit. A large initial deficit may require a substantial depreciation of the currency to bring about a meaningful correction of the trade imbalance.

A depreciation of a deficit country's currency should result in an increase in import prices in domestic currency terms and a decrease in export prices in foreign currency terms. However, empirical studies often find limited pass-through effects of exchange rate changes on traded goods prices. For example, many studies have found that for every 1% decline in a currency's value, import prices rise by only 0.5%—and in some cases by even less—because foreign producers tend to lower their profit margins in an effort to preserve market share. In light of the limited pass-through of exchange rate changes into traded goods prices, the exchange rate adjustment required to narrow a trade imbalance may be far larger than would otherwise be the case.

Many studies have found that the response of import and export demand to changes in traded goods prices is often quite sluggish, and as a result, relatively long lags, lasting several years, can occur between (1) the onset of exchange rate changes, (2) the ultimate adjustment in traded goods prices, and (3) the eventual impact of those price changes on import demand, export demand, and the underlying current account imbalance.

11.1.2 *The Portfolio Balance Channel*

The second mechanism through which current account trends influence exchange rates is the so-called portfolio balance channel. Current account imbalances shift financial wealth from deficit nations to surplus nations. Countries with trade deficits will finance their trade with increased borrowing. This behavior may lead to shifts in global asset preferences, which in turn could influence the path of exchange rates. For example, nations running large current account surpluses versus the United States might find that their holdings of US dollar–denominated assets exceed the amount they desire to hold in a portfolio context. Actions they might take to reduce their dollar holdings to desired levels could then have a profound negative impact on the dollar's value.

11.1.3 *The Debt Sustainability Channel*

The third mechanism through which current account imbalances can affect exchange rates is the so-called debt sustainability channel. According to this mechanism, there should be some upper limit on the ability of countries to run persistently large current account deficits. If a country runs a large and persistent current account deficit over time, eventually it will experience an untenable rise in debt owed to foreign investors. If such investors believe that the deficit country's external debt is rising to unsustainable levels, they are likely to reason that a major depreciation of the deficit country's currency will be required at some point to ensure that the current account deficit narrows significantly and that the external debt stabilizes at a level deemed sustainable.

The existence of persistent current account imbalances will tend to alter the market's notion of what exchange rate level represents the true, long-run equilibrium value. For deficit nations, ever-rising net external debt levels as a percentage of GDP should give rise to steady (but not necessarily smooth) downward revisions in market expectations of the currency's long-run equilibrium value. For surplus countries, ever-rising net external asset levels as a percentage of GDP should give rise to steady upward revisions of the currency's long-run equilibrium value. Hence, one would expect currency values to move broadly in line with trends in debt and/or asset accumulation.

Persistent Current Account Deficits: The US Current Account and the US Dollar

The historical record indicates that the trend in the US current account has been an important determinant of the long-term swings in the US dollar's value but also that there can be rather long lags between the onset of a deterioration in the current account balance and an eventual decline in the dollar's value. For example, the US current account balance deteriorated sharply in the first half of the 1980s, yet the dollar soared over that period. The reason for the dollar's strength over that period was that high US real interest rates attracted large inflows of capital from abroad, which pushed the dollar higher despite the large US external imbalance. Eventually, however, concerns regarding the sustainability of the ever-widening US current account deficit triggered a major dollar decline in the second half of the 1980s.

History repeated itself in the second half of the 1990s, with the US current account balance once again deteriorating while the dollar soared over the same period. This time, the dollar's strength was driven by strong foreign direct investment, as well as both debt- and equity-related flows into the United States. Beginning in 2001, however, the ever-widening US current account deficit, coupled with a decline in US interest rates, made it more difficult for the United States to attract the foreign private capital needed to finance its current account deficit. The dollar eventually succumbed to the weight of ever-larger trade and current account deficits and began a multi-year slide, starting in 2002–2003. Interestingly, the US dollar has undergone three major downward cycles since the advent of floating exchange rates: 1977–1978, 1985–1987, and 2002–2008. In each of those downward cycles, the dollar's slide was driven in large part by concerns over outsized US current account deficits coupled with relatively low nominal and/or real short-term US interest rates, which made it difficult to attract sufficient foreign capital to the United States to finance those deficits.

Exchange Rate Adjustment in Surplus Nations: Japan and China

Japan and, for a number of years, China represent examples of countries with large current account surpluses and illustrate the pressure that those surpluses can bring to bear on currencies. In the case of Japan, its rising current account surplus has exerted persistent upward pressure on the yen's value versus the dollar over time. Part of this upward pressure simply reflected the increase in demand for yen to pay for Japan's merchandise exports. But some of the upward pressure on the yen might also have stemmed from rising commercial tensions between the United States and Japan.

Protectionist sentiment in the United States rose steadily with the rising bilateral trade deficit that the United States ran with Japan in the postwar period. US policymakers contended that the yen was undervalued and needed to appreciate. With the increasing trade imbalance between the two countries contributing to more heated protectionist rhetoric, Japan felt compelled to tolerate steady upward pressure on the yen. As a result, the yen's value versus the dollar has tended to move in sync with the trend in Japan's current account surplus.

12 CAPITAL FLOWS AND THE DETERMINATION OF EXCHANGE RATES AND EQUITY MARKET TRENDS AND EXCHANGE RATES,

j explain how flows in the balance of payment accounts affect currency exchange rates;

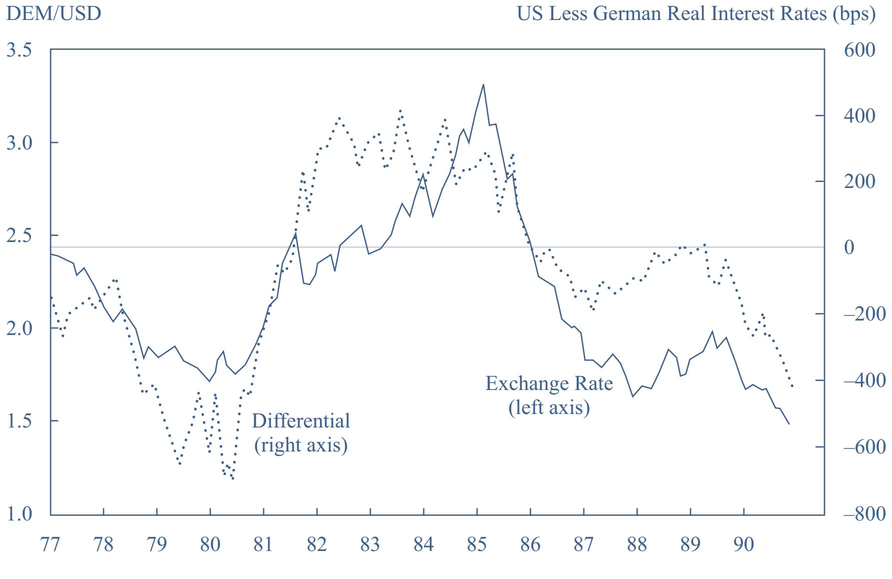
Greater financial integration of the world's capital markets and greater freedom of capital to flow across national borders have increased the importance of global financial flows in determining exchange rates, interest rates, and broad asset price trends. One can cite many examples in which global financial flows either caused or contributed to extremes in exchange rates, interest rates, or asset prices. Two specific examples are given below:

- Yen assets underperformed US dollar assets over much of the 1995–2007 period, a 12-year span of ultra-low Japanese short-term interest rates that gave rise to what became known as the "yen carry trade" as both Japanese and global fund managers borrowed in yen and invested the proceeds in higher-yielding assets in other markets. Such actions helped push the value of the yen significantly lower over time. Periodically, however, such positions became overextended and vulnerable to sudden reversals. In the fall of 1998, a major unwinding of the yen carry trade led to the collapse of several major hedge funds.
- In the first half of the 1980s, a major widening in yield spreads favoring the United States attracted significant amounts of foreign capital to that country, fueling a rise in the dollar. Again in the second half of the 1990s, increasing capital flows to the United States fueled by global demand for US financial assets—specifically US equities during the tech boom—spurred another dramatic rise in the US dollar.

Exhibits 3 and 4 illustrate the pivotal role that real interest rate differentials played in driving the US dollar's value during these periods. As shown in Exhibit 3, the decline of the dollar in the late 1970s, the dramatic rise in its value in the first half of the 1980s, and its subsequent decline in the second half of the 1980s can be explained, to a large extent, by changes in US–foreign real yield spreads. In Exhibits 3 and 4, DEM/USD indicates the number of German Deutsche Marks per US dollar (Germany used Deutsche Marks before the adoption of the euro). Exhibit 4 shows that the dollar's decline in the first half of the 1990s coincided with a significant narrowing in US–foreign real yield spreads, while the dollar's subsequent rise in the second half of the 1990s coincided with a significant widening in US–foreign real yield spreads.

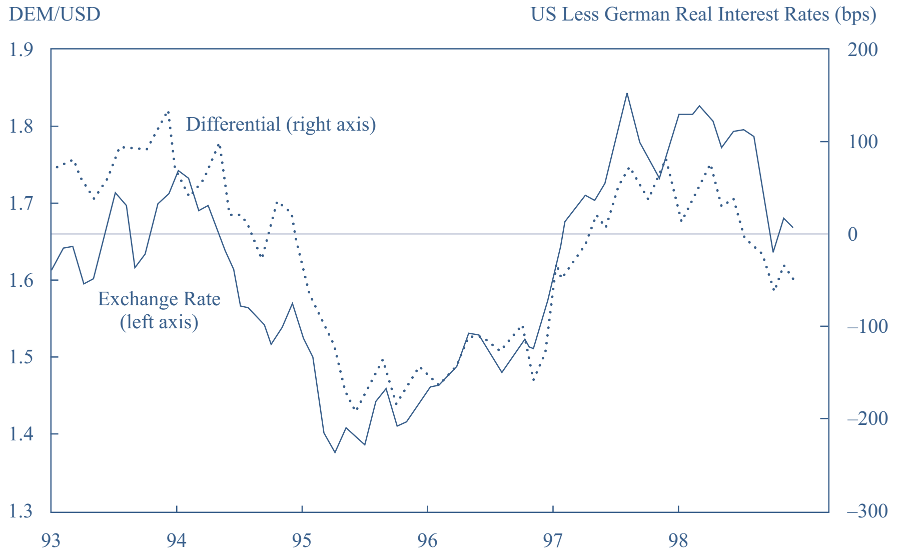
In numerous cases, global capital flows have helped fuel boom-like conditions in emerging market economies for a while before, suddenly and often without adequate warning, those flows reversed. The reversals often caused a major economic downturn, sovereign default, a serious banking crisis, and/or significant currency depreciation. Excessive emerging market capital inflows often plant the seeds of a crisis by contributing to (1) an unwarranted appreciation of the emerging market currency, (2) a huge buildup in external indebtedness, (3) an asset bubble, (4) a consumption binge that contributes to explosive growth in domestic credit and/or the current account deficit, or (5) an overinvestment in risky projects and questionable activities. Governments in emerging markets often resist currency appreciation from excessive capital inflows by using capital controls or selling their currency in the FX market. As an example of the former, in 2016 the Brazilian government was using a tax on foreign exchange transactions to control capital flows and raise government revenue. Note, however, that in general, government control of the exchange rate will not be completely effective because even if a government prohibits investment capital flows, some capital flows will be needed for international trade. In addition, the existence or emergence of black markets for the country's currency will inhibit the ability of the government to fully control the exchange rates for its own currency.

Exhibit 3 Deutsche Mark/US Dollar Exchange Rate and US/German Real Interest Rate Differentials (10-Year Bond Yields less CPI), 1977–1990



Source: Datastream.

Exhibit 4 Deutsche Mark/US Dollar Exchange Rate and US/German Real Interest Rate Differentials (10-Year Bond Yields less CPI), 1993–1998



Source: Datastream.

Sometimes, capital flows due to interest rate spreads have little impact on the trend in exchange rates. Consider the case of the Turkish lira. The lira attracted a lot of interest on the part of global fund managers over the 2002–10 period, in large part because of its attractive yields. Turkish–US short-term yield spreads averaged over 1,000 bps during much of this period. As capital flowed into Turkey, the Turkish authorities intervened in the foreign exchange market in an attempt to keep the lira from appreciating. As a result, international investors were not able to reap the anticipated currency gains over this period. While the return from the movement in the spot exchange rate was fairly small, a long Turkish lira/short US dollar carry trade position generated significant long-run returns, mostly from the accumulated yield spread.

One-sided capital flows can persist for long periods. Consider the case of a high-yield, inflation-prone emerging market country that wants to promote price stability and long-term sustainable growth. To achieve price stability, policymakers in the high-yield economy will initiate a tightening in monetary policy by gradually raising the level of domestic interest rates relative to yield levels in the rest of the world. If the tightening in domestic monetary policy is sustained, inflation expectations for the high-yield economy relative to other economies should gradually decline. The combination of sustained wide nominal yield spreads and a steady narrowing in relative inflation expectations should exert upward pressure on the high-yield currency's value, resulting in carry trade profits over long periods.

Policymakers in high-yield markets can also pursue policies which attract foreign investment; such policies might include tighter fiscal policies, liberalization of financial markets, fewer capital flow restrictions, privatization, and/or a better business environment. Such policies should encourage investors to gradually require a lower risk premium to hold the high-yield currency's assets and revise upward their assessment of the long-run equilibrium value of that country's currency.

The historical evidence suggests that the impact of nominal interest rate spreads on the exchange rate tends to be gradual. Monetary policymakers tend to adjust their official lending rates slowly over time—in part because of the uncertainty that policymakers face and in part because the authorities do not want to disrupt the financial markets. This very gradual change in rates implies a very gradual narrowing of the spread between high-yield and low-yield countries. Similarly, the downward trends in inflation expectations and risk premiums in the higher-yield market also tend to unfold gradually. It often takes several years to determine whether structural economic changes will take root and boost the long-run competitiveness of the higher-yield country. Because these fundamental drivers tend to reinforce each other over time, there may be persistence in capital flows and carry trade returns.

12.1 Equity Market Trends and Exchange Rates

Increasing equity prices can also attract foreign capital. Although exchange rates and equity market returns sometimes exhibit positive correlation, the relationship between equity market performance and exchange rates is not stable. The long-run correlation between the US equity market and the dollar, for example, is very close to zero, but over short to medium periods, correlations tend to swing from being highly positive to being highly negative, depending on market conditions. For instance, between 1990 and 1995, the US dollar fell while the US equity market was strong and the Japanese yen soared while Japanese stocks were weak. In contrast, between 1995 and early 2000, the US dollar soared in tandem with a rising US equity market while the yen weakened in tandem with a decline in the Japanese equity market. *Such instability in the correlation between exchange rates and equity markets makes it difficult to form judgments on possible future currency moves based solely on expected equity market performance.*

Since the global financial crisis, there has been a decidedly negative correlation between the US dollar and the US equity market. Market observers attribute this behavior of the US dollar to its role as a safe haven asset. When investors' appetite for risk is high—that is, when the market is in "risk-on" mode—investor demand for risky assets, such as equities, tends to rise, which drives up their prices. At the same time, investor demand for safe haven assets, such as the dollar, tends to decline, which drives their values lower. The opposite has occurred when the market has been in "risk-off" mode.

EXAMPLE 8

Capital Flows and Exchange Rates

Monique Kwan, a currency strategist at a major foreign exchange dealer, is responsible for formulating trading strategies for the currencies of both developed market (DM) and emerging market (EM) countries. She examines two countries—one DM and one EM—and notes that the DM country has what is considered a low-yield safe haven currency while the EM country has a high-yield currency whose value is more exposed to fluctuations in the global economic growth rate. Kwan is trying to form an opinion about movements in the exchange rate for the EM currency.

1 All else equal, the exchange rate for the EM currency will *most likely* depreciate if the:

 A long-run equilibrium value of the high-yield currency is revised upward.

 B nominal yield spread between the EM and DM countries increases over time.

C expected inflation differential between the EM and DM countries is revised upward.

2 An increase in safe haven demand would *most likely*:

A increase the risk premium demanded by international investors to hold assets denominated in the EM currency.

B raise the return earned on carry trade strategies.

C exert upward pressure on the value of the EM currency.

Kwan notes that the DM country is running a persistent current account deficit with the EM country. To isolate the influence of this chronic imbalance on exchange rates, she focuses only on the bilateral relationship between the EM and DM countries and makes the simplifying assumption that the external accounts of these two countries are otherwise balanced (i.e., there are no other current account deficits).

3 Over time and all else equal, the persistent current account deficit with the EM country would *most likely* lead to:

A a large buildup of the EM country's assets held by the DM country.

B an increase in the trade competitiveness of the EM country.

C an upward revision in the long-run equilibrium EM currency value.

Kwan notes that because of the high yield on the EM country's bonds, international investors have recently been reallocating their portfolios more heavily toward this country's assets. As a result of these capital inflows, the EM country has been experiencing boom-like conditions.

4 Given the current boom-like conditions in the EM economy, in the *near term*, these capital inflows are *most likely* to lead to:

A a decrease in inflation expectations in the EM.

B an increase in the risk premium for the EM.

C an increase in the EM currency value.

5 If these capital inflows led to an unwanted appreciation in the real value of its currency, the EM country's government would *most likely*:

A impose capital controls.

B decrease taxes on consumption and investment.

C buy its currency in the foreign exchange market.

6 If government actions were ineffective and the EM country's bubble eventually burst, this would *most likely* be reflected in an increase in:

A the risk premium for the EM.

B the EM currency value.

C the long-run equilibrium EM currency value.

Finally, Kwan turns to examining the link between the value of the DM country's currency and movements in the DM country's main stock market index. One of her research associates tells her that, in general, the correlation between equity market returns and changes in exchange rates has been found to be highly positive over time.

7 The statement made by the research associate is:

A correct.

B incorrect, because the correlation is highly negative over time.

C incorrect, because the correlation is not stable and tends to converge toward zero in the long run.

Solution to 1:

C is correct. All else equal, an increase in the expected inflation differential should lead to depreciation of the EM currency.

Solution to 2:

A is correct. During times of intense risk aversion, investors will crowd into the safe haven currency. This tendency implies an increased risk premium demanded by investors to hold the EM currency.

Solution to 3:

C is correct. Over time, the DM country will see its level of external debt rise as a result of the chronic current account imbalance. Eventually, this trend should lead to a downward revision of the DM currency's long-run equilibrium level (via the debt sustainability channel). This is equivalent to an *increase* in the EM currency's long-run exchange rate. A is incorrect because the DM country's current account deficit is likely to lead to a buildup in DM country assets held by the EM country. B is incorrect because, at some point, the currency strength should contribute to deterioration in the trade competitiveness of the country with the trade surplus (the EM country).

Solution to 4:

C is correct. Given the current investor enthusiasm for the EM country's assets and the boom-like conditions in the country, it is most likely that in the near term, the EM currency will appreciate. At the same time, expected inflation in the EM country is also likely increasing and—given the enthusiasm for EM assets—the risk premium is likely decreasing.

Solution to 5:

A is correct. To reduce unwanted appreciation of its currency, the EM country would be most likely to impose capital controls to counteract the surging capital inflows. Because these inflows are often associated with overinvestment and consumption, the EM government would not be likely to encourage these activities through lower taxes. Nor would the EM country be likely to encourage further currency appreciation by intervening in the market to *buy* its own currency.

Solution to 6:

A is correct. Episodes of surging capital flows into EM countries have often ended badly (with a rapid reversal of these inflows as the bubble bursts). This is most likely to be reflected in an increase in the EM risk premium. It is much less likely that a bursting bubble would be reflected in an increase in either the EM currency value or its long-term equilibrium value.

Solution to 7:

C is correct. Correlations between equity returns and exchange rates are unstable in the short term and tend toward zero in the long run.

13 MONETARY AND FISCAL POLICIES

k explain the potential effects of monetary and fiscal policy on exchange rates;

As the foregoing discussion indicates, government policies can have a significant impact on exchange rate movements. We now examine the channels through which government monetary and fiscal policies are transmitted.

13.1 The Mundell–Fleming Model

The Mundell–Fleming model describes how changes in monetary and fiscal policy within a country affect interest rates and economic activity, which in turn leads to changes in capital flows and trade and ultimately to changes in the exchange rate. The model focuses only on aggregate demand and assumes there is sufficient slack in the economy to allow increases in output without price level increases.

In this model, expansionary monetary policy affects growth, in part, by reducing interest rates and thereby increasing investment and consumption spending. Given flexible exchange rates and expansionary monetary policy, downward pressure on domestic interest rates will induce capital to flow to higher-yielding markets, putting downward pressure on the domestic currency. The more responsive capital flows are to interest rate differentials, the greater the depreciation of the currency.

Expansionary fiscal policy—either directly through increased spending or indirectly via lower taxes—typically exerts upward pressure on interest rates because larger budget deficits must be financed. With flexible exchange rates and mobile capital, the rising domestic interest rates will attract capital from lower-yielding markets, putting upward pressure on the domestic currency. If capital flows are highly sensitive to interest rate differentials, then the domestic currency will tend to appreciate substantially. If, however, capital flows are immobile and very insensitive to interest rate differentials, the policy-induced increase in aggregate demand will increase imports and worsen the trade balance, creating downward pressure on the currency with no offsetting capital inflows to provide support for the currency.

The specific mix of monetary and fiscal policies in a country can have a profound effect on its exchange rate. Consider first the case of high capital mobility. With floating exchange rates and high capital mobility, a domestic currency will appreciate given a restrictive domestic monetary policy and/or an expansionary fiscal policy. Similarly, a domestic currency will depreciate given an expansionary domestic monetary policy and/or a restrictive fiscal policy. In Exhibit 5, we show that the combination of a restrictive monetary policy and an expansionary fiscal policy is extremely bullish for a currency when capital mobility is high; likewise, the combination of an expansionary monetary policy and a restrictive fiscal policy is bearish for a currency. The effect on the currency of monetary and fiscal policies that are both expansionary or both restrictive is indeterminate under conditions of high capital mobility.

Exhibit 5 Monetary–Fiscal Policy Mix and the Determination of Exchange Rates under Conditions of High Capital Mobility

	Expansionary Monetary Policy	Restrictive Monetary Policy
Expansionary Fiscal Policy	Indeterminate	Domestic currency appreciates
Restrictive Fiscal Policy	Domestic currency depreciates	Indeterminate

Source: Rosenberg (1996, p. 132).

When capital mobility is low, the effects of monetary and fiscal policy on exchange rates will operate primarily through trade flows rather than capital flows. The combination of expansionary monetary *and* fiscal policy will be bearish for a currency. Earlier we said that expansionary fiscal policy will increase imports and hence the trade deficit, creating downward pressure on the currency. Layering on an expansive monetary policy will further boost spending and imports, worsening the trade balance and exacerbating the downward pressure on the currency.

The combination of restrictive monetary *and* fiscal policy will be bullish for a currency. This policy mix will tend to reduce imports, leading to an improvement in the trade balance.

The impact of expansionary monetary and restrictive fiscal policies (or restrictive monetary and expansionary fiscal policies) on aggregate demand and the trade balance, and hence on the exchange rate, is indeterminate under conditions of low capital mobility. Exhibit 6 summarizes these results.

Exhibit 6 Monetary–Fiscal Policy Mix and the Determination of Exchange Rates under Conditions of Low Capital Mobility

	Expansionary Monetary Policy	Restrictive Monetary Policy
Expansionary Fiscal Policy	Domestic currency depreciates	Indeterminate
Restrictive Fiscal Policy	Indeterminate	Domestic currency appreciates

Source: Adapted from Rosenberg (1996, p. 133).

Exhibit 5 is more relevant for the G–10 countries because capital mobility tends to be high in developed economies. Exhibit 6 is more relevant for emerging market economies that restrict capital movement.

A classic case in which a dramatic shift in the policy mix caused dramatic changes in exchange rates was that of Germany in 1990–1992. During that period, the German government pursued a highly expansionary fiscal policy to help facilitate German unification. At the same time, the Bundesbank pursued an extraordinarily restrictive monetary policy to combat the inflationary pressures associated with unification. The expansive fiscal/restrictive monetary policy mix drove German interest rates sharply higher, eventually causing the German currency to appreciate.

13.2 Monetary Models of Exchange Rate Determination

In the Mundell–Fleming model, monetary policy is transmitted to the exchange rate through its impact on interest rates and output. Changes in the price level and/or the inflation rate play no role. Monetary models of exchange rate determination generally take the opposite perspective: Output is fixed and monetary policy affects exchange rates primarily through the price level and the rate of inflation. In this section, we briefly describe two variations of the monetary approach to exchange rate determination.

The monetary approach asserts that an X percent rise in the domestic money supply will produce an X percent rise in the domestic price level. Assuming that purchasing power parity holds—that is, that changes in exchange rates reflect changes in relative inflation rates—a money supply–induced increase (decrease) in domestic prices relative to foreign prices should lead to a proportional decrease (increase) in the domestic currency's value.

One of the major shortcomings of the pure monetary approach is the assumption that purchasing power parity holds in both the short and long runs. Because purchasing power parity rarely holds in either the short or medium run, the pure monetary model may not provide a realistic explanation of the impact of monetary forces on the exchange rate.

To rectify that problem, Dornbusch (1976) constructed a modified monetary model that assumes prices have limited flexibility in the short run but are fully flexible in the long run. The long-run flexibility of the price level ensures that any increase in the domestic money supply will give rise to a proportional increase in domestic prices and thus contribute to a depreciation of the domestic currency in the long run, which is consistent with the pure monetary model. If the domestic price level is assumed to be inflexible in the short run, however, the model implies that the exchange rate is likely to overshoot its long-run PPP path in the short run. With inflexible domestic prices in the short run, any increase in the nominal money supply results in a decline in the domestic interest rate. Assuming that capital is highly mobile, the decline in domestic interest rates will precipitate a capital outflow, which in the short run will cause the domestic currency to depreciate below its new long-run equilibrium level. In the long run, once domestic nominal interest rates rise, the currency will appreciate and move into line with the path predicted by the conventional monetary approach.

Monetary Policy and Exchange Rates: The Historical Evidence

Historically, changes in monetary policy have had a profound impact on exchange rates. In the case of the US dollar, the Federal Reserve's policy of quantitative easing after the global financial crisis resulted in dollar depreciation from mid-2009 to 2011. The subsequent ending of quantitative easing in 2014, along with the anticipation that the United States would raise interest rates before many other countries, played a key role in driving the dollar higher.

Beginning in 2013, Abenomics—fiscal stimulus, monetary easing, and structural reforms—and the use of quantitative easing in Japan led to a steady decline in interest rates and eventually to negative interest rates in 2016. From 2013 to 2015, the value of the yen changed from roughly JPY 90/USD to JPY 120/USD. Likewise, the use of quantitative easing by the European Central Bank in 2015 led to declines in the value of the euro.

Excessively expansionary monetary policies by central banks in emerging markets have often planted the seeds of speculative attacks on their currencies. In the early 1980s, exchange rate crises in Argentina, Brazil, Chile, and Mexico were all preceded by sharp accelerations in domestic credit expansions. In 2012, Venezuela began a period of triple-digit inflation, followed by a massive currency depreciation and an economic crisis.

EXAMPLE 9

Monetary Policy and Exchange Rates

Monique Kwan, the currency strategist at a major foreign exchange dealer, is preparing a report on the outlook for several currencies that she follows. She begins by considering the outlook for the currency of a developed market country with high capital mobility across its borders and a flexible exchange rate. This DM country also has low levels of public and private debt.

Given these conditions, Kwan tries to assess the impact of each of the following policy changes.

1 For the DM currency, increasing the degree of monetary easing (reducing interest rates and increasing money supply) will *most likely*:

- **A** cause the currency to appreciate.
- **B** cause the currency to depreciate.
- **C** have an indeterminate effect on the currency.

2 The pursuit of an expansionary domestic fiscal policy by the DM country will, in the short run, *most likely:*

  A cause the domestic currency's value to appreciate.

  B cause the domestic currency's value to depreciate.

  C have an indeterminate effect on the domestic currency's value.

Next, Kwan turns her attention to an emerging market country that has low levels of public and private debt. Currently, the EM country has a fixed exchange rate but no controls over international capital mobility. However, the country is considering replacing its fixed exchange rate policy with a policy based on capital controls. These proposed controls are meant to reduce international capital mobility by limiting short-term investment flows ("hot money") in and out of its domestic capital markets.

3 To maintain the exchange rate peg while increasing the degree of monetary easing, the EM country will *most likely* have to:

  A tighten fiscal policy.

  B decrease interest rates.

  C buy its own currency in the FX market.

4 After the EM country replaces its currency peg with capital controls, would its exchange rate be unaffected by a tightening in monetary policy?

  A Yes.

  B No, the domestic currency would appreciate.

  C No, the domestic currency would depreciate.

5 After the EM country replaces its currency peg with capital controls, the simultaneous pursuit of a tight monetary policy and a highly expansionary fiscal policy by the EM country will *most likely*:

  A cause the currency to appreciate.

  B cause the currency to depreciate.

  C have an indeterminate effect on the currency.

Solution to 1:

B is correct. A decrease in the policy rate would most likely cause capital to reallocate to higher-yielding markets. This would lead to currency depreciation.

Solution to 2:

A is correct. An expansionary fiscal policy will lead to higher levels of government debt and interest rates, which will attract international capital flows. (In the long run, however, an excessive buildup in debt may eventually cause downward pressure on the domestic currency.)

Solution to 3:

C is correct. The looser monetary policy will lead to exchange rate depreciation. To counter this effect and maintain the currency peg, the central bank will have to intervene in the FX market, buying the country's own currency. A is incorrect because tighter fiscal policy is associated with lower interest rates and is therefore likely to increase rather than mitigate the downward pressure on the domestic currency. Similarly, B is incorrect because a move to lower interest rates would exacerbate the downward pressure on the currency and hence the pressure on the peg.

Solution to 4:

B is correct. In general, capital controls will not completely eliminate capital flows but will limit their magnitude and responsiveness to investment incentives such as interest rate differentials. At a minimum, flows directly related to financing international trade will typically be allowed. The exchange rate will still respond to monetary policy. With limited capital mobility, however, monetary policy's main influence is likely to come through the impact on aggregate demand and the trade balance. A tighter domestic monetary policy will most likely lead to higher interest rates and less domestic demand, including less demand for imported goods. With fewer imports and with exports held constant, there will be modest upward pressure on the currency.

Solution to 5:

C is correct because (1) capital mobility is low, so the induced increase in interest rates is likely to exert only weak upward pressure on the currency; (2) the combined impact on aggregate demand is indeterminate; and (3) if aggregate demand increases, the downward pressure on the currency due to a worsening trade balance may or may not fully offset the upward pressure exerted by capital flows.

13.3 The Portfolio Balance Approach

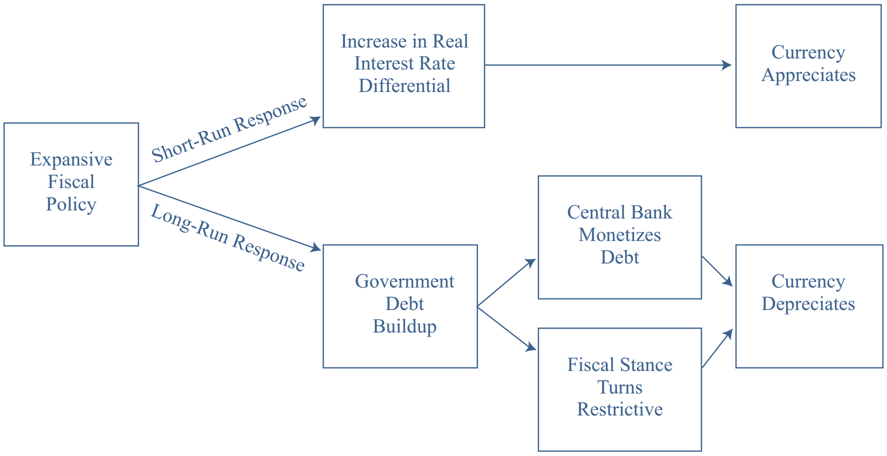
In this section, we re-examine the role fiscal policy plays in determining exchange rates. The Mundell–Fleming model is essentially a short-run model of exchange rate determination. It makes no allowance for the long-term effects of budgetary imbalances that typically arise from sustained fiscal policy actions. The portfolio balance approach to exchange rate determination remedies this limitation. In our previous discussion of the portfolio balance channel, we stated that the currencies of countries with trade deficits will decline over time. We expand that discussion here to more closely examine how exchange rates change over the long term.

In the **portfolio balance approach**, global investors are assumed to hold a diversified portfolio of domestic and foreign assets, including bonds. The desired allocation is assumed to vary in response to changes in expected return and risk considerations. In this framework, a growing government budget deficit leads to a steady increase in the supply of domestic bonds outstanding. These bonds will be willingly held only if investors are compensated in the form of a higher expected return. Such a return could come from (1) higher interest rates and/or a higher risk premium, (2) immediate *depreciation of the currency to a level sufficient to generate anticipation of gains from subsequent currency appreciation*, or (3) some combination of these two factors. The currency adjustments required in the second mechanism are the core of the portfolio balance approach.

One of the major insights one should draw from the portfolio balance model is that *in the long run, governments that run large budget deficits on a sustained basis could eventually see their currencies decline in value.*

The Mundell–Fleming and portfolio balance models can be combined into a single integrated framework in which expansionary fiscal policy under conditions of high capital mobility may be positive for a currency in the short run but negative in the long run. Exhibit 7 illustrates this concept. A domestic currency may rise in value when the expansionary fiscal policy is first put into place. As deficits mount over time and the government's debt obligations rise, however, market participants will begin to wonder how that debt will be financed. If the volume of debt rises to levels that are believed to be unsustainable, market participants may believe that the central bank will eventually be pressured to "monetize" the debt—that is, to buy the government's debt with newly created money. Such a scenario would clearly lead to a rapid reversal of the initial currency appreciation. Alternatively, the market may believe that the government will eventually have to shift toward significant restraint to implement a more restrictive, sustainable fiscal policy over the longer term.

Exhibit 7 The Short- and Long-Run Response of Exchange Rates to Changes in Fiscal Policy



Source: Rosenberg (2003).

EXAMPLE 10

Fiscal Policy and Exchange Rates

Monique Kwan is continuing her analysis of the foreign exchange rate outlook for selected countries. She examines a DM country that has a high degree of capital mobility and a floating-rate currency regime. Kwan notices that although the current outstanding volume of government debt is low, as a percentage of GDP, it is rising sharply as a result of expansionary fiscal policy. Moreover, projections for the government debt-to-GDP ratio point to further increases well into the future.

Kwan uses the Mundell–Fleming and portfolio balance models to form an opinion about both the short-run and long-run implications for the DM country's exchange rate.

1 Over the short run, Kwan is *most likely* to expect:
   - A appreciation of the DM's currency.
   - B an increase in the DM's asset prices.
   - C a decrease in the DM's risk premium.

2 Over the medium term, as the DM country's government debt becomes harder to finance, Kwan would be *most likely* to expect that:
   - A fiscal policy will turn more accommodative.
   - B the mark-to-market value of the debt will increase.
   - C monetary policy will become more accommodative.

3 Assuming that the DM country's government debt becomes harder to finance and there is no change in monetary policy, Kwan is *most likely* to expect that over the longer term, there will be a fiscal policy response that will lead to:

A currency appreciation as yields rise.

B currency depreciation as yields decline.

C an indeterminate impact on the currency, depending on which effect prevails.

Solution to 1:

A is correct. The DM country currently has a low debt load (as a percentage of GDP), and in the short run, its expansionary fiscal policy will lead to higher interest rates and higher real rates relative to other countries. This path should lead to currency appreciation. The higher domestic interest rates will (all else equal) depress local asset prices (so B is incorrect), and the rising debt load is likely to increase rather than decrease the risk premium (so C is incorrect).

Solution to 2:

C is correct. As government debt becomes harder to finance, the government will be tempted to monetize the debt through an accommodative monetary policy. A is incorrect because an inability to finance the debt will make it hard for fiscal policy to become more accommodative. B is incorrect because as investors demand a higher risk premium (a higher return) for holding the DM country's debt, the mark-to-market value of the debt will decline (i.e., bond prices will decrease and bond yields will increase).

Solution to 3:

B is correct. As the DM country's debt ratio deteriorates, foreign investors will demand a higher rate of return to compensate them for the increased risk. Assuming that the central bank will not accommodate (monetize) the rising government debt, the most likely fiscal response is an eventual move toward fiscal consolidation—reducing the public deficit and debt levels that were causing the debt metrics to deteriorate. This policy adjustment would involve issuing fewer government bonds. All else equal, bond yields would decrease, leading to a weaker domestic currency over the longer term.

A is incorrect because currency appreciation is not likely to accompany rising yields when the government is having difficulty financing its deficit. There would be a rising risk premium (a deteriorating investor appetite) for holding DM assets, and hence a currency appreciation would be unlikely despite high DM yields. To avoid paying these high yields on its debt, the DM government would eventually have to take measures to reduce its deficit spending. This approach would eventually help reduce investor risk aversion and DM yields. C is incorrect because given the deterioration in the DM's debt metrics, a depreciation of its exchange rate is likely to be an important part of the restoration of financial market equilibrium.

14 EXCHANGE RATE MANAGEMENT: INTERVENTION AND CONTROLS

l. describe objectives of central bank or government intervention and capital controls and describe the effectiveness of intervention and capital controls;

Capital flows can be both a blessing and a curse. Capital inflows can be a blessing when they increase domestic investment, thereby increasing a country's economic growth and asset values. Currency appreciation often follows, which increases returns to global investors. Capital inflows can be a curse, however, if they fuel boom-like conditions, asset price bubbles, and overvaluation of a country's currency. If capital inflows then reverse, the result may be a major economic downturn, a significant decline in asset prices, and a large depreciation of the currency. Capital inflows often are driven by a combination of "pull" and "push" factors. Pull factors represent a favorable set of developments that encourage foreign capital inflows. These factors may stem from both the public and private sectors. Examples of better economic management by a government include

- a decrease in inflation and inflation volatility,
- more-flexible exchange rate regimes,
- improved fiscal positions,
- privatization of state-owned entities,
- liberalization of financial markets, and
- lifting of foreign exchange regulations and controls.

Ideally, these changes will facilitate strong economic growth in the private sector, which will attract further foreign investment. A healthy export sector will generate improvement in the current account balance and an increase in FX reserves, which can be used by the government as a buffer against future speculative attacks. The returns from the currency and assets should increase, increasing the foreign investor's return.

Push factors driving foreign capital inflows are not determined by the domestic policies but arise from the primary sources of internationally mobile capital, notably the investor base in industrial countries. For example, the pursuit of low interest rate policies in industrial countries since the 2008 financial crisis has encouraged global investors to seek higher returns abroad.

Another important push factor is the long-run trend in asset allocation by industrial country investors. For example, many fund managers have traditionally had underweight exposures to emerging market assets, but with the weight of emerging market equities in broad global equity market indexes on the rise (as of 2019 the EM share of world GDP at current prices is over 40%, up from 17% in the 1960s, according to the IMF), capital flows to EM countries, in the form of increased allocations to EM equities, are likely to rise.

Private capital inflows to emerging markets go through significant changes over time. For example, they rose steadily between 2003 and 2007, posting nearly a six-fold increase over the period. Both push and pull factors contributed to that surge in capital flows. Net private capital flows to emerging markets tumbled in 2008 and 2009 as heightened risk aversion during the global financial crisis prompted investors to unwind some of their EM exposures in favor of US assets. In 2010, capital flows to emerging markets rose as many EM economies weathered the global financial crisis better than many industrial economies. In addition, the pursuit of ultra-low interest rate policies in the United States, the euro area, and Japan encouraged global investors to invest in higher-yielding EM assets.

However beneficial foreign capital is, policymakers must guard against excessive capital inflows that could quickly be reversed. Capital flow surges planted the seeds of three major currency crises in the 1990s—the European Exchange Rate Mechanism (ERM) crisis in 1992–1993, the Mexican peso crisis in late 1994, and the Asian currency and financial crisis in 1997–1998. Each crisis episode was preceded by a surge in capital inflows and a buildup of huge, highly leveraged speculative positions by local as well as international investors in currencies that eventually came under heavy speculative attack. In the run-up to the ERM crisis, investors—believing that European yield convergence would occur as European monetary union approached—took on highly leveraged long positions in the higher-yielding European currencies financed by short positions in the lower-yielding European currencies. Likewise, in the run-up to the Mexican peso crisis, investors and banks were highly leveraged and made extensive use of derivative products in taking on speculative long Mexican peso/short US dollar positions. And in the run-up to the Asian financial crisis, Asian companies and banks were highly leveraged as they took on a huge volume of short-term dollar- and yen-denominated debt to fund local activities. In each case, the sudden unwinding of those leveraged long speculative positions triggered the attacks on the currencies.

Governments resist excessive inflows and currency bubbles by using capital controls and direct intervention (selling their currency) in the foreign exchange market. Capital controls can take many forms. In the Asian financial crisis, many countries, such as Malaysia, prevented their banks from offering currency transactions in which their currency was sold. As mentioned earlier, Brazil has used a tax to limit currency transactions. In 2006, Thailand required a one-year, non-interest-bearing deposit of 30% of an investment's value to reduce new foreign inflows, which had been appreciating the Thai baht. Vietnam has limited the foreign ownership of local financial institutions. In 2015, Ukraine was removed from the MSCI Frontier Markets equity index after its central bank prevented foreign investors from repatriating funds from the sale of Ukrainian stocks. By 2016, Venezuela had instituted capital controls in the form of four different exchange rates, whereby the rate for selling Venezuelan bolivars for US dollars depended on what the dollars were used for. As a result, many Venezuelans used the black market to obtain dollars. Venezuela's capital controls were subsequently loosened in 2018 and 2019.

At one time, capital controls were frowned on as a policy tool for curbing undesired surges in capital inflows. It was generally felt that such controls tended to generate distortions in global trade and finance and that, in all likelihood, market participants would eventually find ways to circumvent the controls. Furthermore, many thought that capital controls imposed by one country could deflect capital flows to other countries, which could complicate monetary and exchange rate policies in those economies. Despite such concerns, the IMF has said that the benefits associated with capital controls may exceed the associated costs. Given the painful lessons that EM policymakers have learned from previous episodes of capital flow surges, some believe that under certain circumstances, capital controls may be needed to prevent exchange rates from overshooting, asset bubbles from forming, and future financial conditions from deteriorating.

Although a case can be made for government intervention and capital controls to limit the potential damage associated with unrestricted inflows of overseas capital, the key issue for policymakers is whether intervention and capital controls will actually work in terms of (1) preventing currencies from appreciating too strongly, (2) reducing the aggregate volume of capital inflows, and (3) enabling monetary authorities to pursue independent monetary policies without having to worry about whether changes in policy rates might attract too much capital from overseas. As an example of the last issue, if a central bank increases interest rates to slow inflation, then capital controls might prevent foreign capital inflows from subsequently depressing interest rates.

Evidence on the effectiveness of direct government intervention suggests that, in the case of industrial countries, the volume of intervention is often quite small relative to the average daily turnover of G–10 currencies in the foreign exchange market. Hence, most studies have concluded that the effect of intervention in developed market economies is limited. For most developed market countries, the ratio of official FX reserves held by the respective central banks to the average daily turnover of foreign exchange trading in that currency is negligible. Most industrial countries hold insufficient reserves to significantly affect the supply of and demand for their currency. Note that if a central bank is intervening in an effort to weaken, rather than strengthen, its own currency, it could (at least in principle) create and sell an unlimited amount of its currency and accumulate a correspondingly large quantity of FX reserves. However, persistent intervention in the FX market can undermine the efficacy of domestic monetary policy.

The evidence on the effectiveness of government intervention in emerging market currencies is more mixed. Intervention appears to contribute to lower EM exchange rate volatility, but no statistically significant relationship has emerged between the level of EM exchange rates and intervention. Some studies have found, however, that EM policymakers might have greater success in controlling exchange rates than their industrial country counterparts because the ratio of EM central bank FX reserve holdings to average daily FX turnover in their domestic currencies is actually quite sizable. With considerably greater firepower in their reserve arsenals, emerging market central banks appear to be in a stronger position than their developed market counterparts to influence the level and path of their exchange rates. What's more, with emerging market central banks' FX reserve holdings expanding at a near-record clip in the past decade, the effectiveness of intervention may be greater now than in the past.

15 WARNING SIGNS OF A CURRENCY CRISIS

m describe warning signs of a currency crisis.

If capital inflows come to a sudden stop, the result may be a financial crisis, in which the economy contracts, asset values plummet, and the currency sharply depreciates. History is filled with examples of currencies that have come under heavy selling pressure within short windows of time. For example, between August 2008 and February 2009, 23 currencies dropped by 25% or more against the US dollar. These included the developed market currencies of Australia, Sweden, and the United Kingdom, which dropped by 35% or more, and the emerging market currencies of Brazil, Russia, and South Korea, which fell by more than 50%.

Currency crises often occur suddenly, with many investors caught by surprise. Once a wave of selling begins, investors and borrowers must immediately reposition their portfolios to avoid excessive capital losses. For example, assume a carry trader had gone long the Brazilian real and borrowed US dollars. Upon an initial depreciation of the real, the trader would be inclined to exit the trade by selling reals and buying dollars. Or consider a Brazilian public or private borrower that had financed in US dollars. The borrower would also be selling reals to buy dollars in order to cover future repayment of the dollar debt. Either of these actions will intensify selling pressure on the depreciated currency. It is this massive liquidation of vulnerable positions, often reinforced by speculative offshore selling, that is largely responsible for the excessive exchange rate movements that occur during currency crises.

Because most crisis episodes have not been adequately anticipated, a great deal of effort has been spent developing early warning systems. One of the problems in developing an early warning system is that views on the underlying causes of currency crises differ greatly. One school of thought contends that currency crises tend to be precipitated by deteriorating economic fundamentals, while a second school contends that currency crises can occur out of the blue, with little evidence of deteriorating fundamentals preceding them.

If, according to the first school of thought, deteriorating economic fundamentals often precede crises and if those economic fundamentals tend to deteriorate steadily and predictably, then it should be possible to construct an early warning system to anticipate when a currency might be vulnerable.

The second school of thought argues that, although evidence of deteriorating economic fundamentals might explain a relatively large number of currency collapses, there might be cases in which economies with relatively sound fundamentals have their currencies come under attack. Clearly, these currency crises would be more difficult to predict. Events that are largely unrelated to domestic economic fundamentals include sudden adverse shifts in market sentiment that become self-fulfilling prophecies and contagion from crises in other markets. A crisis may spread to a country when, for example, the country devalues its currency to keep its exports competitive with those of another country that devalued.

Recognizing that no single model can correctly anticipate the onset of all crisis episodes, an early warning system might nevertheless be useful in assisting investors in structuring and/or hedging their global portfolios. An ideal early warning system would need to incorporate a number of important features. First, it should have a strong record of predicting actual crises but also should not issue false alarms. Second, it should include macroeconomic indicators whose data are available on a timely basis. If data arrive with a long lag, a crisis could be under way before the early warning system starts flashing red. Third, because currency crises tend to be triggered in countries with a number of economic problems, not just one, an ideal early warning system should be broad based, incorporating a wide range of symptoms that crisis-prone currencies might exhibit.

Many studies have been conducted to develop an early warning system for currency crises, typically by constructing a model in which a number of variables constitute the early warning system. Various definitions of currency crises have been used. Although the variables and methodologies differ from one study to the next, the following conditions were identified in one or more studies (Babecký, Havránek, Matějů, Rusnák, Šmídková, and Vašíček 2013 and 2014; Daniels and VanHoose 2018):

1. Prior to a currency crisis, the capital markets have been liberalized to allow the free flow of capital.
2. There are large inflows of foreign capital (relative to GDP) in the period leading up to a crisis, with short-term funding denominated in a foreign currency being particularly problematic.
3. Currency crises are often preceded by (and often coincide with) banking crises.
4. Countries with fixed or partially fixed exchange rates are more susceptible to currency crises than countries with floating exchange rates.
5. Foreign exchange reserves tend to decline precipitously as a crisis approaches.
6. In the period leading up to a crisis, the currency has risen substantially relative to its historical mean.
7. The ratio of exports to imports (known as “the terms of trade”) often deteriorates before a crisis.
8. Broad money growth and the ratio of M2 (a measure of money supply) to bank reserves tend to rise prior to a crisis.
9. Inflation tends to be significantly higher in pre-crisis periods compared with tranquil periods.

These factors are usually interrelated and often feed off one another. For example, in the case of the first five factors, large inflows of foreign capital occur because the financial markets have been liberalized and domestic banks have borrowed abroad. If the borrowing is denominated in a foreign currency and the domestic currency initially depreciates, the bank may have trouble servicing its debt, especially when the debt is of shorter maturity. This scenario may cause foreign investors to withdraw capital and speculators to short the currency, with their actions causing further declines in the currency. If the government is trying to maintain the currency's value, it could increase interest rates to stem capital outflows or defend its currency using direct intervention. The former action may worsen the banking industry's condition and slow down the economy. In the latter approach, the government will have to spend down its foreign currency reserves to buy its own currency in the foreign exchange markets. If the government appears unwilling or unable to defend its currency, then capital outflows and speculative attacks will increase.

The fifth through seventh factors are related because an overvalued currency may make the country's exports less competitive. With fewer exports, the country is not able to earn as much foreign currency. Other interrelationships occur because these factors often coincide.

Models cannot predict every crisis, and they sometimes generate false alarms. Nevertheless, an early warning system can be useful in assessing and preparing for potential negative tail risks. As with any analytical tool, the implementation of an early warning system requires integration with other analysis and judgment that cannot be easily quantified or conceptualized.

Iceland's Currency Crisis of 2008

Iceland, a country with a population of 320,000, had traditionally relied on the fishing, energy, and aluminum industries for economic growth. That began to change in 2001, when the banking industry was liberalized. Three banks dominated the Icelandic banking industry: Glitnir, Kaupthing, and Landsbanki. Given Iceland's small population, these banks sought growth by offering short-term, internet-based deposit accounts to foreign investors. These accounts offered attractive interest rates and were denominated in foreign currencies. In particular, many of the depositors were British, Dutch, and other European citizens who held deposit accounts denominated in pounds and euros.

With government guarantees on their deposit accounts, the banking industry grew rapidly. The largest bank, Kaupthing, experienced asset growth of 30 times between 2000 and 2008. The three banks increased lending rapidly, with many of their loans being long term, resulting in a maturity mismatch of assets and liabilities. The banks' assets were more than 14 times the country's GDP, while foreign debt was five times GDP. The three banks constituted more than 70% of the national stock market capitalization.

The economy expanded at a real growth rate above 20% annually between 2002 and 2005, and many Icelanders left traditional industries to work in the banks. Iceland earned the nickname "Nordic Tiger" as per capita GDP approached USD 70,000 in 2007. The krona increased in value against the US dollar by 40% between 2001 and 2007. By 2007, the unemployment rate was less than 1%. Icelanders went on a shopping spree for consumer goods, in part by using loans tied to the value of foreign currencies, motivated by lower interest rates abroad. A 2002 trade surplus turned into a trade deficit in the years 2003–2007. Iceland's external debt in 2008 was more than 7 times its GDP and 14 times its export revenue. Broad-based monetary aggregates grew at a rate of 14%–35% annually from 2002 to 2007. By the fall of 2008, inflation had reached 14%.

As the global financial crisis unfolded in 2008, interbank lending declined and Icelandic banks were unable to roll over their short-term debt. Anxious foreign depositors began withdrawing their funds. In the first half of 2008, the krona depreciated by more than

40% against the euro. As the Icelandic currency declined in value, it became more difficult for the banks to meet depositors' liquidity demands, while at the same time the banks' depreciating krona-denominated assets could not be used for collateral financing.

The three banks collapsed in 2008. Unfortunately for foreign depositors, because of the relative size of the banks, the government guaranteed only domestic deposits. Iceland's central bank became technically insolvent, as its EUR 2 billion in assets was dwarfed by Iceland's debt to foreign banks of EUR 50 billion. Trading in the stock market was suspended in October 2008. When it reopened several days later, the Icelandic Stock Market Index fell by more than 77% as a result of the elimination of the three banks' equity value.

The government attempted to peg the krona to the euro in October 2008 but abandoned the peg one day later. When trading in the currency was resumed later that month, the currency value fell by more than 60% and trading was eventually suspended. Iceland increased interest rates to 18% to stem outflows of krona and imposed capital controls on the selling of krona for foreign currency. The Icelandic economy contracted, and per capita GDP fell 9.2% in 2009. By the spring of 2009, unemployment was 9%. The country subsequently required a bailout from the IMF and its neighbors of USD 4.6 billion.

Source: Federal Reserve Bank of St. Louis database; Bekaert and Hodrick 2018; Matsangou 2015; Daniels and VanHoose 2017.

EXAMPLE 11

Currency Crises

Monique Kwan now turns her attention to the likelihood of crises in various emerging market currencies. She discusses this matter with a research associate, who tells her that the historical record of currency crises shows that most of these episodes were not very well anticipated by investors (in terms of their positioning), by the bond markets (in terms of yield spreads between countries), or by major credit rating agencies and economists (in terms of the sovereign credit ratings and forecasts, respectively).

1 The research associate is *most likely*:

- **A** correct.
- **B** incorrect, because most credit rating agencies and economists typically change their forecasts prior to a crisis.
- **C** incorrect, because investor positioning and international yield differentials typically shift prior to a crisis.

Kwan delves further into the historical record of currency crises. She concludes that even countries with relatively sound economic fundamentals can fall victim to these crisis episodes and that these attacks can occur when sentiment shifts for reasons unrelated to economic fundamentals.

2 Kwan's conclusion is *most likely*:

- **A** correct.
- **B** incorrect, because there are few historical crises involving currencies of countries with sound economic fundamentals.
- **C** incorrect, because there are few historical episodes in which a sudden adverse shift in market sentiment occurs that is unrelated to economic fundamentals.

To better advise the firm's clients on the likelihood of currency crises, Kwan tries to formulate an early warning system for these episodes. She recognizes that a typical currency crisis tends to be triggered by a number of economic problems, not just one.

3 Kwan's early warning system is *least likely* to indicate an impending crisis when there is:

   A an expansionary monetary policy.

   B an overly appreciated exchange rate.

   C a rising level of foreign exchange reserves at the central bank.

4 Kwan's early warning system would *most likely* be better if it:

   A had a strong record of predicting actual crises, even if it generates a lot of false signals.

   B included a wide variety of economic indicators, including those for which data are available only with a significant lag.

   C started flashing well in advance of an actual currency crisis to give market participants time to adjust or hedge their portfolios before the crisis hits.

Solution to 1:

A is correct. Currency crises often catch most market participants and analysts by surprise.

Solution to 2:

A is correct. Even countries with sound economic fundamentals can be subject to a currency crisis, including instances when market sentiment shifts for non-economic reasons.

Solution to 3:

C is correct. A high level of foreign exchange reserves held by a country typically decreases the likelihood of a currency crisis.

Solution to 4:

C is correct. Early warnings are a positive factor in judging the effectiveness of the system, whereas false signals and the use of lagged data would be considered negative factors.

SUMMARY

Exchange rates are among the most difficult financial market prices to understand and therefore to value. There is no simple, robust framework that investors can rely on in assessing the appropriate level and likely movements of exchange rates.

Most economists believe that there is an equilibrium level or a path to that equilibrium value that a currency will gravitate toward in the long run. Although short- and medium-term cyclical deviations from the long-run equilibrium path can be sizable and persistent, fundamental forces should eventually drive the currency back toward its long-run equilibrium path. Evidence suggests that misalignments tend to build up gradually over time. As these misalignments build, they are likely to generate serious economic imbalances that will eventually lead to correction of the underlying exchange rate misalignment.

We have described how changes in monetary policy, fiscal policy, current account trends, and capital flows affect exchange rate trends, as well as what role government intervention and capital controls can play in counteracting potentially undesirable exchange rate movements. We have made the following key points:

- Spot exchange rates apply to trades for the next settlement date (usually $T + 2$) for a given currency pair. Forward exchange rates apply to trades to be settled at any longer maturity.
- Market makers quote bid and offer prices (in terms of the *price currency*) at which they will buy or sell the *base currency.*
  - The offer price is always higher than the bid price.
  - The counterparty that asks for a two-sided price quote has the option (but not the obligation) to deal at either the bid or offer price quoted.
  - The bid–offer spread depends on (1) the currency pair involved, (2) the time of day, (3) market volatility, (4) the transaction size, and (5) the relationship between the dealer and the client. Spreads are tightest in highly liquid currency pairs, when the key market centers are open, and when market volatility is relatively low.
- Absence of arbitrage requires the following:
  - The bid (offer) shown by a dealer in the interbank market cannot be higher (lower) than the current interbank offer (bid) price.
  - The cross-rate bids (offers) posted by a dealer must be lower (higher) than the implied cross-rate offers (bids) available in the interbank market. If they are not, then a triangular arbitrage opportunity arises.
- Forward exchange rates are quoted in terms of points to be added to the spot exchange rate. If the points are positive (negative), the base currency is trading at a forward premium (discount). The points are proportional to the interest rate differential and approximately proportional to the time to maturity.
- International parity conditions show us how expected inflation, interest rate differentials, forward exchange rates, and expected future spot exchange rates are linked. In an ideal world,
  - relative expected inflation rates should determine relative nominal interest rates,
  - relative interest rates should determine forward exchange rates, and
  - forward exchange rates should correctly anticipate the path of the future spot exchange rate.
- International parity conditions tell us that countries with high (low) expected inflation rates should see their currencies depreciate (appreciate) over time, that high-yield currencies should depreciate relative to low-yield currencies over time, and that forward exchange rates should function as unbiased predictors of future spot exchange rates.
- With the exception of covered interest rate parity, which is enforced by arbitrage, the key international parity conditions rarely hold in either the short or medium term. However, the parity conditions tend to hold over relatively long horizons.

- According to the theory of covered interest rate parity, a foreign-currency-denominated money market investment that is completely hedged against exchange rate risk in the forward market should yield exactly the same return as an otherwise identical domestic money market investment.
- According to the theory of uncovered interest rate parity, the expected change in a domestic currency's value should be fully reflected in domestic–foreign interest rate spreads. Hence, an unhedged foreign-currency-denominated money market investment is expected to yield the same return as an otherwise identical domestic money market investment.
- According to the *ex ante* purchasing power parity condition, expected changes in exchange rates should equal the difference in expected national inflation rates.
- If both *ex ante* purchasing power parity and uncovered interest rate parity held, real interest rates across all markets would be the same. This result is real interest rate parity.
- The international Fisher effect says that the nominal interest rate differential between two currencies equals the difference between the expected inflation rates. The international Fisher effect assumes that risk premiums are the same throughout the world.
- If both covered and uncovered interest rate parity held, then forward rate parity would hold and the market would set the forward exchange rate equal to the expected spot exchange rate: The forward exchange rate would serve as an unbiased predictor of the future spot exchange rate.
- Most studies have found that high-yield currencies do not depreciate and low-yield currencies do not appreciate as much as yield spreads would suggest over short to medium periods, thus violating the theory of uncovered interest rate parity.
- Carry trades overweight high-yield currencies at the expense of low-yield currencies. Historically, carry trades have generated attractive returns in benign market conditions but tend to perform poorly (i.e., are subject to crash risk) when market conditions are highly volatile.
- According to a balance of payments approach, countries that run persistent current account deficits will generally see their currencies weaken over time. Similarly, countries that run persistent current account surpluses will tend to see their currencies appreciate over time.
- Large current account imbalances can persist for long periods of time before they trigger an adjustment in exchange rates.
- Greater financial integration of the world's capital markets and greater freedom of capital to flow across national borders have increased the importance of global capital flows in determining exchange rates.
- Countries that institute relatively tight monetary policies, introduce structural economic reforms, and lower budget deficits will often see their currencies strengthen over time as capital flows respond positively to relatively high nominal interest rates, lower inflation expectations, a lower risk premium, and an upward revision in the market's assessment of what exchange rate level constitutes long-run fair value.
- Monetary policy affects the exchange rate through a variety of channels. In the Mundell–Fleming model, it does so primarily through the interest rate sensitivity of capital flows, strengthening the currency when monetary policy is

tightened and weakening it when monetary policy is eased. The more sensitive capital flows are to the change in interest rates, the greater the exchange rate's responsiveness to the change in monetary policy.

- In the monetary model of exchange rate determination, monetary policy is deemed to have a direct impact on the actual and expected path of inflation, which, via purchasing power parity, translates into a corresponding impact on the exchange rate.
- Countries that pursue overly easy monetary policies will see their currencies depreciate over time.
- In the Mundell–Fleming model, an expansionary fiscal policy typically results in a rise in domestic interest rates and an increase in economic activity. The rise in domestic interest rates should induce a capital inflow, which is positive for the domestic currency, but the rise in economic activity should contribute to a deterioration of the trade balance, which is negative for the domestic currency. The more mobile capital flows are, the greater the likelihood that the induced inflow of capital will dominate the deterioration in trade.
- Under conditions of high capital mobility, countries that simultaneously pursue expansionary fiscal policies and relatively tight monetary policies should see their currencies strengthen over time.
- The portfolio balance model of exchange rate determination asserts that increases in government debt resulting from a rising budget deficit will be willingly held by investors only if they are compensated in the form of a higher expected return. The higher expected return could come from (1) higher interest rates and/or a higher risk premium, (2) depreciation of the currency to a level sufficient to generate anticipation of gains from subsequent currency appreciation, or (3) some combination of the two.
- Surges in capital inflows can fuel boom-like conditions, asset price bubbles, and currency overvaluation.
- Many consider capital controls to be a legitimate part of a policymaker's toolkit. The IMF believes that capital controls may be needed to prevent exchange rates from overshooting, asset price bubbles from forming, and future financial conditions from deteriorating.
- The evidence indicates that government policies have had a significant impact on the course of exchange rates. Relative to developed countries, emerging markets may have greater success in managing their exchange rates because of their large foreign exchange reserve holdings, which appear sizable relative to the limited turnover of FX transactions in many emerging markets.
- Although each currency crisis is distinct in some respects, the following factors were identified in one or more studies:
  1. Prior to a currency crisis, the capital markets have been liberalized to allow the free flow of capital.
  2. There are large inflows of foreign capital (relative to GDP) in the period leading up to a crisis, with short-term funding denominated in a foreign currency being particularly problematic.
  3. Currency crises are often preceded by (and often coincide with) banking crises.
  4. Countries with fixed or partially fixed exchange rates are more susceptible to currency crises than countries with floating exchange rates.
  5. Foreign exchange reserves tend to decline precipitously as a crisis approaches.

6 In the period leading up to a crisis, the currency has risen substantially relative to its historical mean.

7 The terms of trade (exports relative to imports) often deteriorate before a crisis.

8 Broad money growth and the ratio of M2 (a measure of money supply) to bank reserves tend to rise prior to a crisis.

9 Inflation tends to be significantly higher in pre-crisis periods compared with tranquil periods.

REFERENCES

Babecký, Jan, Tomáš Havránek, Jakub Matějů, Marek Rusnák, Kateřina Šmídková, and Bořek Vašíček. 2013. "Leading Indicators of Crisis Incidence: Evidence from Developed Countries." *Journal of International Money and Finance* 35 (June): 1–19.

Babecký, Jan, Tomáš Havránek, Jakub Matějů, Marek Rusnák, Kateřina Šmídková, and Bořek Vašíček. 2014. "Banking, Debt, and Currency Crises in Developed Countries: Stylized Facts and Early Warning Indicators." *Journal of Financial Stability* 15 (December): 1–17.

Bekaert, Geert, and Robert Hodrick. 2018. *International Financial Management*. 3rd ed. Cambridge, UK: Cambridge University Press.

Daniels, Joseph P., and David D. VanHoose. 2017. *Global Economic Issues and Policies*. 4th ed. London: Routledge.

Dimson, Elroy, Paul Marsh, Paul McGinnie, Mike Staunton, and Jonathan Wilmot. 2012. *Credit Suisse Global Investment Returns Yearbook 2012*. Zurich: Credit Suisse Research Institute.

Dornbusch, Rudiger. 1976. "Expectations and Exchange Rate Dynamics." *Journal of Political Economy* 84 (6): 1161–76.

Isard, Peter, Hamid Faruqee, G. Russell Kincaid, and Martin Fetherston. 2001. "*Methodology for Current Account and Exchange Rate Assessments.*" IMF Occasional Paper no. 209 (December).

Matsangou, Elizabeth. 2015. "*Failing Banks, Winning Economy: The Truth about Iceland's Recovery.*" World Finance (15 September).

Rosenberg, Michael R. 1996. *Currency Forecasting: A Guide to Fundamental and Technical Models of Exchange Rate Determination*. Chicago: Irwin Professional Publishing.

Rosenberg, Michael R. 2003. *Exchange Rate Determination: Models and Strategies for Exchange Rate Forecasting*. New York: McGraw-Hill.

APPENDIX

Currency Codes Used

USD	US dollar
EUR	Euro
GBP	UK pound
JPY	Japanese yen
MXN	Mexican peso
CHF	Swiss franc
CAD	Canadian dollar
SEK	Swedish krona
AUD	Australian dollar
KRW	Korean won
NZD	New Zealand dollar

PRACTICE PROBLEMS

The following information relates to Questions 1–5

Ed Smith is a new trainee in the foreign exchange (FX) services department of a major global bank. Smith's focus is to assist senior FX trader Feliz Mehmet, CFA. Mehmet mentions that an Indian corporate client exporting to the United Kingdom wants to estimate the potential hedging cost for a sale closing in one year. Smith is to determine the premium/discount for an annual (360-day) forward contract using the exchange rate data presented in Exhibit 1.

Exhibit 1 Select Currency Data for GBP and INR

Spot (INR/GBP)	79.5093
Annual (360-day) Libor (GBP)	5.43%
Annual (360-day) Libor (INR)	7.52%

Mehmet is also looking at two possible trades to determine their profit potential. The first trade involves a possible triangular arbitrage trade using the Swiss, US, and Brazilian currencies, to be executed based on a dealer's bid/offer rate quote of 0.2355/0.2358 in CHF/BRL and the interbank spot rate quotes presented in Exhibit 2.

Exhibit 2 Interbank Market Quotes

Currency Pair	Bid/Offer
CHF/USD	0.9799/0.9801
BRL/USD	4.1699/4.1701

Mehmet is also considering a carry trade involving the USD and the euro. He anticipates it will generate a higher return than buying a one-year domestic note at the current market quote due to low US interest rates and his predictions of exchange rates in one year. To help Mehmet assess the carry trade, Smith provides Mehmet with selected current market data and his one-year forecasts in Exhibit 3.

© 2017 CFA Institute. All rights reserved.

Exhibit 3 Spot Rates and Interest Rates for Proposed Carry Trade

Today's One-Year Libor		Currency Pair (Price/Base)	Spot Rate Today	Projected Spot Rate in One Year
USD	0.80%	CAD/USD	1.3200	1.3151
CAD	1.71%	EUR/CAD	0.6506	0.6567
EUR	2.20%			

Finally, Mehmet asks Smith to assist with a trade involving a US multinational customer operating in Europe and Japan. The customer is a very cost-conscious industrial company with an AA credit rating and strives to execute its currency trades at the most favorable bid–offer spread. Because its Japanese subsidiary is about to close on a major European acquisition in three business days, the client wants to lock in a trade involving the Japanese yen and the euro as early as possible the next morning, preferably by 8:05 a.m. New York time.

At lunch, Smith and other FX trainees discuss how best to analyze currency market volatility from ongoing financial crises. The group agrees that a theoretical explanation of exchange rate movements, such as the framework of the international parity conditions, should be applicable across all trading environments. They note such analysis should enable traders to anticipate future spot exchange rates. But they disagree on which parity condition best predicts exchange rates, voicing several different assessments. Smith concludes the discussion on parity conditions by stating to the trainees,

> I believe that in the current environment both covered and uncovered interest rate parity conditions are in effect.

1 Based on Exhibit 1, the forward premium (discount) for a 360-day INR/GBP forward contract is *closest* to:

A –1.546.

B 1.546.

C 1.576.

2 Based on Exhibit 2, the *most* appropriate recommendation regarding the triangular arbitrage trade is to:

A decline the trade, because no arbitrage profits are possible.

B execute the trade, buy BRL in the interbank market, and sell BRL to the dealer.

C execute the trade, buy BRL from the dealer, and sell BRL in the interbank market.

3 Based on Exhibit 3, the potential all-in USD return on the carry trade is *closest* to:

A 0.83%.

B 1.23%.

C 1.63%.

4 The factor *least likely* to lead to a narrow bid–offer spread for the industrial company's needed currency trade is the:

A timing of its trade.

B company's credit rating.

C pair of currencies involved.

5 If Smith's statement on parity conditions is correct, future spot exchange rates are *most likely* to be forecast by:

A current spot rates.

B forward exchange rates.

C inflation rate differentials.

The following information relates to Questions 6–12

Connor Wagener, a student at the University of Canterbury in New Zealand, has been asked to prepare a presentation on foreign exchange rates for his international business course. Wagener has a basic understanding of exchange rates but would like a practitioner's perspective, and he has arranged an interview with currency trader Hannah McFadden. During the interview, Wagener asks McFadden,

> Could you explain what drives exchange rates? I'm curious as to why our New Zealand dollar was affected by the European debt crisis in 2011 and what other factors impact it.

In response, McFadden begins with a general discussion of exchange rates. She notes that international parity conditions illustrate how exchange rates are linked to expected inflation, interest rate differences, and forward exchange rates as well as current and expected future spot rates. McFadden makes the following statement:

Statement 1 "Fortunately, the international parity condition most relevant for FX carry trades does not always hold."

McFadden continues her discussion:

> FX carry traders go long (i.e., buy) high-yield currencies and fund their position by shorting—that is, borrowing in—low-yield currencies. Unfortunately, crashes in currency values can occur which create financial crises as traders unwind their positions. For example, in 2008, the New Zealand dollar was negatively impacted when highly leveraged carry trades were unwound. In addition to investors, consumers and business owners can also affect currency exchange rates through their impact on their country's balance of payments. For example, if New Zealand consumers purchase more goods from China than New Zealand businesses sell to China, New Zealand will run a trade account deficit with China.

McFadden further explains,

Statement 2 "A trade surplus will tend to cause the currency of the country in surplus to appreciate, while a deficit will cause currency depreciation. Exchange rate changes will result in immediate adjustments in the prices of traded goods as well as in the demand for imports and exports. These changes will immediately correct the trade imbalance."

McFadden next addresses the influence of monetary and fiscal policy on exchange rates:

> Countries also exert significant influence on exchange rates both through the initial mix of their fiscal and monetary policies and by subsequent adjustments to those policies. Various models have been developed to

identify how these policies affect exchange rates. The Mundell–Fleming model addresses how changes in both fiscal and monetary policies affect interest rates and ultimately exchange rates in the short term.

McFadden describes monetary models by stating,

Statement 3 "Monetary models of exchange rate determination focus on the effects of inflation, price level changes, and risk premium adjustments."

McFadden continues her discussion:

So far, we've touched on balance of payments and monetary policy. The portfolio balance model addresses the impacts of sustained fiscal policy on exchange rates. I must take a client call but will return shortly. In the meantime, here is some relevant literature on the models I mentioned along with a couple of questions for you to consider:

Question 1: Assume an emerging market (EM) country has restrictive monetary and fiscal policies under low capital mobility conditions. Are these policies likely to lead to currency appreciation or currency depreciation or to have no impact?

Question 2: Assume a developed market (DM) country has an expansive fiscal policy under high capital mobility conditions. Why is its currency most likely to depreciate in the long run under an integrated Mundell–Fleming and portfolio balance approach?

Upon her return, Wagener and McFadden review the questions. McFadden notes that capital flows can have a significant impact on exchange rates and have contributed to currency crises in both EM and DM countries. She explains that central banks, such as the Reserve Bank of New Zealand, use FX market intervention as a tool to manage exchange rates. McFadden states,

Statement 4 "Some studies have found that EM central banks tend to be more effective in using exchange rate intervention than DM central banks, primarily because of one important factor."

McFadden continues her discussion:

Statement 5 "I mentioned that capital inflows could cause a currency crisis, leaving fund managers with significant losses. In the period leading up to a currency crisis, I would predict that an affected country's:

Prediction 1 foreign exchange reserves will increase.

Prediction 2 broad money growth will increase.

Prediction 3 exchange rate will be substantially higher than its mean level during tranquil periods."

After the interview, McFadden agrees to meet the following week to discuss more recent events on the New Zealand dollar.

6 The international parity condition McFadden is referring to in Statement 1 is:

A purchasing power parity.

B covered interest rate parity.

C uncovered interest rate parity.

7 In Statement 2, McFadden is *most likely* failing to consider the:

A initial gap between the country's imports and exports.

B effect of an initial trade deficit on a country's exchange rates.

C lag in the response of import and export demand to price changes.

8 The *least* appropriate factor used to describe the type of models mentioned in Statement 3 is:

A inflation.

B price level changes.

C risk premium adjustments.

9 The best response to Question 1 is that the policies will:

A have no impact.

B lead to currency appreciation.

C lead to currency depreciation.

10 The most likely response to Question 2 is a(n):

A increase in the price level.

B decrease in risk premiums.

C increase in government debt.

11 The factor that McFadden is *most likely* referring to in Statement 4 is:

A FX reserve levels.

B domestic demand.

C the level of capital flows.

12 Which of McFadden's predictions in Statement 5 is *least likely to be correct*?

A Prediction 1

B Prediction 2

C Prediction 3

The following information relates to Questions 13–20

Anna Goldsworthy is the chief financial officer of a manufacturing firm headquartered in the United Kingdom. She is responsible for overseeing exposure to price risk in both the commodity and currency markets. Goldsworthy is settling her end-of-quarter transactions and creating reports. Her intern, Scott Underwood, assists her in this process.

The firm hedges input costs using forward contracts that are priced in US dollars (USD) and Mexican pesos (MXN). Processed goods are packaged for sale under licensing agreements with firms in foreign markets. Goldsworthy is expecting to receive a customer payment of JPY 225,000,000 (Japanese yen) that she wants to convert to pounds sterling (GBP). Underwood gathers the exchange rates from Dealer A in Exhibit 1.

Exhibit 1 Dealer A's Spot Exchange Rates

	Spot Exchange Rates		
Currency Pair (Price/Base)	Bid	Offer	Midpoint
JPY/GBP	129.65	129.69	129.67
MXN/USD	20.140	20.160	20.150
GBP/EUR	0.9467	0.9471	0.9469

Exhibit 1 (Continued)

Currency Pair (Price/Base)	Spot Exchange Rates Bid	Offer	Midpoint
USD/EUR	1.1648	1.1652	1.1650
USD/GBP	1.2301	1.2305	1.2303

The firm must also buy USD to pay a major supplier. Goldsworthy calls Dealer A with specific details of the transaction and asks to verify the USD/GBP quote. Dealer A calls her back later with a revised USD/GBP bid–offer quote of 1.2299/1.2307.

Goldsworthy must purchase MXN 27,000,000 to pay an invoice at the end of the quarter. In addition to the quotes from Dealer A, Underwood contacts Dealer B, who provides a bid–offer price of GBP/MXN 0.0403/0.0406. To check whether the dealer quotes are reflective of an efficient market, Underwood examines whether the prices allow for an arbitrage profit.

In three months, the firm will receive EUR 5,000,000 (euros) from another customer. Six months ago, the firm sold EUR 5,000,000 against the GBP using a nine-month forward contract at an all-in price of GBP/EUR 0.9526. To mark the position to market, Underwood collects the GBP/EUR forward rates in Exhibit 2.

Exhibit 2 GBP/EUR Forward Rates

Maturity	Forward Points
One month	4.40/4.55
Three months	14.0/15.0
Six months	29.0/30.0

Goldsworthy also asks for the current 90-day Libors for the major currencies. Selected three-month Libors (annualized) are shown in Exhibit 3. Goldsworthy studies Exhibit 3 and says, "We have the spot rate and the 90-day forward rate for GBP/EUR. As long as we have the GBP 90-day Libor, we will be able to calculate the implied EUR 90-day Libor."

Exhibit 3 90-Day Libor

Currency	Annualized Rate
GBP	0.5800%
JPY	0.0893%
USD	0.3300%

After reading a draft report, Underwood notes, "We do not hedge the incoming Japanese yen cash flow. Your report asks for a forecast of the JPY/GBP exchange rate in 90 days. We know the JPY/GBP spot exchange rate." He asks, "Does the information we have collected tell us what the JPY/GBP exchange rate will be in 90 days?"

Goldsworthy replies, "The JPY/GBP exchange rate in 90 days would be a valuable piece of information to know. An international parity condition can be used to provide an estimate of the future spot rate."

13 Using the quotes in Exhibit 1, the amount received by Goldsworthy from converting JPY 225,000,000 will be *closest* to:

A GBP 1,734,906.

B GBP 1,735,174.

C GBP 1,735,442.

14 Using Exhibit 1, which of the following would be the *best* reason for the revised USD/GBP dealer quote of 1.2299/1.2307?

A A request for a much larger transaction

B A drop in volatility in the USD/GBP market

C A request to trade when both New York and London trading centers are open

15 Using the quotes from Dealer A and B, the triangular arbitrage profit on a transaction of MXN 27,000,000 would be *closest* to:

A GBP 0.

B GBP 5,400.

C GBP 10,800.

16 Based on Exhibits 1, 2, and 3, the mark-to-market gain for Goldsworthy's forward position is *closest* to:

A GBP 19,971.

B GBP 20,500.

C GBP 21,968.

17 Based on Exhibit 2, Underwood should conclude that three-month EUR Libor is:

A below three-month GBP Libor.

B equal to three-month GBP Libor.

C above three-month GBP Libor.

18 Based on the exchange rate midpoint in Exhibit 1 and the rates in Exhibit 3, the 90-day forward premium (discount) for the USD/GBP would be *closest* to:

A –0.0040.

B –0.0010.

C +0.0010.

19 Using Exhibits 1, 2, and 3, which international parity condition would Goldsworthy *most likely* use to calculate the EUR Libor?

A Real interest rate parity

B Covered interest rate parity

C Uncovered interest rate parity

20 The international parity condition Goldsworthy will use to provide the estimate of the future JPY/GBP spot rate is *most likely*:

A covered interest rate parity.

B uncovered interest rate parity.

C relative purchasing power parity.

SOLUTIONS

1 C is correct. The equation to calculate the forward premium (discount) is:

$$F_{f/d} - S_{f/d} = S_{f/d}\left(\frac{\left[\frac{Actual}{360}\right]}{1 + i_d\left[\frac{Actual}{360}\right]}\right)\left(i_f - i_d\right).$$

$S_{f/d}$ is the spot rate with GBP the base currency or d and INR the foreign currency or f. $S_{f/d}$ per Exhibit 1 is 79.5093, i_f is equal to 7.52%, and i_d is equal to 5.43%.

With GBP as the base currency (i.e., the "domestic" currency) in the INR/GBP quote, substituting in the relevant base currency values from Exhibit 1 yields the following:

$$F_{f/d} - S_{f/d} = 79.5093\left(\frac{\left[\frac{360}{360}\right]}{1 + 0.0543\left[\frac{360}{360}\right]}\right)(0.0752 - 0.0543).$$

$$F_{f/d} - S_{f/d} = 79.5093\left(\frac{1}{1.0543}\right)(0.0752 - 0.0543).$$

$$F_{f/d} - S_{f/d} = 1.576.$$

2 B is correct. The dealer is posting a bid rate to buy BRL at a price that is too high. This overpricing is determined by calculating the interbank implied cross rate for the CHF/BRL using the intuitive equation-based approach:

CHF/BRL = CHF/USD × $(\text{BRL/USD})^{-1}$, or

CHF/BRL = CHF/USD × USD/BRL.

Inverting the BRL/USD given the quotes in Exhibit 2 determines the USD/BRL bid–offer rates of 0.23980/0.23982. (The bid of 0.23980 is the inverse of the BRL/USD offer, calculated as 1/4.1702; the offer of 0.23982 is the inverse of the BRL/USD bid, calculated as 1/4.1698.) Multiplying the CHF/USD and USD/BRL bid–offer rates then leads to the interbank implied CHF/BRL cross rate:

Bid: 0.9799 × 0.23980 = 0.2349.

Offer: 0.9801 × 0.23982 = 0.23505.

Since the dealer is willing to buy BRL at 0.2355 but BRL can be purchased from the interbank market at 0.23505, there is an arbitrage opportunity to buy BRL in the interbank market and sell BRL to the dealer for a profit of 0.0045 CHF (0.2355 – 0.23505) per BRL transacted.

3 A is correct. The carry trade involves borrowing in a lower-yielding currency to invest in a higher-yielding one and netting any profit after allowing for borrowing costs and exchange rate movements. The relevant trade is to borrow USD and lend in euros. To calculate the all-in USD return from a one-year EUR Libor deposit, first determine the current and one-year-later USD/EUR exchange rates. Because one USD buys CAD 1.3200 today and one CAD buys EUR 0.6506 today, today's EUR/USD rate is the product of these two numbers: 1.3200 × 0.6506 = 0.8588. The projected rate one year later is 1.3151 × 0.6567 = 0.8636. Accordingly, measured in dollars, the investment return for the unhedged EUR Libor deposit is equal to

$$(1.3200 \times 0.6506) \times (1 + 0.022) \times [1/(1.3151 \times 0.6567)] - 1$$
$$= 0.8588 \times (1.022)(1/0.8636) - 1 = 1.01632 - 1 = 1.632\%.$$

However, the borrowing costs must be charged against this *gross* return to fund the carry trade investment (one-year USD Libor was 0.80%). The *net* return on the carry trade is therefore 1.632% – 0.80% = 0.832%.

4 B is correct. While credit ratings can affect spreads, the trade involves spot settlement (i.e., two business days after the trade date), so the spread quoted to this highly rated (AA) firm is not likely to be much tighter than the spread that would be quoted to a somewhat lower-rated (but still high-quality) firm. The relationship between the bank and the client, the size of the trade, the time of day the trade is initiated, the currencies involved, and the level of market volatility are likely to be more significant factors in determining the spread for this trade.

5 B is correct. By rearranging the terms of the equation defining covered interest rate parity and assuming that uncovered interest rate parity is in effect, the forward exchange rate is equal to the expected future spot exchange rate—$F_{f/d} = S^e_{f/d}$—with the expected percentage change in the spot rate equal to the interest rate differential. Thus, the forward exchange rate is an unbiased forecast of the future spot exchange rate.

6 C is correct. The carry trade strategy is dependent on the fact that uncovered interest rate parity does not hold in the short or medium term. If uncovered interest rate parity held, it would mean that investors would receive identical returns from either an unhedged foreign currency investment or a domestic currency investment because the appreciation/depreciation of the exchange rate would offset the yield differential. However, during periods of low volatility, evidence shows that high-yield currencies do not depreciate enough and low-yield currencies do not appreciate enough to offset the yield differential.

7 C is correct. McFadden states that exchange rates will *immediately* correct the trade imbalance. She is describing the flow supply/demand channel, which assumes that trade imbalances will be corrected as the deficit country's currency depreciates, causing its exports to become more competitive and its imports to become more expensive. Studies indicate that there can be long lags between exchange rate changes, changes in the prices of traded goods, and changes in the trade balance. In the short run, exchange rates tend to be more responsive to investment and financing decisions.

8 C is correct. Risk premiums are more closely associated with the portfolio balance approach. The portfolio balance approach addresses the impact of a country's net foreign asset/liability position. Under the portfolio balance approach, investors are assumed to hold a diversified portfolio of assets including foreign and domestic bonds. Investors will hold a country's bonds as long as they are compensated appropriately. Compensation may come in the form of higher interest rates and/or higher risk premiums.

9 B is correct. The currency is likely to appreciate. The emerging market country has both a restrictive monetary policy and a restrictive fiscal policy under conditions of low capital mobility. Low capital mobility indicates that interest rate changes induced by monetary and fiscal policy will not cause large changes in capital flows. Implementation of restrictive policies should result in an improvement in the trade balance, which will result in currency appreciation.

10 C is correct. Expansionary fiscal policies result in currency depreciation in the long run. Under a portfolio balance approach, the assumption is that investors hold a mix of domestic and foreign assets including bonds. Fiscal stimulus policies result in budget deficits, which are often financed by debt. As the debt level rises, investors become concerned as to how the ongoing deficit will be financed. The country's central bank may need to create more money in order to purchase the debt, which would cause the currency to depreciate. Or the government could adopt a more restrictive fiscal policy, which would also depreciate the currency.

11 A is correct. EM countries are better able to influence their exchange rates because their reserve levels as a ratio of average daily FX turnover are generally much greater than those of DM countries. This means that EM central banks are in a better position to affect currency supply and demand than DM countries, where the ratio is negligible. EM policymakers use their foreign exchange reserves as a kind of insurance to defend their currencies, as needed.

12 A is correct. Prediction 1 is least likely to be correct. Foreign exchange reserves tend to decline precipitously, not increase, as a currency crisis approaches. Broad money growth tends to rise in the period leading up to a currency crisis, and the exchange rate is substantially higher than its mean level during tranquil periods.

13 A is correct. Goldsworthy has been given a bid–offer spread. Because she is buying the base currency—in this case, GBP—she must pay the offer price of JPY 129.69 per GBP.

$$\frac{\text{JPY } 225{,}000{,}000}{129.69 \text{ JPY/GBP}} = \text{GBP } 1{,}734{,}906.$$

14 A is correct. Posted quotes are typically for transactions in 1 million units of the base currency. Larger transactions may be harder for the dealer to sell in the interbank market and would likely require the dealer to quote a wider spread (lower bid price and higher offer price).

15 A is correct. Using quotes from Dealer A, she can find

$$\frac{\text{MXN}}{\text{GBP}} = \frac{\text{MXN}}{\text{USD}} \times \frac{\text{USD}}{\text{GBP}}.$$

The bid from Dealer A for MXN/GBP is effectively

$$\left(\frac{\text{MXN}}{\text{GBP}}\right)_{bid} = \left(\frac{\text{MXN}}{\text{USD}}\right)_{bid} \times \left(\frac{\text{USD}}{\text{GBP}}\right)_{bid}$$
$$= 20.140 \times 1.2301 = 24.7742.$$

The offer from Dealer A is

$$\left(\frac{\text{MXN}}{\text{GBP}}\right)_{offer} = \left(\frac{\text{MXN}}{\text{USD}}\right)_{offer} \times \left(\frac{\text{USD}}{\text{GBP}}\right)_{offer}$$
$$= 20.160 \times 1.2305 = 24.8069.$$

To compare with Dealer B's quote, she must take the inverse of MXN/GBP, so she has an offer to sell MXN at a rate of 1/24.7742 = GBP 0.0404 and a bid to purchase MXN at a rate of 1/24.8069 = GBP 0.0403. Dealer A is effectively quoting MXN/GBP at 0.0403/0.0404. Although she can effectively buy pesos more cheaply from Dealer A (GBP 0.0404 from Dealer A, versus GBP 0.0406 from Dealer B), she cannot resell them to Dealer B for a higher price than GBP 0.0403. There is no profit from triangular arbitrage.

16 A is correct. Marking her nine-month contract to market six months later requires buying GBP/EUR three months forward. The GBP/EUR spot rate is 0.9467/0.9471, and the three-month forward points are 14.0/15.0. The three-month forward rate to use is 0.9471 + (15/10000) = 0.9486. Goldsworthy sold EUR 5,000,000 at 0.9526 and bought at 0.9486. The net cash flow at the settlement date will equal EUR 5,000,000 × (0.9526 – 0.9486) GBP/EUR = GBP 20,000. This cash flow will occur in three months, so we discount at the three-month GBP Libor rate of 58 bps:

$$\frac{\text{GBP } 20{,}000}{1 + 0.0058\left(\frac{90}{360}\right)} = \text{GBP } 19{,}971.04.$$

17 A is correct. The positive forward points for the GBP/EUR pair shown in Exhibit 2 indicate that the EUR trades at a forward premium at all maturities, including three months. Covered interest rate parity,

$$F_{f/d} = S_{f/d}\left(\frac{1 + i_f\left[\frac{\text{Actual}}{360}\right]}{1 + i_d\left[\frac{\text{Actual}}{360}\right]}\right),$$

suggests a forward rate greater than the spot rate requires a non-domestic risk-free rate (in this case, the GBP Libor) greater than the domestic risk-free rate (EUR Libor). When covered interest rate parity is violated, traders can step in and conduct arbitrage.

18 B is correct. Using covered interest rate parity, the forward rate is

$$F_{f/d} = S_{f/d}\left(\frac{1 + i_f\left[\frac{\text{Actual}}{360}\right]}{1 + i_d\left[\frac{\text{Actual}}{360}\right]}\right)$$

$$= 1.2303\left(\frac{1 + 0.0033\left[\frac{90}{360}\right]}{1 + 0.0058\left[\frac{90}{360}\right]}\right) = 1.2295.$$

Because the domestic rate (Libor) is higher than the non-domestic rate, the forward rate will be less than the spot rate, giving a forward discount of

$$F_{f/d} - S_{f/d} = 1.2295 - 1.2303 = -0.0008.$$

19 B is correct. The covered interest rate parity condition,

$$F_{f/d} = S_{f/d}\left(\frac{1 + i_f\left[\frac{\text{Actual}}{360}\right]}{1 + i_d\left[\frac{\text{Actual}}{360}\right]}\right), \text{(Equation 1)}$$

specifies the forward exchange rate that must hold to prevent arbitrage given the spot exchange rate and the risk-free rates in both countries. If the forward and spot exchange rates, as well as one of the risk-free rates, are known, the other risk-free rate can be calculated.

20 B is correct. According to uncovered interest rate parity,

$$\%\Delta S^{e}_{f/d} = i_f - i_d, \text{(Equation 2)}$$

the expected change in the spot exchange rate should reflect the interest rate spread between the two countries, which can be found in Exhibit 3. Given the spot exchange rate (from Exhibit 1) and the expected future change, she should be able to estimate the future spot exchange rate.

READING

7

Economic Growth

by Paul R. Kutasovic, PhD, CFA

Paul R. Kutasovic, PhD, CFA, is at New York Institute of Technology (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** compare factors favoring and limiting economic growth in developed and developing economies;
☐	**b.** describe the relation between the long-run rate of stock market appreciation and the sustainable growth rate of the economy;
☐	**c.** explain why potential GDP and its growth rate matter for equity and fixed income investors;
☐	**d.** contrast capital deepening investment and technological progress and explain how each affects economic growth and labor productivity;
☐	**e.** demonstrate forecasting potential GDP based on growth accounting relations;
☐	**f.** explain how natural resources affect economic growth and evaluate the argument that limited availability of natural resources constrains economic growth;
☐	**g.** explain how demographics, immigration, and labor force participation affect the rate and sustainability of economic growth;
☐	**h.** explain how investment in physical capital, human capital, and technological development affects economic growth;
☐	**i.** compare classical growth theory, neoclassical growth theory, and endogenous growth theory;
☐	**j.** explain and evaluate convergence hypotheses;
☐	**k.** describe the economic rationale for governments to provide incentives to private investment in technology and knowledge;
☐	**l.** describe the expected impact of removing trade barriers on capital investment and profits, employment and wages, and growth in the economies involved.

© 2020 CFA Institute. All rights reserved.

1 AN INTRODUCTION TO GROWTH IN THE GLOBAL ECONOMY: DEVELOPED VS. DEVELOPING ECONOMIES

Forecasts of long-run economic growth are important for global investors. Equity prices reflect expectations of the future stream of earnings, which depend on expectations of future economic activity. This dynamic means that in the long term, the same factors that drive economic growth will be reflected in equity values. Similarly, the expected long-run growth rate of real income is a key determinant of the average real interest rate level in the economy, and therefore the level of real returns in general. In the shorter term, the relationship between actual and potential growth (i.e., the degree of slack in the economy) is a key driver of fixed-income returns. Therefore, in order to develop global portfolio strategies and investment return expectations, investors must be able to identify and forecast the factors that drive long-term sustainable growth trends. Based on a country's long-term economic outlook, investors can then evaluate the long-term investment potential and risk of investing in the securities of companies located or operating in that country.

In contrast to the short-run fluctuations of the business cycle, the study of economic growth focuses on the long-run trend in aggregate output as measured by potential GDP. Over long periods, the actual growth rate of GDP should equal the rate of increase in potential GDP because, by definition, output in excess of potential GDP requires employing labor and capital beyond their optimum levels. Thus, the growth rate of potential GDP acts as an upper limit to growth and determines the economy's sustainable rate of growth. Increasing the growth rate of potential GDP is the key to raising the level of income, the level of profits, and the living standard of the population. Even small differences in the growth rate translate into large differences in the level of income over time.

What drives long-run growth? What distinguishes the "winners" from the "losers" in the long-run growth arena? Will poor countries catch up with rich countries over time? Can policies have a permanent effect on the sustainable growth rate? If so, how? If not, why not? These and other key questions are addressed in detail in this reading.

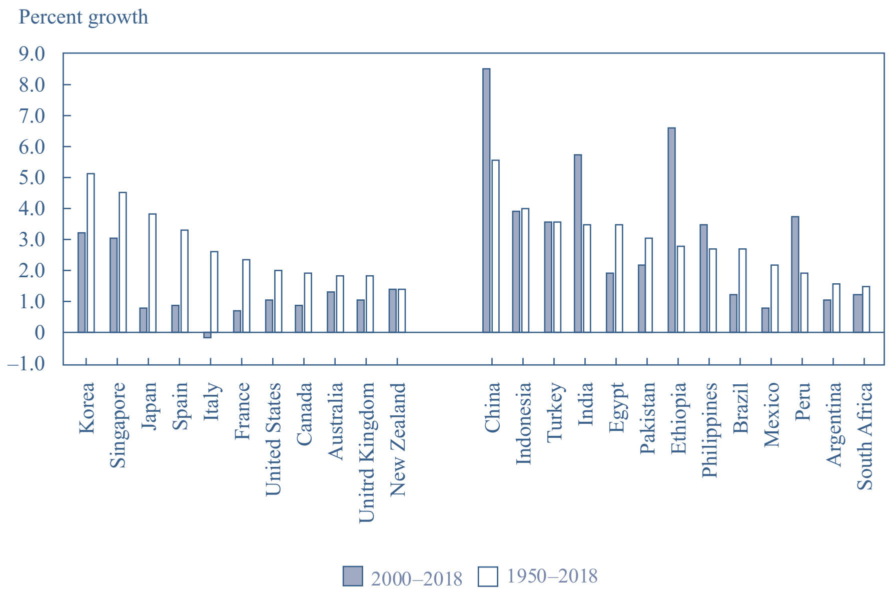
We first examine the long-term growth record, focusing on the extent of growth variation across countries and across decades. We then discuss the importance of economic growth to global investors and examine the relationship between investment returns and economic growth. We next turn to the factors that determine long-run economic growth before presenting the classical, neoclassical, and endogenous growth models. We also discuss whether poorer countries are converging to the higher income levels of the richer countries. Finally, we look at the impact of international trade on economic growth.

1.1 Growth in the Global Economy: Developed vs. Developing Economies

The first step in our study of long-term growth is to compare the economic performance of countries. GDP and per capita GDP are the best indicators economists have for measuring a country's standard of living and its level of economic development. Economic growth is calculated as the annual percentage change in real GDP or in real per capita GDP. Growth in real GDP measures how rapidly the total economy is expanding. Real per capita GDP reflects the average standard of living in each country—essentially the average level of material well-being. Growth in real GDP per capita (i.e., real GDP growing faster than the population) implies a rising standard of living.

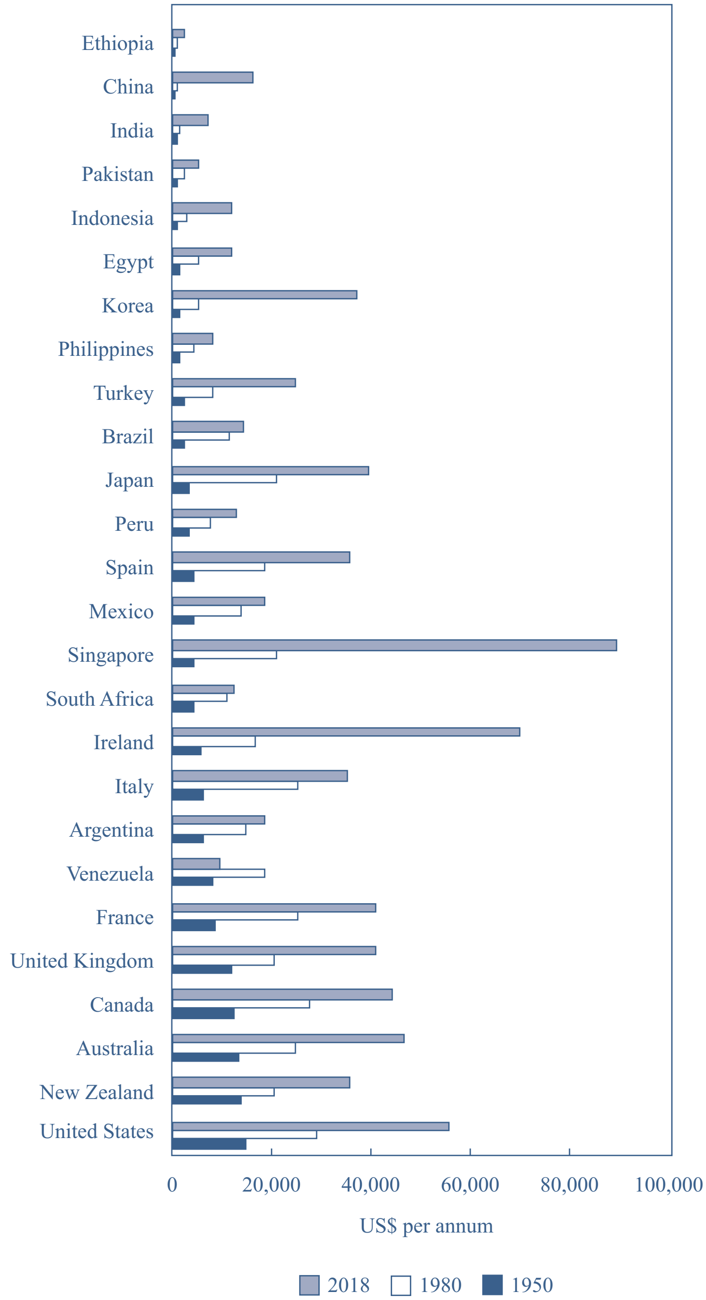
Exhibit 1 presents data on the level of per capita GDP and the growth rate of GDP for a selection of economies. Because each economy reports its data in its own currency, each one's data must be converted into a common currency, usually the US dollar. One can convert the GDP data into dollars using either current market exchange rates or the exchange rates implied by **purchasing power parity (PPP)**. Purchasing power parity is the idea that exchange rates move to equalize the purchasing power of different currencies. At the exchange rates implied by PPP, the cost of a typical basket of goods and services is the same across all economies. In other words, exchange rates should be at a level where you can buy the same goods and services with the equivalent amount of any economy's currency.

Exhibit 1 Divergent Real GDP Growth among Selected Economies and Real GDP per Capita, in US$



(continued)

Exhibit 1 (Continued)



Note: The measure of GDP per capita is in constant US dollar market prices for 2011 and adjusted for cross-economy differences in the relative prices of goods and services using PPP.
Sources: International Monetary Fund, World Economic Outlook database for growth rates, and Conference Board, Total Economy Database (September 2019).

In general, the simple method of taking a country's GDP measured in its own currency and then multiplying by the current exchange rate to express it in another currency is not appropriate. Using market exchange rates has two problems. First, market exchange rates are very volatile. Changes in the exchange rate could result in large swings in measured GDP even if there is little or no growth in the country's economy. Second, market exchange rates are determined by financial flows and flows in tradable goods and services. This dynamic ignores the fact that much of global consumption is for non-tradable goods and services. Prices of non-traded goods and services differ by country. In particular, non-traded goods are generally less expensive in developing countries than in developed countries. For example, because labor is cheaper in Mexico City than in London, the prices of labor-intensive products, such as haircuts or taxi rides, are lower in Mexico City than in London. Failing to account for differences in the prices of non-traded goods and services across countries tends to understate the standard of living of consumers in developing countries. To compare standards of living across time or across countries, we need to use a common set of prices among a wide range of goods and services. Thus, cross-country comparisons of GDP should be based on purchasing power parity rather than current market exchange rates.

The economies in Exhibit 1 are divided into two categories: developed (or advanced) economies and developing economies. Developed economies tend to be those with high per capita GDP. There are no universally agreed-upon criteria, however, for classifying economies as advanced or developing. The International Monetary Fund (IMF) classifies 39 economies as advanced and 155 as developing. It says that "this classification is not based on strict criteria, economic or otherwise, and has evolved over time" (IMF 2019). Developed countries include the United States, Canada, Australia, Japan, and major economies in Europe. Growth in the large, developed economies generally slowed over the last few decades, with US growth exceeding that of Europe and Japan. Also included in this group are countries such as Singapore, Ireland, and Spain, which were poor in the 1950s but now have relatively high per capita real GDPs because of high growth rates over the past 50 years.

The second group of countries is the developing countries of Africa, Asia, and Latin America. Per capita GDP in these countries is lower than in the advanced countries, but GDP is generally growing faster than in the developed countries. Although the growth rates of the developing countries exceed those of the advanced countries, there is significant variation in economic performance among the developing countries. China and India have been growing at a rapid rate. Meanwhile, growth in Latin America, Africa, and the Middle East has lagged behind Asia.

What explains the diverse experiences among the developing countries and between the developed and developing ones? Singapore, for example, had less than half the per capita GDP of the United States in 1970 but now has per capita GDP that exceeds that of the United States. In contrast, such countries as Ethiopia and Kenya have remained poor, with little growth in per capita GDP. The literature on economic growth focuses primarily on the role of capital and labor resources and the use of technology as sources of growth. In addition to these purely economic drivers, developed and developing countries differ with respect to the presence or absence of appropriate institutions that support growth. These institutions enable developing countries to raise their standards of living and eventually move into the ranks of the developed countries. We now examine some of the key institutions and requirements for growth.

FACTORS FAVORING AND LIMITING ECONOMIC GROWTH

2

a compare factors favoring and limiting economic growth in developed and developing economies;

One of the major problems for some of the developing countries is a low level of capital per worker. Countries accumulate capital through private and public sector (e.g., infrastructure) investment. But increasing the investment rate may be difficult in developing countries because low levels of disposable income can make it difficult to generate significant saving. The low saving rate contributes to a vicious cycle of poverty: Low savings lead to low levels of investment, which leads to slow GDP growth, which implies persistently low income and savings. Therefore, it is very difficult to design policies to increase domestic saving and investment rates in developing countries. The good news is that the savings of domestic residents are not the only source of investment funds. A developing country can break out of the cycle of low savings by attracting foreign investment.

2.1 Financial Markets and Intermediaries

In addition to the saving rate, growth depends on how efficiently saving is allocated within the economy. A role of the financial sector in any economy is to channel funds from savers to investment projects. Financial markets and intermediaries, such as banks, can promote growth in at least three ways. First, by screening those who seek funding and monitoring those who obtain funding, the financial sector channels financial capital (savings) to projects that are likely to generate the highest risk-adjusted returns. Second, the financial sector may encourage savings and assumption of risk by creating attractive investment instruments that facilitate risk transfer and diversification and enhance liquidity. Finally, the existence of well-developed financial markets and intermediaries can mitigate the credit constraints that companies might otherwise face in financing capital investments. For example, banks can aggregate small amounts of savings into a larger pool enabling them to finance larger projects that can exploit economies of scale. Evidence suggests that countries with better-functioning financial markets and intermediaries grow at a faster rate (Levine, 2005). Not all financial sector developments promote economic growth, however. Financial sector intermediation that results in declining credit standards and/or increasing leverage will increase risk and not necessarily increase long-run growth.

2.2 Political Stability, Rule of Law, and Property Rights

Stable and effective government, a well-developed legal and regulatory system, and respect for property rights are key ingredients for economic growth. Property rights are the legal arrangements that govern the protection of private property, including intellectual property. Clearly established property rights create the incentive for domestic households and companies to invest and save. A legal system—substantive and procedural laws—is needed to establish and protect these rights. Substantive law focuses on the rights and responsibilities of entities and relationships among entities, and procedural law focuses on the protection and enforcement of the substantive laws. In developed countries these rights and arrangements are well established, but they may be lacking or ineffective in developing countries.

In addition, economic uncertainty increases when wars, military coups, corruption, and other sources of political instability are widespread. These factors raise investment risk, discourage foreign investment, and weaken growth. In many developing countries, especially those in Africa, the first priority in trying to enhance growth is to enact a legal system that establishes, protects, and enforces property rights.

2.3 Education and Health Care Systems

Inadequate education at all levels is a major impediment to growth for many developing countries. Many workers are illiterate, and few workers have the skills needed to use the latest technology. At the same time, many developing countries also suffer from a "brain drain," in which the most highly educated individuals leave the developing country for more-advanced countries. Basic education raises the skill level of the workforce and thus contributes to the country's potential for growth. In addition, because physical capital and human capital are often complementary, education can raise growth by increasing the productivity of existing physical capital. Thus, improving education, through both formal schooling and on-the-job training, is an important component of a sustainable growth strategy for a developing country. China and India are investing large amounts in education and have successfully graduated large numbers of students majoring in engineering and technology-related areas of study. This effort is significantly improving the quality of their workforces.

Empirical studies show that the allocation of education spending among different types and levels (primary, secondary, and post-secondary) of education is a key determinant of growth, especially in comparing growth in developed countries with growth in developing ones. The impact of education spending depends on whether the country is on the leading edge of technology and fostering innovation or simply relying on imitation as a source of growth. Typically, developed countries, such as the United States, Japan, and western European nations, are on the leading edge of technology and need to invest in post-secondary education to encourage innovation and growth. For these countries, incremental spending on primary and secondary education will have a relatively small impact on growth. In contrast, the developing countries, which largely apply technology developed elsewhere, should emphasize primary and secondary education. Such spending will improve growth by improving the countries' ability to absorb new technologies and to organize existing tasks more efficiently.

Poor health is another obstacle to growth in the developing countries. Life expectancy rates are substantially lower in many developing countries than in developed ones. In Africa, tropical diseases are rampant and AIDS has had a devastating impact. The GDP growth rate in Botswana, a huge success story in the 1970s and 1980s, slowed dramatically during the following two decades as a result, at least in part, of the AIDS epidemic.

2.4 Tax and Regulatory Systems

Tax and regulatory policies have an important impact on growth and productivity, especially at the company level. Analysis suggests that limited regulations encourage entrepreneurial activity and the entry of new companies. There is also a strong positive correlation between the entry of new companies and average productivity levels. Studies by the Organisation for Economic Co-Operation and Development (OECD) indicate that low administrative startup cost is a key factor encouraging entrepreneurship (OECD 2003).

2.5 Free Trade and Unrestricted Capital Flows

Opening an economy to capital and trade flows has a major impact on economic growth. In an open economy, world savings can finance domestic investment. As a potential source of funds, foreign investment can break the vicious cycle of low income, low domestic savings, and low investment. Foreign investment can occur in two ways:

- Foreign companies can invest directly in a domestic economy (so-called foreign direct investment, or FDI) by building or buying property, plant, and equipment.
- Foreign companies and individuals can invest indirectly in a domestic economy by purchasing securities (equity and fixed income) issued by domestic companies.

Both of these forms of foreign investment will potentially increase the developing economy's physical capital stock, leading to higher productivity, employment and wages, and perhaps even increased domestic savings. This suggests that developing countries would benefit from policies that encourage investment from abroad, such as eliminating high tariffs on imports (especially capital goods) and removing restrictions on foreign direct and indirect investments.

Brazil and India are examples of developing countries that have benefited from foreign investment. Foreign companies directly invested $48.5 billion in Brazil in 2010, an important source of investment spending for the Brazilian economy (see Exhibit 19). Foreign direct investment also provides developing countries with access to advanced technology developed and used in the advanced countries. In 1999, India enacted new regulations that liberalized direct and indirect foreign investments in Indian companies. Foreign institutional and venture capital investors were given greater flexibility to invest directly in Indian entities as well as in the Indian capital markets. These changes also made it easier for foreign companies to invest in plant and equipment. These developments contributed to the acceleration in India's economic growth over the last decade (see Exhibit 1).

Capital flows are just one way that the international economy affects economic growth. The other is through trade in goods and services. In general, free trade benefits an economy by providing its residents with more goods at lower costs. With free trade, domestic companies face increased competition, which limits their price discretion, but they also obtain access to larger markets. The evidence of the benefits of open markets is discussed later in the reading.

2.6 Summary of Factors Limiting Growth in Developing Countries

Developing countries differ significantly from developed countries in terms of their institutional structure and their legal and political environments. Lack of appropriate institutions and poor legal and political environments restrain growth in the developing economies and partially explain why these countries are poor and experience slow growth. Factors limiting growth include the following:

- Low rates of saving and investment
- Poorly developed financial markets
- Weak, or even corrupt, legal systems and failure to enforce laws
- Lack of property rights and political instability
- Poor public education and health services
- Tax and regulatory policies discouraging entrepreneurship
- Restrictions on international trade and flows of capital

Although these factors are not necessarily absent in developed countries, they tend to be more prevalent in developing countries. Policies that correct these issues, or mitigate their impact, enhance the potential for growth. In addition to these institutional restraints, as we will see later, growth in developing countries may be limited by a lack of physical, human, and public capital, as well as little or no innovation.

EXAMPLE 1

Why Growth Rates Matter

In 1950, Argentina and Venezuela were relatively wealthy countries with per capita GDP levels of \$6,164 and \$8,104, respectively. Per capita GDPs in these Latin American countries were well above those of Japan, South Korea, and Singapore, which had per capita GDPs of \$3,048, \$1,185, and \$4,299, respectively. By 2018, however, a dramatic change occurred in the relative GDPs per capita of these countries.

Real GDP Per Capita in US Dollars

	Venezuela	Argentina	Singapore	Japan	South Korea
1950	\$8,104	\$6,164	\$4,299	\$3,048	\$1,185
2018	\$9,487	\$18,255	\$89,196	\$39,313	\$36,756

1. Calculate the annual growth rate in per capita GDP for each of the five countries over the period 1950–2018.
2. Explain the implication of the growth rates for these countries.
3. Suppose that GDP per capita in Argentina had grown at the same rate as in Japan from 1950 to 2018. How much larger would real per capita GDP have been in Argentina in 2018?
4. Venezuela plans to stimulate growth in its economy by substantially increasing spending on infrastructure, education, and health care. Nevertheless, foreign investment is discouraged, and reforms such as strengthening the legal system and encouraging private ownership have been largely ignored. Explain whether the measures described here could lead to faster economic growth.

Solution to 1:

The annual growth rates for the five countries are calculated as follows:

Argentina	$[(\$18{,}255/\$6{,}164)^{1/68}] - 1 = 1.6\%$
Venezuela	$[(\$9{,}487/\$8{,}104)^{1/68}] - 1 = 0.2\%$
Japan	$[(\$39{,}313/\$3{,}048)^{1/68}] - 1 = 3.8\%$
Singapore	$[(\$89{,}196/\$4{,}299)^{1/68}] - 1 = 4.6\%$
South Korea	$[(\$36{,}756/\$1{,}185)^{1/68}] - 1 = 5.2\%$

Solution to 2:

Differences in GDP growth rates sustained over a number of decades will significantly alter the relative incomes of countries. Nations that experience sustained periods of high growth will eventually become high-income countries and move up the income ladder. In contrast, countries with slow growth will experience relative declines in living standards. This dynamic is well illustrated in this example by a historical comparison of growth in Argentina and Venezuela with Japan, Singapore, and South Korea. In 1950, Argentina and Venezuela were relatively wealthy countries with per capita GDP levels well above those of Japan, South Korea, and Singapore. Over the next 60 years, however, the rate of growth in per capita GDP was significantly slower in Venezuela and Argentina in comparison to the three Asian countries. This disparity resulted in a dramatic change in these countries' relative incomes. The per capita GDP of the three Asian countries rose sharply as each joined the ranks of developed countries. In contrast, Argentina and Venezuela stagnated and moved from the ranks of developed countries to developing country status. By 2018, per capita income in Singapore was more than eight times higher than in Venezuela.

Over the long run, the economic growth rate is an extremely important variable. Even small differences in growth rates matter because of the power of compounding. Thus, policy actions that affect the long-term growth rate even by a small amount will have a major economic impact.

Solution to 3:

Assuming Argentina had grown at the same rate as Japan since 1950, its GDP per capita in 2018 would have been $(\$6,164)(1+ 0.038)^{68} = (\$6,164)(12.63) =$ \$77,854, versus \$18,255 from Exhibit 1.

If Argentina had grown at the same rate as Japan, it would have had one of the highest standards of living in the world in 2018. The question is why the growth rates in Argentina and Venezuela diverged so much from the three Asian countries.

Solution to 4:

The preconditions for economic growth are well-functioning financial markets, clearly defined property rights and rule of law, open international trade and flows of capital, an educated and healthy population, and tax and regulatory policies that encourage entrepreneurship. Investment in infrastructure would increase Venezuela's stock of physical capital, which would raise labor productivity and growth. Better education and health care would increase human capital and also increase productivity and growth. These measures would raise Venezuela's growth prospects. Missing, however, are a legal system that could better enforce property rights, openness to international trade and foreign investment, and well-functioning capital markets. Without changes in these preconditions, a significant improvement in Venezuela's growth is unlikely to occur. The following table summarizes these preconditions:

Preconditions for Growth	Impact of Planned Policy Action in Venezuela
Saving and investment	Improve growth potential
Developed financial markets	No impact
Legal systems	No impact
Property rights and political stability	No impact
Education and health	Improve growth potential
Tax and regulatory polices discouraging entrepreneurship	No impact
Restrictions on international trade and flows of capital	No impact

It should be noted that the global economy is evolving rapidly and past trends may or may not be sustained. Nonetheless, in order to provide concrete answers that do not require the reader to bring in additional information, our exercise solutions must assume past patterns are indicative of the future.

WHY POTENTIAL GROWTH MATTERS TO INVESTORS

3

b describe the relation between the long-run rate of stock market appreciation and the sustainable growth rate of the economy;

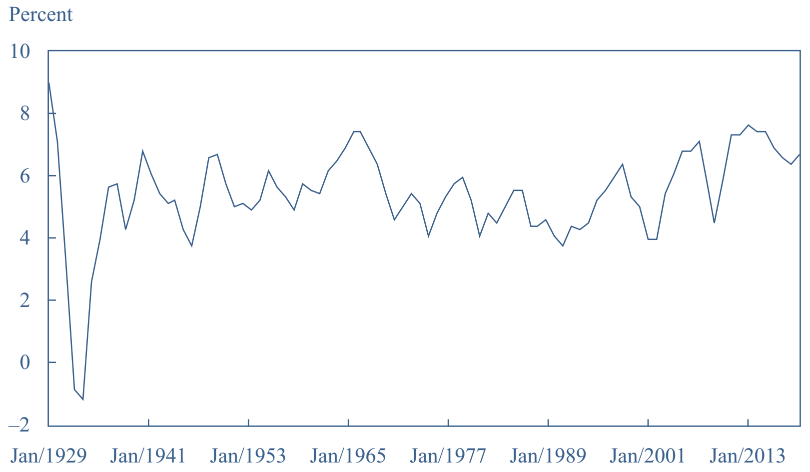
c explain why potential GDP and its growth rate matter for equity and fixed income investors;

The valuations of both equity and fixed-income securities are closely related to the growth rate of economic activity. Anticipated growth in aggregate earnings is a fundamental driver of the equity market. Growth in an economy's productive capacity, measured by **potential GDP**, places a limit on how fast the economy can grow. The idea is that potential GDP is the maximum amount of output an economy can sustainably produce without inducing an increase in the inflation rate. A key question for equity investors, therefore, is whether earnings growth is also bounded or limited by the growth rate of potential GDP.

For earnings growth to exceed GDP growth, the ratio of corporate profits to GDP must trend upward over time. It should be clear that the share of profits in GDP cannot rise forever. At some point, stagnant labor income would make workers unwilling to work and would also undermine demand, making further profit growth unsustainable. Thus, in the long run, real earnings growth cannot exceed the growth rate of potential GDP. Note that earnings growth for the overall national economy can differ from EPS growth in a country's equity market composites. This difference results from the presence of new businesses that are not yet included in the equity indexes and are typically growing at a faster rate than the mature companies that make up the composites. Thus, the earnings growth rate of companies making up the composites should be lower than the overall economy's earnings growth rate.

Exhibit 2 illustrates the long-run stability of after-tax profits as a share of GDP using US data derived from the National Income and Product Accounts (NIPA). The chart shows that since 1947, after-tax profits have ranged between 3.1% and 10.1% of GDP and have averaged around 6% of GDP. Note that there is neither an upward trend in the ratio of after-tax profits to GDP nor a move to a permanent increase in the ratio. The share of profits in 1947, at 8.5%, was essentially equal to the 9.9% share at the end of the period in 2018. Because there is no trend in the ratio, the same factors that limit economic growth also set the upper limit or bound on the long-run growth of aggregate earnings.

Exhibit 2 US After-Tax Corporate Profits as a Percentage of GDP



Source: FRED database, Federal Reserve Bank of St. Louis.

To examine the relationship between economic growth and stock prices, it is useful to express the aggregate value of the stock market as the product of key ratios. Letting P represent the aggregate value (price) of equities and E represent aggregate earnings, we can write the following:

$$P = GDP\left(\frac{E}{GDP}\right)\left(\frac{P}{E}\right)$$

This equation represents the aggregate value of equities as the product of GDP, corporate earnings as a share of GDP, and the price-to-earnings ratio for the market. Note that GDP may be interpreted as either real or nominal with a corresponding real or nominal interpretation of the other variables.

This equation can be expressed in terms of logarithmic rates of change over a time horizon T:

$$(1/T)\%\Delta P = (1/T)\%\Delta GDP + (1/T)\%\Delta(E/GDP) + (1/\mathrm{T})\%\Delta(P/E)$$

Thus, the percentage change in stock market value equals the percentage change in GDP plus the percentage change in the share of earnings (profit) in GDP plus the percentage change in the price-to-earnings multiple (issuance and repurchases not incorporated). Over short to immediate horizons, all three of these factors contribute to appreciation or depreciation of the stock market. In the long run, however, the growth rate of GDP must dominate. As noted earlier, the ratio of earnings to GDP cannot rise forever. It cannot decline forever, either, because unprofitable businesses will disappear. Hence, the second term in the preceding equation must be approximately zero over long horizons (T). Similarly, the price-to-earnings ratio cannot grow or contract forever because investors will not pay an arbitrarily large price for a unit of earnings, nor will they give away earnings for nothing. Hence, the third term must also be approximately zero over long horizons. The conclusion is that the drivers of potential GDP are ultimately the drivers of stock market price performance.

Exhibit 3 shows the close relationship between economic growth and equity market appreciation over long horizons. Over the period 1946–2007, the S&P 500 Index returned 10.82% per year, of which 7.15% per year came from price appreciation. The price appreciation was almost exactly equal to the 6.95% growth rate of US nominal GDP (real GDP growth plus inflation). Changes in the earnings-to-GDP and price-to-earnings ratios contributed only a combined 0.20% per year. As shown in the last column of the exhibit, these two ratios contributed much more to the volatility of the market than to its return. Note that the period chosen for the study features endpoints that correspond to fairly normal economic and market conditions. Selecting endpoints that correspond to crisis or bubble conditions would distort the role played by the various components of return.

Exhibit 3 Decomposition of S&P 500 Index Returns: Log Returns, 1946–2007

	Annual Return/Growth Rate	Standard Deviation
S&P 500 return	10.82%	15.31%
Real GDP growth	3.01	2.97
Inflation	3.94	3.29
EPS/GDP	−0.12	17.62
P/E	0.32	23.80
Dividend yield	3.67	1.49
Total	10.82	

Source: Stewart, Piros, and Heisler (2011).

Estimates of potential GDP and its growth rate are widely available. For example, both the OECD and the International Monetary Fund (IMF) provide such estimates as a basis for their intermediate-term and long-term forecasts of economic growth by country. In addition, central banks regularly make projections of potential GDP. The methods used to estimate potential GDP are examined later in the reading. Simply extrapolating past GDP growth into the future may produce an incorrect forecast. A country's GDP growth rate can and does change over time. GDP growth can either slow down, as was the case for Japan (the pre-1990 growth rate is much higher than the post-1990 rate), or speed up, as was the case for Brazil for more than a decade after 1999. Factors or policies that cause potential growth to increase or decrease by even a small amount will have a large impact on living standards and the future level of economic activity. The effect is analogous to the rate of return on a portfolio, in which small differences in return compounded over many years result in a substantially higher or lower portfolio value. Being able to recognize these changes is critical for the global investor.

Estimates of an economy's growth potential are also relevant for global fixed income investors. One of the uses of potential GDP is to gauge inflationary pressures in the economy. Actual GDP growth above (below) the potential growth rate puts upward (downward) pressure on inflation, which puts corresponding pressure on nominal interest rates and bond prices. Note that this is an argument about cyclical variations in growth and inflation around the economy's long-term potential growth rate. It does not imply that there is a long-run trade-off between growth and inflation.

The growth rate of potential GDP is also an important determinant of the level of real interest rates, and therefore real asset returns in general, in the economy. The real interest rate is essentially the real return that consumers/savers demand in exchange for postponing consumption. Faster growth in potential GDP means that consumers expect their real income to rise more rapidly. This dynamic implies that an extra unit of future income/consumption is less valuable than it would be if income were expected to grow more slowly. Hence, all else equal, the real interest rate will have to be higher in order to induce the savings needed to fund required capital accumulation. Thus, higher rates of potential GDP growth translate into higher real interest rates and higher expected real asset returns in general.

Potential GDP and its growth rate enter into fixed-income analysis in other ways as well. Among them are the following:

1. A higher rate of potential GDP growth improves the general credit quality of fixed-income securities because most such securities are ultimately backed by a flow of income even if the lender has a claim on specific underlying assets.
2. Central banks frequently explain their monetary policy decisions by referring to the level of "resource utilization" and the degree of "slack in the economy." In other words, monetary policy decisions are affected by the difference between an economy's estimated potential output and its actual operating level (referred to as the output gap) and by growth of actual GDP relative to the sustainable growth rate. Thus, fixed income investors need to closely monitor the output gap and growth rates of actual and potential GDP to assess the likelihood of a change in central bank policy.
3. Credit rating agencies use the growth rate of potential GDP as an input in evaluating the credit risk of sovereign debt or government-issued debt. All else equal, slower estimated potential GDP growth raises the perceived risk of these bonds.
4. Government budget deficits typically increase during recessions and decrease during expansions. In examining fiscal policy, actual fiscal positions are often judged relative to structural or cyclically adjusted deficits—the budgetary balance that would exist if the economy were operating at potential GDP.

EXAMPLE 2

Impact on Equity and Fixed Income Investors

Your firm subscribes to asset class risk and return estimates generated by a large pension consultant. The equity market return estimates are based primarily on long-term average index returns. Following a multi-year period of very high equity returns driven by unusually high earnings growth and expanding P/E multiples, capital's share of total income as well as valuation multiples are near all-time highs. Based on the latest data, the vendor projects that your domestic equity market will return 13.5% per year—11% annual appreciation and 2.5% dividend yield—forever.

Your firm also subscribes to a macroeconomic forecasting service that provides, in addition to shorter-term projections, estimates of the long-term growth rate of potential GDP and the long-term inflation rate. This service forecasts 3.25% real growth in the future and 3.75% inflation, down from 4.0% and 5.0%, respectively, over the last 75 years.

1. Why might you have greater confidence in the macroeconomic service's forecasts than in the pension consultant's equity market return forecast?
2. Assuming the macroeconomic forecasts are accurate, what implicit assumptions underlie the pension consultant's forecast of 11% equity market appreciation?

3 Assuming the macroeconomic forecasts are accurate, what would be a more reasonable forecast for long-term equity returns?

4 In addition to its long-term potential GDP forecast, the macroeconomic forecasting service estimates sluggish 1.5% GDP growth for the next year. Based on this short-term GDP forecast, the bond analyst at your firm recommends that the firm increase its fixed-income investments. What assumptions underlie the bond analyst's forecast?

Solution to 1:

High volatility makes equity returns very hard to predict based on their own history. As illustrated in Exhibit 3, the high volatility of equity returns results from the underlying volatility of earnings as a share of GDP and valuation ratios. Long-term real GDP growth rates tend to be far less volatile, especially for developed economies such as the United States or the euro area, because long-term potential growth is governed by fundamental economic forces that tend to evolve slowly over time. Similarly, for countries with prudent monetary policies, inflation rates are much less volatile than stock prices. Thus, one could reasonably place much higher confidence in forecasts of long-term real and nominal (real growth plus inflation) GDP growth than in equity market return forecasts based on historical equity returns.

Solution to 2:

We can decompose the equity market appreciation rate into components resulting from (a) nominal GDP growth, (b) expansion/contraction of the share of profits in GDP, and (c) expansion/contraction of the P/E. The macroeconomic forecast indicates that nominal GDP will grow at 7% (3.25% real + 3.75% inflation). So the pension consultant's forecast of 11% equity market appreciation implies a 4% annual combined contribution from expansion in the P/E multiple and/or the profit share of GDP—*forever*.

Solution to 3:

Neither the P/E nor the profit share of GDP can grow at a non-negligible rate forever. A much more reasonable forecast of long-term equity market appreciation would be the projected 7% growth rate of nominal GDP.

Solution to 4:

With forecasted actual GDP growth well below the growth in potential GDP, the bond analyst assumes a growing output gap or slack in the economy. This slack may place downward pressure on inflation and reduce inflationary expectations. To close this gap, the central bank may need to lower short-term interest rates and ease policy. In such an environment, bond prices should rise.

4 DETERMINANTS OF ECONOMIC GROWTH: PRODUCTION FUNCTION AND GROWTH ACCOUNTING

d contrast capital deepening investment and technological progress and explain how each affects economic growth and labor productivity;

e demonstrate forecasting potential GDP based on growth accounting relations;

What are the forces driving long-run economic growth? The following sections discuss labor, physical and human capital, technology, and other factors, such as natural resources and public infrastructure, as inputs to economic growth and production functions and how changes in such inputs affect growth. We begin the discussion by presenting one of the simplest useful models of the production function.

4.1 Production Function

A production function is a model of the quantitative link between the inputs (factors of production), technology, and output. A two-factor aggregate production function with labor and capital as the inputs can be represented as

$$Y = AF(K,L), \tag{1}$$

where Y denotes the level of aggregate output in the economy, L is the quantity of labor or number of workers or hours worked in the economy, and K is an estimate of the capital services provided by the stock of equipment and structures used to produce goods and services. The function $F(\,)$ embodies the fact that capital and labor can be used in various combinations to produce output.

In the production function above, A is a multiplicative scale factor referred to as **total factor productivity (TFP)**. Note that an increase in TFP implies a proportionate increase in output for any combination of inputs. Hence, TFP reflects the general level of productivity or technology in the economy. The state of technology embodies the cumulative effects of scientific advances, applied research and development, improvements in management methods, and ways of organizing production that raise the productive capacity of factories and offices.

It is worth noting that both the function $F(\,)$ and the scale factor A reflect technology. An innovation that makes it possible to produce the same output with the same amount of capital but fewer workers would be reflected in a change in the function $F(\,)$ because the relative productivity of labor and capital has been altered. In contrast, an increase in TFP does not affect the relative productivity of the inputs. As is standard in the analysis of economic growth, *unless stated otherwise, the level of "technology" should be interpreted as referring to TFP*.

To obtain concrete results, it is useful to use a specific functional form for the production function. The **Cobb–Douglas production function**, given by

$$F(K,L) = K^{\alpha}L^{1-\alpha}, \tag{2}$$

is widely used because it is easy to analyze and does a good job of fitting the historic data relating inputs and output. The parameter α determines the shares of output (factor shares) paid by companies to capital and labor and is assumed to have a value between 0 and 1. The reason for this follows from basic microeconomics. In a competitive economy, factors of production are paid their marginal product. Profit maximization requires that the marginal product of capital equal the **rental price of capital** and the marginal product of labor equal the (real) wage rate. In the case of capital, the marginal product of capital (MPK) for the Cobb–Douglas production function is[1]

$$\text{MPK} = \alpha AK^{\alpha-1}L^{1-\alpha} = \alpha Y/K.$$

Setting the MPK equal to the rental price (r) of capital,

$$\alpha Y/K = r.$$

1 The marginal product of capital is simply the derivative of output with respect to capital. This relationship can be approximated as $\Delta Y/\Delta K \approx [A(K + \Delta K)^{\alpha}L^{1-\alpha} - AK^{\alpha}L^{1-\alpha}]/\Delta K \approx [A\alpha K^{\alpha-1}\Delta KL^{1-\alpha}]/\Delta K = A\alpha K^{\alpha-1}L^{1-\alpha} = \alpha Y/K$. The approximation becomes exact for very small increments, ΔK.

If we solve this equation for α, we find that it equals the ratio of capital income, rK, to output or GDP, Y. Thus, α *is the share of GDP paid out to the suppliers of capital*. A similar calculation shows that $1 - \alpha$ is the share of income paid to labor. This result is important because it is easy to estimate α for an economy by simply looking at capital's share of income in the national income accounts.

The Cobb–Douglas production function exhibits two important properties that explain the relationship between the inputs and the output. First, the Cobb–Douglas production function exhibits **constant returns to scale**. This means that if all the inputs into the production process are increased by the same percentage, then output rises by that percentage. Under the assumption of constant returns to scale, we can modify the production function (Equation 1) and examine the determinants of the quantity of output per worker. Multiplying the production function by $1/L$ gives

$$Y/L = AF(K/L,L/L) = AF(K/L,1).$$

Defining $y = Y/L$ as the output per worker or (average) **labor productivity** and $k = K/L$ as the capital-to-labor ratio, the expression becomes

$$y = AF(k,1).$$

Specifying the Cobb–Douglas production function in output per worker terms, where again lower case letters denote variables measured on a per capita basis, we get

$$y = Y/L = A(K/L)^{\alpha}(L/L)^{1-\alpha} = Ak^{\alpha}. \quad (3)$$

This equation tells us that the amount of goods a worker can produce (labor productivity) depends on the amount of capital available for each worker (capital-to-labor ratio), technology or TFP, and the share of capital in GDP (α). It is important to note that this equation contains two different measures of productivity or efficiency. Labor productivity measures the output produced by a unit of labor, dividing the output (GDP) by the labor input used to produce that output ($y = Y/L$). TFP is a scale factor that multiplies the impact of the capital and labor inputs. Changes in TFP are estimated using a growth accounting method discussed in the next section.

A second important property of the model is the relation between an individual input and the level of output produced. The Cobb–Douglas production function exhibits **diminishing marginal productivity** with respect to each individual input. Marginal productivity is the extra output produced from a one-unit increase in an input keeping the other inputs unchanged. It applies to any input as long as the other inputs are held constant. For example, if we have a factory of a fixed size and we add more workers to the factory, the marginal productivity of labor measures how much additional output each additional worker will produce. Diminishing marginal productivity means that at some point, the extra output obtained from each additional unit of the input will decline. To continue our example, if we hire more workers at the existing factory (fixed capital input in this case), each additional worker adds less to output than the previously hired worker does, and average labor productivity (y) falls.

The significance of diminishing marginal returns in the Cobb–Douglas production function depends on the value of α. *A value of α close to zero means diminishing marginal returns to capital are very significant and the extra output made possible by additional capital declines quickly as capital increases.* In contrast, a value of α close to one means that the next unit of capital increases output almost as much as the previous unit of capital. In this case, diminishing marginal returns still occur but the impact is relatively small. Note that the exponents on the K and L variables in the Cobb–Douglas production function sum to one, indicating constant returns to scale—that is, there are no diminishing marginal returns if both inputs are increased proportionately.

4.2 Growth Accounting

Since the publication of Solow's seminal work (Solow 1957), growth accounting has been used to analyze the performance of economies. The growth accounting equation is essentially the production function written in the form of growth rates. It starts with the Cobb–Douglas production function and decomposes the percentage change in output into components attributable to capital, labor, and technology:

$$\Delta Y/Y = \Delta A/A + \alpha\Delta K/K + (1 - \alpha)\Delta L/L \tag{4}$$

The **growth accounting equation** states that the growth rate of output equals the rate of technological change plus α times the growth rate of capital plus (1 − α) multiplied by the growth rate of labor. Because a 1% increase in capital leads to an α% increase in output, α is the elasticity of output with respect to capital. Similarly, (1 − α) is the elasticity of output with respect to labor. Thus, in the Cobb–Douglas production function, the exponents α and (1 − α) play dual roles as both output elasticities and the shares of income paid to each factor. Note that the impact of any unspecified inputs (e.g., natural resources) is subsumed into the TFP component.

Data on output, capital, labor, and the elasticities of capital and labor are available for most developed countries. The rate of technological change is not directly measured and must therefore be estimated. The elasticities of capital and labor in the growth accounting equation are the relative shares of capital (α) and labor (1 − α) in national income and are estimated from the GDP accounts. For the United States, the relative shares of labor and capital are approximately 0.7 and 0.3, respectively. This means that an increase in the growth rate of labor will have a significantly larger impact—roughly double—on potential GDP growth than will an equivalent increase in the growth rate of capital, holding all else equal. For example, because capital's share in GDP in the US economy is 0.3, a 1% increase in the amount of capital available for each worker increases output by only 0.3%. An equivalent increase in the labor input would boost growth by 0.7%.

The growth accounting equation has a number of uses in studying an economy. First, Solow used the equation to estimate the contribution of technological progress to economic growth. Solow estimated the growth in TFP as a residual in the preceding equation by plugging in $\Delta Y/Y$, $\Delta K/K$, $\Delta L/L$, and α and solving for $\Delta A/A$. This residual measures the amount of output that cannot be explained by growth in capital or labor and can thus be regarded as progress in TFP.

Second, the growth accounting equation is used to empirically measure the sources of growth in an economy. In such studies, the growth accounting equation is used to quantify the contribution of each factor to long-term growth in an economy and answer such questions as the following: How important are labor and demographic factors to growth? What is the contribution of capital, and how important is capital deepening as a source of growth? What is the impact of TFP? The growth accounting equation can be expanded by considering different forms of capital and labor inputs, such as human capital and knowledge capital, and by considering the quality of the inputs as well.

Finally, the growth accounting equation is used to measure potential output. Potential GDP is estimated using Equation 4 with trend estimates of labor and capital and α estimated as one minus the labor share of GDP. The difficult task is estimating the growth rate of TFP, which, by definition, is a residual in the growth accounting equation TFP is computed as the growth in output less the growth in the factor inputs. These inputs include labor and capital in the traditional Solow two-factor production model. If the production function is expanded by including more inputs, the weighted growth rates of these inputs would also be subtracted from the growth in output. The standard methodology treats TFP as exogenous and estimates its growth rate using various time-series models.

An alternative method of measuring potential GDP is the **labor productivity growth accounting equation**. It is very similar to the Solow approach but is simpler and models potential GDP as a function of the labor input and the productivity of the labor input. It avoids the need to estimate the capital input and the difficulty associated with computing total factor productivity. The disadvantage is that it incorporates both capital deepening and TFP progress in the productivity term in a way that can be difficult to analyze and to predict over long periods. Under this approach, the equation for estimating potential GDP is

$$\begin{aligned}\text{Growth rate in potential GDP} &= \text{Long-term growth rate of labor force}\\ &\quad + \text{Long-term growth rate in labor productivity}\end{aligned} \tag{5}$$

Thus, potential GDP growth is a combination of the long-term growth rate of the labor force and the long-term growth rate of labor productivity. If the labor force is growing at 1% per year and productivity per worker is rising at 2% per year, then potential GDP is rising at 3% per year.

4.3 Extending the Production Function

As a simplification, the production function in Equation 1 focused on only the labor and capital inputs. A more complete specification of the production function expands the list of inputs to include the following:

- Raw materials: natural resources such as oil, lumber, and available land (N)
- Quantity of labor: the number of workers in the country (L)
- Human capital: education and skill level of these workers (H)
- Information, computer, and telecommunications (ICT) capital: computer hardware, software, and communication equipment (K_{IT})
- Non-ICT capital: transport equipment, metal products and plant machinery other than computer hardware and communications equipment, and non-residential buildings and other structures (K_{NT})
- Public capital: infrastructure owned and provided by the government (K_P)
- Technological knowledge: the production methods used to convert inputs into final products, reflected by total factor productivity (A)

The expanded production function is expressed mathematically as

$Y = AF(N,L,H,K_{IT},K_{NT},K_P)$.

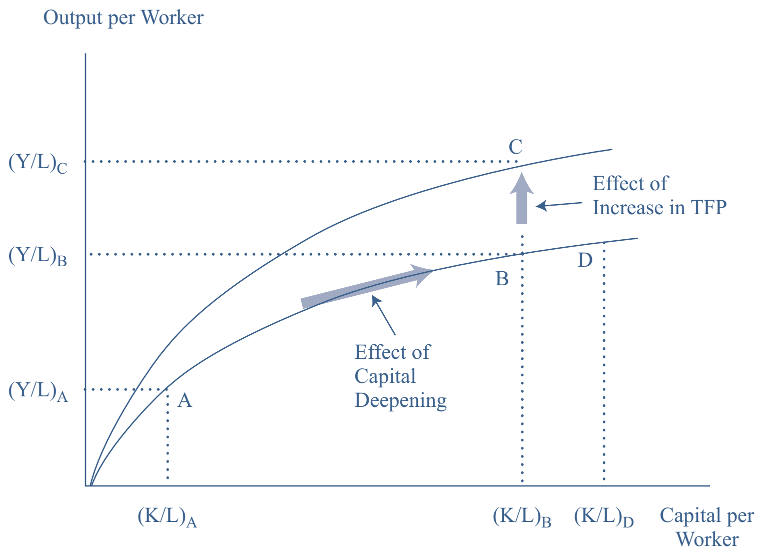
The impact of each of these inputs on economic growth is addressed in the following sections.

5 CAPITAL DEEPENING VS. TECHNOLOGICAL PROGRESS

d contrast capital deepening investment and technological progress and explain how each affects economic growth and labor productivity;

The property of diminishing marginal returns plays an important role in assessing the contribution of capital and technology to economic growth. Exhibit 4 shows the relationship between per capita output and the capital-to-labor ratio. It shows that adding more and more capital to a fixed number of workers increases per capita output but at a decreasing rate. Looking at Equation 3 and Exhibit 4, we can think of growth in per capita output coming from two sources: capital deepening and an improvement in technology, often referred to as technological progress.

Exhibit 4 Per Capita Production Function Capital Deepening vs. Technological (TFP) Progress



Capital deepening, an increase in the capital-to-labor ratio, is reflected in the exhibit by a move along the production function from point A to point B. The increase in the capital-to-labor ratio reflects rising investment in the economy. The ratio will increase as long as the growth rate of capital (net investment) exceeds the growth rate of labor. Once the capital-to-labor ratio becomes very high, however, as at point B, further additions to capital have relatively little impact on per capita output (e.g., moving to point D). This dynamic occurs because the marginal product of capital declines as more capital is added to the labor input.

At the point at which the marginal product of capital equals its marginal cost, profit maximizing producers will stop adding capital (i.e., stop increasing the capital-to-labor ratio). As we will discuss later, this point is very significant in the neoclassical model of growth because per capita growth in the economy will come to a halt. Once the economy reaches this steady state, capital deepening cannot be a source of sustained growth in the economy. Only when the economy is operating below the steady state and when the marginal product of capital exceeds its marginal cost can capital deepening raise per capita growth. Note that once technological progress (TFP growth) is introduced, the capital-to-labor ratio will have to keep increasing just to keep the marginal productivity of capital equal to its marginal cost. But the point remains: Once that equality is attained, companies will not increase the capital-to-labor ratio faster than is necessary to maintain that equality.

The neoclassical model's stark implication that more-rapid capital accumulation—that is, higher rates of investment—cannot result in a permanently higher rate of per capita growth is somewhat disappointing. As we will see in our discussion of endogenous growth, capital accumulation can result in a permanently higher growth rate if the investment results not just in *more* capital (i.e., pure capital deepening) but also in new, innovative products and processes. That is, if the additional capital embodies new, more efficient methods of production or previously unavailable products, then more rapid capital accumulation can result in a permanently higher growth rate of per capita output.

In contrast to moves along a given production function, an improvement in TFP causes a proportional upward shift in the entire production function. As a result, the economy can produce higher output per worker for a given level of capital per worker. This dynamic is shown in Exhibit 4 by the move from B to C. Technological progress also increases the marginal product of capital relative to its marginal cost. This increase makes additional capital investments profitable and tends to mitigate the limits imposed on growth by diminishing marginal returns. In addition, continued growth in per capita output is possible even in the steady state as long as there is ongoing technological progress (increases in TFP). In summary, *sustained growth in per capita output requires progress in TFP.*

EXAMPLE 3

Capital Deepening vs. Technological Progress

One of main differences between developed and developing countries is the amount of capital available for each worker. Country A is an advanced economy with $100,000 of capital available for each worker and thus a high capital-to-labor ratio. In contrast, Country B is a developing country with only $5,000 of capital available for each worker. What impact will the following developments have on the growth rate of potential GDP?

1. An increase in business investment in both countries
2. An increase in the amount of spending on university research in both countries
3. An elimination of restrictions in Country B on the inflow of foreign investment

Solution to 1:

An increase in business investment will raise the capital-to-labor ratio in both countries. It results in capital deepening and a movement along the per worker production function. The impact on growth, however, will be significantly different for the two countries. Country B will experience an increase in output per worker and thus in the growth rate of potential GDP. This is because Country B operates at a low level of capital per worker, at a point like A in Exhibit 4. Diminishing returns to capital are small, so any addition to capital has a major impact on growth. Country A operates at a point like B in Exhibit 4, so additions to capital have little impact on growth because of diminishing returns.

Solution to 2:

An increase in spending on university research will increase TFP and cause an upward shift in the production function in both countries. This can be seen in the move from point B to point C in Exhibit 4. The shift in the production function will raise growth in both countries and offset the negative impact of diminishing returns. This result shows that developing countries have the potential to grow through both capital deepening and technological progress, whereas improvement in potential GDP growth in developed countries is largely driven by technological progress.

Solution to 3:

The elimination of restrictions will result in higher foreign investment, which has the same impact as an increase in domestic business investment. This is again a movement along the production function such as from point A to B in Exhibit 4. With diminishing returns insignificant at low levels of capital to labor, the higher level of foreign investment will boost growth of potential GDP in Country B.

6 NATURAL RESOURCES

f explain how natural resources affect economic growth and evaluate the argument that limited availability of natural resources constrains economic growth;

Raw materials, including everything from available land to oil to water, are an essential input to growth. There are two categories of natural resources:

1 **Renewable resources** are those that can be replenished, such as a forest. For example, if a tree is cut, a seedling can be planted and a new forest harvested in the future.

2 **Non-renewable resources** are finite resources that are depleted once they are consumed. Oil and coal are examples.

Although it seems intuitive that countries with more natural resources will be wealthier, the relation between resource endowment and growth is not so straightforward. Natural resources do account for some of the differences in growth among countries. Today, Middle Eastern countries and such countries as Brazil and Australia have relatively high per capita incomes because of their resource base. Countries in the Middle East have large pools of oil. Brazil has an abundance of land suitable for large-scale agricultural production, allowing it to be a major exporter of coffee, soybeans, and beef.

Even though *access* to natural resources (e.g., via trade) is important, *ownership and production of natural resources is not necessary for a country to achieve a high level of income*. Countries in East Asia, such as South Korea, have experienced rapid economic growth but have few natural resources. In contrast, both Venezuela and Saudi Arabia have large oil reserves and are major producers of oil, yet both countries have experienced subpar growth compared with the natural-resource-poor countries of Singapore and South Korea. As was examined earlier, economic growth in Venezuela over the last 60 years was well below that of Singapore, Japan, and South Korea.

For some countries, the presence of natural resources may even restrain growth, resulting in a "resource curse." Venezuela and Nigeria are two examples of countries blessed with resources yet with sluggish economic growth. There are two main reasons why this may occur. First, countries rich in natural resources may fail to develop the economic institutions necessary for growth. Second, countries rich in resources may suffer the **Dutch disease**, where currency appreciation driven by strong export demand for resources makes other segments of the economy, in particular manufacturing, globally uncompetitive. The name for this phenomenon comes from the experience of the Netherlands: Following the discovery of large natural gas fields in the Netherlands, the Dutch guilder (the nation's currency at that time) appreciated and the manufacturing sector contracted. Because of this contraction, the resource-rich country does not participate in the TFP progress that occurs in countries with more vigorous manufacturing sectors.

In contrast, there is a longstanding concern that non-renewable natural resources will eventually limit growth. The idea is that a combination of rapid economic growth and a fixed stock of resources will cause resource depletion as the available pool of resources is used up. These concerns are probably overstated. Technological progress (TFP from all sources) enables the economy to use fewer resources per unit of output and to develop substitutes. The growing scarcity of specific resources will increase their price and encourage a shift toward more plentiful substitutes. Finally, the share of national income going to land and resources has been declining for most countries, especially as the composition of output in the global economy shifts toward the use of more services.

EXAMPLE 4

Impact of Natural Resources

The table below shows the share of world proved oil reserves as of 1990 for a selection of countries, along with the growth rate of real per capita GDP from 1990 to 2018. The simple correlation between the share of oil reserves and subsequent growth is not statistically different from zero.

	Percentage of World Proved Oil Reserves, 1990	Avg. Real Per Capita GDP Growth (%) 1990–2018		Percentage of World Proved Oil Reserves, 1990	Avg. Real Per Capita GDP Growth (%) 1990–2018
Saudi Arabia	25.75	0.4	Germany	0.04	1.4
Venezuela	5.85	2.8	France	0.02	1.1
Mexico	5.64	1.3	New Zealand	0.01	1.6
United States	2.62	1.5	Pakistan	0.01	2.1
China	2.40	9.1	Japan	0.01	0.9
Nigeria	1.60	2.1	Spain	0.00	1.6
Indonesia	0.82	3.7			
India	0.75	5.2	Botswana	0.00	2.5
Canada	0.61	1.3	Ethiopia	0.00	4.6
Egypt	0.45	2.1	Ireland	0.00	4.6
United Kingdom	0.43	1.5	Kenya	0.00	0.9
Brazil	0.28	1.1	Singapore	0.00	3.7
Argentina	0.23	2.0	South Africa	0.00	2.7
Australia	0.17	1.8	South Korea	0.00	4.4

Sources: US Energy Information Administration (www.eia.gov) and IMF.

What might account for the fact that real per capita GDP growth appears to be unrelated to oil reserves, perhaps the single most economically important natural resource (aside from water)?

Solution:

Energy is a vital input for any economy. Thus, *access* to energy resources is critical. *Ownership* of raw energy resources, however, is not. Countries that are not self-sufficient in oil or other resources acquire what they need through trade. It should be noted that countries that lack oil may possess other types of energy resources, such as natural gas, coal, hydropower, or geothermal energy.

In addition, countries can grow by emphasizing less energy intensive products, especially services, and adopting more energy efficient production methods. In sum, natural resources are important but not necessary for growth.

7 LABOR SUPPLY

g explain how demographics, immigration, and labor force participation affect the rate and sustainability of economic growth;

As noted earlier, economic growth is affected by increases in inputs, mainly labor and capital. Growth in the number of people available for work (quantity of workforce) is an important source of economic growth and partially accounts for the superior growth performance of the United States among the advanced economies—in particular, relative to Europe and Japan. Most developing countries, such as China, India, and Mexico, have a large potential labor supply. We can measure the potential size of the labor input as the total number of hours available for work. This, in turn, equals the labor force times the average hours worked per worker. The **labor force** is defined as the working age population (ages 16 to 64) that is either employed or available for work but not working (i.e., unemployed). Thus, growth in the labor input depends on four factors: population growth, labor force participation, net migration, and average hours worked.

7.1 Population Growth

Long-term projections of the labor supply are largely determined by the growth of the working age population. Population growth is determined by fertility rates and mortality rates. Population growth rates are significantly lower in the developed countries than in the developing countries. As a result, there is an ongoing decline in the developed countries' share of the world's population. Note that although population growth may increase the growth rate of the overall economy, it has no impact on the rate of increase in *per capita* GDP.

The age mix of the population is also important. The percentage of the population over the age of 65 and the percentage below the age of 16 are key considerations. Some of the developed countries, especially European countries, Japan, and South Korea, are facing a growing demographic burden as the portion of non-working elders (over 65) grows as a share of the population. In contrast, growth in many developing countries will receive a demographic boost as the fraction of the population below the age of 16 begins to decline. Interestingly, China is similar to the advanced economies, with a growing proportion of the population over age 65.

Exhibit 5 Population Data for Selected Countries (in millions, except growth rate)

	2000	2009	2018	Annual Growth (%), 2000–2018
France	59.1	64.4	66.9	0.69
Germany	82.2	81.9	82.9	0.05
Ireland	3.8	4.5	4.9	1.42
Spain	40.3	46.4	46.7	0.82

Exhibit 5 (Continued)

	2000	2009	2018	Annual Growth (%), 2000–2018
United Kingdom	58.9	62.3	66.4	0.67
Russia	146.7	141.9	144.5	–0.08
Japan	126.9	128.0	126.4	-0.02
United States	282.2	306.8	327.2	0.83
Mexico	98.4	112.1	125.3	1.35
China	1,267.4	1,352.1	1,415.0	0.61
India	1,024.3	1,214.3	1,354.1	1.56

Source: OECD.Stat.

7.2 Labor Force Participation

In the short run, the labor force growth rate may differ from population growth because of changes in the participation rate. The **labor force participation rate** is defined as the percentage of the working age population in the labor force. It has trended upward in most countries over the last few decades because of rising participation rates among women. In contrast to population, an increase in the participation rate may raise the growth of per capita GDP. In many southern European countries, such as Greece and Italy, the participation rate among women is well below the rates in the United States and northern European countries (see Exhibit 6). Thus, rising participation rates among women in these countries could increase growth in the labor force and in potential GDP. This has been the case for Spain, where the female labor force participation rate rose from 52.0% in 2000 to 67.9% in 2018. It should be noted, however, that rising or falling labor force participation is likely to represent a transition to a new higher or lower level of participation rather than a truly permanent rate of change. Thus, although trends in participation may contribute to or detract from potential growth for substantial periods, one should be cautious in extrapolating such trends indefinitely.

Exhibit 6 Labor Force Data for Selected Countries (2018)

	Percentage of Population under Age 15	Percentage of Population over Age 65	Participation Rate: Male	Participation Rate: Female
France	18.0%	19.9%	75.4%	69.7%
Germany	13.5	21.5	83.1	76.5
Greece	14.4	21.8	75.6	59.7
Ireland	20.8	13.6	77.3	66.6
Italy	13.2	22.7	73.9	55.6
Spain	14.5	19.3	77.7	67.9
Sweden	17.8	19.9	84.9	83.2
United Kingdom	17.9	18.3	81.9	74.2
Japan	12.2	28.1	71.2	51.4

(continued)

Exhibit 6 (Continued)

	Percentage of Population under Age 15	Percentage of Population over Age 65	Participation Rate: Male	Participation Rate: Female
United States	18.6	16.0	76.2	68.3
Mexico	26.5	7.2	81.8	47.3
Turkey	23.5	8.6	78.6	38.4

Source: OECD Stat Extracts.

EXAMPLE 5

Impact of the Age Distribution on Growth: Mexico vs. Germany

Exhibits 5 and 6 provide population data for selected countries. The data show that the rate of population growth and the age composition vary significantly among countries. Thus, demographic factors can be expected to have a significant impact on relative growth rates across countries. This effect is very clear in the cases of Mexico and Germany. There was essentially zero growth in Germany's population from 2000 to 2018, while Mexico's population increased by 1.35% annually. The age composition of the two countries is also very different. How will the age distribution impact growth over the next decade?

Solution:

What is important for growth is the number of workers available to enter the workforce. Over the next decade, Mexico will receive a demographic benefit because of the high percentage of young people entering the workforce. This is because 26.5% of the population in 2018 was below the age of 15. In contrast, only 13.5% of the German population was below the age of 15. In addition, Germany is facing a demographic challenge given the high and growing share of its population over the age of 65. In Mexico, only 7.2% of the population is above the age of 65, compared with 21.5% in Germany. In sum, the lack of population growth and a rapidly aging population in Germany will limit its potential rate of growth. Germany must rely on high labor productivity growth, increase its workforce participation rate, or encourage immigration if it is to increase its near-term potential rate of growth. Meanwhile, potential GDP growth in Mexico should receive a boost from its favorable population trends.

7.3 Net Migration

Another factor increasing economic and population growth, especially among the developed countries, is immigration. Heightened immigration is a possible solution to the slowing labor force growth being experienced by many developed countries with low birthrates within the native population. The growth rate of the labor force in Ireland, Spain, the United Kingdom, and the United States has increased between 2000 and 2010 because of immigration, although it slowed substantially in the 2010–2018 period. Focusing on the decade starting in 2000, Exhibit 5 shows the population growth rates for Ireland and Spain at 1.71% and 1.35%, respectively. The population growth rates were well above the population growth rates in other European countries. As shown in Exhibit 7, this is because of the impact of immigration. The open-border policies of both countries led to a significant population of immigrants that contributed to a large increase in labor input for both countries. As a consequence, both countries experienced GDP growth above the European average during this period (see Exhibit 1).

Exhibit 7 Ireland and Spain: Net Migration

	2000–2007	2008	2009	2010	Total 2000–2010	Total 2011–2016
Ireland	357,085	38,502	−7,800	−12,200	375,587	186,724
Spain	4,222,813	460,221	181,073	111,249	4,975,356	1,243,375

Source: OECD Stat Extracts

EXAMPLE 6

Potential Growth in Spain: Labor Input

The scenario below is set in early 2011. The Investment Policy Committee of Global Invest Inc. reviewed a report on the growth prospects for Spain and noted that, with total hours worked growing at a 1.2% annual rate between 2000 and 2010, labor input had been a major source of growth for the economy. Some members expect the growth rate of labor to slow considerably given projection from the OECD and IMF that immigration into Spain will fall to essentially zero over the next few years. A research assistant at the firm gathered demographic data on Spain from Exhibits 5–7 and other sources. The data are presented in the following table:

	2000	2010	Annual Growth (2000–2010)
Population (millions)	40.3	46.1	1.35%
Immigration since 2000 (millions)		4.975	
Percentage of population under 15		15.0%	
Percentage of population over 65		17.0%	
Male labor force participation rate		80.4%	
Female labor force participation rate		66.1%	
Unemployment rate		20.1%	

Using this information for Spain and Exhibits 5 and 6 for relevant comparison data, determine the following:

1 Whether a change in the trend growth rate of the labor input is likely over the next few years.

2 How the high unemployment rate of 20.1% is likely to affect the growth rate of the labor force.

Solution to 1:

The growth in the labor input depends on a number of factors, including the population growth rate, the labor force participation rate, and the percentage of the population below the age of 15. The labor force in Spain expanded sharply between 2000 and 2010 mainly because of a 5.8 million person increase in the population, going from 40.3 million in 2000 to 46.1 million in 2010. Looking ahead, growth in the labor force is set to slow substantially for a number of reasons:

- The population increase between 2000 and 2010 is very misleading because it is not likely to be repeated in the future. Between 2000 and 2010, immigration raised the population of Spain by nearly 5 million people. Without the immigrants, the population would have grown by only about 825,000 people during this period, or at an annual rate of 0.2%. With immigration, the population growth rate was 1.35%. The pace of immigration that occurred between 2000 and 2010 is not sustainable and is likely to slow, which will result in slower growth in both the population and the labor force.
- In the short run, the growth rate of the labor force may differ from population growth because of changes in the participation rate. Looking at the data, the male participation rate in Spain, at 80.4%, is very high and, as shown in Exhibit 6, is above the male participation rates in France, Greece, and Italy and slightly below that of Germany. The female participation rate is low in comparison to northern European countries, such as Sweden. But it is higher than in Italy, which is probably a better comparison. Thus, little increase is likely in the male or female participation rates.
- Only 15% of the Spanish population is below the age of 15. The comparable figure from Exhibit 6 for the United Kingdom is 17.9%, for France 18.0%, for the United States 18.6%, and for Mexico 26.5%. Thus, Spain does not appear poised for a notable surge in young adults entering the labor force.

In summary, growth in the labor input in Spain should slow over the next few years, and the growth rate of potential GDP should do the same.

Solution to 2:

Reducing the unemployment rate would mitigate some of the negative demographic factors because a reduction in the number of unemployed workers would boost utilization of the existing labor supply. This shift would represent a transition to a higher level of employment rather than a permanent increase in the potential growth rate. Nonetheless, it could boost potential growth for a substantial period.

7.4 Average Hours Worked

The contribution of labor to overall output is also affected by changes in the average hours worked per worker. Average hours worked is highly sensitive to the business cycle. The long-term trend in average hours worked, however, has been toward a shorter work week in the advanced countries. This development is the result of legislation, collective bargaining agreements, the growth of part-time and temporary work, and the impact of both the “wealth effect” and high tax rates on labor income, which cause workers in high-income countries to value leisure time relatively more highly than labor income.

Exhibit 8 provides data on average hours worked per year per person in the labor force for selected years since 1995. For most countries, the average number of hours worked per year has been declining. There is also a significant difference in hours worked across countries. In 2018, average hours worked per year in South Korea, at 1,993 hours, were 46.1% more than the 1,363 average hours worked per year in Germany. The increase in female labor force participation rates may be contributing to the shorter average workweek because female workers disproportionately take on part-time, rather than full-time, jobs.

Exhibit 8 Average Hours Worked per Year per Person in Selected Countries

	1995	2005	2018
France	1,651	1,559	1,520
Germany	1,534	1,435	1,363
Greece	2,123	2,081	1,956
Ireland	1,875	1,654	1,782
Italy	1,859	1,819	1,723
Spain	1,733	1,688	1,701
Sweden	1,609	1,607	1,474
United Kingdom	1,743	1,676	1,538
Japan	1,884	1,775	1,680
South Korea	2,658	2,364	1,993
Canada	1,761	1,738	1,708
United States	1,840	1,795	1,786
Mexico	1,857	1,909	2,148
Turkey	1,876	1,918	1,832

Source: OECD data.

7.5 Labor Quality: Human Capital

In addition to the quantity of labor, the quality of the labor force is an important source of growth for an economy. **Human capital** is the accumulated knowledge and skills that workers acquire from education, training, or life experience. In general, better-educated and more-skilled workers will be more productive and more adaptable to changes in technology or other shifts in market demand and supply.

An economy's human capital is increased through investment in education and on-the-job training. Like physical capital, investment in education is costly, but studies show that there is a significant return on that investment. That is, people with more education earn higher wages. In addition, education may also have a spillover or externality impact. Increasing the educational level of one person raises not only the output of that person but also the output of those around that person. The spillover effect operates through the link between education and advances in technology. Education not only improves the quality of the labor force, and thus the stock of human capital, but also encourages growth through innovation. Importantly, increased education, obtained both formally and via on-the-job training, could result in a permanent increase in the growth rate of an economy if the more educated workforce results in more innovations and a faster rate of technological progress. Investment in the population's health is also a major contributor to human capital, especially in developing countries.

8 CAPITAL: ICT & NON-ICT, TECHNOLOGY AND PUBLIC INFRASTRUCTURE

h explain how investment in physical capital, human capital, and technological development affects economic growth;

The physical capital stock increases from year to year as long as net investment (gross investment less the depreciation of the capital) is positive. Thus, countries with a higher rate of investment should have a growing physical capital stock and a higher rate of GDP growth. Note that the impact on growth of per capita GDP will be somewhat smaller if the population is growing because a proportion of net investment simply provides the capital needed to maintain the capital-to-labor ratio. Exhibit 9 shows the level of gross non-residential investment as a share of GDP. The exhibit shows significant variation across countries, with the investment share in the United States being low in comparison to other developed countries.

Exhibit 9 Business Investment as a Percentage of GDP

	Investment Percentage of GDP		
	2000	**2008**	**2018**
Developed Countries			
France	19.5	24.1	22.8
Germany	21.5	20.9	21.2
Ireland	23.9	24.4	24.5
Italy	20.3	21.8	18.0
Spain	26.2	29.6	21.9
United Kingdom	17.1	17.2	17.2
Australia	22.0	28.4	24.2
Japan	25.4	24.5	24.4
South Korea	30.6	33.0	30.2
Singapore	33.1	30.5	27.0
Canada	19.2	24.1	23.0
United States	19.9	21.1	21.1
Developing Countries			
Brazil	18.3	21.9	15.4
China	35.1	47.9	44.2
India	24.3	36.5	31.6
Mexico	25.5	22.8	23.0
South Africa	15.1	19.5	17.9

Source: IMF.

The correlation between economic growth and investment is high. Countries that devote a large share of GDP to investment, such as China, India, and South Korea, have high growth rates. Some of the fastest-growing countries in Europe in the 1990s and for long periods since the year 2000, including Ireland and Spain, have the some of the highest investment-to-GDP ratios. Countries that devote a smaller share of GDP to investment, such as Brazil and Mexico, have slower growth rates. The data show why the Chinese economy has expanded at such a rapid rate: annual GDP growth rate in excess of 10% over long periods. Investment spending in China on new factories, equipment, and infrastructure as a percentage of GDP is the highest in the world, at more than 40% of GDP.

As we discussed earlier, long-term sustainable growth cannot rely on pure capital deepening. How can we reconcile this notion with the strong correlation between investment spending and economic growth across countries? First, although diminishing marginal productivity will eventually limit the impact of capital deepening, investment-driven growth may last for a considerable period, especially in countries that start with relatively low levels of capital per worker.

A second, and closely related, explanation is that the impact of investment spending on available capital depends on the existing physical capital stock. As with the share of GDP devoted to investment, the stock of capital available per worker varies significantly across countries. In 2000, the average US worker had $148,091 worth of capital, compared with $42,991 in Mexico and $6,270 in India (Heston, Summers, and Aten 2009). The wide difference in physical capital per worker suggests that the positive impact of changes in the physical capital stock on growth is very significant in developing countries. Mexican workers have relatively little access to machinery or equipment, so adding even a little can make a big percentage difference. In developed countries, such as the United States, Japan, Germany, France, and the United Kingdom, the physical capital stock is so large that positive net investment in any given year has only a small percentage effect on the accumulated capital stock. For the developed countries, a sustained high level of investment over many years is required to have a meaningful relative impact on the physical capital stock even though the absolute size of the increase in any given year is still larger than in the developing countries.

Third, because physical capital is not really homogeneous, the composition of investment spending and the stock of physical capital matters for growth and productivity. Insights obtained from the endogenous theory of growth (discussed later) and from studies attempting to obtain a more accurate measure of TFP show that the composition of the physical capital stock is very important. These studies suggest that capital spending could be separated into two categories. The first is spending on information, computers, and telecommunications equipment (ICT investment). Capital spending on these goods is a measure of the impact of the information technology sector on economic growth. One of the key drivers of growth in the developed countries over the last decade has been the IT sector. Growth in the IT sector has been driven by technological innovation that has caused the price of key technologies, such as semiconductors, to fall dramatically. The steep decline in the price of high-technology capital goods has encouraged investment in IT at the expense of other assets.

The IT sector has grown very rapidly and has made a significant contribution to increasing the rate of economic and productivity growth. The greater use of IT equipment in various industries has resulted in **network externalities**. Computers allow people to interconnect through the internet and by email, enabling them to work more productively. *The more people in the network, the greater the potential productivity gains.* The effects of the network externalities are largely captured in TFP rather than observed as a distinct, direct effect. The share of ICT investment in GDP tends to be in the 3%–5% range for most developed economies. The IT sector is still relatively small in most countries, and IT spending actually declined as a share of GDP between 2000 and 2008 because the early 2000s recession disproportionately affected high-technology spending.

The other category of investment, non-ICT capital spending, includes non-residential construction, transport equipment, and machinery. High levels of capital spending for this category should eventually result in capital deepening and thus have less impact on potential GDP growth. In contrast, a growing share of ICT investments in the economy, through their externality impacts, may actually boost the growth rate of potential GDP.

It is worthwhile to note that there have been important "transformational technologies" at various stages of history. One need only consider the impact of the steam engine, the internal combustion engine, powered flight, atomic energy, vaccination, and so on, to realize that revolutionary advances are not unique to information, computers, and telecommunications. All of these are, to some extent, "general purpose technologies" (GPT) that affect production and/or innovation in many sectors of the economy. ICT capital clearly embodies this GPT characteristic. Nanotechnology could well become the next "super GPT," at which point investing in ICT may begin to look like mere capital deepening.

8.1 Technology

The most important factor affecting growth of per capita GDP is technology, especially in developed countries. Technology allows an economy to overcome some of the limits imposed by diminishing marginal returns and results in an upward shift in the production function, as we noted in Exhibit 4. Technological progress makes it possible to produce more and/or higher-quality goods and services with the same resources or inputs. It also results in the creation of new goods and services. Technological progress can also be one of the factors improving how efficiently businesses are organized and managed.

Technological change can be embodied in human capital (knowledge, organization, information, and experience base) and/or in new machinery, equipment, and software. Therefore, high rates of investment are important, especially investment in ICT goods. Countries can also innovate through expenditures, both public and private, on research and development (R&D). Expenditures on R&D and the number of patents issued, although not directly measuring innovation, provide some useful insight into innovative performance. Exhibit 10 shows R&D spending as a share of GDP for various countries. The developed countries spend the highest percentage of GDP on R&D because they must rely on innovation and the development of new products and production methods for growth. In contrast, developing countries spend less on R&D because these countries can acquire new technology through imitation or copying the technology developed elsewhere. The embodiment of technology in capital goods can enable relatively poor countries to narrow the gap relative to the technology leaders. It should also be noted that the relationship between economic growth and R&D spending is not clear-cut. Although technological innovation resulting from high R&D spending raises output and productivity in the long run, it may result in a cyclical slowing of growth as companies and workers are displaced by the new technologies. This is the Schumpeterian concept of creative destruction, which captures the double-edged nature of technological innovation.

Exhibit 10 Research and Development as a Percentage of GDP in Selected Countries

	1990	2009	2016
France	2.3	2.2	2.2
Germany	2.6	2.8	2.9
Ireland	0.8	1.8	1.2

Exhibit 10 (Continued)

	1990	2009	2016
Italy	1.2	1.3	1.3
Spain	0.8	1.4	1.2
United Kingdom	2.1	1.9	1.7
Australia	1.3	2.2	1.9
Japan	3.0	3.4	3.1
South Korea	1.7	3.1	4.2
Singapore	1.1	2.9	2.2
Canada	1.5	2.0	1.6
United States	2.6	2.9	2.7
China	NA	1.7	2.1
India	NA	0.8	0.8
Mexico	NA	0.4	0.5

Source: OECD.

The state of technology, as reflected by total factor productivity, embodies the cumulative effects of scientific advances, applied research and development, improvements in management methods, and ways of organizing production that raise the productive capacity of factories and offices. Because it is measured as a residual, TFP estimates are very sensitive to the measurements of the labor and capital inputs. Empirical work at the Conference Board and the OECD accounts for changes in the composition and quality of both the labor and capital inputs. The resulting measure of TFP should capture the technological and organizational improvements that increase output for a given level of inputs. Exhibit 11 provides data for the periods 1995–2005 and 2005–2018 on the growth rate in labor productivity and total factor productivity. Labor productivity growth depends on both capital deepening and technological progress. The contribution of capital deepening can be measured as the difference between the growth rates of labor productivity and total factor productivity. For example, from 2005 to 2018, South Korea's labor productivity grew by 3.3% per year, of which 2.5% (3.3% − 0.8%) came from capital deepening, with the rest coming from changes in TFP (note that rounding causes minor discrepancies in the calculations in the exhibit). The larger the difference between the productivity growth measures, the more important capital deepening is as a source of economic growth. As we discussed previously, however, growth in per capita income cannot be sustained perpetually by capital deepening.

Exhibit 11 Labor and Total Factor Productivity

	Growth in Hours Worked (%)	Growth in Labor Prod. (%)	Growth in TFP (%)	Growth from Capital Deepening (%)	Growth in GDP (%)	Productivity Level 2018; GDP per Hour Worked ($)
Germany						70
1995–2005	−0.3	1.6	0.9	0.7	1.3	
2005–2018	0.8	0.8	0.2	0.7	1.6	
Ireland						84
1995–2005	3.2	4.1	1.7	2.4	7.3	

(continued)

Exhibit 11 (Continued)

	Growth in Hours Worked (%)	Growth in Labor Prod. (%)	Growth in TFP (%)	Growth from Capital Deepening (%)	Growth in GDP (%)	Productivity Level 2018; GDP per Hour Worked ($)
2005–2018	0.6	2.9	0.1	2.7	3.3	
United States						73
1995–2005	0.9	2.4	0.9	1.5	3.3	
2005–2018	0.9	1.2	0.0	1.3	1.9	
Japan						47
1995–2005	–1.0	2.1	0.4	1.7	1.1	
2005–2018	0.0	1.0	–0.1	1.1	1.0	
South Korea						39
1995–2005	0.0	4.3	2.4	1.9	4.3	
2005–2018	0.1	3.3	0.8	2.5	3.3	
China						15
1995–2005	1.1	6.7	1.5	5.2	7.8	
2005–2018	0.2	9.2	4.3	5.0	9.0	
India						9
1995–2005	2.1	4.2	1.9	2.3	6.3	
2005–2018	1.2	6.3	1.8	4.5	7.2	
Brazil						19
1995–2005	2.1	0.3	–0.3	0.6	2.4	
2005–2018	0.8	1.3	–0.7	1.9	2.0	
Mexico						21
1995–2005	2.2	1.4	0.4	1.0	3.6	
2005–2018	2.1	0.1	–0.2	0.4	2.2	

Source: Conference Board Total Economy Database.

Exhibit 11 also provides data on the *level* of labor productivity or the amount of GDP produced per hour of work. The level of productivity depends on the accumulated stock of human and physical capital and is much higher among the developed countries. For example, China has a population of more than 1.3 billion people, compared with slightly more than 300 million people in the United States. Although the United States has significantly fewer workers than China because of its smaller population, its economy as measured by real GDP is much larger. This is because US workers have historically been more productive than Chinese workers as measured by GDP per hour worked, as shown in Exhibit 11. In contrast to the *level* of productivity, the *growth rate* of productivity will typically be higher in the developing countries, where human and physical capital are scarce but growing rapidly and the impact of diminishing marginal returns is relatively small.

An understanding of productivity trends is critical for global investors. A permanent increase in the rate of labor productivity growth will increase the sustainable rate of economic growth and raise the upper boundary for earnings growth and the potential return on equities. In contrast, a low growth rate of labor productivity, if it persists over a number of years, suggests poor prospects for equity prices. A slowdown in productivity growth lowers both the long-run potential growth rate of the economy and the upper limit for earnings growth. Such a development would be associated with slow growth in profits and correspondingly low equity returns.

EXAMPLE 7

Why the Sluggish Growth in the Japanese Economy?

Annual growth in real GDP in Japan averaged about 1% since 1990. This growth is in sharp contrast to the 4.2% annual growth rate experienced from 1971 to 1990. The sluggish growth in Japan should not be surprising. Japan's economy is growing at its potential rate of growth, which is limited by the following:

1 The labor input is not growing. Population growth has been essentially zero since 2000 (Exhibit 5), and average hours worked per year per person is declining (Exhibit 8).
2 There has been a lack of technological innovation. The lack of growth in the labor input could be offset through higher productivity derived from innovation and more efficient use of available inputs. However, this is not occurring in Japan. Total factor productivity (Exhibit 11) increased at a sluggish 0.4% annual rate from 1995 to 2005 and declined slightly between 2005 and 2018.
3 Diminishing returns to capital are very significant. Despite the negative growth in TFP, labor productivity growth remained relatively high. This means that all the growth in labor productivity in Japan resulted from capital deepening (Exhibit 11). The problem for Japan, as discussed in earlier, is that once the capital-to-labor ratio becomes high, further additions to capital have little impact on per capita output. Thus, the growth in labor productivity should slow.

Use the data for 2005–2018 and the labor productivity growth accounting equation to estimate the growth rate in potential GDP for Japan.

Solution:

To estimate the growth rate in potential GDP, we use Equation 5, given by

$$\text{Growth rate of potential GDP} = \text{Long-term growth rate of labor force} + \text{Long-term growth rate in labor productivity}$$

To use this equation, we need to project the growth rate in the labor input and labor productivity.

The hours worked data in Exhibit 11 are a potential source to use to estimate the growth rate of the labor input. Exhibit 11 shows the labor input for Japan unchanged between 2005 and 2018. This was partly caused by the negative impact of the global recession on hours worked. As an alternative, the labor input should grow at the same rate as the population plus the net change in immigration. The population data in Exhibit 5 show essentially zero population growth in Japan for the period 2000–2018. This trend is likely to continue. Thus, a reasonable estimate for potential GDP growth in Japan is around 1%. We get this estimate by assuming no growth in the labor input and a 1% annual increase in labor productivity (using data from Exhibit 11 for 2005–2018).

8.2 Public Infrastructure

The final expansion of the definition of the capital input is public infrastructure investment. Roads, bridges, municipal water, dams and, in some countries, electric grids are all examples of public capital. They have few substitutes and are largely complements to the production of private sector goods and services. Ashauer (1990) found that infrastructure investment is an important source of productivity growth and should be included as an input in the production function. As with R&D spending, the full impact of government infrastructure investment may extend well beyond the direct benefits of the projects because improvements in the economy's infrastructure generally boost the productivity of private investments.

9 SUMMARY OF ECONOMIC GROWTH DETERMINANTS

d contrast capital deepening investment and technological progress and explain how each affects economic growth and labor productivity;

e demonstrate forecasting potential GDP based on growth accounting relations;

g explain how demographics, immigration, and labor force participation affect the rate and sustainability of economic growth;

h explain how investment in physical capital, human capital, and technological development affects economic growth;

Long-term sustainable growth is determined by the rate of expansion of real potential GDP. Expansion of the supply of factors of production (inputs) and improvements in technology are the sources of growth. The factors of production include human capital, ICT and non-ICT capital, public capital, labor, and natural resources. Data for the sources of growth are available from the OECD and the Conference Board. Exhibit 12 provides data from the Conference Board on the sources of output growth for various countries. These estimates are based on the growth accounting formula.

Exhibit 12 Sources of Output Growth

	Contribution from:					
	Labor Quantity (%)	Labor Quality (%)	Non-ICT Capital (%)	ICT Capital (%)	TFP (%)	Growth in GDP (%)
Germany						
1995–2005	−0.2	0.1	0.3	0.2	0.9	1.3
2005–2018	0.4	0.1	0.6	0.3	0.2	1.6
Ireland						
1995–2005	2.0	0.3	2.6	0.7	1.7	7.3
2005–2018	0.0	0.3	2.5	0.3	0.1	3.3
United States						
1995–2005	0.6	0.3	0.7	0.8	0.9	3.3
2005–2018	0.4	0.3	0.7	0.5	0.0	1.9
Japan						
1995–2005	−0.6	0.4	0.6	0.3	0.4	1.1
2005–2018	0.0	0.3	0.5	0.3	−0.1	1.0

Exhibit 12 (Continued)

	Contribution from:					
	Labor Quantity (%)	Labor Quality (%)	Non-ICT Capital (%)	ICT Capital (%)	TFP (%)	Growth in GDP (%)
South Korea						
1995–2005	–0.5	0.8	1.1	0.5	2.4	4.3
2005–2018	0.0	0.1	1.9	0.5	0.8	3.3
China						
1995–2005	0.5	0.2	4.5	1.1	1.5	7.8
2005–2018	0.1	0.3	3.9	0.4	4.3	9.0
India						
1995–2005	1.0	0.2	2.7	0.5	1.9	6.3
2005–2018	0.7	0.6	3.4	0.8	1.8	7.2
Brazil						
1995–2005	0.8	0.1	1.1	0.7	–0.3	2.4
2005–2018	0.4	0.8	1.2	0.4	–0.7	2.0
Mexico						
1995–2005	1.2	0.2	1.4	0.4	0.4	3.6
2005–2018	1.0	0.1	1.1	0.2	–0.2	2.2

Source: Conference Board Total Economy Database.
Notes: A standard growth accounting model (expanded version of Equation 4) is used to compute the contribution of each input to aggregate output (GDP) growth. The inputs include both the quantity and quality of labor and ICT and non-ICT capital. Each input is weighted by its share in national income, and TFP captures all sources of growth that are left unexplained by the labor and capital inputs. Rounding is used throughout.

EXAMPLE 8

The Irish Economy

As shown in Exhibit 1, economic growth in Ireland since 1980 has been significantly higher than that experienced in the major European economies of Germany, France, and the United Kingdom. In 1970, the per capita GDP of Ireland, at $9,869, was 45.2% below the per capita GDP of the United Kingdom. By 2010, per capita GDP in Ireland caught up with or exceeded most other developed European countries. Like most of the global economy, Ireland fell into a deep recession in 2009, with GDP contracting by more than 7%, before staging a recovery and reaching annual growth of more than 5% for several years in the 2010–2018 period. To understand the factors driving the Irish economy and the prospects for future equity returns, use the data in Exhibits 11 and 12 and the following population data to address these questions:

1 Using the growth accounting framework data, evaluate the sources of growth for the Irish economy starting from 1995.

2 What is likely to happen to the potential rate of growth for Ireland? What are the prospects for equity returns?

	2000	2010	2016	Avg. Annual Growth Rate
Population (millions)	3.8	4.5	4.9	1.6%
Net immigration total (2000–2010)		0.38m		
Net immigration total (2011–2016)		0.19m		
Population less immigrants (millions)	3.8	4.1	4.3	0.8%

Solution to 1:

The sources of growth for an economy include labor quantity, labor quality, non-ICT capital, ICT capital, and TFP. The growth accounting data in Exhibit 12 indicate that economic growth in Ireland from 1995 to 2018 is explained by the following factors:

Input	Contribution: 1995–2005		Contribution: 2005–2018	
Labor		**2.3%**		**0.3%**
Labor quantity	2.0%		0.0%	
Labor quality	0.3%		0.3%	
Capital/Investment		**3.3%**		**2.8%**
Non-ICT capital	2.6%		2.5%	
ICT capital	0.7%		0.3%	
TFP		**1.7%**		**0.1%**
Total: GDP growth		**7.3%**		**3.3%**

In sum, the main driver of growth for the Irish economy since 1995 has been capital spending. It accounted for more than 45% of growth in 1995–2005 and has been the dominant factor contributing to growth in the Irish economy since 2005. Another way to look at growth in Ireland for the period 2005–2018 is that all the growth is through capital deepening. As shown in Exhibit 11, capital deepening added 2.7% to growth, which caused an increase in labor productivity of 2.9%.

Solution to 2:

If we look forward, prospects for the economy are not as favorable as in the past. To estimate the growth rate in potential GDP, we use Equation 5, given by

$$\begin{aligned}\text{Growth rate of potential GDP} &= \text{Long-term growth rate of labor force}\\ &\quad + \text{Long-term growth rate in labor}\\ &\quad \text{productivity}\end{aligned}$$

To use this equation, we need to project the growth rate in the labor input and labor productivity. The total hours worked data in Exhibit 11 are one potential source to use to estimate the growth rate of the labor input. Exhibit 11 shows the labor input increasing by 0.6% annually between 2005 and 2018. The problem here is that the decline in hours worked is overstated because of the negative impact of the 2008–2009 recession on hours worked. As an alternative, the labor input should grow at the same rate as the population plus the net change resulting from immigration. The population data for Ireland (given above) show that more than half of the population growth between 2000 and 2010 resulted from immigration. Since 2009, however, outward migration has replaced inward migration for a short period, and the rate of growth in labor input declined. Thus, a more reasonable, perhaps somewhat conservative, estimate for labor force growth is 0.3%. We also assume the following:

1 There is no increase in labor productivity coming from capital deepening as investment slows (resulting in essentially no growth in net investment and the physical capital stock).
2 TFP growth reverts to its average growth rate of 1.7% in the 1995–2005 period (see Exhibit 11).
3 Labor productivity grows at the same rate as TFP.

Thus, growth in potential GDP is 0.3% + 1.7% = 2.0%.

In summary, despite the projected rebound in TFP growth, overall potential growth in Ireland is likely to decline because labor input growth and capital deepening no longer contribute to overall growth. As discussed earlier, slower growth in potential GDP will limit potential earnings growth and equity price appreciation.

EXAMPLE 9

Investment Outlook for China and India

The Investment Policy Committee at Global Invest Inc. is interested in increasing the firm's exposure to either India or China because of their rapid rates of economic growth. Economic growth in China has been close to 9% over the last few years, and India has grown more than 7%. You are asked by the committee to do the following:

1 Determine the sources of growth for the two economies and review the data on productivity and investment using information from Exhibits 5, 9, 10, 11, and 12. Which of the two countries looks more attractive based on the sources of growth?
2 Estimate the long-term sustainable earnings growth rate using data from 1995 to 2018.
3 Make an investment recommendation.

Solution to 1:

The sources of economic growth include size of labor force, quality of labor force (human capital), ICT and non-ICT capital, natural resources, and technology. Looking at the sources of growth in Exhibit 12, we determine the following:

Input	Percent Contribution: 1995–2005	Percent Contribution: 2005–2018
	India	
Labor quantity	1.0	0.7
Labor quality	0.2	0.6
Non-ICT capital	2.7	3.4
ICT capital	0.5	0.8

(continued)

Input	Percent Contribution: 1995–2005	Percent Contribution: 2005–2018
	India	
TFP	1.9	1.8
Total: GDP growth	6.3	7.2
	China	
Labor quantity	0.5	0.1
Labor quality	0.2	0.3
Non-ICT capital	4.5	3.9
ICT capital	1.1	0.4
TFP	1.5	4.3
Total: GDP growth	7.8	9.0

- The contribution of the labor quantity input is more important to growth in India than in China. Labor quantity contributed 1% to India's GDP growth over 1995–2005 and 0.7% over 2005–2018. The equivalent numbers for China are 0.5% and 0.1%, respectively. Looking ahead, we can project that labor is likely to be a major factor adding to India's growth. The population of India (Exhibit 5) is growing at a faster rate than that of China. The annual growth rate in population from 2005 to 2018 was 1.34% in India versus 0.50% in China. Also, hours worked in India (Exhibit 11) are growing at a faster rate than in China. Therefore, the workforce and labor quantity input should grow faster in India. The edge here goes to India.
- The contribution to GDP made by the quality of the labor force is essentially identical in the two countries (0.2% in China versus 0.2% in India between 1995 and 2005 and 0.3% in China and 0.6% in India between 2005 and 2018). This factor is stronger in India.
- The contribution of non-ICT capital investment is significantly higher in China (4.5% in China versus 2.7% in India between 1995 and 2005 and 3.9% in China and 3.4% in India between 2005 and 2018). The edge goes to China.
- The contribution of ICT capital investment was significantly higher in China (1.1% in China versus 0.5% in India between 1995 and 2005). Since 2005, it has contributed 0.4% to growth in China and 0.8% in India, which has an edge.
- Both countries spend a high percentage of GDP on capital investment (Exhibit 9). In 2018, investment spending as a percentage of GDP was 44% in China and 32% in India. The Chinese share is higher, and this provides China with an edge unless diminishing marginal returns to capital deepening become an issue. This scenario is not likely for a while, however, given the relatively low level of capital per worker in China. China and India still have a way to go to converge with the developed economies. The advantage goes to China.
- The contribution of technological progress is measured by TFP. Comparing the two countries, we find that TPF growth was higher in India over the period 1995–2005 (1.9% in India versus 1.5% in China). For the period 2005–2018, however, TFP growth was significantly higher

in China (4.3% versus 1.8%). In addition, expenditures on R&D for 2016 (Exhibit 10) as a percentage of GDP were higher in China (2.1% in China and 0.8% in India). The edge here goes to China.

- Finally, growth in overall labor productivity (Exhibit 11) is considerably higher in China than India (9.2% in China versus 6.3% in India between 2005 and 2018). This dynamic resulted from both a greater increase in the capital-to-labor ratio in China (because of the high rate of investment, the physical capital stock is growing faster than the labor input) and faster technological progress in China. The edge here goes to China.

In sum, based on the sources of growth, China appears to be slightly better positioned for growth in the future.

Solution to 2:

Estimates of potential GDP using the inputs from Exhibit 11 for China and India are

Growth rate in potential GDP = Long-term growth rate of labor force (equals growth in hours worked in Exhibit 11) + Long-term growth rate in labor productivity.

China (using 1995–2018)

Growth in potential GDP = 0.6% + 7.9% = 8.5% (calculated as geometric mean growth rates using data for the 1995–2005 and 2005–2018 subperiods).

India (using 1995–2018)

Growth in potential GDP = 1.6% + 5.2% = 6.8%

Solution to 3:

Growth prospects in both countries are very attractive. China's growth potential is higher, however, because of its greater level of capital spending and the greater contribution of technological progress toward growth. Long-term earnings growth is closely tied to the growth rate in potential GDP. Therefore, based on the previous calculations, earnings in China would be projected to grow at an annual rate of 8.5%, compared with 6.8% in India. Over the next decade, ignoring current valuation, the Chinese equity market would be projected to outperform the Indian market as its higher rate of sustainable growth translates into a higher rate of appreciation in equity values. Note that the global economy is evolving rapidly, and past trends may or may not be sustained. This is especially true of China and India. To provide concrete answers that do not require the reader to bring in additional information, our exercise solutions must assume that past patterns are indicative of the future.

10 THEORIES OF GROWTH, CLASSICAL AND NEOCLASSICAL ECONOMIC MODELS AND BALANCED OR STEADY STATE RATE OF GROWTH

i. compare classical growth theory, neoclassical growth theory, and endogenous growth theory;

The factors that drive long-term economic growth and determine the rate of sustainable growth in an economy are the subject of much debate among economists. The academic growth literature includes three main paradigms with respect to per capita growth in an economy—the classical, neoclassical, and endogenous growth models. Per capita economic growth under the classical model is only temporary because an exploding population with limited resources brings growth to an end. In the neoclassical model, long-run per capita growth depends solely on exogenous technological progress. The final model of growth attempts to explain technology within the model itself—thus the term endogenous growth.

10.1 Classical Model

Classical growth theory was developed by Thomas Malthus in his 1798 publication *Essay on the Principle of Population.* Commonly referred to as the Malthusian theory, it is focused on the impact of a growing population in a world with limited resources. The concerns of resource depletion and overpopulation are central themes within the Malthusian perspective on growth. The production function in the classical model is relatively simple and consists of a labor input with land as a fixed factor. The key assumption underlying the classical model is that population growth accelerates when the level of per capita income rises above the subsistence income, which is the minimum income needed to maintain life. This means that technological progress and land expansion, which increase labor productivity, translate into higher population growth. But because the labor input faces diminishing marginal returns, the additional output produced by the growing workforce eventually declines to zero. Ultimately, the population grows so much that labor productivity falls and per capita income returns back to the subsistence level.

The classical model predicts that in the long run, the adoption of new technology results in a larger but not richer population. Thus, the standard of living is constant over time even with technological progress, and there is no growth in per capita output. As a result of this gloomy forecast, economics was labeled the "dismal science."

The prediction from the Malthusian model failed for two reasons:

1 The link between per capita income and population broke down. In fact, as the growth of per capita income increased, population growth slowed rather than accelerating as predicted by the classical growth model.
2 Growth in per capita income has been possible because technological progress has been rapid enough to more than offset the impact of diminishing marginal returns.

Because the classical model's pessimistic prediction never materialized, economists changed the focus of the analysis away from labor to capital and to the neoclassical model.

10.2 Neoclassical Model

Robert Solow devised the mainstream neoclassical theory of growth in the 1950s (Solow 1957). The heart of this theory is the Cobb–Douglas production function discussed earlier. As before, the potential output of the economy is given by

$$Y = AF(K,L) = AK^{\alpha}L^{1-\alpha},$$

where K is the stock of capital, L is the labor input, and A is total factor productivity. In the neoclassical model, both capital and labor are variable inputs each subject to diminishing marginal productivity.

The objective of the neoclassical growth model is to determine the long-run growth rate of output per capita and relate it to (a) the savings/investment rate, (b) the rate of technological change, and (c) population growth.

10.2.1 *Balanced or Steady-State Rate of Growth*

As with most economic models, the neoclassical growth model attempts to find the equilibrium position toward which the economy will move. In the case of the Solow model, this equilibrium is the balanced or **steady-state rate of growth** that occurs when the output-to-capital ratio is constant. Growth is balanced in the sense that capital per worker and output per worker grow at the same rate.

We begin the analysis by using the per capita version of the Cobb–Douglas production function given earlier in Equation 3:

$$Y = Y/L = Ak^{\alpha},$$

where $k = K/L$. Using their definitions, the rates of change of capital per worker and output per worker are given by

$$\Delta k/k = \Delta K/K - \Delta L/L$$

and

$$\Delta y/y = \Delta Y/Y - \Delta L/L.$$

From the production function, the growth rate of output per worker is also equal to

$$\Delta y/y = \Delta A/A + \alpha\Delta k/k. \tag{6}$$

Note that these and other rate of change equations are exact only for changes over arbitrarily short periods ("continuous time").

The physical capital stock in an economy will increase because of gross investment (I) and decline because of depreciation. In a closed economy, investment must be funded by domestic saving. Letting s be the fraction of income (Y) that is saved, gross investment is given by $I = sY$. Assuming the physical capital stock depreciates at a constant rate, δ, the change in the physical capital stock is given by

$$\Delta K = sY - \delta K.$$

Subtracting labor supply growth, $\Delta L/L \equiv n$, and rearranging gives

$$\Delta k/k = sY/K - \delta - n. \tag{7}$$

In the steady state, the growth rate of capital per worker is equal to the growth rate of output per worker. Thus,

$$\Delta k/k = \Delta y/y = \Delta A/A + \alpha\Delta k/k,$$

from which we get

$$\Delta y/y = \Delta k/k = (\Delta A/A)/(1 - \alpha).$$

Letting θ denote the growth rate of TFP (i.e., $\Delta A/A$), we see that the equilibrium sustainable growth rate of output per capita (= Growth rate of capital per worker) is a constant that depends only on the growth rate of TFP (θ) and the elasticity of output with respect to capital (α). Adding back the growth rate of labor (n) gives the sustainable growth rate of output.

$$\text{Growth rate of output per capita} = \frac{\theta}{1-\alpha} \tag{8}$$

$$\text{Growth rate of output} = \frac{\theta}{1-\alpha} + n$$

This is the key result of the neoclassical model. Note that $\theta/(1 - \alpha)$ is the steady-state growth rate of labor productivity, so Equation 8 is consistent with the labor productivity growth accounting equation discussed earlier.

Substituting $\theta/(1 - \alpha)$ into the left-hand side of Equation 7 and rearranging gives the equilibrium output-to-capital ratio, denoted by the constant Ψ.

$$\frac{Y}{K} = \left(\frac{1}{s}\right)\left[\left(\frac{\theta}{1-\alpha}\right) + \delta + n\right] \equiv \Psi \tag{9}$$

In the steady state, the output-to-capital ratio is constant and the capital-to-labor ratio (k) and output per worker (y) grow at the same rate, given by $\theta/(1 - \alpha)$. On the steady-state growth path, the marginal product of capital is also constant and, given the Cobb–Douglas production function, is equal to $\alpha(Y/K)$. The marginal product of capital is also equal to the real interest rate in the economy. Note that even though the capital-to-labor ratio (k) is rising at rate $\theta/(1 - \alpha)$ in the steady state, the increase in the capital-to-labor ratio (k) has no impact on the marginal product of capital, which is not changing. *Capital deepening is occurring, but it has no effect on the growth rate of the economy or on the marginal product of capital once the steady state is reached.*

EXAMPLE 10

Steady-State Rate of Growth for China, Japan, and Ireland

Earlier examples generated estimates of potential growth for China (8.5%), Japan (1.0%), and Ireland (2.0%). Given the following data, address these questions:

1 Calculate the steady-state growth rates from the neoclassical model for China, Japan, and Ireland.

2 Compare the steady-state growth rates to the growth rates in potential GDP estimated in Examples 7–9 and explain the results.

	Labor Cost in Total Factor Cost (%)	TFP Growth (%)	Labor Force Growth (%)
China	56.1	2.9	1.2
Japan	53.8	0.2	0.0
Ireland	57.4	0.9	0.3

Sources: Conference Board Total Economy Database; labor cost based on 2008-2018, TFP growth based on 1995–2018. Labor force growth are assumptions and estimates from earlier examples.

Solution to 1:

Using Equation 8, the steady-state growth rate in the neoclassical model is given by

$$\Delta Y/Y = (\theta)/(1 - \alpha) + n = \text{Growth rate of TFP scaled by labor factor share} \\ + \text{Growth rate in the labor force}$$

Using the preceding equation and data, we can estimate steady-state growth rates for the three countries as follows:

China: The labor share of output (1 − α) is given by the average of the labor cost as a percentage of total factor cost, which is equal to 0.561 for China. The growth rate in the labor force is 1.2%, and the growth rate of TFP is 2.9%.

Steady-state growth rate = 2.9%/0.561 + 1.2% = 6.37%

Japan: The labor share of output (1 − α) for Japan is 0.538. The growth rate in the labor force is 0.0%, and TFP growth is 0.2%.

Steady-state growth rate = 0.2%/0.538 + 0.0% = 0.37%

Ireland: The labor share of output (1 − α) is 0.574% for Ireland. The growth rate in the labor force is 0.3%, and TFP growth is 0.9%.

Steady-state growth rate = 0.9%/0.574 + 0.3% = 1.87%

Solution to 2:

The growth rate in potential GDP for China (8.5%, estimated in Example 9) is significantly above the estimated 6.37% steady-state growth rate. The reason for this is that the economy of China is still in the process of converging to the higher income levels of the United States and the major economies in Europe. The physical capital stock is below the steady state, and capital deepening is a significant factor increasing productivity growth (see Exhibit 11) and the growth in potential GDP.

This is not the case for Japan and Ireland. Both countries are operating at essentially the steady state. The estimated growth rate in potential GDP for Japan (1.0%, from Example 7) is only slightly above its 0.37% steady-state growth rate. Likewise, the estimated growth rate in potential GDP for Ireland (2.0%, from Example 8) is effectively equal to its estimated steady-state growth rate of 1.87%. Operating close to the steady state means that capital investment in these countries, which results in an increasing capital-to-labor ratio, has no significant effect on the growth rate of the economy. Only changes in the growth rates of TFP and labor and in the labor share of output have an impact on potential GDP growth.

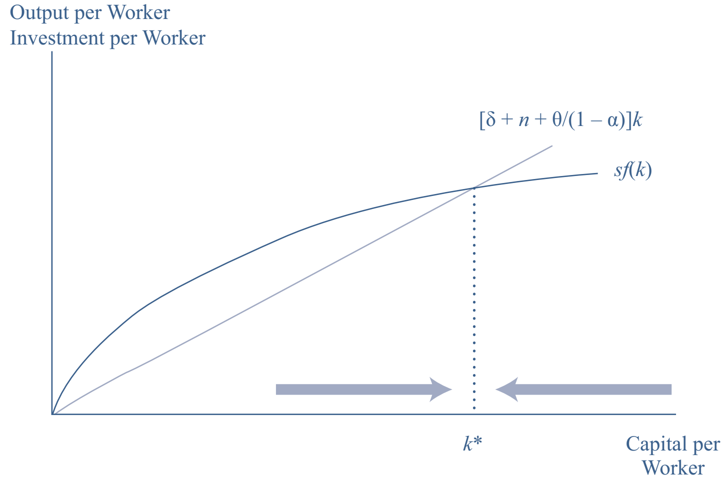
An intuitive way to understand the steady-state equilibrium given in Equation 9 is to transform it into a savings/investment equation:

$$sy = \left[\left(\frac{\theta}{1-\alpha}\right) + \delta + n\right]k$$

Steady-state equilibrium occurs at the output-to-capital ratio where the savings and actual gross investment per worker generated in the economy (sy) are just sufficient to (1) provide capital for new workers entering the workforce at rate n, (2) replace plant and equipment wearing out at rate δ, and (3) deepen the physical capital stock at the rate $\theta/(1 - \alpha)$ required to keep the marginal product of capital equal to the rental price of capital.

Exhibit 13 shows the steady-state equilibrium graphically. The straight line in the exhibit indicates the amount of investment required to keep the physical capital stock growing at the required rate. Because the horizontal axis is capital per worker, the slope of the line is given by $[\delta + n + \theta/(1 - \alpha)]$. The curved line shows the amount of actual investment per worker and is determined by the product of the saving rate and the production function. It is curved because of diminishing marginal returns to the capital input in the production function. The intersection of the required investment and actual investment lines determines the steady state. Note that *this exhibit is a snapshot at a point in time*. Over time, the capital-to-labor ratio rises at rate $[\theta/(1 - \alpha)]$ as the actual saving/investment curve $[sf(k)]$ shifts upward because of TFP growth, and *the equilibrium moves upward and to the right along the straight line.*

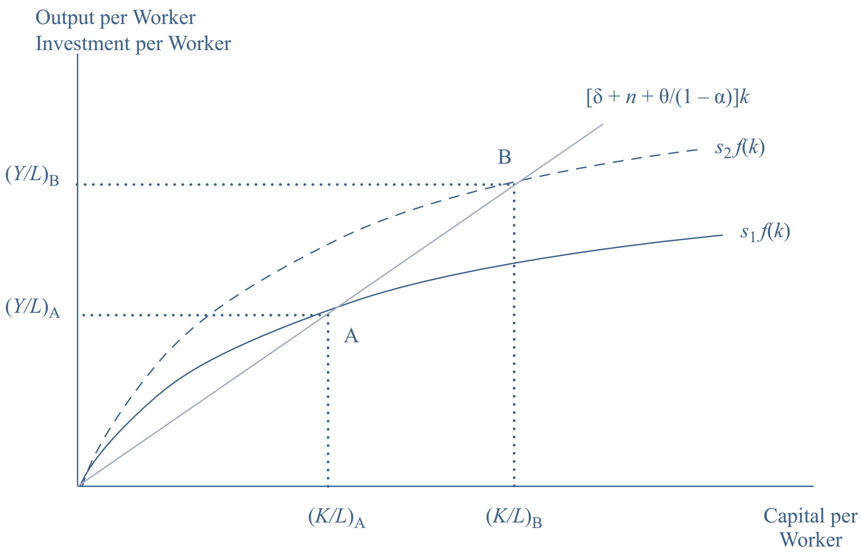
Exhibit 13 Steady State in the Neoclassical Model



The impact of the various parameters in the model on the steady state can also be seen in the exhibit. At any point in time when the economy is on its steady-state growth path, the exogenous factors—labor supply and TFP—are fixed. We would like to know what effect each of the parameters in the model has on the steady-state capital-to-labor ratio and therefore on output per worker. For example, if there are two economies that differ only with respect to one parameter, what does that imply about their per capita incomes? All else the same, we can say the following regarding the impact of the parameters:

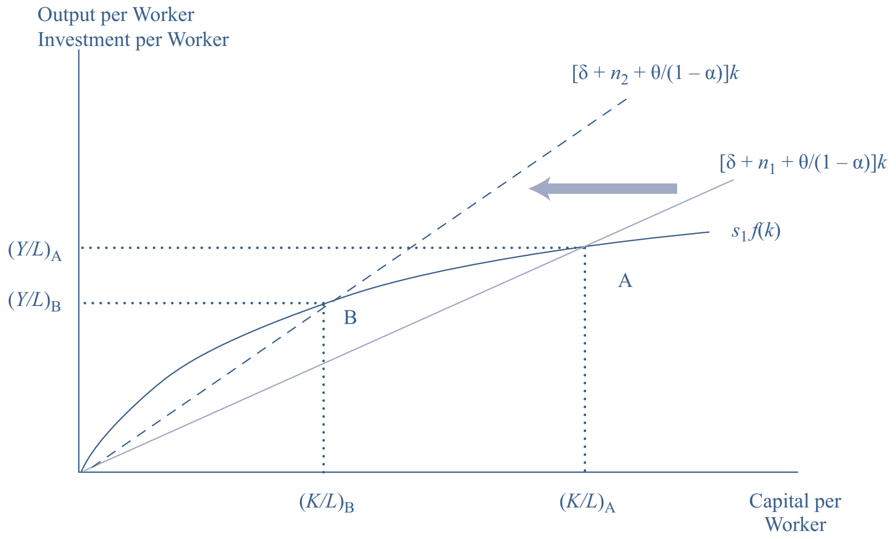
- *Saving rate (s)*: An increase in the saving rate implies a higher capital-to-labor ratio (k) and higher output per worker (y) because a higher saving rate generates more saving/investment at every level of output. In Exhibit 14, the saving/investment curve $[sf(k)]$ shifts upward from an initial steady-state equilibrium at point A to a new equilibrium at point B. At the new equilibrium point, it intersects the required investment line $[\delta + n + \theta/(1 - \alpha)]$ at higher capital-to-labor and output per worker ratios. Note that although the higher saving rate increases both k and y, it has no impact on the steady-state growth rates of output per capita or output (Equation 8).

Exhibit 14 Impact on the Steady State: Increase in the Saving Rate



- *Labor force growth (n)*: An increase in the labor force growth rate reduces the equilibrium capital-to-labor ratio because a corresponding increase in the steady-state growth rate of capital is required. Given the gross saving/investment rate, this can be achieved only at a lower capital-to-labor ratio. Output per worker is correspondingly lower as well. In Exhibit 15, the higher population growth rate increases the slope of the required investment line. This shifts the steady-state equilibrium from point A to point B, where it intersects the supply of saving/investment curve at lower capital-to-labor and output per worker ratios.
- *Depreciation rate (δ)*: An increase in the depreciation rate reduces the equilibrium capital-to-labor and output per worker ratios because a given rate of gross saving generates less net capital accumulation. Graphically, it increases the slope of the required investment line and affects the steady-state equilibrium in the same way as labor force growth (Exhibit 15).
- *Growth in TFP (θ)*: An increase in the growth rate of TFP reduces the steady-state capital-to-labor ratio and output per worker for given levels of labor input and TFP. This result must be interpreted with care. Raising the growth rate of TFP means that output per worker will grow faster in the future (Equation 8), but at a given point in time, a given supply of labor, and a given *level* of TFP, output per worker is lower than it would be with a slower TFP growth rate. In effect, the economy is on a steeper trajectory off a lower base of output per worker. Graphically, it is identical to Exhibit 15 in that faster TFP growth steepens the required investment line (increases the slope), which intersects with the available saving/investment curve at lower capital-to-labor and investment per worker ratios.

Exhibit 15 Impact on the Steady State: Increase in Labor Force Growth



In sum, such factors as the saving rate, the growth rate of the labor force, and the depreciation rate change the *level* of output per worker but do not permanently change the *growth rate* of output per worker. A permanent increase in the growth rate in output per worker can occur only if there is a change in the growth rate of TFP.

So far we have focused on the steady-state growth path. What happens if the economy has not yet reached the steady state? During the transition to the steady-state growth path, the economy can experience either faster or slower growth relative to the steady state. Using Equations 6, 7, and 9, we can write the growth rates of output per capita and the capital-to-labor ratio as, respectively,

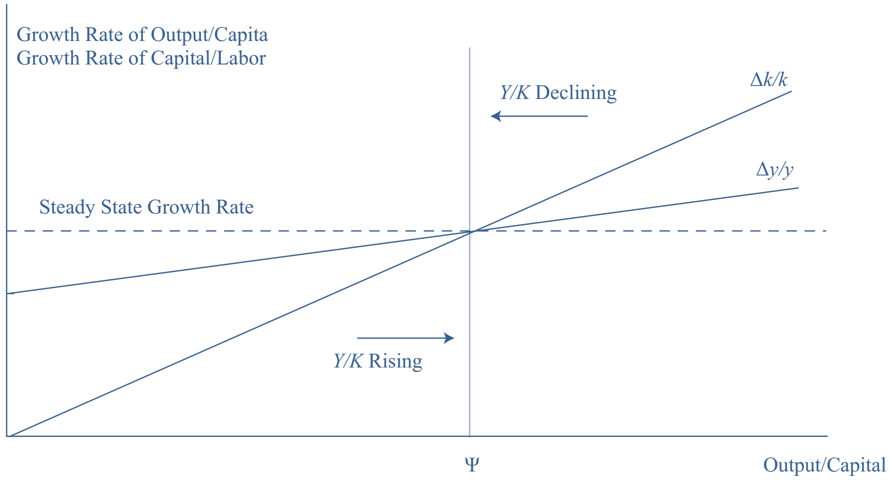
$$\frac{\Delta y}{y} = \left(\frac{\theta}{1-\alpha}\right) + \alpha s\left(\frac{Y}{K} - \Psi\right) = \left(\frac{\theta}{1-\alpha}\right) + \alpha s\left(y/k - \Psi\right) \quad (10)$$

and

$$\frac{\Delta k}{k} = \left(\frac{\theta}{1-\alpha}\right) + s\left(\frac{Y}{K} - \Psi\right) = \left(\frac{\theta}{1-\alpha}\right) + s\left(y/k - \Psi\right), \quad (11)$$

where the second equality in each line follows from the definitions of y and k, which imply $(Y/K) = y/k$. These relationships are shown in Exhibit 16.

Exhibit 16 Dynamics in the Neoclassical Model



If the output-to-capital ratio is above its equilibrium level (ψ), the second term in Equations 10 and 11 is positive and the growth rates of output per capita and the capital-to-labor ratio are above the steady-state rate $\theta/(1 - \alpha)$. This corresponds to a situation in which actual saving/investment exceeds required investment and above-trend growth in per capita output is driven by an above-trend rate of capital deepening. This situation usually reflects a relatively low capital-to-labor ratio but could, at least in principle, arise from high TFP. Because $\alpha < 1$, capital is growing faster than output and the output-to-capital ratio is falling. Over time, the growth rates of both output per capita and the capital-to-labor ratio decline to the steady-state rate.

Of course, the converse is true if the output-to-capital ratio is below its steady-state level. Actual investment is insufficient to sustain the trend rate of growth in the capital-to-labor ratio, and both output per capita and the capital-to-labor ratio grow more slowly. This situation usually corresponds to a relatively high and unsustainable capital-to-labor ratio, but it could reflect relatively low TFP and hence relatively low output. Over time, output grows faster than capital, the output-to-capital ratio rises, and growth converges to the trend rate.

IMPLICATIONS OF NEOCLASSICAL MODEL 11

i. compare classical growth theory, neoclassical growth theory, and endogenous growth theory;

There are four major groups of conclusions from the neoclassical model:

1 Capital Accumulation

- **a** Capital accumulation affects the level of output but not the growth rate in the long run.
- **b** Regardless of its initial capital-to-labor ratio or initial level of productivity, a growing economy will move to a point of steady-state growth.
- **c** In a steady state, the growth rate of output equals the rate of labor force growth plus the rate of growth in TFP scaled by labor's share of income [$n + \theta/(1 - \alpha)$]. The growth rate of output does not depend on the accumulation of capital or the rate of business investment. Those familiar with the

"labor-augmenting" technical change formulation of the neoclassical model should note that in that formulation, the rate of labor-augmenting technical change is also the growth rate of labor productivity. In our formulation, the growth rate of labor productivity is $\theta/(1 - \alpha)$. So both formulations imply that long-run growth equals the growth rate of the labor supply (n) plus a constant growth rate of labor productivity.

2 Capital Deepening vs. Technology

a Rapid growth that is above the steady-state rate of growth occurs when countries first begin to accumulate capital; but growth will slow as the process of accumulation continues (see Exhibit 16).

b Long-term sustainable growth cannot rely solely on capital deepening investment—that is, on increasing the stock of capital relative to labor. If the capital-to-labor ratio grows too rapidly (i.e., faster than labor productivity), capital becomes less productive, resulting in slower rather than faster growth.

c More generally, increasing the supply of some input(s) too rapidly relative to other inputs will lead to diminishing marginal returns and cannot be the basis for sustainable growth.

d In the absence of improvements in TFP, the growth of labor productivity and per capita output would eventually slow.

e Because of diminishing marginal returns to capital, the only way to sustain growth in potential GDP per capita is through technological change or growth in total factor productivity. This results in an upward shift in the production function—the economy produces more goods and services for any given mix of labor and capital inputs.

3 Convergence

a Given the relative scarcity and hence high marginal productivity of capital and potentially higher saving rates in developing countries, the growth rates of developing countries should exceed those of developed countries.

b As a result, there should be a convergence of per capita incomes between developed and developing countries over time.

4 Effect of Savings on Growth

a The initial impact of a higher saving rate is to temporarily raise the rate of growth in the economy. Mathematically, this can be seen as follows: Equation 9 indicates that an increase in the saving rate (s) reduces the steady-state output-to-capital ratio (ψ). This makes the last term in Equations 10 and 11 positive, raising the growth rates of output per capita (y) and the capital-to-labor ratio (k) above the steady-state rate. In response to the higher saving rate, growth exceeds the steady-state growth rate during a transition period. However, the economy returns to the balanced growth path after the transition period.

b During the transition period, the economy moves to a higher level of per capita output and productivity.

c Once an economy achieves steady-state growth, the growth rate does not depend on the percentage of income saved or invested. Higher savings cannot permanently raise the growth rate of output.

d Countries with higher saving rates, however, will have a higher level of per capita output, a higher capital-to-labor ratio, and a higher level of labor productivity.

EXAMPLE 11

Comparative Statics and Transitional Growth in the Neoclassical Model

Beginning in steady-state equilibrium, an economy's saving rate suddenly increases from 20% of income to 30% of income. Other key parameters describing the economy are as follows:

Growth rate of TFP (θ)	= 0.02
Income share of capital (α)	= 0.35
Depreciation rate (δ)	= 0.10
Labor force growth rate (n)	= 0.01

The following table shows the output-to-capital ratio that will prevail in this economy at various points in time after the increase in the saving rate.

Years after Saving Rate Increase	Output-to-Capital Ratio
5	0.5947
10	0.5415
25	0.4857
50	0.4708
100	0.4693
New steady state	??

By rearranging the Cobb–Douglas production function (Equation 3), the proportional impact of the saving rate change on the capital-to-labor ratio can be expressed in terms of the proportional impact on the output-to-capital ratio. The proportional impact on per capita income can then be determined from the production function (Equation 3). Labeling the paths with and without the change in saving rate as "new" and "old" respectively, at each date we have:[a]

$$\frac{k_{new}}{k_{old}} = \left[\frac{\left(Y/K\right)_{new}}{\left(Y/K\right)_{old}}\right]^{\frac{1}{\alpha-1}}$$

and

$$\frac{y_{new}}{y_{old}} = \left(\frac{k_{new}}{k_{old}}\right)^{\alpha}$$

1 Using Equations 8 and 9, calculate the steady-state growth rate of per capita income and the steady-state output-to-capital ratio both before and after the change in the saving rate. What happens to the capital-to-labor ratio and output per capita?

2 Use the output-to-capital ratios given in the preceding table along with Equation 10 and your answers to Question 1 to determine the growth rate of per capita income that will prevail immediately following the change in the saving rate and at each of the indicated times after the change. Explain the pattern of growth rates.

3 Using the output-to-capital ratios given in the preceding table, calculate the proportional impact of the increased saving rate on the capital-to-labor ratio and on per capita income over time. With respect to these variables, how will the new steady state compare with the old steady state?

Solution to 1:

From Equation 8, the steady-state growth rate of per capita income, both before and after the increase in the saving rate, is $\Delta y/y = \theta/(1 - \alpha) = 0.02/(1 - 0.35) = 0.0308$, or 3.08%. From Equation 9, the steady-state output-to-capital ratio is

$$\frac{Y}{K} = \left(\frac{1}{s}\right)\left[\left(\frac{\theta}{1-\alpha}\right) + \delta + n\right] \equiv \Psi$$

Using the parameter values given, $\theta/(1 - \alpha) + \delta + n = 0.0308 + 0.10 + 0.01 = 0.1408$, so the steady-state output-to-capital ratio is 0.1408/0.2 = 0.7040 with the initial 20% saving rate and 0.1408/0.30 = 0.4693 with the new 30% saving rate. As shown in Exhibit 14, both the capital-to-labor ratio and output per worker are at higher *levels* in the new steady state. But once the new steady state is achieved, they do not grow any faster than they did in the steady state with the lower saving rate.

Solution to 2

According to Equation 10, the growth rate of per capita income is given by

$$\frac{\Delta y}{y} = \left(\frac{\theta}{1-\alpha}\right) + \alpha s\left(y/k - \Psi\right)$$

Immediately following the increase in the saving rate, the relevant value of ψ becomes the new steady-state output-to-capital ratio (0.4693). The actual output-to-capital ratio does not change immediately, so y/k is initially still 0.7040. Plugging these values into the foregoing growth equation gives the growth rate of per capita income:

$\Delta y/y = 0.0308 + (0.35)(0.30)(0.7040 - 0.4693) = 0.0554$, or 5.54%

Similar calculations using the output-to-capital ratios in the preceding table give the following:

Years after Saving Rate Increase	Output-to-Capital Ratio	Growth Rate of Per Capita Income (%)
0	0.7040	5.54
5	0.5947	4.39
10	0.5415	3.84
25	0.4857	3.25
50	0.4708	3.09
100	0.4693	3.08
New steady state	0.4693	3.08

The growth rate "jumps" from the steady-state rate of 3.08% to 5.54% when the saving rate increases because the increase in saving/investment results in more rapid capital accumulation. Over time, the growth rate slows because the marginal productivity of capital declines as the capital-to-labor ratio increases. In addition, as the capital-to-labor ratio increases and the output-to-capital ratio declines, a greater portion of savings is required to maintain the capital-to-labor ratio, leaving a smaller portion for continued capital deepening. Roughly two-thirds of the growth acceleration has dissipated after 10 years.

Solution to 3:

Using the output-to-capital ratio that will prevail five years after the saving rate increase, the proportional impact on the capital-to-labor ratio and on per capita income will be

$$\frac{k_{new}}{k_{old}} = \left[\frac{\left(Y/K\right)_{new}}{\left(Y/K\right)_{old}}\right]^{\frac{1}{\alpha-1}} = \left[\frac{0.5947}{0.7040}\right]^{\frac{-1}{0.65}} = 1.2964$$

and

$$\frac{y_{new}}{y_{old}} = \left(\frac{k_{new}}{k_{old}}\right)^{\alpha} = 1.2964^{0.35} = 1.0951$$

Thus, after five years, the capital-to-labor ratio will be 29.64% higher than it would have been without the increase in the saving rate, and per capita income will be 9.51% higher. Similar calculations for the other periods give the following:

Years after Saving Rate Increase	Proportionate Increase (%) in: Capital-to-Labor Ratio	Proportionate Increase (%) in: Per Capita Income
0	0.00	0.00
5	29.64	9.51
10	49.74	15.18
25	77.01	22.12
50	85.71	24.19
100	86.68	24.42
New steady state	86.68	24.42

In the new steady state, the capital-to-labor ratio will be 86.68% higher at every point in time than it would have been in the old steady state. Per capita income will be 24.42% higher at every point in time. Both variables will be growing at the same rate (3.08%) as they would have been in the old steady state.

[a] Note that the output-to-capital ratio would have been constant on the original steady state path. Because of the impact of total factor productivity, the capital-to-labor ratio and output per capita are not constant even in steady state. In comparing "paths" for these variables, we isolate the impact of the saving rate change by canceling out the effect of TFP growth. Mathematically, we cancel out A in Equation 3 to produce the equations shown here.

EXTENSION OF NEOCLASSICAL MODEL

12

i. compare classical growth theory, neoclassical growth theory, and endogenous growth theory;

Solow (1957) used the growth accounting equation to determine the contributions of each factor to economic growth in the United States for the period 1909–1949. He reached the surprising conclusion that more than 80% of the per capita growth in the United States resulted from TFP. Denison (1985) authored another study examining US growth for the period 1929–1982 using the Solow framework. His findings were similar to Solow's, with TFP explaining nearly 70% of US growth. The problem with these findings is that the neoclassical model provides no explicit explanation of the economic determinants of technological progress or how TFP changes over time. Because technology is determined outside the model (i.e., exogenously), critics argue that the neoclassical model ignores the very factor driving growth in the economy. Technology is simply the residual or the part of growth that cannot be explained by other inputs, such as capital and labor. This lack of an explanation for technology led to growing dissatisfaction with the neoclassical model.

The other source of criticism of the neoclassical model is the prediction that the steady-state rate of economic growth is unrelated to the rate of saving and investment. Long-run growth of output in the Solow model depends only on the rates of growth of the labor force and technology. Higher rates of investment and savings have only a transitory impact on growth. Thus, an increase in investment as a share of GDP from 10% to 15% of GDP will have a positive impact on the near-term growth rate but will not have a permanent impact on the ultimately sustainable percentage growth rate. This conclusion makes many economists uncomfortable. Mankiw (1995) provided evidence rebutting this hypothesis and showed that saving rates and growth rates are positively correlated across countries. Finally, the neoclassical model predicts that in an economy where the stock of capital is rising faster than labor productivity, the return to investment should decline with time. For the advanced countries, the evidence does not support this argument because returns have not fallen over time.

Critiques of the neoclassical model led to two lines of subsequent research on economic growth. The first approach, originated by Jorgenson (1966, 2000), is termed the augmented Solow approach. It remains in the neoclassical tradition in that diminishing marginal returns are critical and there is no explanation for the determinants of technological progress. Instead, this approach attempts to reduce empirically the portion of growth attributed to the unexplained residual labeled technological progress (TFP). The idea is to develop better measures of the inputs used in the production function and broaden the definition of investment by including human capital, research and development, and public infrastructure. In addition, the composition of capital spending is important. Higher levels of capital spending on high-technology goods will boost productivity more than spending on machine tools or structures.

By adding inputs such as human capital to the production function, the augmented Solow model enables us to more accurately measure the contribution of technological progress to growth. The economy still moves toward a steady-state growth path, however, because even broadly defined capital is assumed to eventually encounter diminishing marginal returns. In essence, this line of research uses the growth accounting methodology and increases the number of inputs in the production function in order to provide a more accurate measure of technological progress. The second approach is the endogenous growth theory, which we examine in the next section.

13 ENDOGENOUS GROWTH MODEL

i. compare classical growth theory, neoclassical growth theory, and endogenous growth theory;

The alternative to the neoclassical model is a series of models known as endogenous growth theory. These models focus on explaining technological progress rather than treating it as exogenous. In these models, self-sustaining growth emerges as a natural consequence of the model and the economy does not necessarily converge to a steady-state rate of growth. Unlike the neoclassical model, there are *no diminishing marginal returns to capital for the economy as a whole* in the endogenous growth models. So increasing the saving rate permanently increases the rate of economic growth. These models also allow for the possibility of increasing returns to scale.

Romer (1986) provided a model of technological progress and a rationale for why capital does not experience diminishing marginal returns. He argued that capital accumulation is the main factor accounting for long-run growth, once the definition of capital is broadened to include such items as human or knowledge capital and research and development (R&D). R&D is defined as investment in new knowledge that improves the production process. In endogenous growth theory, knowledge or human capital and R&D spending are factors of production, like capital and labor, and have to be paid for through savings.

Companies spend on R&D for the same reason they invest in new equipment and build new factories: to make a profit. R&D spending is successful if it leads to the development of a new product or method of production that is successful in the marketplace. There is a fundamental difference, however, between spending on new equipment and factories and spending on R&D. The final product of R&D spending is ideas. These ideas can potentially be copied and used by other companies in the economy. Thus, R&D expenditures have potentially large positive externalities or spillover effects. This means that spending by one company has a positive impact on other companies and increases the overall pool of knowledge available to all companies. Spending by companies on R&D and knowledge capital generates benefits to the economy as a whole that exceed the private benefit to the individual company making the R&D investment. Individual companies cannot fully capture all the benefits associated with creating new ideas and methods of production. Some of the benefits are external to the company, and so are the social returns associated with the investment in R&D and human capital.

This distinction between the private and social returns or benefits to capital is important because it solves an important microeconomic issue. The elimination of the assumption of diminishing marginal returns to capital implies constant returns to capital and increasing returns to all factors taken together. If individual companies could capture these scale economies, then all industries would eventually be dominated by a single company—a monopoly. There is simply no empirical evidence to support this implication. Separating private returns from social returns solves the problem. If companies face constant returns to scale for all private factors, there is no longer an inherent advantage for a company being large. But the externality or social benefit results in increasing returns to scale across the entire economy as companies benefit from the private spending of the other companies.

The role of R&D spending and the positive externalities associated with this spending have important implications for economic growth. In the endogenous growth model, the economy does not reach a steady growth rate equal to the growth of labor plus an exogenous rate of labor productivity growth. Instead, saving and investment decisions can generate self-sustaining growth at a permanently higher rate. This situation is in sharp contrast to the neoclassical model, in which only a transitory increase in growth above the steady state is possible. The reason for this difference is that because of the externalities on R&D, diminishing marginal returns to capital do not set in. The production function in the endogenous growth model is a straight line given by

$$y_e = f(k_e) = ck_e, \tag{12}$$

where output per worker (y_e) is proportional to the stock of capital per worker (k_e), c is the (constant) marginal product of capital in the aggregate economy, and the subscript e denotes the endogenous growth model. In contrast, the neoclassical production function is a curved line that eventually flattens out (see Exhibit 4).

To understand the significance of introducing constant returns to aggregate capital accumulation, note that in this model the output-to-capital ratio is fixed (= c) and therefore output per worker (y_e) always grows at the same rate as capital per worker (k_e). Thus, faster or slower capital accumulation translates one for one into faster or slower growth in output per capita. Substituting Equation 12 into Equation 7 gives an equation for the growth rate of output per capita in the endogenous growth model:

$$\Delta y_e / y_e = \Delta k_e / k_e = sc - \delta - n$$

Because all the terms on the right-hand side of this equation are constant, this is both the long-run and short-run growth rate in this model. Examination of the equation shows that *a higher saving rate (s) implies a permanently higher growth rate.* This is the key result of the endogenous growth model.

The positive externalities associated with spending on R&D and knowledge capital suggest that spending by private companies on these inputs may be too low from an overall societal point of view. This is an example of a market failure wherein private companies under-invest in the production of these goods. In this case, there may be a role for government intervention to correct for the market failure by direct government spending on R&D and/or providing tax breaks and subsidies for private production of knowledge capital. Higher levels of spending on knowledge capital could translate into faster economic growth even in the long run. Finally, according to the endogenous growth theory, there is *no reason why the incomes of developed and developing countries should converge.* Because of constant or even increasing returns associated with investment in knowledge capital, the developed countries can continue to grow as fast as, or faster than, the developing countries. As a result, there is no reason to expect convergence of income over time. We now turn to the convergence debate in more detail.

EXAMPLE 12

Neoclassical vs. Endogenous Growth Models

Consider again an economy with per capita income growing at a constant 3.08% rate and with a 20% saving rate, an output-to-capital ratio (c in the endogenous growth model, Equation 12) of 0.7040, a depreciation rate (δ) of 10%, and a 1% labor force growth (n).

1 Use the endogenous growth model to calculate the new steady-state growth rate of per capita income if the saving rate increases to 23.5%.

2 How much higher will per capita income be in 10 years because of the higher saving rate? How does this compare with the impact calculated in Example 11 using the neoclassical model? What accounts for the difference?

3 In an effort to boost growth, the government is considering two proposals. One would subsidize all private companies that increase their investment spending. The second would subsidize only investments in R&D and/or implementation of new technologies with potential for network externalities. Interpret these proposals in terms of the neoclassical and endogenous growth models and assess their likely impact on growth. (Focus only on "supply-side" considerations here.)

Solution to 1:

In the endogenous growth model, the new growth rate of per capita income is

$$\Delta y_e/y_e = sc - \delta - n = (0.235)(0.7040) - 0.10 - 0.01 = 0.0554, \text{ or } 5.54\%.$$

This is the same as the growth rate immediately following the increase in the saving rate (to 30% in that case) in the earlier example using the neoclassical model (Example 11). Unlike in the neoclassical model, in the endogenous growth model this higher growth rate will be sustained.

Solution to 2:

According to the endogenous growth model, per capita income will grow 2.46% (= 5.54% − 3.08%) faster with the higher saving rate. After 10 years, the cumulative impact of the faster growth rate will be

$$\exp(0.0246 \times 10) = \exp(0.246) = 1.2789.$$

So, per capita income will be almost 28% higher than it would have been at the lower saving rate. This increase is substantially larger than the 15.18% cumulative increase after 10 years found in Example 11 assuming a much larger increase in the saving rate (to 30% instead of 23.5%) in the neoclassical model. The difference arises because the endogenous growth model assumes that capital accumulation is not subject to diminishing returns. Therefore, the growth rate is permanently, rather than temporarily, higher.

Solution to 3:

Subsidizing all private investment would tend to have a significant, pure capital deepening component. That is, companies would be encouraged to buy more, but not necessarily better, plant and equipment. The neoclassical model indicates that this scenario is likely to result in a temporary surge in growth, but even if the higher rate of investment/saving is sustained, growth will again decline over time. On the positive side, this proposal is very likely to succeed, at least for a while, because it does not require investment in unproven technologies or ill-defined network effects. The impact of the other proposal is more uncertain but potentially much more powerful. If the investments in R&D and/or new technologies lead to new knowledge, greater efficiency, new products and methods, and/or network externalities, then the endogenous growth model suggests that growth is likely to be permanently enhanced.

CONVERGENCE HYPOTHESES 14

j explain and evaluate convergence hypotheses;

As is evident in Exhibit 1, a wide gap separates the living standards in developed and developing nations. The question is, will this difference persist forever or will the per capita income levels of the developing countries converge to those of the developed countries? Convergence means that countries with low per capita incomes should grow at a faster rate than countries with high per capita incomes. Thus, over time the per capita income of developing countries should converge toward that of the developed countries. Whether convergence occurs has major implications for the future growth prospects of developed versus developing countries. It also has important investment implications.

Neoclassical growth theory predicts two types of convergence: absolute convergence and conditional convergence. **Absolute convergence** means that developing countries, regardless of their particular characteristics, will eventually catch up with the developed countries and match them in per capita output. The neoclassical model assumes that all countries have access to the same technology. As a result, per capita income in all countries should eventually grow at the same rate. Thus, the model implies convergence of per capita *growth rates* among all countries. It does not, however, imply that the *level* of per capita income will be the same in all countries regardless of underlying characteristics; that is, it does not imply absolute convergence.

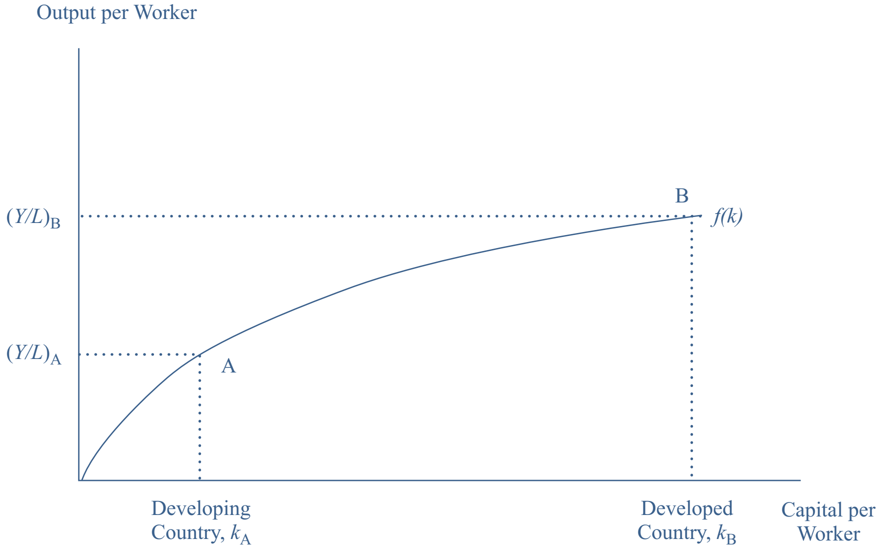
Conditional convergence means that convergence is conditional on the countries having the same saving rate, population growth rate, and production function. If these conditions hold, the neoclassical model implies convergence to the same *level* of per capita output as well as the same steady-state growth rate. In terms of Exhibit 13, these economies would have the same k^* and thus the same steady state. If they start with different capital-to-labor ratios, their growth rates will differ in the transition to the steady state. The economy with a lower capital-to-labor ratio will experience more rapid growth of productivity and per capita income, but the differential will diminish until they finally converge. Countries with different saving rates or population growth rates and thus different steady-state values for k^* will have different steady-state *levels* of per capita income, but their growth rates of per capita output will still converge.

The data (see Exhibit 18) indicate that some of the poorer countries are diverging rather than converging to the income levels of the developed countries. Thus, in addition to the first two convergence concepts, we have the notion of **club convergence**, where only rich and middle-income countries that are members of the club are converging to the income level of the world's richest countries. This means that the countries with the lowest per capita income in the club grow at the fastest rate. In contrast, countries outside the club continue to fall behind. Poor countries can join the club if they make appropriate institutional changes, such as those summarized earlier in our discussion of factors limiting growth in developing economies. Finally, countries may fall into a **non-convergence trap** if they do not implement necessary institutional reforms. For example, failure to reform labor markets has undermined growth in some European countries that have experienced weak growth in employment and high rates of unemployment over the last two decades. Certain institutional arrangements that initially enhance growth may later generate non-convergence traps if maintained too long. Import substitution policies enabled the Latin American countries to grow rapidly in the 1950s and 1960s but caused them to stagnate in the 1970s and 1980s.

If convergence, and especially club convergence, does occur, investing in countries with lower per capita incomes that are members of the club should, over long periods, provide a higher rate of return than investing in higher-income countries. Convergence means that the rate of growth of potential GDP should be higher in developing countries that have made the institutional changes that are a precondition for growth and that enable these countries to become members of the convergence club. With higher long-term growth in these economies, corporate profits should also grow at a faster rate. Given the faster rate of growth in earnings, stock prices may also rise at a faster rate. Of course, risk is also likely to be higher in these markets. Nonetheless, it is reasonable to conclude that long-term investors should allocate a risk-tolerance-appropriate portion of their assets to those developing economies that have become members of the convergence club.

Convergence between the developed and developing countries can occur in two ways. First, convergence takes place through capital accumulation and capital deepening. Exhibit 17 illustrates the difference between developed and developing countries using the per capita neoclassical production function. The developed countries operate at point B, so increases in capital have almost no impact on productivity. In contrast, developing countries operate at point A, where increases in capital significantly boost labor productivity.

Exhibit 17 Per Capita Production Function Developed vs. Developing Countries



A second source of convergence is that developing countries can imitate or adopt technology already widely utilized in the advanced countries. Developing countries can learn from advanced countries as scientific and management practices spread with globalization. By importing technology from the advanced countries, the developing countries can achieve faster economic growth and converge to the income of the advanced countries. Technology transfers will narrow the income gap between developed and developing countries only if the poor countries invest the resources to master the technology and apply it to their economies. This spending is similar to R&D spending and allows the country to join the convergence club. The steady-state rate of growth for members of the convergence club will be determined by the global rate of technological progress. Without such spending, the country will be left out and will continue to fall behind the developed countries.

In contrast to the neoclassical model, the endogenous growth model makes no prediction that convergence should occur. This model allows for countries that start with high per capita income and more capital to grow faster and stay ahead of the developing countries. If the externalities associated with knowledge and human capital are large, the higher-income country can maintain its lead through high rates of investment in these capital inputs.

If the convergence hypothesis is correct, there should be an inverse relation between the initial level of per capita real GDP and the growth rate in per capita GDP. Exhibit 18 shows the countries in Exhibit 1 in descending order of per capita income in 1950. If incomes are converging across countries, the poor countries in 1950 should have a higher growth rate between 1950 and 2018 than the rich countries.

Exhibit 18 Real Per Capita GDP by Selected Economy

	Real GDP Per Capita (in US dollars)				Avg. Annual Growth in Per Capita GDP (%)			
	1950	1980	2000	2018	1950–80	1980–2000	2000–18	1950–2018
United States	14,559	29,136	45,640	55,650	2.3	2.3	1.1	2.0
New Zealand	13,795	20,526	27,514	35,676	1.3	1.5	1.5	1.4
Australia	13,219	24,403	36,469	46,555	2.1	2.0	1.4	1.9
Canada	12,053	27,356	37,555	44,135	2.8	1.6	0.9	1.9
United Kingdom	11,602	20,547	33,531	40,627	1.9	2.5	1.1	1.9
France	8,266	24,901	35,778	40,689	3.7	1.8	0.7	2.4
Venezuela	8,104	18,247	14,469	9,487	2.7	–1.2	–2.3	0.2
Argentina	6,164	14,710	15,011	18,255	2.9	0.1	1.1	1.6
Italy	5,954	24,937	36,085	35,233	4.9	1.9	–0.1	2.6
Ireland	5,496	16,707	39,345	70,032	3.8	4.4	3.3	3.8
South Africa	4,361	10,781	9,715	12,156	3.1	–0.5	1.3	1.5
Singapore	4,299	20,626	51,748	89,196	5.4	4.7	3.1	4.6
Mexico	4,180	13,546	15,811	18,313	4.0	0.8	0.8	2.2
Spain	3,964	18,353	30,347	35,679	5.2	2.5	0.9	3.3
Peru	3,464	7,314	6,498	12,644	2.5	–0.6	3.8	1.9
Japan	3,048	20,769	33,875	39,313	6.6	2.5	0.8	3.8
Brazil	2,365	11,372	11,470	14,360	5.4	0.0	1.3	2.7
Turkey	2,327	7,990	13,258	24,850	4.2	2.6	3.6	3.5
Philippines	1,296	4,390	4,277	7,943	4.2	–0.1	3.5	2.7
Korea	1,185	5,084	20,757	36,756	5.0	7.3	3.2	5.2
Egypt	1,132	5,228	8,452	11,881	5.2	2.4	1.9	3.5
Indonesia	804	2,911	5,863	11,760	4.4	3.6	3.9	4.0
Pakistan	666	2,080	3,406	5,049	3.9	2.5	2.2	3.0
India	658	1,297	2,546	6,999	2.3	3.4	5.8	3.5
China	402	722	3,682	16,098	2.0	8.5	8.5	5.6
Ethiopia	314	727	653	2,073	2.8	–0.5	6.6	2.8

Source: IMF.
Note: GDP per capita, constant prices at 2011 in US dollars, adjusted for PPP.

The results for the convergence hypothesis are mixed. The economies with the highest per capita income in 1950 were the United States, New Zealand, Australia, and Canada. The markets with the fastest growth rate over the period 1950–2018 were China and Korea, each growing at a rate above 5%. This result strongly supports convergence because the per capita incomes of these economies in 1950 were well below that of the United States. In addition, the results for Japan, Singapore, Spain and Korea showed a convergence to the level of income in the advanced economies. In total, 17 of the 27 economies in our sample grew faster than the United States during the period. However, South Africa, Argentina, Venezuela, and New Zealand were among those that fell further behind the United States. Interestingly, since 2000, convergence has been relatively strong overall, with 16 countries in our sample (60%) growing faster than the United States—including Ethiopia, the Philippines, Peru, Turkey, South Africa, Australia, and New Zealand—but has not continued among the most advanced economies: France, Japan, and Italy all lagged the United States, Canada, and Australia.

The evidence seems to suggest that poorer countries may converge if they develop the appropriate legal, political, and economic institutions. In addition, trade policy is an important factor, which we address in the next section.

GROWTH IN AN OPEN ECONOMY

15

k describe the economic rationale for governments to provide incentives to private investment in technology and knowledge;

l describe the expected impact of removing trade barriers on capital investment and profits, employment and wages, and growth in the economies involved.

The Solow model discussed earlier assumed a closed economy in which domestic investment equals domestic savings and there is no international trade or capital flows. Opening up the economy to trade and financial flows can significantly affect the rate of growth in an economy for the following reasons:

1. A country can borrow or lend funds in global markets, and domestic investment can be funded by global savings. Thus, investment is not constrained by domestic savings.
2. Countries can shift resources into industries in which they have a comparative advantage and away from industries in which they are relatively inefficient, thereby increasing overall productivity.
3. Companies have access to a larger, global market for their products, allowing them to better exploit any economies of scale and increasing the potential reward for successful innovation.
4. Countries can import technology, thus increasing the rate of technological progress.
5. Global trade increases competition in the domestic market, forcing companies to produce better products, improve productivity, and keep costs low.

According to the neoclassical model, convergence should occur more quickly if economies are open and there is free trade and international borrowing and lending. Opening up the economy should increase the rate at which countries' capital-to-labor ratios converge. The dynamic adjustment process can be described as follows:

1. Developing countries have less capital per worker, and as a result, the marginal product of capital is higher. Thus, the rate of return on investments should be higher in countries with low capital-to-labor ratios and lower in countries with high capital-to-labor ratios.
2. Global savers, seeking higher returns on investments, will invest in the capital-poor countries. In an open economy, capital should flow from countries with high capital-to-labor ratios to those that are capital poor.
3. Because of the capital inflows, the physical capital stock in the developing countries should grow more rapidly than in rich countries even if the saving rate is low in the poorer countries. Faster capital growth will result in higher productivity growth, causing per capita incomes to converge.
4. Because capital flows must be matched by offsetting trade flows, capital-poor countries will tend to run a trade deficit as they borrow globally to finance domestic investment. In contrast, the developed countries will tend to run trade surpluses as they export capital.

5 During the transition to the new steady state, the inflows of capital will temporarily raise the rate of growth in the capital-poor country above the steady-state rate of growth. At the same time, growth in the capital-exporting countries will be below the steady state.
6 Over time, the physical capital stock will rise in the capital-poor country, reducing the return on investments. As a result, the rate of investment and size of the country's trade deficit will decline. Growth will slow and approach the steady-state rate of growth. If investment falls below the level of domestic savings, the country will eventually shift from a trade deficit to a trade surplus and become an exporter of capital.
7 In the Solow model, after the reallocation of world savings, there is no permanent increase in the rate of growth in an economy. Both the developed and developing countries grow at the steady-state rate of growth.

In contrast to the neoclassical model, endogenous growth models predict that a more open trade policy will permanently raise the rate of economic growth. In these models, international trade increases global output through the following:

1 A selection effect, whereby increased competition from foreign companies forces less efficient domestic companies to exit and more efficient ones to innovate and raises the efficiency of the overall national economy.
2 A scale effect that allows producers to more fully exploit economies of scale by selling to a larger market.
3 A backwardness effect arising from less advanced countries or sectors of an economy catching up with the more advanced countries or sectors through knowledge spillovers.

Open trade also affects the innovation process by encouraging higher levels of spending on R&D and on human capital as companies invest to take advantage of access to larger markets and the greater flow of ideas and knowledge among countries. The rate of return to new investment increases, as does the rate of economic growth. In general, most countries gain from open trade, with the scale effect benefiting smaller countries and the backwardness effect benefiting the poorer, less developed countries. But trade can also retard growth in some cases, especially in small countries that lag behind the technology leaders. Opening these countries to trade may discourage domestic innovation because companies will recognize that, even if they innovate, they may lose out to more efficient foreign companies.

EXAMPLE 13

The Entry of China and India into the Global Economy

China and India effectively entered the global economy in the 1980s as they shifted toward more market-oriented policies and opened up to global trade. Their impact on global growth was significant. In 2018, according to the IMF, China and India accounted for 20% and 8% of world GDP (based on PPP), respectively, whereas the two countries combined for only 4.2% of global output in 1980. The entry of these two countries significantly increased the global supply of skilled and unskilled labor receiving relatively lower wages. As a result of the surge in available labor, global potential GDP increased sharply. Economic theory suggests that the supply-side increase in the global capacity to produce goods and services would increase global output and put downward pressure on prices.

The neoclassical model of growth can provide us with some further insights into the impact of China and India entering the global economy. At the time, China and India had relatively lower wages and capital compared with the United States and Europe. One would expect that the rate of return on capital would be higher in China and India and that capital would flow from the developed countries to China and India. Hence, both China and India would be expected to run trade deficits. This has been the case for India but, contrary to the model's prediction, China has run trade surpluses. These surpluses stem mainly from China's very high domestic saving rate.

Nonetheless, China has experienced large foreign direct investment (see Exhibit 19) inflows, which have reinforced its already high private investment rate. As China and India accumulate capital, their capital-to-labor ratios, real wage levels, and per capita income should converge toward those of the advanced economies. Depending on global aggregate demand conditions, wages might even have to fall in the developed countries in the process of shifting wealth and income to the developing economies. Because of the surge in the global supply of labor, the overall share of labor in global income should decline relative to capital. In addition, global productivity should rise as China and India account for a rising share of global output. In sum, over the long run, the growing share of global GDP going to China and India will benefit the global economy as more efficient utilization of resources allows global potential GDP to grow more rapidly for an extended period.

Although both the neoclassical and endogenous models of growth show the benefits of open markets, over the last 50 years developing countries have pursued two contrasting strategies for economic development:

- *Inward-oriented policies* attempt to develop domestic industries by restricting imports. Instead of importing goods and services, these policies encourage the production of domestic substitutes, despite the fact that it may be more costly to do so. These policies are also called import substitution policies.
- *Outward-oriented policies* attempt to integrate domestic industries with those of the global economy through trade and make exports a key driver of growth.

Many African and Latin American countries pursued inward-oriented policies from the 1950s to the 1980s that resulted in poor GDP growth and inefficient industries producing low-quality goods. In contrast, many East Asian countries, such as Singapore and South Korea, pursued outward-oriented polices during this same period, which resulted in high rates of GDP growth and convergence with developed countries. These countries also benefited from the positive effects of foreign direct investment, which suggests that more open and trade-oriented economies will grow at a faster rate. The evidence strongly supports this case.

In Example 1, we compared the economic performance of Argentina and Venezuela with that of Japan, South Korea, and Singapore. In 1950, the per capita GDP of the two Latin American countries was well above that of the three East Asian countries. By 2010, however, the per capita GDPs of the three Asian countries was well above those of Argentina and Venezuela. The difference in the growth rates between Argentina and Venezuela and the three Asian countries is explained largely by the openness of their economies. Argentina and Venezuela were relatively closed economies, whereas the Asian countries relied on foreign investment and open markets to fuel growth.

Many African and Latin American countries have removed trade barriers and are pursuing more outward-oriented policies, and they experienced better growth. Brazil is a good example. Exports of goods and services increased from $64.6 billion in 2000 to $218 billion in 2018, an increase of more than 237%. As shown in Exhibit 19, exports as a share of GDP rose from 10.2% to 14.8% over this period.

Exhibit 19 Exports and Foreign Direct Investment of Selected Countries

	1980	1990	2000	2018
Brazil				
Exports as a percentage of GDP	9.1%	8.2%	10.2%	14.8%
Inflows of foreign direct investment ($ billions)	NA	NA	$32.8	$61.2
China				
Exports as a percentage of GDP	5.9%	13.6%	20.9%	19.5%
Inflows of foreign direct investment ($ billions)	NA	NA	$38.4	$203.5
India				
Exports as a percentage of GDP	6.1%	7.1%	13.0%	19.7%
Inflows of foreign direct investment ($ billions)	NA	NA	$3.6	$42.1
Ireland				
Exports as a percentage of GDP	44.3%	54.6%	94.5%	122.3%
Inflows of foreign direct investment ($ billions)	NA	NA	$25.8	$28.1
Mexico				
Exports as a percentage of GDP	10.1%	18.7%	25.4%	39.2%
Inflows of foreign direct investment ($ billions)	NA	NA	$18.0	$32.7
South Africa				
Exports as a percentage of GDP	34.3%	23.5%	27.2%	30.1%
Inflows of foreign direct investment ($ billions)	NA	NA	$0.9	$5.3
South Korea				
Exports as a percentage of GDP	28.5%	25.3%	35.1%	44.0%
Inflows of foreign direct investment ($ billions)	NA	NA	$9.3	$14.5
United States				

Exhibit 19 (Continued)

	1980	1990	2000	2018
Exports as a percentage of GDP	9.8%	9.3%	10.7%	12.2%
Inflows of foreign direct investment ($ billions)	NA	NA	$159.2	$268.4

Source: OECD (2019).

EXAMPLE 14

Why Some Countries Converge and Others Do Not

As evident from the high rates of growth between 1950 and 2018 shown in Exhibit 18, China and South Korea are converging toward the income levels of the advanced economies. In contrast, the economies of Mexico and South Africa have not converged. Using the data in Exhibits 9 and 19, give some reasons why this has occurred.

Solution:

Two reasons largely account for the difference. First, growth in the Chinese and South Korean economies has been driven by high rates of business investment. As shown in Exhibit 9, investment as a share of GDP in 2018 was 44% in China, almost double the rate of 23.0% in Mexico and more than double the rate of 18% in South Africa. Although investment as a share of GDP in South Korea is lower than in China, it is well above that of Mexico and South Africa.

Second, both China and South Korea have pursued an aggressive export-driven, outward-oriented policy focusing on manufactured goods. In 2018, exports were 44% of GDP for South Korea and 19.5% of GDP for China (Exhibit 19). In addition, foreign direct investment is a major factor underlying growth in China.

On the other hand, South Africa's more inward-oriented economy attracted only $5.3 billion in foreign direct investment in 2018, significantly less than that of Ireland—a smaller but much wealthier and very open country—and the $203 billion inflow of foreign investment into China. These trends are changing, however, as many African and Latin American countries are increasingly relying on growing exports and foreign investment to increase GDP growth.

EXAMPLE 15

Investment Prospects in Spain: Estimating the Sustainable Growth Rate

You are a financial analyst at Global Invest Inc., an investment management firm that runs a number of global mutual funds with a significant exposure to Spain. The IBEX 35 Index, which reached a crisis-induced low of 6,065 in May 2012, remains far below its October 2007 peak of almost 16,000. The members of the investment policy committee at the firm believe the equity market in Spain is attractive and is currently being depressed by temporary problems in the banking and real estate markets of Spain, which they feel are overstated. They believe that higher profits will ultimately drive the market higher but are concerned about the long-term prospects and the sustainable rate of growth for Spain. One of the research assistants at the firm gathers the data shown in Exhibit 20 from the OECD and the Conference Board.

Exhibit 20 Growth Data for Spain

	GDP in Billions of USD Adjusted for PPP	Gross Capital Spending as Percentage of GDP	Consumption of Fixed Capital (percent of GDP)	Labor Cost as Percentage of Total Factor Cost	Total Hours Worked (millions)	Output per Hour Worked in 2011 USD Adjusted for PPP	Growth in Total Factor Productivity (%)
2006	1,390	31.3	14.9	61.4	35,358	48	–0.5
2007	1,481	31.3	15.1	60.5	36,259	48	–0.4
2008	1,527	29.6	15.6	60.2	36,519	48	–1.6
2009	1,484	24.6	16.4	60.5	34,371	50	–1.7
2010	1,501	23.5	16.8	59.1	33,591	51	–0.1
2011	1,517	21.9	17.4	59.1	32,788	51	–0.8
2012	1,501	20.0	18.0	57.3	31,204	52	–1.2
2013	1,501	18.7	17.8	56.5	30,250	53	–1.0
2014	1,551	19.5	17.8	56.1	30,569	53	–0.3
2015	1,625	20.4	17.5	56.4	31,527	54	–0.9

Sources: OECD Stat Extracts and the Conference Board Total Economy Database.

According to the Conference Board website, the physical capital stock for Spain was estimated at $1,808 billion (adjusted for purchasing power parity) in 2005. The research analyst calculated the physical capital stock (*K*) for Spain for the years 2006–2015 using the following equation:

$$K_t = K_{t-1} + I - D,$$

where *I* is gross investment or gross capital spending and *D* is the depreciation or the consumption of fixed capital. So for 2006 and 2007, the physical capital stock is calculated as follows:

$$K_{2006} = \$1{,}808 + \$1{,}390\ (0.313 - 0.149) = \$2{,}036 \text{ billion}$$

$$K_{2007} = \$2{,}036 + \$1{,}481\ (0.313 - 0.151) = \$2{,}276 \text{ billion}$$

The physical capital stock for the remaining years is calculated in the same way and given by Exhibit 21.

Exhibit 21 Estimated Physical Capital Stock (USD billions)

2006	$2,036
2007	2,276
2008	2,490
2009	2,611
2010	2,713

Exhibit 21 (Continued)

2011	2,782
2012	2,812
2013	2,826
2014	2,851
2015	2,898

The investment policy committee asks you to use the preceding data to address the following:

1. Calculate the potential growth rate of the Spanish economy using the production function or growth accounting method (Equation 4), and determine the amount of growth attributed to each source.
2. Calculate the potential growth rate of the Spanish economy using the labor productivity method (Equation 5).
3. How significant are capital deepening and technology in explaining growth for Spain?
4. What is the steady-state growth rate for Spain according to the neoclassical model?
5. Assess the implications of the growth analysis for future economic growth and equity prices in Spain.

Solution to 1:

The production function or growth accounting method estimates the growth in GDP using Equation 4:

$$\text{Growth in potential GDP} = \alpha\Delta K/K + (1-\alpha)\Delta L/L + \Delta A/A$$

The annual growth rate in capital is calculated from Exhibit 21 as

$$(2{,}898/2{,}036)^{1/9} - 1 = 4.0\%.$$

The labor input is measured by the growth rate in total hours worked in the economy (Exhibit 20) and given by

$$(31{,}527/35{,}358)^{1/9} - 1 = -1.27\%.$$

The growth rate in total factor productivity (Exhibit 20) is calculated by using a geometric average of the growth rates for 2000–2009 and is equal to −0.68%. Finally, the labor share of output is given by the average of the labor cost as a percentage of total factor cost, which is 58.7% for 2006–2015 (Exhibit 20). Thus, the share of capital (α) is 1 − 0.587 = 41.3%.

Using these numbers, the growth in potential GDP is

$$\begin{aligned}\text{Growth in potential GDP} &= \alpha\Delta K/K + (1-\alpha)\Delta L/L + \Delta A/A\\ &= (0.413)0.04 + (0.587)(-0.0127) + (-0.0068)\\ &= 0.23\%\end{aligned}$$

Sources of growth for Spain over the period 2006–2015 were as follows:

Capital	(0.413) × (0.04) = 1.65%
Labor	(0.587) × (−0.0127) = −0.75%
TFP	= −0.68%

Solution to 2:

The labor productivity method estimates the growth in GDP using Equation 5:

Growth rate in potential GDP = Long-term growth rate of labor force + Long-term growth rate in labor productivity

As before, we use the growth in total hours worked to measure the growth in the labor force. The growth in labor productivity per hour worked is as follows:

$(54/48)^{1/9} - 1 = 1.32\%$

Growth in potential GDP = -1.27% + 1.32% = 0.05%

Note that the estimate of potential GDP growth using the labor productivity approach is broadly similar to that obtained from the growth accounting method. In general, the two methods are likely to give somewhat different estimates because they rely on different data inputs. The growth accounting method requires measurements of the physical capital stock and TFP. As discussed earlier, TFP is estimated using various time-series or econometric models of the component of growth that is not accounted for by the explicit factors of production. As a result, the estimate of TFP reflects the average (or "smoothed") behavior of the growth accounting residual. The labor productivity approach is simpler, and it avoids the need to estimate the capital input and TFP. In contrast to the estimated value of TFP, labor productivity is measured as a pure residual; that is, it is the part of GDP growth that is not explained by the labor input (and only the labor input). The cost of the simplification is that the labor productivity approach does not allow a detailed analysis of the drivers of productivity growth.

Solution to 3:

Capital deepening occurs in an economy when there is an increase in the capital-to-labor ratio. The labor input for Spain is measured in terms of total hours worked in the economy. Thus, the capital-to-labor ratio for Spain is calculated by dividing the physical capital stock in Exhibit 21 by total hours worked in Exhibit 20. The results, shown in Exhibit 22, indicate that capital deepening was very significant in Spain: The amount of capital per hour worked increased from $57.6 in 2006 to $91.9 in 2015. In terms of the growth rate, the capital-to-labor ratio increased at an annual rate of 5.3%.

The contribution of TFP is measured by the growth in total factor productivity. In contrast to capital deepening, TFP made a negative contribution to growth; the average rate of growth for TFP from 2006 to 2015 was –0.68%. However, TFP is estimated using various statistical techniques, and given the uncertainty around these estimates, it should be viewed with some caution.

Exhibit 22	Estimated Capital-to-Labor Ratio ($/hour worked)
2006	$57.6
2007	62.8
2008	68.2
2009	76.0
2010	80.8
2011	84.8
2012	90.1

Exhibit 22 (Continued)	
2013	93.4
2014	93.3
2015	91.9

Solution to 4:

The steady-state growth rate in the neoclassical model is estimated by the following (see Equation 8):

$\Delta Y/Y = (\theta)/(1 - \alpha) + n$ = Growth rate of TFP scaled by labor factor share + Growth rate in the labor force

Steady-state growth rate = −0.68%/(1 − 0.413) + (−1.27%) = −2.4%

As expected, the growth rate in potential GDP (calculated as in the solutions to 1 and 2) is above the steady-state growth rate. The reason for this is that Spain's economy is still in the process of converging to the higher income levels of the United States and the major economies in Europe. The physical capital stock is below the steady state, and capital deepening is a significant factor increasing productivity growth and the growth in potential GDP. Steady-state growth may be somewhat underestimated in our analysis given that TFP growth is likely to revert to the 1% annual rate of increase exhibited in other major developed economies. This shift is likely to be offset by a lower growth rate in the labor input (see Example 6). It should also be noted that the negative growth in the labor force used in the calculation is based on a period whose start coincides with high level of hours worked and ends in a year when the hours worked were particularly low. The hours worked actually rose subsequent to the 2006–2015 period.

Solution to 5:

The results suggest that potential GDP growth in Spain is close to 0%. As we saw in Exhibit 1, the growth rate of actual GDP since early 2000 has been 0.91% per year, close to the previous estimate of potential but well above the steady state. The problem is that all the growth in potential GDP results from the increase in the labor and capital inputs, with capital deepening being very significant as the capital-to-labor ratio is increasing at a 5.3% annual rate. The neoclassical model suggests that the impact of capital deepening will decline over time and the economy will move toward a steady-state rate of growth. Thus, growth based on capital deepening should not be sustainable over time.

The other major question raised is whether the labor input can continue to decline at an annual rate of 1.3%. We examined this question in Example 6. In sum, potential GDP growth is likely to be negatively influenced over time by Spain's reliance on capital deepening. A positive impact may come from increasing labor input. The reversion of TFP growth to levels more typical of other European economies should also be a positive factor. Even if TFP does rebound, relatively slow growth in potential GDP in Spain will likely restrain future stock price increases.

SUMMARY

This reading focuses on the factors that determine the long-term growth trend in the economy. As part of the development of global portfolio equity and fixed-income strategies, investors must be able to determine both the near-term and the sustainable rates of growth within a country. Doing so requires identifying and forecasting the factors that determine the level of GDP and that determine long-term sustainable trends in economic growth.

- The sustainable rate of economic growth is measured by the rate of increase in the economy's productive capacity or potential GDP.
- Growth in real GDP measures how rapidly the total economy is expanding. Per capita GDP, defined as real GDP divided by population, measures the standard of living in each country.
- The growth rate of real GDP and the level of per capita real GDP vary widely among countries. As a result, investment opportunities differ by country.
- Equity markets respond to anticipated growth in earnings. Higher sustainable economic growth should lead to higher earnings growth and equity market valuation ratios, all other things being equal.
- The best estimate for the long-term growth in earnings for a given country is the estimate of the growth rate in potential GDP.
- In the long run, the growth rate of earnings cannot exceed the growth in potential GDP. Labor productivity is critical because it affects the level of the upper limit. A permanent increase in productivity growth will raise the upper limit on earnings growth and should translate into faster long-run earnings growth and a corresponding increase in stock price appreciation.
- For global fixed-income investors, a critical macroeconomic variable is the rate of inflation. One of the best indicators of short- to intermediate-term inflation trends is the difference between the growth rate of actual and potential GDP.
- Capital deepening, an increase in the capital-to-labor ratio, occurs when the growth rate of capital (net investment) exceeds the growth rate of labor. In a graph of output per capita versus the capital-to-labor ratio, it is reflected by a move along the curve (i.e., the production function).
- An increase in total factor productivity causes a proportional upward shift in the entire production function.
- One method of measuring sustainable growth uses the production function and the growth accounting framework developed by Solow. It arrives at the growth rate of potential GDP by estimating the growth rates of the economy's capital and labor inputs plus an estimate of total factor productivity.
- An alternative method measures potential growth as the long-term growth rate of the labor force plus the long-term growth rate of labor productivity.
- The forces driving economic growth include the quantity and quality of labor and the supply of non-ICT and ICT capital, public capital, raw materials, and technological knowledge.
- The labor supply is determined by population growth, the labor force participation rate, and net immigration. The physical capital stock in a country increases with net investment. The correlation between long-run economic growth and the rate of investment is high.

- Technological advances are discoveries that make it possible to produce more or higher-quality goods and services with the same resources or inputs. Technology is a major factor determining TFP. TFP is the main factor affecting long-term, sustainable economic growth rates in developed countries and also includes the cumulative effects of scientific advances, applied research and development, improvements in management methods, and ways of organizing production that raise the productive capacity of factories and offices.
- Total factor productivity, estimated using a growth accounting equation, is the residual component of growth after accounting for the weighted contributions of all explicit factors (e.g., labor and capital).
- Labor productivity is defined as output per worker or per hour worked. Growth in labor productivity depends on capital deepening and technological progress.
- The academic growth literature is divided into three theories —the classical view, the neoclassical model, and the new endogenous growth view.
- In the classical model, growth in per capita income is only temporary because an exploding population with limited resources brings per capita income growth to an end.
- In the neoclassical model, a sustained increase in investment increases the economy's growth rate only in the short run. Capital is subject to diminishing marginal returns, so long-run growth depends solely on population growth, progress in TFP, and labor's share of income.
- The neoclassical model assumes that the production function exhibits diminishing marginal productivity with respect to any individual input.
- The point at which capital per worker and output per worker are growing at equal, sustainable rates is called the steady state or balanced growth path for the economy. In the steady state, total output grows at the rate of labor force growth plus the rate of growth of TFP divided by the elasticity of output with respect to labor input.
- The following parameters affect the steady-state values for the capital-to-labor ratio and output per worker: saving rate, labor force growth, growth in TFP, depreciation rate, and elasticity of output with respect to capital.
- The main criticism of the neoclassical model is that it provides no quantifiable prediction of the rate or form of TFP change. TFP progress is regarded as exogenous to the model.
- Endogenous growth theory explains technological progress within the model rather than treating it as exogenous. As a result, self-sustaining growth emerges as a natural consequence of the model and the economy does not converge to a steady-state rate of growth that is independent of saving/investment decisions.
- Unlike the neoclassical model, where increasing capital will result in diminishing marginal returns, the endogenous growth model allows for the possibility of constant or even increasing returns to capital in the aggregate economy.
- In the endogenous growth model, expenditures made on R&D and for human capital may have large positive externalities or spillover effects. Private spending by companies on knowledge capital generates benefits to the economy as a whole that exceed the private benefit to the company.
- The convergence hypothesis predicts that the rates of growth of productivity and GDP should be higher in the developing countries. Those higher growth rates imply that the per capita GDP gap between developing and developed economies should narrow over time. The evidence on convergence is mixed.

- Countries fail to converge because of low rates of investment and savings, lack of property rights, political instability, poor education and health, restrictions on trade, and tax and regulatory policies that discourage work and investing.
- Opening an economy to financial and trade flows has a major impact on economic growth. The evidence suggests that more open and trade-oriented economies will grow at a faster rate.

REFERENCES

Ashauer, David. 1990. "Why Is Infrastructure Important?" In Is There a Shortfall in Public Capital Investment? Edited by Alicia Munnell. Federal Reserve Bank of Boston Conference Series No. 34.

Denison, Edward. 1985. Trends in American Growth. Washington, DC: Brookings Institution.

Heston, Alan, Robert Summers, and Bettina Aten. 2009. Penn World Table Version 6.3. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania (August).

IMF. 2019. "Statistical Appendix." In April 2019. Washington, DC: International Monetary Fund.

Jorgenson, Dale. 1966. Technology in Growth Theory. Technology and Growth, Federal Reserve Bank of Boston Conference Series.

Jorgenson, Dale. 2000. Raising the Speed Limit: US Economic Growth in the Information Age. Brooking Papers on Economic Activity.

Levine, R. 2005. "Finance and Growth: Theory and Evidence." In Handbook of Economic Growth, ed. Philippe Aghion and Steven Durlauf. Amsterdam: Elsevier, B.V.

Mankiw, Gregory. 1995. "The Growth of Nations." *Brookings Papers on Economic Activity*.

OECD. 2003. The Sources of Economic Growth in the OECD Countries. Paris: Organisation for Economic Co-Operation and Development.

Romer, Paul. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94 (5): 1002–37.

Solow, Robert. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39 (3): 312–20.

Stewart, Scott, Christopher Piros, and Jeffrey Heisler. 2011. Running Money: Professional Portfolio Management. McGraw-Hill/Irwin.

PRACTICE PROBLEMS

The following information refers to Questions 1–6

Hans Schmidt, CFA, is a portfolio manager with a boutique investment firm that specializes in sovereign credit analysis. Schmidt's supervisor asks him to develop estimates for GDP growth for three countries. Information on the three countries is provided in Exhibit 1.

Exhibit 1 Select Economic Data for Countries A, B, and C

Country	Economy	Capital per Worker
A	Developed	High
B	Developed	High
C	Developing	Low

After gathering additional data on the three countries, Schmidt shares his findings with colleague, Sean O'Leary. After reviewing the data, O'Leary notes the following observations:

Observation 1 The stock market of Country A has appreciated considerably over the past several years. Also, the ratio of corporate profits to GDP for Country A has been trending upward over the past several years and is now well above its historical average.

Observation 2 The government of Country C is working hard to bridge the gap between its standard of living and that of developed countries. Currently, the rate of potential GDP growth in Country C is high.

Schmidt knows that a large part of the analysis of sovereign credit is to develop a thorough understanding of the potential GDP growth rate for a particular country and the region in which the country is located. Schmidt is also doing research on Country D for a client of the firm. Selected economic facts on Country D are provided in Exhibit 2.

Exhibit 2 Select Economic Facts for Country D

- Slow GDP Growth
- Abundant Natural Resources
- Developed Economic Institutions

Prior to wrapping up his research, Schmidt schedules a final meeting with O'Leary to see if he can provide any other pertinent information. O'Leary makes the following statements to Schmidt:

© 2012 CFA Institute. All rights reserved.

Statement 1	Many countries that have the same population growth rate, savings rate, and production function will have growth rates that converge over time.
Statement 2	Convergence between countries can occur more quickly if economies are open and there is free trade and international borrowing and lending; however, there is no permanent increase in the rate of growth in an economy from a more open trade policy.

1 Based on Exhibit 1, the factor that would *most likely* have the greatest positive impact on the per capita GDP growth of Country A is:

A free trade.

B technology.

C saving and investment.

2 Based on Observation 1, in the long run the ratio of profits to GDP in Country A is *most likely* to:

A remain near its current level.

B increase from its current level.

C decrease from its current level.

3 Based on Observation 2, Country C is *most likely* to have:

A relatively low real asset returns.

B a relatively low real interest rate.

C a relatively high real interest rate.

4 Based on Exhibit 2, the *least likely* reason for the current pace of GDP growth in Country D is:

A a persistently strong currency.

B strong manufacturing exports.

C strong natural resource exports.

5 The type of convergence described by O'Leary in Statement 1 is *best* described as:

A club convergence.

B absolute convergence.

C conditional convergence.

6 Which of the following growth models is *most* consistent with O'Leary's Statement 2?

A Classical

B Endogenous

C Neoclassical

The following information relates to Questions 7–15

Victor Klymchuk, the chief economist at ECONO Consulting (EC), is reviewing the long-term GDP growth of three countries. Klymchuk is interested in forecasting the long-term change in stock market value for each country. Exhibit 1 presents current country characteristics and historical information on select economic variables for the three countries.

Exhibit 1 Select Country Factors and Historical Economic Data

		2009–2019			
	Country Factors	Growth in Hours Worked (%)	Growth in Labor Productivity (%)	Growth in TFP (%)	Growth in GDP (%)
Country A	■ High level of savings and investment ■ Highly educated workforce ■ Low tariffs on foreign imports ■ Limited natural resources	0.9	2.4	0.6	3.3
Country B	■ Developed financial markets ■ Moderate levels of disposable income ■ Significant foreign direct and indirect investments ■ Significant natural resources	–0.3	1.6	0.8	1.3
Country C	■ Politically unstable ■ Limited property rights ■ Poor public education and health ■ Significant natural resources	1.8	0.8	–0.3	2.6

Klymchuk instructs an associate economist at EC to assist him in forecasting the change in stock market value for each country. Klymchuk reminds the associate of the following:

Statement 1 "Over short time horizons, percentage changes in GDP, the ratio of earnings to GDP, and the price-to-earnings ratio are important factors for describing the relationship between economic growth and stock prices. However, I am interested in a long-term stock market forecast."

A client is considering investing in the sovereign debt of Country A and Country B and asks Klymchuk his opinion of each country's credit risk. Klymchuk tells the client the following:

Statement 2 "Over the next 10 years, I forecast higher potential GDP growth for Country A and lower potential GDP growth for Country B. The capital per worker is similar and very high for both countries, but per capita output is greater for Country A."

The client tells Klymchuk that Country A will offer 50-year bonds and that he believes the bonds could be a good long-term investment given the higher potential GDP growth. Klymchuk responds to the client as follows:

Statement 3 "After the next 10 years, I think the sustainable rate of economic growth for Country A will be affected by a growing share of its population over the age of 65, a declining percentage under age 16, and minimal immigration."

The client is surprised to learn that Country C, a wealthy, oil-rich country with significant reserves, is experiencing sluggish economic growth and asks Klymchuk for an explanation. Klymchuk responds by stating:

Statement 4 "Although countries with greater access to natural resources are often wealthier, the relationship between resource abundance and economic growth is not clear. My analysis shows that the presence of a dominant natural resource (oil) in Country C is constraining growth. Interestingly, Country A has few natural resources but is experiencing a strong rate of increase in per capita GDP growth."

Klymchuk knows that growth in per capita income cannot be sustained by pure capital deepening. He asks the associate economist to determine how important capital deepening is as a source of economic growth for each country. Klymchuk instructs the associate to use the data provided in Exhibit 1.

Klymchuk and his associate debate the concept of convergence. The associate economist believes that developing countries, irrespective of their particular characteristics, will eventually equal developed countries in per capita output. Klymchuk responds as follows:

Statement 5 "Poor countries will only converge to the income levels of the richest countries if they make appropriate institutional changes."

7 Based on the country factors provided in Exhibit 1, the country *most likely* to be considered a developing country is:

- **A** Country A.
- **B** Country B.
- **C** Country C.

8 Based on Exhibit 1, capital deepening as a source of growth was *most* important for:

- **A** Country A.
- **B** Country B.
- **C** Country C.

9 Based on Statement 1, over the requested forecast horizon, the factor that will *most likely* drive stock market performance is the percentage change in:

A GDP.

B the earnings to GDP ratio.

C the price-to-earnings ratio.

10 Based solely on the predictions in Statement 2, over the next decade Country B's sovereign credit risk will *most likely*:

A increase.

B decrease.

C not change.

11 Based on Statement 2, the difference in per capita output between Country A and Country B *most likely* results from differences in:

A capital deepening.

B capital per worker.

C total factor productivity.

12 Based on Statement 3, after the next 10 years, the growth rate of potential GDP for Country A will *most likely* be:

A lower.

B higher.

C unchanged.

13 Based on Statement 4 and Exhibit 1, the sluggish economic growth in Country C is *least likely* to be explained by:

A limited labor force growth.

B export driven currency appreciation.

C poorly developed economic institutions.

14 Based on Statement 4, the higher rate of per capita income growth in Country A is *least likely* explained by the:

A rate of investment.

B growth of its population.

C application of information technology.

15 The type of convergence described by Klymchuk in Statement 5 is *best* described as:

A club convergence.

B absolute convergence.

C conditional convergence.

The following information relates to Questions 16–21

At an international finance and economics conference in Bamako, Mali, Jose Amaral of Brazil and Lucinda Mantri of India are discussing how to spur their countries' economic growth. Amaral believes that growth can be bolstered by removing institutional impediments and suggests several possibilities for Brazil: launching a rural literacy program, clarifying property rights laws, and implementing a new dividend tax on foreign investors.

Mantri responds that for India, capital deepening will be more effective, and she has proposed the following ideas: building a group of auto and textile factories in the southern states, developing a north–south and east–west highway network, and sponsoring a patent initiative.

In response, Amaral says to Mantri:

> "Based on endogenous growth theory, one of those proposals is more likely to raise total factor productivity than result in pure capital deepening."

Although Mantri recognizes that India lacks the significant natural resources that Brazil has, she states that India can overcome this challenge by bolstering long-term growth through three channels:

Channel 1	Deepening the capital base
Channel 2	Making investments in technology
Channel 3	Maintaining a low rupee exchange rate

Each country's basic economic statistics were presented at the conference. Selected data for Brazil and India are presented in Exhibit 1. Adama Kanté, a fund manager based in Mali, is planning to increase the fund's allocation to international equities and, after some preliminary analysis, has determined the new allocation will be to Brazilian or Indian equities. After reviewing the data in Exhibit 1, Kanté decides that the allocation will be to Indian equities.

Exhibit 1 Economic Statistics, Brazil and India

Economic Statistic	Brazil	India
GDP per capita, 2018	$14,360	$6,999
GDP per capita growth, 2000–2018	1.3%	5.8%
GDP growth, 2005–2018	2.0%	7.2%
- Growth resulting from labor productivity component	1.3%	6.3%
- Growth resulting from capital deepening component	1.9%	4.5%

Kanté is concerned about the low standard of living in Mali and its large informal sector. To improve per capita GDP, Kanté is considering five specific strategies:

Strategy 1	Lower the country's tax rate.
Strategy 2	Introduce policies that encourage the return of highly educated Malian emigrants.
Strategy 3	Build daycare centers to increase women's participation in the workforce.
Strategy 4	Impose high tariffs on imports to protect the country's nascent industries.
Strategy 5	Use economic development bank loans to improve the country's transport and manufacturing infrastructure.

16 Which of Amaral's initiatives is *least likely* to achieve his stated growth objective?

A Dividend tax

B Rural literacy

C Property rights

17 Which proposal for India is Amaral *most likely* referring to in his response to Mantri?

A Patent initiative

B Highway network

C Auto and textile factories

18 The channel that is *least likely* to help India overcome its challenge of lacking significant natural resources is:

A Channel 1.

B Channel 2.

C Channel 3.

19 Based on Exhibit 1, which Indian economic statistic *least likely* supports Kanté's international equity allocation preference?

A GDP per capita

B Growth resulting from labor productivity

C Growth resulting from capital deepening

20 The strategy that is *least likely* to improve per capita GDP in Mali is:

A Strategy 1.

B Strategy 2.

C Strategy 3.

21 Which of the following strategies being considered by Kanté is *most likely* to undermine or delay convergence with developed economies?

A Strategy 2

B Strategy 4

C Strategy 5

SOLUTIONS

1 B is correct. Country A is a developed country with a high level of capital per worker. Technological progress and/or more intensive use of existing technology can help developed countries increase productivity and thereby increase per capita GDP. Most developed countries have reasonably low trade barriers; thus, somewhat freer trade is likely to have only an incremental, and probably transitory, impact on per capita GDP growth. Also, because the country already has a high capital-to-labor ratio, increased saving/investment is unlikely to increase the growth rate substantially unless it embodies improved technology.

2 C is correct. The ratio of profits to GDP for Country A has been trending upward over the past several years and is now well above its historical average. The ratio of profits to GDP cannot rise forever. At some point, stagnant labor income would make workers unwilling to work without an increase in wages and would also undermine demand, making further expansion of profit margins unsustainable. Thus, the ratio of profits to GDP will likely decline in the long run toward its historical average.

3 C is correct. A high growth rate of potential GDP would cause real incomes to rise more rapidly and also translate into higher real interest rates and higher expected/required real asset returns. The real interest rate is essentially the real return that consumers/savers demand in exchange for postponing consumption. Faster growth in potential GDP means that consumers expect their real income to rise more rapidly. This implies that an extra unit of future income/consumption is less valuable than it would be if income were expected to grow more slowly. All else the same, the real interest rate will have to be relatively high in order to induce the savings required to fund required/desired capital accumulation.

4 B is correct. Country D is a country with abundant resources and has developed the economic institutions necessary for growth, yet the country is experiencing slow economic growth. It is likely that Country D is experiencing the Dutch Disease, whereby currency appreciation driven by strong export demand for natural resources makes other segments of the economy, in particular manufacturing, globally uncompetitive. Strong manufacturing exports indicate that Country D is globally competitive and likely to have adopted leading edge technology. Thus, it is unlikely that the slow growth reflects inability to maintain productivity growth. Similarly, strong exports suggest adequate demand for its products. Thus, strong exports are unlikely to be the cause of slow growth.

5 C is correct. Conditional convergence means that convergence is conditional on the countries having the same savings rate, population growth rate, and production function. If these conditions hold, the neoclassical model implies convergence to the same *level* of per capita output as well as the same steady-state growth rate.

6 C is correct. According to the neoclassical model, convergence should occur more quickly if economies are open and there is free trade and international borrowing and lending. Opening up the economy should increase the rate at which the capital-to-labor ratio converges among countries. In the neoclassical Solow model, however, after the reallocation of world savings, there is no permanent increase in the rate of growth in an economy. Both the developed and developing countries eventually grow at the same steady-state rate.

7 C is correct. Country C is the most likely to be a developing economy. Political instability, limited property rights, and poor public education and health are all factors that limit economic growth and thereby contribute to a relatively low standard of living.

8 A is correct. The associate economist can measure the effect of pure capital deepening by measuring the difference of the growth rates of labor productivity and total factor productivity. The larger the difference, the more important capital deepening is as a source of economic growth. From 2000–2010, Country A's labor productivity grew by 2.4% per year, of which 0.6% came from TFP growth and 1.8% from capital deepening (2.4% – 0.6% = 1.8%).

9 A is correct. In the long run, the GDP growth rate is the most important driver of stock market performance. Therefore, the associate economist should focus on the drivers of long-run potential GDP growth. The ratio of earnings to GDP cannot increase indefinitely because that would imply that profit would eventually absorb all of GDP. This ratio cannot shrink forever, either, because unprofitable companies will go out of business. Thus, the annualized growth rate of the earnings to GDP ratio must be approximately zero over long time horizons, and this ratio should not be a dominant factor in forecasting long-term stock market performance. Similarly, the price-to-earnings ratio cannot grow or contract at a finite rate forever because investors will not pay an excessive price for each dollar of earnings, nor will they give away earnings for free. Therefore the rate of change in the price-to-earnings ratio must be approximately zero over long time horizons and should not be a dominant factor in the forecast of long-term stock market performance.

10 A is correct. Credit rating agencies consider the growth rate of potential GDP when evaluating the credit risk of sovereign debt. The chief economist's expectation for lower potential GDP growth for Country B over the next decade increases the perceived credit risk of its sovereign bonds.

11 C is correct. The higher per capita output for Country A is most likely the result of differences in the cumulative impact of technological progress embodied in total factor productivity. Technological progress raises a country's productive capacity and causes an upward shift in the entire production function, resulting in higher output per worker for a given level of capital per worker.

12 A is correct. Demographic factors can positively or negatively contribute to a country's sustainable rate of economic growth. After the next 10 years, Country A is expected to experience a growing share of the population over the age of 65 and a declining percentage of the population under the age of 16. All else the same, this implies slower growth of the labor force and hence slower growth of potential GDP. Immigration could offset these demographic challenges. However, Statement 3 indicates that Country A is expected to experience minimal immigration.

13 A is correct. Country C is an example of a country endowed with an abundant natural resource yet experiencing slow economic growth. Although labor force growth is an important source of economic growth, it is the least likely explanation of the sluggish economic growth in Country C. As shown in Exhibit 1, growth in total hours worked has accounted for most of Country C's growth. Furthermore, export driven currency appreciation and poorly developed economic institutions are both likely causes of sluggish growth in countries with abundant natural resources.

14 B is correct. Population growth can increase the growth rate of the overall economy but does not affect the rate of increase in *per capita* GDP. Therefore, population growth does not explain Country A's higher rate of per capita income growth. An increase in labor force participation could, however, raise the growth rate of per capita GDP.

15 A is correct. Klymchuk is referring to the concept of club convergence. The basic premise is that lower-income members of the club are converging to the income levels of the richest countries. This implies that the countries with the lowest per capita income in the club grow at the fastest rate. Countries outside the club, however, continue to fall behind.

16 A is correct. Amaral's initiative to implement a new dividend tax is likely to impede inflows of equity capital by making equity investment in Brazil less attractive for foreign investors. Capital flows, or lack thereof, have a major impact on economic growth because, in an open economy, world savings can finance domestic investment. As a potential source of funds, foreign investment breaks the vicious cycle of low income, low domestic savings, and low investment.

17 A is correct. Mantri's proposal to sponsor a patent initiative, which is likely to result in technology investment and improvement, is likely to cause a proportional upward shift in the entire production function, allowing the economy to produce higher output per worker for a given level of capital per worker. Technological progress also increases the marginal product of capital relative to its marginal cost.

18 C is correct. Maintaining a low currency exchange rate is a policy aimed at maintaining demand for the country's exports. It would have little direct impact on the potential growth rate of aggregate supply. It might boost long-term capacity growth indirectly, however, by encouraging adoption of leading edge technology. Nonetheless, it would not be expected to be as powerful as capital deepening and/or investment in technology.

19 A is correct. Kanté's decision to invest in equities in India is supported by the country's strong economic growth. For global investors, economic growth is important because equity composite valuations depend to a great extent on both the level of economic output (GDP per capita and GDP overall) and on the rate of economic growth. Relative to Brazil, India's growth rate in per capita GDP has been much higher, and furthermore, the growth rate in GDP has also been much higher than that of Brazil. In contrast to the growth rate, the relatively low *level* of GDP per capita in India is less likely to indicate attractive equity investment opportunities. Low per capita GDP suggests that India may lack sufficient industrial and financial infrastructure to support some types of industries. It also indicates that domestic purchasing power is relatively limited, decreasing the potential for higher-margin, domestically oriented businesses.

20 A is correct. With Mali's low standard of living, i.e., GDP per capita and large informal workforce, the tax rate is unlikely to be an impediment to growth, so lowering the tax rate is not likely to be a major contributor to growth.

21 B is correct. The strategy for Mali to impose high tariffs (trade restrictions) on imports is likely to undermine rather than enhance growth and therefore is not supportive of convergence with developed economies. Freer trade (fewer trade restrictions) tends to enhance growth by, for example, inducing a shift of resources into industries in which the country has a comparative thereby increasing overall productivity; forcing less efficient domestic companies to exit and more efficient ones to innovate; allowing domestic producers to more

fully exploit economies of scale by selling to a larger market; and enabling less advanced sectors of an economy to catch up with more-advanced countries or sectors through knowledge spillovers.

READING

8

Economics of Regulation

by Chester S. Spatt, PhD

Chester S. Spatt, PhD, is at Tepper School of Business at Carnegie Mellon University (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** describe the economic rationale for regulatory intervention;
☐	**b.** explain the purposes of regulating commerce and financial markets;
☐	**c.** describe anticompetitive behaviors targeted by antitrust laws globally and evaluate the antitrust risk associated with a given business strategy;
☐	**d.** describe classifications of regulations and regulators;
☐	**e.** describe uses of self-regulation in financial markets;
☐	**f.** describe regulatory interdependencies and their effects;
☐	**g.** describe tools of regulatory intervention in markets;
☐	**h.** describe benefits and costs of regulation;
☐	**i.** describe the considerations when evaluating the effects of regulation on an industry.

INTRODUCTION

1

Regulation can be described as a form of government intervention in markets that involves rules and their enforcement. It is an important topic because regulation has potential effects not only at the macro level on the economy but also at the micro level on companies and individuals. Regulation may develop either proactively in anticipation of consequences of changes in the market environment or reactively in response to some occurrence(s). For example, changes that resulted from technological advances in the markets because of new means of communication and applications of computers have led to a variety of regulation, both proactive and reactive. Regulation has also developed in response to financial crises and undesirable behaviors or actions

© 2020 CFA Institute. All rights reserved.

that have occurred in the past. Regulations are necessary because in some situations market solutions are inadequate. In other words, regulations exist to protect end users from market failings.

A significant challenge for financial regulators is how to deal with systemic risk (the risk of failure of the financial system) and the consequences of risk taking by financial institutions. Issues such as labor regulation, environmental regulation, and electronic privacy are also receiving increased attention.

How regulations are developed and applied can have significant impacts on businesses. Changes in regulatory framework and regulatory uncertainty can also have substantial effects on business decisions. So, one of the significant challenges facing professionals in the finance industry is to anticipate and understand the consequences of potential changes in the regulatory environment and of specific regulations.

In the following sections, we describe the economic rationale of regulation, including how regulation improves fairness in markets and addresses the danger to society of financial system failure. We also provide an overview of regulators, the tools at their disposal, and how the work of regulators around the globe is interdependent. Lastly, we describe the assessment of costs and benefits of regulation and highlight practical issues that arise from the implementation of regulation.

2 ECONOMIC RATIONALE FOR REGULATION

a describe the economic rationale for regulatory intervention;

Regulations are necessary because market solutions are not adequate for all market situations. Conceptually, this need can be understood best using ideas from economic theory. One of the basic principles in economics is the "fundamental theorem of welfare economics." Assuming constant returns to scale, no frictions (such as costs for or restraints on trading and asymmetrical information), and no externalities, competitive market (equilibrium) allocations are efficient, or *Pareto optimal*. Note that market (equilibrium) allocations are ones in which i) agents maximize utility given relative prices and ii) markets clear. In such a scenario, there is no way to redistribute resources and make some agents better off without making others worse off. (Note that if resources can be redistributed such that any one agent can be made better off without making any other agent worse off, then the original allocation would not be *Pareto optimal.*) Furthermore, any efficient allocation of resources can be sustained as a market equilibrium for an appropriate set of prices. Hence, absent frictions and externalities, the market solution will be economically efficient and regulation would be needed only to ensure consumer protection and privacy rights.

The case for regulatory intervention rests on the presence of **informational frictions**, externalities, weak competition, and social objectives. Informational frictions are market inefficiencies that lead to sub-optimal outcomes. They include lack of access to information and inadequate information. Such frictions result in a variety of issues that regulators attempt to address. These issues include "adverse selection" (private information in the hands of some, but not all, market participants that allows the holder of that information to gain at the expense of others) and "moral hazard" (incentive conflicts that arise from the delegation of decision making to agents or from contracts that will affect the behavior of one party to the contract to the detriment of the other party). Asymmetrical information, in general, may give one entity an inherent advantage over another entity with which interaction occurs. The resulting regulation focuses on establishing rights and responsibilities of entities and on seeking to establish a level playing field in the dissemination of information in the market.

Externalities are spillover effects of production and consumption activities onto others who are not directly involved in a particular transaction, activity, or decision. A positive externality provides a spillover benefit, and a negative externality generates a spillover cost. Systemic risk posed by failures of financial institutions is an example of an externality, as is environmental pollution. Both can have far-reaching consequences for the public. One example of a positive externality would be home improvements, whereby neighbors may benefit from increases in their home values resulting from renovations to a nearby house, even though they have expended no resources to improve their own properties.

Weak competition can also give rise to regulatory intervention. Weak competition is considered to be detrimental to consumers owing to high prices, less choice, and lack of innovation. It is associated with scenarios in which a dominant firm has significant market power or in which firms collude and agree to keep prices high.

Social objectives are typically resolved by the provision of public goods that the market would not otherwise provide. An important feature of public goods is that consumption by one individual does not reduce the availability of the good for others. Usually funded by the government, public goods include defense, police protection, and education. Alternatively, social objectives may be achieved by placing regulatory obligations on firms—for example, by requiring energy companies to give discounts on energy bills to vulnerable customers or by requiring telecommunication companies to provide service to remote customers who would otherwise not be served because of the additional costs of providing such service.

It is difficult, if not impossible, to think of an area of life unaffected by regulation. Regulations address a broad range of issues and can be classified by their objectives. These include the following:

- Safety (for example, food and products)
- Privacy (for example, financial information)
- Protection (for example, intellectual property)
- Environmental (for example, pollution)
- Labor or employment (for example, workers' rights and employment practices)
- Commerce or trade (for example, consumers' rights and protection, investors' protection, and antitrust)
- Financial system (for example, prudential supervision of institutions, capital requirements, and insider trading)

2.1 Rationale for the Regulation of Financial Markets

The regulation of securities markets and financial institutions is essential because of the consequences to society of failures in the financial system. These consequences could be experienced at both micro and macro levels. Potential consequences include individual financial losses experienced by individuals, an overall loss of confidence, and disruption to commerce. These consequences were evident in the 2008 global financial crisis. Securities regulation focuses on such goals as protecting investors, creating confidence in markets, and encouraging capital formation. Although it is difficult to define precisely how regulation enhances confidence in the financial system, increasing confidence is cited as one of the motives for securities regulation. Many of the rules oriented toward transparency, equitable access to information (which, in turn, encourages capital formation), and protecting small investors implicitly serve to promote confidence in the markets.

Among the objectives of many financial regulators is the protection of consumers and investors, the safety and soundness of financial institutions, the smooth operation of payment systems, and access to credit. Other (macroeconomic) concerns of financial regulators, particularly central banks, include price stability, levels of employment/unemployment, and economic growth.

A key focus of regulators is maintaining the integrity of markets, ensuring that they operate efficiently and that consumers and investors are informed and not exploited. This role is distinct from financial stability regulation, which is focused on specific outcomes. In addition to securities registration requirements, disclosure requirements are important to facilitate and support the marketplace and the confidence of investors. Disclosures allow investors to use available information to assess the consequences for investing in and valuing financial instruments and to allow markets to operate. Securities market disclosures occur at various levels, in various forms, and with varied and sometimes unexpected consequences. For example, in the European Union, the Markets in Financial Instruments Directive II (MiFID II), implemented in 2018, focuses on improving transparency in financial markets, including in fixed income, derivatives, and other over-the-counter markets in which prices and volumes were not previously publicly disclosed. The opacity of these markets meant that the buildup of risks prior to the 2008 financial crisis went largely undetected by regulators and market participants.

Disclosures are wide ranging and have high importance. They include financial reporting requirements and accounting standards, prospectus disclosure requirements in conjunction with both securities offerings and annual reports, disclosure requirements in the context of proxy proposals and contests, mutual fund disclosure rules, and financial market price transparency rules. Disclosure requirements tend to be oriented toward the protection of investors and the provision of information to investors (either to investors directly or to their service providers).

Many of the regulations governing securities markets are designed to mitigate agency problems that arise through delegation to intermediaries. For many financial transactions, parties need to act through others (agents), leading to the potential for agency conflicts. Among examples of regulations addressing potential agency conflicts are those related to mutual fund fees and governance, the governance of listed companies, rules for proxy voting in companies, best execution requirements for broker/dealers, and treatment of inducements (commissions and other non-monetary benefits) that arise in the provision of investment advice and in portfolio management. For example, MiFID II requires advisers to disclose all costs and charges, including all one-off and ongoing charges, transaction costs associated with the financial instrument, all third-party payments received, and the total combined costs of these three categories. These disclosures must also be accompanied by an illustration that shows the cumulative effect of the overall costs and charges on the return to investors.

Historically, securities regulators have tended to focus primarily on protecting retail investors (individual investors with modest resources and less investment expertise than professional investors). This tendency has resulted in a lesser focus on financial regulation of hedge funds, private equity, and venture capital funds because of the type of investors (institutional and affluent individual investors) that invest in these funds. For these larger investors, regulators have taken more of a "buyer beware" stance. For larger investors, it is more difficult to define suitability standards. One approach is to require a more modest range of disclosure requirements related to offering memorandums for a variety of different types of transactions, alongside basic antifraud rules.

Regulations related to prudential supervision of financial institutions and financial stability are critical because of the cost that failure of a financial institution can impose on the economy, capital markets, and society. Prudential supervision is regulation and monitoring of the safety and soundness of financial institutions in order to promote financial stability, reduce system-wide risks, and protect customers of financial institutions. The failure of a bank can result in loss of savings and access to credit. The failure of an insurance company can result in unanticipated losses to those insured. If government-sponsored entities provide protection against these losses or the government chooses to cover all or a portion of these losses, the losses can be spread across a broader section of society than simply those directly affected. Additionally, the resulting loss of confidence in the financial system can have far-reaching consequences. Note that prudential supervision is part of prudential regulation, which includes setting capital adequacy and liquidity standards for such financial institutions as banks and insurers.

Types of prudential supervision include those that focus on diversifying assets, managing and monitoring risk taking, and ensuring adequate capitalization. In addition, regulators may set up funds to provide insurance against losses and mandate premiums or fees to be paid into these funds. Some regulators, such as those in the European Union, may also require that designated investment firms have in place appropriate recovery plans and resolution plans to be applied if they encounter financial distress.

The benefits of regulation, however, generally come with associated costs. For example, regulations that require certain entities or individuals to use insurance when undertaking certain activities may create moral hazard and result in greater risk-taking incentives. Similarly, regulations that increase capital-holding requirements can reduce the amount of capital available to be distributed in the market.

EXAMPLE 1

Rationale for Regulation

1 Which of the following is least likely to be a reason for the use of regulation?
   - A Systemic risk posed by the financial services industry
   - B Informational frictions in the form of private information
   - C Extensive disclosure of operating and financial information by companies seeking to attract investors' attention

2 Prudential supervision is primarily concerned with:
   - A treatment of inducements in the provision of investment advice.
   - B provision of information about financial products to retail investors.
   - C safety and soundness of the financial system.

Solution to 1:

C is correct. Extensive disclosures are not a reason for the use of regulation. They may be the result of regulation or a reflection of good business practices. Presence of systemic risk and informational frictions give rise to the need for regulation.

Solution to 2:

C is correct. The primary objective of prudential regulation is to ensure safety and soundness of the financial system.

3 REGULATION OF COMMERCE

b explain the purposes of regulating commerce and financial markets;

Government regulation in certain areas of commerce, such as consumer protection, commercial law, and antitrust, is critical to setting out an underlying framework for the operation of private markets and facilitating business decisions that involve a considerable degree of coordination.

Issues pertaining to externalities and public goods (for example, national defense and transportation infrastructure) are critical to the operation of national and global economies and are essential considerations for the work of investment analysts. The relevant decisions arise at a number of levels. Many of these decisions would be within the domain of national governments, but some of the relevant externalities are global.

Although common examples involve local environmental issues, such as pollution, global externalities—such as nuclear waste storage and global warming—occur across countries. So it is important to have international mechanisms to facilitate the coordination and acceptance of responsibilities among national governments (typically, national governments are best able to coordinate decisions within their respective countries). Some of these externalities have long-term consequences (costs) and implications. In most cases, these long-run consequences may be difficult to fully quantify and assess.

Government policy can be important for promoting commerce locally, nationally, regionally, and globally. Trade agreements are important to global commerce. At the national, regional, or local level, governments can facilitate basic features of the business environment, such as establishing the legal framework for contracting and setting standards. Regulation is also central to fundamental aspects of labor markets, such as workers' and employers' rights and responsibilities, as well as workplace safety. Immigration issues are also handled through regulation. Fundamental safety regulations with respect to drugs, food products, medical devices, and pollution are significant too.

Several issues have emerged as particularly relevant in the context of globalization and the internet. One issue is the recognition and protection of intellectual property. Government policies regulate intellectual property, prescribing standards and processes that define and govern patents, trademarks, and copyrights. The legal standards are country specific, and although most countries recognize the importance of protecting intellectual property, lack of enforcement and protection of intellectual property at a global level has emerged as an issue. Setting common technical standards is another global issue, given the focus on ensuring higher levels of interoperability between the technology and electronic tools used in commerce. Even something as basic as establishing domain names and the related standard setting requires some appropriate delegation of authority.

Technological change, including a shift toward digitization, is leading to an increasing amount of data being collected, processed, shared, and used in digital form at lower cost and on a larger scale than ever before. "Big Data" gives rise to potential market opportunities, such as the development of data-based business models that rely on the sharing of data and the extraction of commercial value from data. However, it is also giving rise to concerns about privacy and data protection. Privacy is particularly important with respect to medical, financial, academic, and employment records. New regulations, such as the General Data Protection Regulation (GDPR) and the proposed e-Privacy Regulation in the European Union, require entities, including businesses and governments, to apply certain protections and safeguards to personal data in their possession and maintain appropriate security procedures. The internet raises a broad set of issues involving privacy because of the breadth of information potentially available about a person's situation (financial and personal), activities, interactions, and purchases. How internet software navigates these privacy concerns will influence both the perceptions and actions of regulators, as well as the acceptance of software innovations and business models in the marketplace.

An effective legal environment is also crucial for the successful operation of commerce. Clearly defined rules governing contracts, their interpretation, and each party's legal rights under a contract are necessary. A framework for financial liability and dealing with bankruptcy is also necessary as an incentive to enter into economic contracts, particularly those that require long-term commitments. Such activities as construction projects, energy exploration, and extraction projects—and even mundane commercial activities, such as relocation decisions—involve significant long-term, dynamic commitments. Pre-commitment by society to a well-defined set of rules and standards is crucial to facilitating the willingness of market participants to engage in long-term commercial activities.

For example, consider the situation in which a company needs to incur significant costs to start a project. These costs are unrecoverable if the project does not progress; in other words, these are sunk costs. Without a strong legal framework to guarantee that the party will recover these initial costs, the party paying the sunk costs would be reluctant to incur them because of the potential of a "holdout" problem in which the other side exploits the fact that the sunk costs have been incurred to force a renegotiation of the deal. Such contractual difficulties would destabilize the operation of businesses and weaken the economy.

4 ANTITRUST REGULATION AND FRAMEWORK

c describe anticompetitive behaviors targeted by antitrust laws globally and evaluate the antitrust risk associated with a given business strategy;

In a global context, an implicit regulatory goal of government may be to restrict competition from other countries. In a domestic context, a regulatory goal often pursued is to promote competition in most economic sectors (this goal can alternatively be viewed as monitoring and preventing activities that restrict or distort competition). There are several dimensions to this goal. Regulatory approval or notification is typically required for mergers and acquisition of major companies in a specific market. When a merger or acquisition is expected to substantially reduce competition, regulators can block the merger or acquisition or suggest remedies to resolve a perceived issue (for example, divestiture of particular segments of the businesses to resolve an antitrust issue). When there are competing bids, the regulator can effectively decide the outcome based on its assessment of each bid's effects. Considering the potential response of competition or antitrust agencies is a central aspect to the evaluation of mergers and acquisitions.

Competition and antitrust laws also typically prohibit anticompetitive arrangements or practices, such as price collusion or exchanging certain information, and anticompetitive behavior by companies that dominate a market. Types of behavior that are problematic when undertaken by a dominant company (beyond mergers that substantially lessen competition) include exclusive dealings and refusals to deal, price discrimination, and engaging in predatory pricing. In response to antitrust issues, regulators not only may impose monetary sanctions but also may require companies to change their business (for example, divest portions or change operating/marketing practices). In some jurisdictions, such as the United States, the United Kingdom, Germany, Denmark, Ireland, France, and Australia, individuals can also face imprisonment for engaging in a cartel.

There has been an increasing focus on applying antitrust laws to the technology sector, which includes investigations of Google (owned by a holding company, Alphabet Inc.), Apple, Intel, and Microsoft. In Europe, for example, Google was found to have abused its dominant position in the internet search market by favoring its own comparison shopping service over those of its rivals. Similarly, Intel was found to have abused its dominant position by engaging in exclusive dealing with certain computer equipment manufacturers and retailers. Both companies challenged the claims on the basis that they have brought considerable benefits to consumers in terms of low prices and ever-improving quality. Using competition laws to challenge rivals can also represent a business strategy. An example of such a challenge is Microsoft's challenge in Europe that Google is unfairly impeding competition in the search engine market.

A significant issue that companies need to face in addressing antitrust (lack of competition) issues is that in many cases, they need to satisfy simultaneously a range of regulators across multiple jurisdictions. For example, a company may have to satisfy both the US Department of Justice and the European Union if it plans to use a common product and market strategy across jurisdictions. Despite language and cultural differences, it often is advantageous to adopt a unified strategy around the globe because of business imperatives and likely overlapping views among regulators of competition.

EXAMPLE 2

Antitrust Regulation

Which of the following issues is least likely to be the subject of antitrust rules?

A Privacy and data protection

B Anticompetitive behavior by dominant companies in a market

C Mergers and acquisitions by major companies

Solution:

A is correct. Privacy and data protection issues are regulated, but only as part of regulations besides antitrust rules.

5 CLASSIFICATION OF REGULATIONS AND REGULATORS

d describe classifications of regulations and regulators;

e describe uses of self-regulation in financial markets;

Regulations are sometimes enacted by legislative bodies (often these regulations are laws) but more typically arise from the determination of regulatory bodies.

5.1 Classification of Regulations and Regulators

Broadly speaking, regulators can be either sanctioned by the government or created by an industry on a voluntary basis.

Government-backed regulatory bodies can be either governmental departments and agencies or independent regulators, which derive their power and authority from the state. Government-backed regulatory bodies have legal authority to enact and enforce regulations within the parameters of the mandate given to them. In many instances, a legislative body enacts a statute at a broad level, leaving it to regulatory bodies to implement and apply the detail of the regulation.

Illustration of Regulatory Process

This description by the US Securities and Exchange Commission (SEC) is illustrative of how the process works: "Rulemaking is the process by which federal agencies implement legislation passed by Congress and signed into law by the President. Major pieces of legislation, such as the Securities Act of 1933, the Securities Exchange Act of 1934, and the Investment Company and Investment Adviser Acts of 1940, provide the framework for the SEC's oversight of the securities markets. These statutes are broadly drafted, establishing basic principles and objectives. To ensure that the intent of Congress is carried out in specific circumstances—and as the securities markets evolve technologically, expand in size, and offer new products and services—the SEC engages in rulemaking." (www.sec.gov/about/whatwedo.shtml)

Courts play an important role in regulation as well—helping interpret regulations and laws, defining permitted and proscribed regulatory practices, and, in some instances, imposing sanctions for regulatory violations. State-backed regulations can, therefore, be classified as comprising

- laws enacted by legislative bodies (**statutes**),
- rules issued by government agencies or other regulators (**administrative regulations or administrative law**), and
- interpretations of courts (**judicial law**).

Although government departments and agencies make many regulations, **independent regulators** can also make regulations in accordance with their powers and objectives. The authority of independent regulators, such as the Financial Conduct Authority in the United Kingdom, comes from their recognition, autonomy, and powers given to them by a statute, government department, or government agency, but they are not government agencies per se. One distinction between government agencies and independent regulators is that the latter typically do not rely on government funding and are often given a degree of autonomy in terms of decision making. Some argue that an advantage of independent regulators is that they are to some extent immune from political influence and pressure and can, therefore, take a more technical and long-term view of policies, which would achieve the objectives they have been created to pursue.

In contrast to state-backed government agencies or independent regulators, industry **self-regulatory bodies** are private organizations that both represent and regulate their members. Although these organizations are independent of the government and to an extent are isolated from political pressure, they may be subject to pressure from their members. Industry self-regulatory bodies derive authority from their members, who agree to comply with the organization's rules and standards and their enforcement. This authority does not have the force of law, but industry self-regulatory bodies do have the power to exclude or expel parties from being members. To ensure minimum standards are maintained, certain entry requirements (such as training or ethical standards) may be imposed.

Some industry self-regulatory bodies, particularly in the securities industry, are known as **self-regulating organizations (SROs)**. SROs differ from standard industry self-regulatory bodies in that they are given recognition and authority, including enforcement power, by a government body or agency. SROs are funded independently, rather than by the government. For example, the US SEC, the government agency that regulates the securities markets in the United States, allocates some regulatory responsibilities to the Financial Industry Regulatory Authority (FINRA), which is an SRO. It has the authority to enforce industry rules and federal securities laws.

FINRA

On its website, FINRA states:

FINRA is dedicated to protecting investors and safeguarding market integrity in a manner that facilitates vibrant capital markets.

FINRA plays a critical role in ensuring the integrity of America's financial system—all at no cost to taxpayers. Working under the supervision of the Securities and Exchange Commission, we:

- *Write and enforce rules governing the ethical activities of all registered broker-dealer firms and registered brokers in the U.S.;*
- *Examine firms for compliance with those rules;*
- *Foster market transparency; and*
- *Educate investors.*

(http://www.finra.org/about)

The role of SROs varies among countries. In some countries, such as the United States, SROs have specific regulatory authority, and in other countries, self-regulating organizations are rarely or never recognized as independent regulators. For example, the Australian financial regulator, ASIC, has stated, "One of the many significant recent legislative amendments that was introduced in Australia with the Financial Services Reform Act 2001 was the removal of the official regulatory standing of SROs. SROs, whether they are exchanges, industry associations, or some other form of peer group, have traditionally set standards of behavior or codes of conduct for market participants" (https://asic.gov.au/media/1339352/integration-financial-regulatory-authorities.pdf). According to the World Bank, the role of self-regulation in Europe, with the exception of the United Kingdom, was limited because of civil law systems and the resulting reliance on government supervision. In the United Kingdom and other countries with common law systems, reliance on self-regulation has been more extensive. The roles of SROs in regulation in these countries range from non-existent to having some regulatory authority. Regulators are concerned with the corporate governance of SROs and the management of their conflicts of interest. The extent of the concern is a factor in deciding the regulatory role, if any, of the SRO in question.

The relatively simple classification of regulators (legislative bodies, government agencies, independent regulators, and courts) and regulations (statutes, administrative regulations, and judicial law) is useful but does not reflect the complexities and nuances that exist. In some cases, the classification of a regulator is clear, and in other cases, it is ambiguous. For example, the Public Company Accounting Oversight Board (PCAOB) is a non-profit corporation, established by the US Congress to oversee the audits of public companies. Previously, the audit profession was self-regulated. The

PCAOB is funded primarily through annual fees paid by public companies, brokers, and dealers. The SEC oversees the PCAOB. The PCAOB is an independent regulator rather than a government agency, but it is not an SRO.

In Singapore, statutory boards are entities separate from the government, with specific legislation governing their operations. Most, if not all, statutory boards impose charges for some or all of their services. Those statutory boards that generate insufficient revenue to meet their expenses receive grants from the government to finance their operations. The grants are funded from the government's annual budget. The statutory boards are described as separate from the government, yet they are subject to specific legislation governing their operations and may receive government funding.

Whether Singapore's statutory boards are government agencies or independent regulators is ambiguous. The Singapore Economic Development Board (EDB), one such statutory board, describes itself as "a government agency under the Ministry of Trade and Industry…responsible for strategies that enhance Singapore's position as a global centre for business, innovation, and talent" (www.edb.gov.sg/en/about-edb/who-we-are.html). Another statutory board, the Accounting and Corporate Regulatory Authority (ACRA), describes itself as "the national regulator of business entities, public accountants and corporate service providers in Singapore" (www.acra.gov.sg/who-we-are/overview-of-acra). Although EDB clearly identifies itself as a government agency, it is less clear whether ACRA, given the description of a statutory board and the description of itself, should be classified as a government agency or an independent regulator.

Classifying regulatory bodies that exist in unions, such as the Union of South American Nations and the European Union, can also present challenges. For example, the European Commission, which has a mission to promote the general interest of the EU, can initiate legislation in the form of directives and regulations, which are subject to debate and approval by the European Parliament and the European Council (the co-legislators). The directives and regulations passed by the European Parliament and the European Council are jointly referred to as "EU law." Regulations have binding legal force in every EU member state on a par with national laws. Directives identify desired results and require national authorities to put laws in place to achieve them. Decisions are binding laws addressed to specific parties and are the result of specific cases (https://ec.europa.eu/info/law/law-making-process/types-eu-law_en). Regulations appear to have the characteristics of administrative regulations. Directives appear to have the characteristics of statutes; they are at a broad level, and another body needs to fill in the implementation details. Decisions appear similar to judicial law. Regardless of how a regulatory body is classified, it is important to identify the regulators and regulations that might affect the industry or company being analyzed.

EXAMPLE 3

Classification of Regulators

1. The media devotes considerable coverage to a regulatory body that has been given autonomy by the government and is empowered by statute. The regulatory body has recently raised the fees charged to the companies it regulates. The regulatory body in question is most likely to be a(n):
   - **A** self-regulatory organization.
   - **B** government agency.
   - **C** independent regulator.
2. Which of the following is least likely to be a characteristic of a self-regulatory body?

A It represents and regulates its members.

B It carries out government policy.

C It can discipline members that violate its rules and principles.

Solution to 1:

C is correct. Independent regulators are given authority by the government and are empowered by statute. Unlike government agencies, they are funded by fees that they collect from the firms they regulate.

Solution to 2:

B is correct. Self-regulating bodies do not carry out government policy. They are meant to be independent from government and immune to its influence.

Regulatory authorities may reference the work of outside bodies in their regulations. Examples of these outside bodies are accounting standard–setting bodies, such as the International Accounting Standards Board (IASB) and the Financial Accounting Standards Board (FASB), and credit-rating agencies. Regulatory authorities have the legal authority to enforce any regulation that references the work of these bodies. In the case of accounting standard–setting bodies—which are typically private sector, non-profit, self-regulated organizations—the requirement to prepare financial reports in accordance with specified accounting standards is the responsibility of regulatory authorities. The standard-setting bodies may set the standards, but the regulatory authorities recognize and enforce the standards. Ratings by credit-rating agencies—which are typically private sector, profit-oriented entities—were often referenced in regulations related to acceptable holdings by certain entities. Issues with conflicts of interest when the agencies were paid by the firms they rated, however, have resulted in efforts to reduce references to credit-rating agencies in regulations.

Although much of the focus of this reading is on the rules themselves and their development, impact, and implementation, regulatory enforcement and sanctions also play an important role. This division between development and enforcement of regulation also represents a possible way to classify laws or regulation. **Substantive law** focuses on the rights and responsibilities of entities and relationships among entities, and **procedural law** focuses on the protection and enforcement of the substantive laws. Regulators typically have responsibility for both substantive and procedural aspects of their regulations.

6 REGULATORY INTERDEPENDENCIES

f describe regulatory interdependencies and their effects;

An interesting facet of regulation is how regulated entities view the regulation, which is often context specific. Although there are many examples in which regulated companies fight against new proposed regulations, an outright opposition is relatively rare. Regulated company efforts to fight particular regulations tend to attract more public attention than when the companies are sympathetic to the proposed regulations. Even more fundamentally, academics have argued that regulation can sometimes enhance and work to the benefit of the interests of the regulated. This argument is often called the "**regulatory capture**" theory (see Stigler 1971). For example, regulatory actions and determinations can restrict potential competition (for example, by limiting entry) or effectively coordinate the choices of rivals (by imposing certain quality standards or price controls). In the interactions between regulated entities and their regulators, the regulated entities may possess considerable expertise and knowledge, and some of the individual regulators may have worked in the industry or aspire to be in the industry in which the regulated entities operate. These interactions may reinforce regulatory capture.

Regulatory differences across jurisdictions can lead to shifts in location and behavior of entities because of **regulatory competition** and **regulatory arbitrage**. Regulators may compete to provide a regulatory environment designed to attract certain entities (regulatory competition). As a result, companies may engage in regulatory arbitrage; for example, they may identify and use some aspect of regulations that allows them to exploit differences in economic substance and regulatory interpretation in foreign and domestic regulatory regimes to the companies' benefit.

Interdependence in the actions of regulators dealing with the same issues and activities is important in the international arena. Many regulatory issues are relatively similar around the globe. This commonality reflects both similarities in the challenges confronting different countries and the diffusion of the underlying problems around the globe. Such issues as financial systemic risk, terrorism financing, money laundering, and climate change reflect global concerns and, therefore, are well suited to an approach based on regulatory cooperation and coordination. For other issues, however, domestic regulators in specific jurisdictions often adopt different perspectives or face different trade-offs when developing and applying regulations in their jurisdiction. These varying perspectives can lead to differences in regulatory treatments of the same issue across countries. For example, some jurisdictions have significantly greater disclosure requirements to protect investors than other jurisdictions have. Although such differences are often justified, "regulatory competition" can reduce the effectiveness of regulation in particular countries. Regulatory competition can lead to what is sometimes referred as a "race to the bottom," in which countries continually reduce their regulatory standards to attract as many companies as possible to their jurisdiction.

Consider issues related to global warming and pollution. How should governments manage and coordinate efforts around the globe? The relevant externality is not simply within countries but, rather, extends beyond country borders. One of the challenging aspects of this issue is that countries differ in how much they contribute to climate change and in terms of how they are affected by it. Put simply, the countries most affected by climate change may not be the ones that are contributing the most to it. What are the institutional and governance mechanisms that would be appropriate to address this issue on a global basis? Although an economist's solution to the problem of pollution externalities might be to tax pollution or to introduce an emission trading system (or a cap and trade system) in order to allocate the pollution to the parties that can absorb the cost, the practical application of the solution may be complicated. How should one allocate "permits" to pollute among countries? Should countries have the "right" to pollute related to their past pollution? If not, how would one accommodate differences in living standards? How should one address the equity issues associated with low wealth and developing countries' having a potential comparative advantage in absorbing pollution?

The point of this overall discussion of interdependencies among jurisdictions is not to suggest the existence of global governance or a global regulator but, rather, is to recognize the reality and implications of diverse trade-offs and preferences among regional, national, and local regulators. To a degree, the presence of diverse and competing jurisdictions influences the stances of national and regional regulators. Evidence that governments recognize the necessity for global regulatory cooperation and coordination on some issues exists. For example, the Basel Accords established and promote internationally consistent capital requirements and risk management practices for large international banks. The Basel Committee on Banking Supervision has evolved into a standard setter for bank supervision, among other functions. Another example is the International Organization of Securities Commissions (IOSCO), a self-regulating organization but not a regulatory authority. Its members (national regulators) regulate a significant portion of the world's capital markets. This organization has established objectives and principles to guide securities and capital market regulation, to which its members agree to adhere.

How IOSCO Enhances Regulatory Cooperation

The member agencies of IOSCO have resolved, through its permanent structures,

- "to cooperate in developing, implementing and promoting adherence to internationally recognized and consistent standards of regulation, oversight and enforcement in order to protect investors, maintain fair, efficient and transparent markets, and seek to address systemic risks;
- to enhance investor protection and promote investor confidence in the integrity of securities markets, through strengthened information exchange and cooperation in enforcement against misconduct and in supervision of markets and market intermediaries; and
- to exchange information at both global and regional levels on their respective experiences in order to assist the development of markets, strengthen market infrastructure and implement appropriate regulation" (www.iosco.org/about/?subsection=about_iosco).

IOSCO is a standard setter and an establisher of best practices for securities regulators, and it has developed a framework of matters to be addressed in the domestic laws of a jurisdiction to facilitate effective securities legislation (IOSCO, "Methodology for Assessing Implementation of the IOSCO Objectives and Principles of Securities Regulation," May 2017). This framework is shown in Exhibit 1.

The framework also serves as a useful, but by no means exhaustive, list of areas of regulation relevant to an analyst. Labor, consumer protection, and environmental, health, and safety laws, which are not included in the list, may also significantly affect a business or industry.

Awareness of the basic types of laws and regulations that affect economies, financial systems, industries, and businesses is useful to an analyst. This knowledge will help the analyst to identify areas of concern and to consider proactively potential effects of regulations, existing and anticipated.

Exhibit 1 IOSCO's Objectives and Principles of Securities Regulation

Effective securities regulation depends on an appropriate legal framework. The matters to be addressed in the domestic laws of a jurisdiction include the following:

1 **Company Law**
   - **1.1** company formation
   - **1.2** duties of directors and officers
   - **1.3** regulation of takeover bids and other transactions intended to effect a change in control
   - **1.4** laws governing the issue and offer for sale of securities
   - **1.5** disclosure of information to security holders to enable informed voting decisions

Exhibit 1 (Continued)

- **1.6** disclosure of material shareholdings

2 Commercial Code/Contract Law

- **2.1** private right of contract
- **2.2** facilitation of securities lending and hypothecation
- **2.3** property rights, including rights attaching to securities, and the rules governing the transfer of those rights

3 Taxation Laws

- **3.1** clarity and consistency, including, but not limited to, the treatment of investments and investment products

4 Bankruptcy and Insolvency Laws

- **4.1** rights of security holders on winding up
- **4.2** rights of clients on insolvency of intermediary
- **4.3** netting

5 Competition Law

- **5.1** prevention of anticompetitive practices
- **5.2** prevention of unfair barriers to entry
- **5.3** prevention of abuse of a market dominant position

6 Banking Law

7 Dispute Resolution System

- **7.1** a fair and efficient judicial system (including the alternative of arbitration or other alternative dispute resolution mechanisms)
- **7.2** enforceability of court orders and arbitration awards, including foreign orders and awards

Source: International Organization of Securities Commissions, "Methodology For Assessing Implementation of the IOSCO Objectives and Principles of Securities Regulation" (May 2017): Appendix 1, "The Legal Framework." www.iosco.org/library/pubdocs/pdf/IOSCOPD562.pdf

Even within countries, the objectives of diverse government regulators can differ and potentially lead to regulations that seem inconsistent. Bank supervisors (whether as a function of the central bank, another entity, or a combination of entities) generally focus on **prudential supervision**—regulation and monitoring of the safety and soundness of financial institutions in order to promote financial stability, reduce system-wide risks, and protect customers of financial institutions. The objectives of securities commissions or regulators are typically to protect investors; ensure that markets are fair, efficient, and transparent; and reduce systemic risk. In some situations, the goals of the bank supervisor and securities regulator can be in tension, resulting in conflicting objectives. For example, on the one hand, the bank supervisor may be reluctant or even unwilling to release the results of stress tests of financial institutions in order to promote financial stability and avoid systemic risk because of the potential loss of confidence. On the other hand, a securities regulator might advocate for the release of information that might be relevant to investor decision making and act to protect investors (see Spatt 2009).

A general conclusion is that regulation by different regulators, even with seemingly similar objectives, can lead to very different regulatory outcomes.

EXAMPLE 4

Regulatory Interdependencies

1 A country's securities regulator is looking to attract a higher number of smaller companies to its capital markets. It proposes to ease hurdles that companies face when preparing to list shares on the country's stock exchange. The proposals include the lowering of the frequency of financial reporting, reducing the extent of disclosures required, and reducing the minimum size of company that can be accepted on the market.

   This is an example of:

   A regulatory competition.

   B regulatory coordination.

   C regulatory capture.

2 Regulatory capture is most likely to be a concern where there is reliance on:

   A SROs.

   B government agencies.

   C government departments.

Solution to 1:

A is correct. Regulatory competition occurs when regulators compete to provide a regulatory environment designed to attract certain entities.

Solution to 2:

A is correct. Regulatory capture has been a concern when SROs are used.

7 REGULATORY TOOLS

g describe tools of regulatory intervention in markets;

Regulatory and government policies should be predictable as well as effective in achieving objectives. It is very difficult for any entity to function with confidence and success in an environment where the rules are unclear or in a state of flux (in other words, where there is considerable regulatory uncertainty). Regulatory choices or government policies that will be consistent over time are desirable. If these choices occur, the regulatory environment is likely to be stable despite the fact that, in many countries, governmental decision makers (with diverse political preferences) change on a regular basis. It is helpful to use regulatory tools that are consistent with maintaining a stable regulatory environment. Regulatory tools and government interventions in markets include the use of price mechanisms, such as taxes and subsidies; regulatory mandates and restrictions on behaviors, including establishing rights and responsibilities; provision of public goods; and public financing of private projects.

The issue of how to address pollution is a classic example in regulation. By taxing polluters (or subsidizing those who do not pollute, by using a suitable baseline), one can create a system in which marginal incentives are equated across economic agents. The advantage of such an arrangement is that, theoretically, the rights to pollute are redistributed in an "efficient" manner relative to a fixed allocation. In particular, the structure of the regulation allows market incentives to redistribute the pollution rights to those for whom they are the most valuable at the margin. There are important issues, however, about how to initially establish and distribute the amount of acceptable total pollution. In some situations, historical usage (amount of pollution produced) is used to allocate pollution rights One problem is that marginal incentives may be altered in anticipation of this allocation. In other situations, the allocation is the outcome of a political process, which can lead to considerable lobbying. At the heart of this example is the use of a price mechanism to create the appropriate marginal incentives and an efficient allocation of resources. The Coase theorem states that if an externality can be traded and there are no transaction costs, then the allocation of property rights will be efficient and the resource allocation will not depend on the initial assignment of property rights.

Governments can intervene in markets in ways other than through the price mechanism. These include restricting some activities (e.g., insider trading and short selling), mandating some activities (e.g., capital requirements for banks and registration with a securities commission for certain activities), providing public goods (e.g., national defense and transportation infrastructure), and financing private projects (e.g., loans to individuals or companies for specified activities that the government deems desirable to encourage). The extent of government provision of public goods and government financing of private projects depends on a number of factors, including the political philosophy of the country and/or government in power, the structure of the government, and the country's gross domestic product. The problem of **systemic risk** (the risk of failure of the financial system) as a result of the failure of a major financial institution has emerged as an issue in many countries around the world in the aftermath of the 2008 global financial crisis. Systemic risk and **financial contagion** (a situation in which financial shocks spread from their place of origin to other regions; in essence, a faltering economy infects other, healthier economies) are examples of negative externalities. In the EU, the European Systemic Risk Board, formed in December 2010, is an advisory EU body within the European Central Bank tasked with advising national macroprudential bodies to take steps to address risks.

Exhibit 2 focuses on how "bail-in" tools can mitigate systemic risk.

Exhibit 2 Bail-In Tools

Among the regulatory tools introduced following the 2008 financial crisis are so-called bail-in powers, which were endorsed by the Financial Stability Board and the G–20. A bail-in tool is seen as improving the toolkit for dealing with the failure of large, globally systemic banks. Bail-in involves shareholders of a failing institution being divested of their shares and creditors of the institution having their claims canceled or reduced to the extent necessary to restore the institution to financial viability. Bail-in policies are intended to ensure that shareholders and creditors of the failed institution, rather than taxpayers, pay the costs of the failure. This situation contrasts with how many governments dealt with the financial crisis of 2008, when banks and insurers were rescued by taxpayers in a number of countries. Such policies also allow the failed institution to continue to operate, so that it can introduce restructuring measures to address the cause of the failure. These policies can limit disruption to the customers of the institution and help maintain confidence in the banking system.

It is difficult to assess the extent to which the new approaches and tools, such as bail-in policies, will reduce systemic risk. There are a number of reasons for this difficulty. The types and sources of future crises are likely to be different from those of the past, so regulations designed with a prior crisis in mind may not prevent a future crisis. It can be difficult to assess the potential effectiveness of regulatory actions before an event and even after the fact. The mere fact that a crisis does not occur is not necessarily evidence that regulations prevented one. It is also plausible that some regulatory responses have the unintended consequence of mitigating one source of risk while increasing another source of risk. All these issues make effective regulation challenging to design.

Generally, more than one regulatory approach or policy is feasible and worthy of consideration in a specific situation. Two examples that illustrate a range of possible regulatory responses are (1) conflict-of-interest policies and (2) trading restrictions on insiders, which are explored in Exhibits 3 and 4, respectively.

Exhibit 3 Regulating Conflicts of Interest

Consider the hypothetical scenario in which a potential employee of a regulator has some degree of financial exposure to a regulated company. Such exposure could come about in many ways (for example, spousal employment, a marketable position in an investment portfolio, or an illiquid position resulting from past employment) and at a variety of financial levels.

What types of regulatory policies might be appropriate to mitigate these risks?

Among the potential regulatory responses are the following: The individual could be barred from employment at the regulatory agency or from working on specific (or all) projects involving the regulated company in question. The individual could sell the position; the sale could be voluntary or mandated. The individual could be required to disclose the nature of his potential conflict to higher-level decision makers to whom he will be providing recommendations. Other potential policies include a bar on involvement, resolution of the conflict, or disclosure of it.

Exhibit 4 Regulating Corporate Insiders

Turning to the case of corporate insiders, there are potential regulatory and corporate restrictions. Examples of regulatory responses are a ban from trading on non-public information and a requirement that when they do trade, the insiders disclose the trades. The company may impose a blackout period during which insiders are banned from trading on the company's stock (these periods often precede earnings announcements and continue for a short while afterward). The appropriate remedy depends on the underlying facts and circumstances, and arguably the appropriate standards would reflect the specific context.

An important aspect of effective regulation is the potential ability to impose sanctions on violators of the regulations; in other words, it is important to be able to enforce the regulations. IOSCO clearly identifies this aspect as one of the agreed-on principles of securities regulation: "The regulator should have comprehensive enforcement powers" (IOSCO, "Objectives and Principles of Securities Regulation," 2017). Enforcement of securities regulations and regulations on businesses may include sanctions for the violating corporation (business or company), the individual violator(s), or both. Corporate sanctions may be appropriate if the company caused harm to others. The sanctions often involve monetary fines/fees/settlement, and in the case of individuals, the sanctions may also involve prison terms. In some situations, such as in cases of accounting fraud, shareholders may actually be the victims. In such instances, if the stockholders were harmed by the wrongdoing, the case for sanctions, such as fines, against the company is often far from compelling. The sanctions may simply redistribute funds from current shareholders to the stockholders who were the specific victims, and the company incurs real resource costs.

For various reasons, it can be difficult to prosecute or achieve settlements with individual violators. First, it often is difficult to detect violations and to identify exactly which individuals were at fault. Second, the individuals possess strong incentives to fight in order to protect their reputation and livelihood. Indeed, individuals are often able to fight using corporate resources because of indemnification provisions in their employment contract. The intent of these provisions may be to protect risk-averse executives against inadvertent liability and to potentially align their interests with those of the stockholders, but the provisions instead may result in protecting executives to the detriment of the stockholders. The incentive to fight individual sanctions may be especially strong because of not only financial costs but also other costs, such as reputational costs.

EXAMPLE 5

Regulatory Tools

1. Globalization of capital markets is *most likely* to result in increased concerns about:
   - **A** financial contagion.
   - **B** regulatory competition.
   - **C** both contagion and regulatory competition.
2. The regulatory tools *least likely* to be used by self-regulating organizations are:
   - **A** price mechanisms.
   - **B** restrictions on behaviors.
   - **C** provision of public goods.

Solution to 1:

C is correct. Globalization is likely to result in increased concerns about contagion and regulatory competition. It is easier for a financial shock to spread. Governments may use their regulatory environment to attract entities from around the world.

Solution to 2:

A is correct. SROs are least likely to use price mechanisms. They typically regulate behaviors and often provide public goods in the form of standards.

COST-BENEFIT ANALYSIS

8

h describe benefits and costs of regulation;

The effects of regulation can range from macro effects that impact large parts of the economy to micro effects on an individual business. Section 8 introduces the concept of cost–benefit analysis carried out by regulators. Section 9 illustrates how an analyst could approach the task of assessing the effect of regulation on a particular industry. Because regulations can evolve in response to market, technological, and societal changes, it is important to monitor issues of concern to regulators and ongoing developments to evaluate the implications of potential changes in regulation. Understanding the regulatory process will help an analyst recognize the types of challenges that regulators and policymakers face and formulate expectations of regulatory outcomes.

8.1 Basic Concepts of Cost–Benefit Analysis

In assessing regulation and regulatory outcomes, it is common practice for regulators to assess the overall benefits and costs of regulatory proposals to assess the trade-offs associated with a particular regulatory action and to assess alternative solutions. Regulators, guided by economic principles, strive to develop techniques to enhance the measurement of the costs and benefits of regulations. The general benefits of regulation as discussed in previous sections may be clear, but the measurement of the full impact of the regulation (both benefits and costs) can be challenging. In conducting cost–benefit analysis of regulation, it often is easier to assess the costs of regulation, although doing so can also be challenging.

Regulatory burden refers to the costs of regulation for the regulated entity; these costs are sometimes viewed as the private costs of regulation or government burden. **Net regulatory burden** is the private costs of regulation less the private benefits of regulation. Many regulators focus narrowly on the implementation costs of regulation (for example, how many compliance lawyers will need to be hired—and at what cost), but in many instances, the most significant costs are the indirect ones that relate to the ways in which economic decisions and behavior are altered and market allocations are changed.

Regulators view some of the costs associated with regulations as "unintended," but it is important to distinguish between two types of such costs. There may be implementation costs that were unanticipated (for example, if it turns out more compliance lawyers need to be hired than originally thought) and indirect costs because of unintended consequences. It is important for regulators to recognize that their evaluation of potential regulations should reflect indirect costs as well as the consequences that were the direct objective of the rule making. Furthermore, in some cases, regulatory filings and consultations in response to proposed regulations identify at least some of the "unintended consequences" prior to the implementation of the regulations. In these circumstances, it is difficult to argue that such consequences were unanticipated and unintended if they were identified prior to the implementation of the regulation. Unintended consequences are reflective of underlying policy risk and may result in high unanticipated costs.

Regulatory costs and benefits are especially difficult to assess on a prospective basis relative to a retrospective basis. An after-the-fact analysis allows a comparison of the item(s) of interest before and after the regulation occurs. This comparison allows for a more informed assessment of a regulation because the actual costs and benefits may be identifiable. In some instances, a trial or pilot analysis may be appropriate and helpful to more fully understand the potential impacts in advance of a proposed regulation. A potentially feasible and relevant approach in the context of an environment with frequent trading is to use natural experiments and trial phase-ins to generate data suitable for careful cost–benefit analysis. This approach facilitates the assessment of statistical evidence to evaluate the effects prior to the full implementation of the proposed regulation. Among the contexts in which US securities regulators have used such techniques are rules involving short sales, post-trade price reporting, and the tick size increment for trading. Such approaches are more feasible for a trading rule in a market with high trading frequency that will generate considerable data and run little risk of disrupting the real economy. Similar approaches, sometimes known as "regulatory sandboxes," are being introduced by regulators in such countries as the United Kingdom, Singapore, Australia, the United Arab Emirates, and Malaysia.

In the United States, administrative law requires that federal regulatory agencies conduct a cost–benefit analysis to assess the consequences of their actions. Court rulings have struck down regulatory actions because cost–benefit analyses performed were deemed inadequate. For example, the US Circuit Court of Appeals overturned the 2004 SEC rule requiring that mutual funds have independent chairs and at least 75% independent directors on such grounds (see *Chamber of Commerce v. SEC*, 412 F.3d 133 [D.C. Cir. 2005] and 443 F.3d 890 [D.C. Cir. 2006]). In 2011, as reported in the *Wall Street Journal*, "Striking a blow to the shareholder rights movement, a federal appeals court threw out a controversial new Securities and Exchange Commission regulation that would give investors more power to oust corporate directors.... The court issued a harsh rebuke to the SEC, saying it didn't adequately analyze the costs to U.S. companies of fighting in contested board elections" (*The Wall Street Journal*, 23 July 2011, "Court Deals Blow to SEC, Activists," Jessica Holzer). Requirements to undertake cost–benefit analyses (also known as "impact assessments") prior to the introduction of new regulations exist in a number of other jurisdictions, including in the EU and Australia.

Ideally, regulatory judgments should reflect economic principles and full consideration of the economic costs and benefits, rather than the preferences of current decision makers. Although the potential failure of the fundamental theorem of welfare economics suggests the potential relevance of regulation, it is important to use economic principles to identify and assess alternative remedies and specific actions.

Illustration of Cost–Benefit Analysis

To illustrate the issues that may be relevant to a regulator when considering a cost–benefit analysis, we focus here on a proposal for the introduction of new regulation in Europe to remove the red tape and compliance burdens faced by small and medium-sized enterprises (SMEs) when seeking an IPO of their shares or for issuing bonds on public markets. The proposed regulation is intended to address three problems. First, the "one-size-fits-all" approach to some areas of financial regulation has led to a perception that the costs of listing outweigh the benefits. Second, there has been a progressive decline in the number of smaller brokers and investment firms that specialize in trading shares of smaller companies. A small, local brokerage ecosystem is necessary to support smaller companies with the listing process. The third problem is the lack of investment in shares and bonds of small companies.

The objective of the proposed regulation would, therefore, be to revive IPOs and bond offerings of small companies. The cost–benefit assessment would examine two proposals: first, whether some of the existing regulations, such as MiFID II and the Market Abuse Regulation (MAR), could be adapted to accommodate smaller issuers. The costs and benefits of (1) targeted changes and clarifications to the existing regulations or (2) an overhaul of the provisions would be examined relative to a "baseline scenario" where the existing rules under MiFID II and MAR do not change. The second proposal would examine the cost and benefits of the introduction of new provisions. These could include simpler delisting rules, rules enabling easy transfer of a listing from less regulated "small company growth markets" to regulated markets, and less stringent free float requirements to make it more attractive for issuers, investors, and market operators.

A regulator carrying out cost–benefit analysis would need to assess the positive impact on the companies' investment and growth rates as a result of easier access to capital. The costs the regulator would have to consider include the additional regulatory oversight over the small companies in question and the consequences of any impact of corporate failures among the listed smaller firms on investor confidence and, consequently, the remaining firms' cost of capital.

EXAMPLE 6

Cost–Benefit Analysis

An investment adviser is discussing a client's portfolio exposure to the electric utilities sector. The sector's regulator has outlined series of proposals for new regulation, on which it is carrying out cost–benefit analysis. The adviser makes two statements to the client about the regulator's cost–benefit analysis.

Statement 1 "The regulator will assess and take into account as part of its cost–benefit analysis only the indirect costs of new regulation arising from changed economic decisions and behaviors."

Statement 2 "Regulatory costs and benefits are easier to assess on a retrospective, after-the-fact basis."

Which statement is correct?

A Only Statement 1 is correct.

B Only Statement 2 is correct.

C Both statements are correct.

Solution:

B is correct. Statement 2 is correct because actual costs and benefits may be available during retrospective analysis, allowing a more informed assessment of regulation. Statement 1 is incorrect because both indirect and implementation costs will be taken into account.

9 ANALYSIS OF REGULATION

i. describe the considerations when evaluating the effects of regulation on an industry.

Christopher Decker (University of Oxford) contributed content for this discussion.

In the previous section, we considered cost–benefit analysis of new regulations from the point of view of the regulator that is in the process of developing new regulations or is analyzing the impact of regulatory intervention retrospectively, after the event. In this section, we focus on the considerations that an analyst or investor could take into account when evaluating the effects of a specific regulation on a particular industry or company for the purpose of making an investment recommendation. In-depth coverage of industry and company analysis is featured elsewhere in the CFA Program curriculum.

The fact that rules and regulations can take different forms and can affect industries and individual companies in different ways adds to the complexity of the task. Analysts need to understand not just how regulation affects companies and industries at present; they should also be able to understand and anticipate the impact of proposed new or changing regulations on the future prospects for companies and industries. Having assessed the impact of regulations on the company and its prospects, analysts can then use suitable valuation tools to establish fair values for the business and make investment recommendations. Although no framework or template is adequate for all the possible scenarios, there are certain steps an analyst can take that are common to most circumstances.

Assessment of the likelihood of regulatory change

The analyst will need to assess the likelihood of the proposed regulation actually being implemented. Understanding the regulator's intentions, the cost–benefit analysis framework used by the regulator, and the extent of engagement with the regulated companies will help the analyst draw conclusions about the likelihood that the proposed regulation will be implemented. Where relevant, public and political pressure may also play a role in determining the likelihood that regulatory intervention will materialize.

Assessment of the impact of regulatory change on a sector

Industry and company analysis performed by an analyst (explored in depth elsewhere) will incorporate the analyst's or investor's opinion on the impact of regulations. The following text describes some, but not all, of the effects that regulations may have.

Impact on revenues

Regulatory bodies sometimes introduce limits on prices, tariffs, rents, or fees that companies may charge, usually to protect consumers. Alternatively, certain products or services may be banned by the regulators, or companies may be required to provide product descriptions that discourage their consumption (food or tobacco product labeling). The analyst would need to estimate the impact of such regulatory interventions on the companies' turnover, noting that not all entities in the sector would be affected equally. For example, telecommunication and utilities companies in Europe have been subject to caps on prices and tariffs in the last few decades. Another example is the former biannual pricing revision process in Japan's pharmaceutical sector. The analyst would also need to be aware that if certain charges or fees are no longer allowed by the regulator, the companies may find alternative ways of generating revenues, helping to offset the negative impact of the regulation. For instance, the ban on commissions from financial product providers that financial advisory firms used to receive in parts of Europe led firms instead to charge their clients fees for giving financial advice. In this way, they could recoup some of the revenues they lost through the ban on commissions.

In some scenarios, pricing or charging may not be regulated or limited by the regulator in any way, but companies may be required to increase their pricing transparency and provide a detailed breakdown of their fees. For example, power utilities in the United Kingdom must provide transparent monthly bills and, if applicable, suggest that their customers switch to a different tariff if the customer could benefit from a lower monthly payment by switching.

Another form of regulation that relates to the revenue line is the arrangement in which utility companies, often natural monopolies, are allowed to earn only a certain return on their assets.

Cost impact

Compliance with some regulations results in additional costs for companies. These costs could take the form of higher operating expenses if, for example, manufactured products need to incorporate new safety features. Alternatively, these costs could take the form of higher capital expenditure if additional or new equipment is required. Analysts may also need to consider additional costs related to minimum wages, increased information disclosure, and data protection requirements. Companies may also incur additional costs resulting from the need to use (or not to use) certain raw materials, introduce product features, or subject themselves to regular inspections that make products or services safer for consumers. There may be costs related to expensive wastewater treatment for certain industries. The analyst will try to estimate such costs and incorporate their impact into forecasts of a company's future profitability. When quantification of specific additional costs is not possible or relevant, the analyst will need to take into account the potentially reduced flexibility of the company's operations. It may also be possible for some companies to pass on some additional costs to their customers. The analyst will need to understand the competitive position of the industry.

Examples of regulations that companies may be subjected to include labor laws that may impose limits on hours of work, such as the maximum working week in parts of Europe or, more globally, limits on the number of work hours for pilots and cabin crew in the aviation industry. Of course, some companies that adhere to high environmental, social, and governance standards would incur some of those costs regardless of government regulation. In the financial services industry, the increasing requirements in relation to record keeping, data protection, risk control, and prevention of money laundering often result in significant additional personnel, training, and information technology infrastructure expenses.

Business risk

Many industries have seen greater regulator involvement in the form of fines, requirements to pay compensation to customers, or bans on certain activities. Such events are difficult to forecast, and their impact may be hard to quantify and incorporate into future cash flow or growth forecasts. Analysts should take these types of events into account either by attempting to assign probabilities to them or by reflecting the risk in the discount rate used to value the company. For example, companies prone to a particular regulatory risk may deserve to trade on lower valuation multiples relative to peers or other industries. Or, when future earnings or cash flows are discounted to present values, the discount rate should reflect an additional risk premium.

Example of Regulatory Analysis

The scenario outlined earlier (in "Illustration of Cost–Benefit Analysis") considered proposed changes to the regulations concerning the IPO market for smaller companies. An analyst may wish to evaluate the impact of such changes to the rules in order to analyze prospects for a particular company (issuer) or to assess the portfolio fund flows into or out of the smaller companies segment. The analyst may also want to understand the impact on companies in the brokerage business involved in that segment of the market.

The analyst might consider a series of questions. One set of questions could concern the design of the proposal: Are there any exemptions from the new regulation, or will it apply to all small companies? What thresholds, if any, are used to determine whether a company is designated as small? Will the regulation be applied identically across all EU member states, or will each jurisdiction have an ability to tailor it to their own conditions? Will the regulation be subject to review or withdrawal later?

Another set of questions that an analyst might consider relates to the potential market and participant impacts: Who will benefit most from the regulation? Will it benefit high-growth companies in particular sectors or countries? What is the scale of the potential benefit associated with the regulation? What costs are associated with the regulation, and what is their scale? Will all participants face the same costs, or will they differ by market segment? Are the costs likely to be one-off in nature or recurring? Is the regulation likely to lead to market entry, expansion, or innovation by certain SMEs? Could the regulation lead to the potential exit of some existing providers of brokerage services or potentially crowd out other means of supplying capital and finance? How might the regulation change the behavior of SMEs? For example, will it reduce their reliance on bank loans?

Finally, the analyst might consider any wider impacts of the regulation on the market, industry, and society: Will the regulation affect financial stability or resilience? Could the regulation widen the opportunities for other investors, allowing them to better diversify their portfolios? Could there be potential spillover effects, allowing firms, for example, to shift between "junior markets" and more regulated markets? Could the impacts on financial markets be greater in some EU member states than in others? What are the possible macroeconomic impacts of the regulation? For example, could it improve capital inflows?

EXAMPLE 7

Analysis of Regulation

1 Jessica Wong, CFA, is an equity analyst responsible for the materials and industrial sectors in Europe. The regulatory authorities are preparing new rules on transportation, further limiting the age and exhaust emissions of the trucks used by industrial companies. What is likely to be of greatest concern to the analyst when evaluating the impact of the new rules on companies?

   A Changes to costs related to the acquisition, operation, and maintenance of trucks

   B Positive impact of reduced pollution on public health and subsequent health care cost savings in the wider society

   C The methods and techniques used by the regulator during its cost–benefit analysis

2 Which of the following questions would be the least relevant for the analyst to ask?

   A Will the costs be one-off in nature, or will they be recurring?

   B Will all companies in the sector be affected equally?

   C What is the legal status of the regulator? Is it a government department, a government agency, or a self-regulating organization?

Solution to 1:

A is correct. The analyst will want to understand and analyze the impact of the proposed regulations on the performance of companies she covers. Answers B and C relate to cost–benefit analysis carried out by the regulatory authority.

Solution to 2:

C is correct. The status of the regulator is the least important question. Answers A and B represent items that the analyst will find relevant when evaluating the impact of new regulations on the companies under coverage.

Regulators and the Regulated

How may regulation affect the economics of businesses?

One example is the effect of the SEC's Regulation National Market System (NMS) on competition among equity trading platforms in the United States. Regulation NMS, adopted in 2005, was intended to reflect technological advances and achieve the objectives of efficient, competitive, fair, and orderly markets. Since the 2005 adoption of Regulation NMS, the market share of the trading floor of the NYSE has fallen substantially to account for a fraction of the overall NYSE volumes. Prior to Regulation NMS, NYSE "specialists" or market makers could take up to 30 seconds to react to orders sent by other platforms. The other platforms were checking whether the NYSE would execute at a more favorable price than the original platform had quoted. This process provided considerable opportunity for an NYSE specialist to observe subsequent pricing and to exploit the implicit optionality in the process. This process also made it hard for the rival platform to compete. Consequently, the NYSE could position itself to attract and concentrate much of the market liquidity, and so it came to resemble a natural monopoly. After Regulation NMS, which the NYSE had endorsed, the advantage to the NYSE diminished (see the discussion of the impact of Regulation NMS in Angel, Harris, and Spatt 2011)). Because of the change in regulation, many new trading platforms developed and trading execution fragmented. Clearly, the structure of regulation plays a crucial role with respect to the viability of different order tactics and even the viability of the business models underlying different trading platforms.

The history of the money market mutual fund industry is another example of how regulation can affect business models. Money market mutual funds in the United States first arose in the early 1970s in response to Regulation Q, which imposed a ceiling on the interest rates paid by banks for various types of bank deposits. When market interest rates rose above the ceiling, there was considerable migration from bank deposits toward marketed fixed-income instruments, such as Treasury bills and notes. Money market mutual funds developed in response to the binding Regulation Q rate ceilings. During the 2008 global financial crisis, the collapse of a major US money market mutual fund (the Reserve Fund) led to a run until the government launched a short-term insurance program to protect money market mutual fund balances. Government policy (motivated by an attempt to stabilize the financial system) helped protect this product. In response to resulting pressures from banks and the new advantage that the money market fund industry obtained, however, the Federal Deposit Insurance Corporation subsequently raised its insurance limit from \$100,000 to \$250,000. As this example illustrates, regulatory constraints have played a major role in the organization of short-term deposits in the United States. Changes in the effective regulatory structure have led to dramatic changes in the competitive landscape.

Government regulation can affect the structure of the industry. The issues can be seen in the pricing of joint products in the utility industries. For example, it can be difficult to separate fully the underlying economics associated with the production, transmission, and distribution of such services as electricity, telecommunications, and water. Suppose, for example, that there is a natural monopoly with respect to transportation (transmission and distribution) of a utility service but that there is competition in complementary activities (such as gas or electricity production or retail competition in telecommunications). How much should the provider of the natural monopoly services be able to obtain from the consumer or other companies providing upstream services, such as an energy product or access to a communication network? Although for some products there is increased and vigorous competition, these issues are still important with respect to the returns available from building various types of infrastructure. Although the market can sort out the allocation of profits and pricing across stages when there is vigorous competition at each stage, these issues are challenging in the case of a natural monopoly. Monopoly power is at the root of one of the most important traditional uses of regulation—to set pricing and returns at utility providers. In many jurisdictions, a government regulator sets or approves public utility prices because a utility provider has a monopolistic position.

SUMMARY

Knowledge of regulation is important because regulation has potentially far-reaching and significant effects. These effects can range from macro-level effects on the economy to micro-level effects on individual entities and securities.

Regulation originates from a variety of sources and in a variety of areas. A framework that includes types of regulators and regulation as well as areas of regulation that may affect the entity of interest (including the economy as an entity) is useful. The framework will help in assessing possible effects of new regulation. It can also help in assessing the effects of regulation on various entities.

More than one regulator may develop regulations in response to a particular issue. Each of the relevant regulators may have different objectives and choose to address the issue using different regulatory tools.

In developing regulations, the regulator should consider costs and benefits. In the analysis, the net regulatory burden (private costs less private benefits of regulation) may also be relevant. Potential costs and benefits, regardless of the perspective, may be difficult to assess. A critical aspect of regulatory analysis, however, is assessing the costs and benefits of regulation.

The following are some key points of the reading.

- The existence of informational frictions and externalities creates a need for regulation. Regulation is expected to have societal benefits and should be assessed using cost–benefit analysis.
- The regulation of securities markets and financial institutions is extensive and complex because of the consequences of failures in the financial system. These consequences include financial losses, loss of confidence, and disruption of commerce.
- The focus of regulators in financial markets includes prudential supervision, financial stability, market integrity, and economic growth.
- Regulatory competition is competition among different regulatory bodies to use regulation in order to attract certain entities.
- The breadth of regulation of commerce necessitates the use of a framework that identifies potential areas of regulation. This framework can be referenced to identify specific areas of regulation, both existing and anticipated, that may affect the entity of interest.
- Legislative bodies, regulatory bodies, and courts typically enact regulation.
- Regulatory bodies include government agencies and independent regulators granted authority by a government or governmental agency. Some independent regulators are self-regulating organizations.
- Typically, legislative bodies enact broad laws or statutes. Regulatory bodies issue administrative regulations, often implementing statutes. Courts interpret statutes and administrative regulations; these interpretations may result in judicial law.
- Interdependence in the actions and potentially conflicting objectives of regulators is an important consideration for regulators, regulated entities, and those assessing the effects of regulation.
- Regulation that arises to enhance the interests of regulated entities reflects regulatory capture.
- Regulators have responsibility for both substantive and procedural laws. The former focuses on rights and responsibilities of entities and relationships among entities. The latter focuses on the protection and enforcement of the former.
- Regulatory arbitrage is the use of regulation by an entity to exploit differences in economic substance and regulatory interpretation or in regulatory regimes to the entity's benefit.
- There are many regulatory tools available to regulators, including regulatory mandates and restrictions on behaviors, provision of public goods, and public financing of private projects.

- The choice of regulatory tool should be consistent with maintaining a stable regulatory environment. "Stable" does not mean unchanging but, rather, refers to desirable attributes of regulation, including predictability, effectiveness in achieving objectives, time consistency, and enforceability.
- In assessing regulation and regulatory outcomes, regulators should conduct ongoing cost–benefit analyses, develop techniques to enhance the measurement of these outcomes, and use economic principles to guide them.
- Net regulatory burden to the entity of interest is an important consideration for analysts.

REFERENCES

Angel, James, Lawrence Harris, and Chester Spatt. 2011. "Equity Trading in the 21st Century." *Quarterly Journal of Finance* 1 (1): 1–53.

International Organization of Securities Commissions (IOSCO)."*Objectives and Principles of Securities Regulation*," May 2017

IOSCO."*Methodology for Assessing Implementation of the IOSCO Objectives and Principles of Securities Regulation*," May 2017

Spatt, Chester. 2009. "Regulatory Conflict: Market Integrity vs. Financial Stability." *University of Pittsburgh Law Review. University of Pittsburgh. School of Law* 71 (3): 625–39.

Stigler, George J. 1971. "The Economic Theory of Regulation." *Bell Journal of Economics* 2 (1): 3–21.

PRACTICE PROBLEMS

The following information relates to Questions 1–4

Tiu Asset Management (TAM), a hypothetical financial services firm, recently hired Jonna Yun. A member of TAM's global equity portfolio team, Yun is assigned the task of analyzing the effects of regulation on the financial services sector of a particular country. In her first report to the team, Yun makes the following statements:

Statement 1	The country's regulator, a government agency, concerned about systemic risk, is calling for an accelerated adoption of centralized derivatives settlement (as opposed to bilateral settlement between two counterparties)—a more stringent rule—ahead of other major countries that are considering a similar move.
Statement 2	Regulators use various tools to intervene in the financial services sector.
Statement 3	Regulations may bring benefits to the economy, but they may also have unanticipated costs.
Statement 4	The country's regulatory authorities are considering a regulation that is similar to Regulation Q in the United States, which imposed a ceiling on interest rates paid by banks for certain bank deposits.

1 What is the *most likely* basis for the concerns noted in Statement 1?

A Externalities

B Regulatory arbitrage

C Informational friction

2 The tools *least likely* to be used by regulators to intervene in financial markets owing to informational frictions are:

A blackout periods.

B capital requirements.

C insider-trading restrictions.

3 Which of the following is *most likely* an unanticipated effect of regulation?

A Hiring compliance lawyers

B Setting legal standards for contracts

C Establishing employers' rights and responsibilities

4 After Regulation Q was imposed, the demand for money market funds *most likely*:

A increased.

B decreased.

C remained unchanged.

© 2019 CFA Institute. All rights reserved.

The following information relates to Questions 5–11

Cate Stephenson is an analyst in the economics research division of an international securities firm. She is conducting research on the regulatory environment in certain European countries. Stephenson begins with an analysis of a hypothetical country, Genovia.

Genovia has recently introduced a new accounting statute. In Genovia, there is an independent regulator—"Le régulateur." Le régulateur is not a self-regulating organization (SRO). There is also an SRO—"L'organisation." L'organisation is not an independent regulator.

In her research report, Stephenson makes the following statements:

Statement 1 Le régulateur has been given legal authority by the government to enforce the new statute.

Statement 2 L'organisation issues administrative regulations related to the new statute using government funding.

Statement 3 L'organisation has member companies that accept the authorization of L'organisation to set and enforce rules and standards.

Stephenson and her supervisor discuss the intended and unintended effects of implementing the new statute, and Stephenson makes two comments.

Comment 1 It is likely that some unintended consequences will be identified in regulatory filings prior to implementation of the new legislation.

Comment 2 Indirect costs arise because of unintended consequences and may result in high unanticipated costs.

Stephenson reads a report titled "International Trade," which has three sections about Genovia's policies and regulations.

- The first section of the report discusses policies that legislators may implement to accomplish Genovia's objective of promoting free trade on industrial goods.
- The second section of the report covers corporate domicile. Stephenson learns that regulators in Genovia recently amended regulations to encourage foreign businesses to move their corporate domicile to Genovia.
- The third section of the report reviews the regulation of commerce. Genovia's goal is to establish an environment that encourages foreign businesses to increase trade with domestic businesses. Stephenson considers two features of Genovia's regulation of commerce.

Feature 1 Recent court decisions have upheld financial liability and bankruptcy laws.

Feature 2 A legal structure that governs contracts and each party's rights is in place.

Stephenson then reviews two initiatives by Genovia to improve domestic policies and regulations.

- The first initiative by Genovia is its passage of conflict-of-interest regulations. Regulators implement regulatory restrictions and regulatory mandates that apply to employees of securities firms. One of Stephenson's research colleagues writes reports on a company in which he owns shares.

- The second initiative by Genovia is to reduce pollution and promote renewable electricity generation. Two years ago, the government implemented taxes on fossil fuels and subsidies on hydropower and other renewables. Stephenson reviews the changes in sources of electricity production since the policies were introduced, shown in Exhibit 1.

Exhibit 1 Genovia's Domestic Electricity Generation Production

Sector	Year 0	Year 1	Year 2
Fossil fuels	462	446	426
Hydropower	186	231	273
Other renewables	97	120	154
Total	745	797	853

Note: Amounts are in terawatt hours (TWh).

5 Which of Stephenson's statements regarding Le régulateur and L'organisation is correct?

A Only Statement 1 is correct.

B Only Statement 2 is correct.

C Both Statement 1 and Statement 2 are correct.

6 Is Stephenson's Statement 3 correct?

A Yes

B No, because L'organisation is given the authority to enforce regulations by a government agency

C No, because pressure from its member companies prevents L'organisation from enforcing its rules and standards

7 Which of Stephenson's comments to her supervisor is most likely correct?

A Only Comment 1 is correct.

B Only Comment 2 is correct.

C Both Comment 1 and Comment 2 are correct.

8 Which of the following policies would *best* address Genovia's objective of promoting free trade on industrial goods?

A Imposing tariffs on foreign-produced goods

B Allowing a floating currency

C Providing subsidies to domestic companies

9 By amending regulations to encourage foreign businesses to change their corporate domicile, regulators are engaging in regulatory:

A capture.

B arbitrage.

C competition.

10 Which feature discussed in the third section of "International Trade" will *most likely* help Genovia achieve its goal of encouraging foreign businesses to increase trade with domestic businesses?

A Only Feature 1

B Only Feature 2

C Both Feature 1 and Feature 2

11 Based on Exhibit 1, which government policy has been *most effective* in helping Genovia achieve its second initiative?

A Tax on fossil fuels

B Subsidy on hydropower

C Subsidy on other renewables

SOLUTIONS

1 B is correct. Firms based in the country are likely to be concerned because of the earlier timing of the application of new (more stringent) regulations in the country than in other large countries. With more stringent regulations, some business may flow to less stringent regulatory environments or jurisdictions.

2 A is correct. Blackout periods are established by *companies* in response to concerns about insider trading. Thus, blackout periods are not a tool used by regulators to intervene in the financial services sector. Capital requirements are used by government regulators to reduce systemic risk and financial contagion. Insider-trading restrictions are used by regulators concerned about insiders using their greater knowledge to the disadvantage of others; insider-trading restrictions respond to informational frictions.

3 A is correct. The hiring of more lawyers to deal with compliance is an example of an "unintended" implementation cost. Establishing legal standards for contracts and employers' rights and responsibilities are objectives (intended consequences) of some regulation.

4 A is correct. Regulation Q set a ceiling on the interest rates paid by banks for various types of deposits, which resulted in investors' shifting funds to money market funds.

5 A is correct. Le régulateur, as an independent regulator but not an SRO, has legal authority from the Genovia government to regulate. Therefore, Le régulateur both enacts and enforces regulations related to the new accounting statute in Genovia.

6 A is correct. L'organisation is an SRO but not an independent regulator, so it is a private entity that is not affiliated with Genovia's government. SROs that are not independent regulators receive authority from their members, who agree to comply with the organization's rules and standards and its enforcement thereof.

7 C is correct. Comment 1 is correct because regulatory filings, in response to proposed regulations, often identify at least some of the unintended consequences prior to the implementation of the regulation. Comment 2 is correct because the cost of unintended consequences, including both indirect costs and unanticipated implementation costs, can be high.

8 B is correct. A floating currency allows international trade in Genovia to be market based. International disputes about whether a country is manipulating or fixing its currency price often center on issues related to competitiveness.

9 C is correct. Regulatory competition describes actions by regulators to encourage behaviors. Regulators may compete to provide a regulatory environment designed to attract certain entities (regulatory competition). By amending regulations, Genovia's regulators seek to encourage foreign companies to change their corporate domicile.

10 C is correct. Genovia needs unambiguous laws concerning financial liability and bankruptcy to encourage foreign businesses to enter into contracts, particularly those that are long term and may involve sunk costs. The court decisions help Genovia achieve its goal. Also, clearly defined rules governing contracts, their interpretation, and each party's legal rights under a contract are necessary. Thus, both features help Genovia achieve its goal.

11 C is correct. At the end of Year 2, the compound annual growth rate (CAGR) for each sector is calculated as follows: (Year 2/Year 0)$^{0.5}$ – 1.

$$\text{Fossil fuels: } (426/462)^{0.5} - 1 = -4\%$$
$$\text{Hydropower: } (273/186)^{0.5} - 1 = 21\%$$
$$\text{Other renewables: } (154/97)^{0.5} - 1 = 26\%$$

The CAGR indicates that the 26% increase in production from the subsidy on other renewables has been more effective than the 4% decrease in production from the tax on fossil fuels or the 21% increase in production from the subsidy on hydropower. Thus, the subsidy on other renewables of 26% is the highest, indicating that this policy has been the most effective in helping Genovia achieve its second initiative.

APPENDICES

© 2019 CFA Institute. All rights reserved.

Appendix A
Cumulative Probabilities for a Standard Normal Distribution
P(*Z* ≤ x) = *N*(x) for x ≥ 0 or *P*(*Z* ≤ z) = *N*(z) for z ≥ 0

x or z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.50	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.60	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.70	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.80	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.90	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

For example, to find the *z*-value leaving 2.5 percent of the area/probability in the upper tail, find the element 0.9750 in the body of the table. Read 1.90 at the left end of the element's row and 0.06 at the top of the element's column, to give 1.90 + 0.06 = 1.96. *Table generated with Excel.*

Quantitative Methods for Investment Analysis, Second Edition, by Richard A. DeFusco, CFA, Dennis W. McLeavey, CFA, Jerald E. Pinto, CFA, and David E. Runkle, CFA. Copyright © 2004 by CFA Institute.

Appendix A (continued)
Cumulative Probabilities for a Standard Normal Distribution
$P(Z \le x) = N(x)$ for $x \le 0$ or $P(Z \le z) = N(z)$ for $z \le 0$

x or z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
−0.10	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.20	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.30	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.40	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.50	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.70	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.80	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.30	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.40	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.50	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.60	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.70	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.80	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.90	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−2.00	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−2.10	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.20	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.30	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.40	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.50	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.60	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.70	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.80	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.90	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−3.00	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−3.10	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.20	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.30	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.40	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.50	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
−3.60	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
−3.70	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
−3.80	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
−3.90	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
−4.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

For example, to find the z-value leaving 2.5 percent of the area/probability in the lower tail, find the element 0.0250 in the body of the table. Read −1.90 at the left end of the element's row and 0.06 at the top of the element's column, to give −1.90 − 0.06 = −1.96. *Table generated with Excel.*

Appendix B
Table of the Student's *t*-Distribution (One-Tailed Probabilities)

df	$p = 0.10$	$p = 0.05$	$p = 0.025$	$p = 0.01$	$p = 0.005$	df	$p = 0.10$	$p = 0.05$	$p = 0.025$	$p = 0.01$	$p = 0.005$
1	3.078	6.314	12.706	31.821	63.657	**31**	1.309	1.696	2.040	2.453	2.744
2	1.886	2.920	4.303	6.965	9.925	**32**	1.309	1.694	2.037	2.449	2.738
3	1.638	2.353	3.182	4.541	5.841	**33**	1.308	1.692	2.035	2.445	2.733
4	1.533	2.132	2.776	3.747	4.604	**34**	1.307	1.691	2.032	2.441	2.728
5	1.476	2.015	2.571	3.365	4.032	**35**	1.306	1.690	2.030	2.438	2.724
6	1.440	1.943	2.447	3.143	3.707	**36**	1.306	1.688	2.028	2.434	2.719
7	1.415	1.895	2.365	2.998	3.499	**37**	1.305	1.687	2.026	2.431	2.715
8	1.397	1.860	2.306	2.896	3.355	**38**	1.304	1.686	2.024	2.429	2.712
9	1.383	1.833	2.262	2.821	3.250	**39**	1.304	1.685	2.023	2.426	2.708
10	1.372	1.812	2.228	2.764	3.169	**40**	1.303	1.684	2.021	2.423	2.704
11	1.363	1.796	2.201	2.718	3.106	**41**	1.303	1.683	2.020	2.421	2.701
12	1.356	1.782	2.179	2.681	3.055	**42**	1.302	1.682	2.018	2.418	2.698
13	1.350	1.771	2.160	2.650	3.012	**43**	1.302	1.681	2.017	2.416	2.695
14	1.345	1.761	2.145	2.624	2.977	**44**	1.301	1.680	2.015	2.414	2.692
15	1.341	1.753	2.131	2.602	2.947	**45**	1.301	1.679	2.014	2.412	2.690
16	1.337	1.746	2.120	2.583	2.921	**46**	1.300	1.679	2.013	2.410	2.687
17	1.333	1.740	2.110	2.567	2.898	**47**	1.300	1.678	2.012	2.408	2.685
18	1.330	1.734	2.101	2.552	2.878	**48**	1.299	1.677	2.011	2.407	2.682
19	1.328	1.729	2.093	2.539	2.861	**49**	1.299	1.677	2.010	2.405	2.680
20	1.325	1.725	2.086	2.528	2.845	**50**	1.299	1.676	2.009	2.403	2.678
21	1.323	1.721	2.080	2.518	2.831	**60**	1.296	1.671	2.000	2.390	2.660
22	1.321	1.717	2.074	2.508	2.819	**70**	1.294	1.667	1.994	2.381	2.648
23	1.319	1.714	2.069	2.500	2.807	**80**	1.292	1.664	1.990	2.374	2.639
24	1.318	1.711	2.064	2.492	2.797	**90**	1.291	1.662	1.987	2.368	2.632
25	1.316	1.708	2.060	2.485	2.787	**100**	1.290	1.660	1.984	2.364	2.626
26	1.315	1.706	2.056	2.479	2.779	**110**	1.289	1.659	1.982	2.361	2.621
27	1.314	1.703	2.052	2.473	2.771	**120**	1.289	1.658	1.980	2.358	2.617
28	1.313	1.701	2.048	2.467	2.763	**200**	1.286	1.653	1.972	2.345	2.601
29	1.311	1.699	2.045	2.462	2.756	∞	1.282	1.645	1.960	2.326	2.576
30	1.310	1.697	2.042	2.457	2.750						

To find a critical t-value, enter the table with df and a specified value for α, the significance level. For example, with 5 df, $\alpha = 0.05$ and a one-tailed test, the desired probability in the tail would be $p = 0.05$ and the critical t-value would be $t(5, 0.05) = 2.015$. With $\alpha = 0.05$ and a two-tailed test, the desired probability in each tail would be $p = 0.025 = \alpha/2$, giving $t(5, 0.025) = 2.571$.
Table generated using Excel.

Quantitative Methods for Investment Analysis, Second Edition, by Richard A. DeFusco, CFA, Dennis W. McLeavey, CFA, Jerald E. Pinto, CFA, and David E. Runkle, CFA. Copyright © 2004 by CFA Institute.

Appendix C
Values of χ^2 (Degrees of Freedom, Level of Significance)

Degrees of Freedom	Probability in Right Tail								
	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.000157	0.000982	0.003932	0.0158	2.706	3.841	5.024	6.635	7.879
2	0.020100	0.050636	0.102586	0.2107	4.605	5.991	7.378	9.210	10.597
3	0.1148	0.2158	0.3518	0.5844	6.251	7.815	9.348	11.345	12.838
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
100	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.170

To have a probability of 0.05 in the right tail when df = 5, the tabled value is $\chi^2(5, 0.05) = 11.070$.

Quantitative Methods for Investment Analysis, Second Edition, by Richard A. DeFusco, CFA, Dennis W. McLeavey, CFA, Jerald E. Pinto, CFA, and David E. Runkle, CFA. Copyright © 2004 by CFA Institute.

Appendix D
Table of the *F*-Distribution

Panel A. Critical values for right-hand tail area equal to 0.05 — **Numerator: df_1 and Denominator: df_2**

	df1:1	2	3	4	5	6	7	8	9	10	11	12	15	20	21	22	23	24	25	30	40	60	120	∞
df2: 1	161	200	216	225	230	234	237	239	241	242	243	244	246	248	248	249	249	249	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.70	8.66	8.65	8.65	8.64	8.64	8.63	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.86	5.80	5.79	5.79	5.78	5.77	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.55	4.54	4.53	4.53	4.52	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.94	3.87	3.86	3.86	3.85	3.84	3.83	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.51	3.44	3.43	3.43	3.42	3.41	3.40	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.22	3.15	3.14	3.13	3.12	3.12	3.11	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.01	2.94	2.93	2.92	2.91	2.90	2.89	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.85	2.77	2.76	2.75	2.75	2.74	2.73	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.72	2.65	2.64	2.63	2.62	2.61	2.60	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.62	2.54	2.53	2.52	2.51	2.51	2.50	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.53	2.46	2.45	2.44	2.43	2.42	2.41	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.46	2.39	2.38	2.37	2.36	2.35	2.34	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.40	2.33	2.32	2.31	2.30	2.29	2.28	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.35	2.28	2.26	2.25	2.24	2.24	2.23	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.31	2.23	2.22	2.21	2.20	2.19	2.18	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.27	2.19	2.18	2.17	2.16	2.15	2.14	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.23	2.16	2.14	2.13	2.12	2.11	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.20	2.12	2.11	2.10	2.09	2.08	2.07	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.18	2.10	2.08	2.07	2.06	2.05	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.15	2.07	2.06	2.05	2.04	2.03	2.02	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.13	2.05	2.04	2.02	2.01	2.01	2.00	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.11	2.03	2.01	2.00	1.99	1.98	1.97	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.09	2.01	2.00	1.98	1.97	1.96	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.01	1.93	1.92	1.91	1.90	1.89	1.88	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.92	1.84	1.83	1.81	1.80	1.79	1.78	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.84	1.75	1.73	1.72	1.71	1.70	1.69	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.75	1.66	1.64	1.63	1.62	1.61	1.60	1.55	1.50	1.43	1.35	1.25
Infinity	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.67	1.57	1.56	1.54	1.53	1.52	1.51	1.46	1.39	1.32	1.22	1.00

Appendix D (continued)
Table of the *F*-Distribution

Panel B. Critical values for right-hand tail area equal to 0.025 — Numerator: df_1 and Denominator: df_2

	df1: 1	2	3	4	5	6	7	8	9	10	11	12	15	20	21	22	23	24	25	30	40	60	120	∞
df2: 1	648	799	864	900	922	937	948	957	963	969	973	977	985	993	994	995	996	997	998	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.41	39.43	39.45	39.45	39.45	39.45	39.46	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.37	14.34	14.25	14.17	14.16	14.14	14.13	14.12	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75	8.66	8.56	8.55	8.53	8.52	8.51	8.50	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52	6.43	6.33	6.31	6.30	6.29	6.28	6.27	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	5.37	5.27	5.17	5.15	5.14	5.13	5.12	5.11	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71	4.67	4.57	4.47	4.45	4.44	4.43	4.41	4.40	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24	4.20	4.10	4.00	3.98	3.97	3.96	3.95	3.94	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	3.87	3.77	3.67	3.65	3.64	3.63	3.61	3.60	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66	3.62	3.52	3.42	3.40	3.39	3.38	3.37	3.35	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47	3.43	3.33	3.23	3.21	3.20	3.18	3.17	3.16	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	3.28	3.18	3.07	3.06	3.04	3.03	3.02	3.01	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.20	3.15	3.05	2.95	2.93	2.92	2.91	2.89	2.88	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.09	3.05	2.95	2.84	2.83	2.81	2.80	2.79	2.78	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01	2.96	2.86	2.76	2.74	2.73	2.71	2.70	2.69	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.93	2.89	2.79	2.68	2.67	2.65	2.64	2.63	2.61	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.87	2.82	2.72	2.62	2.60	2.59	2.57	2.56	2.55	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.81	2.77	2.67	2.56	2.54	2.53	2.52	2.50	2.49	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.76	2.72	2.62	2.51	2.49	2.48	2.46	2.45	2.44	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	2.68	2.57	2.46	2.45	2.43	2.42	2.41	2.40	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.68	2.64	2.53	2.42	2.41	2.39	2.38	2.37	2.36	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.65	2.60	2.50	2.39	2.37	2.36	2.34	2.33	2.32	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.62	2.57	2.47	2.36	2.34	2.33	2.31	2.30	2.29	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59	2.54	2.44	2.33	2.31	2.30	2.28	2.27	2.26	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.56	2.51	2.41	2.30	2.28	2.27	2.26	2.24	2.23	2.18	2.12	2.05	1.98	1.91
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	2.41	2.31	2.20	2.18	2.16	2.15	2.14	2.12	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	2.29	2.18	2.07	2.05	2.03	2.02	2.01	1.99	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22	2.17	2.06	1.94	1.93	1.91	1.90	1.88	1.87	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.10	2.05	1.94	1.82	1.81	1.79	1.77	1.76	1.75	1.69	1.61	1.53	1.43	1.31
Infinity	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.99	1.94	1.83	1.71	1.69	1.67	1.66	1.64	1.63	1.57	1.48	1.39	1.27	1.00

Appendix D (continued)
Table of the *F*-Distribution

Panel C. Critical values for right-hand tail area equal to 0.01 — Numerator: df_1 and Denominator: df_2

	df1: 1	2	3	4	5	6	7	8	9	10	11	12	15	20	21	22	23	24	25	30	40	60	120	∞
df2: 1	4052	5000	5403	5625	5764	5859	5928	5982	6023	6056	6083	6106	6157	6209	6216	6223	6229	6235	6240	6261	6287	6313	6339	6366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1	26.9	26.7	26.7	26.6	26.6	26.6	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.5	14.4	14.2	14.0	14.0	14.0	13.9	13.9	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	10.0	9.89	9.72	9.55	9.53	9.51	9.49	9.47	9.45	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.56	7.40	7.37	7.35	7.33	7.31	7.30	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.31	6.16	6.13	6.11	6.09	6.07	6.06	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.52	5.36	5.34	5.32	5.30	5.28	5.26	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	4.96	4.81	4.79	4.77	4.75	4.73	4.71	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.56	4.41	4.38	4.36	4.34	4.33	4.31	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.25	4.10	4.08	4.06	4.04	4.02	4.01	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.01	3.86	3.84	3.82	3.80	3.78	3.76	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.82	3.66	3.64	3.62	3.60	3.59	3.57	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.66	3.51	3.48	3.46	3.44	3.43	3.41	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.52	3.37	3.35	3.33	3.31	3.29	3.28	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.41	3.26	3.24	3.22	3.20	3.18	3.16	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.31	3.16	3.14	3.12	3.10	3.08	3.07	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.23	3.08	3.05	3.03	3.02	3.00	2.98	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.15	3.00	2.98	2.96	2.94	2.92	2.91	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.09	2.94	2.92	2.90	2.88	2.86	2.84	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.03	2.88	2.86	2.84	2.82	2.80	2.79	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	2.98	2.83	2.81	2.78	2.77	2.75	2.73	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.93	2.78	2.76	2.74	2.72	2.70	2.69	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.89	2.74	2.72	2.70	2.68	2.66	2.64	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.85	2.70	2.68	2.66	2.64	2.62	2.60	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.70	2.55	2.53	2.51	2.49	2.47	2.45	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.52	2.37	2.35	2.33	2.31	2.29	2.27	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.35	2.20	2.17	2.15	2.13	2.12	2.10	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.19	2.03	2.01	1.99	1.97	1.95	1.93	1.86	1.76	1.66	1.53	1.38
Infinity	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.04	1.88	1.85	1.83	1.81	1.79	1.77	1.70	1.59	1.47	1.32	1.00

Appendix D (continued)
Table of the *F*-Distribution

Panel D. Critical values for right-hand tail area equal to 0.005 — **Numerator: df_1 and Denominator: df_2**

	df1: 1	2	3	4	5	6	7	8	9	10	11	12	15	20	21	22	23	24	25	30	40	60	120	∞
df2: 1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24222	24334	24426	24630	24836	24863	24892	24915	24940	24959	25044	25146	25253	25359	25464
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	200
3	55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.68	43.52	43.39	43.08	42.78	42.73	42.69	42.66	42.62	42.59	42.47	42.31	42.15	41.99	41.83
4	31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14	20.97	20.82	20.70	20.44	20.17	20.13	20.09	20.06	20.03	20.00	19.89	19.75	19.61	19.47	19.32
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.49	13.38	13.15	12.90	12.87	12.84	12.81	12.78	12.76	12.66	12.53	12.40	12.27	12.14
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.13	10.03	9.81	9.59	9.56	9.53	9.50	9.47	9.45	9.36	9.24	9.12	9.00	8.88
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.27	8.18	7.97	7.75	7.72	7.69	7.67	7.64	7.62	7.53	7.42	7.31	7.19	7.08
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.10	7.01	6.81	6.61	6.58	6.55	6.53	6.50	6.48	6.40	6.29	6.18	6.06	5.95
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.31	6.23	6.03	5.83	5.80	5.78	5.75	5.73	5.71	5.62	5.52	5.41	5.30	5.19
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.75	5.66	5.47	5.27	5.25	5.22	5.20	5.17	5.15	5.07	4.97	4.86	4.75	4.64
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.32	5.24	5.05	4.86	4.83	4.80	4.78	4.76	4.74	4.65	4.55	4.45	4.34	4.23
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.99	4.91	4.72	4.53	4.50	4.48	4.45	4.43	4.41	4.33	4.23	4.12	4.01	3.90
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.72	4.64	4.46	4.27	4.24	4.22	4.19	4.17	4.15	4.07	3.97	3.87	3.76	3.65
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.51	4.43	4.25	4.06	4.03	4.01	3.98	3.96	3.94	3.86	3.76	3.66	3.55	3.44
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.33	4.25	4.07	3.88	3.86	3.83	3.81	3.79	3.77	3.69	3.59	3.48	3.37	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.18	4.10	3.92	3.73	3.71	3.68	3.66	3.64	3.62	3.54	3.44	3.33	3.22	3.11
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	4.05	3.97	3.79	3.61	3.58	3.56	3.53	3.51	3.49	3.41	3.31	3.21	3.10	2.98
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.94	3.86	3.68	3.50	3.47	3.45	3.42	3.40	3.38	3.30	3.20	3.10	2.99	2.87
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.84	3.76	3.59	3.40	3.37	3.35	3.33	3.31	3.29	3.21	3.11	3.00	2.89	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.76	3.68	3.50	3.32	3.29	3.27	3.24	3.22	3.20	3.12	3.02	2.92	2.81	2.69
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.68	3.60	3.43	3.24	3.22	3.19	3.17	3.15	3.13	3.05	2.95	2.84	2.73	2.61
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.61	3.54	3.36	3.18	3.15	3.12	3.10	3.08	3.06	2.98	2.88	2.77	2.66	2.55
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.55	3.47	3.30	3.12	3.09	3.06	3.04	3.02	3.00	2.92	2.82	2.71	2.60	2.48
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.50	3.42	3.25	3.06	3.04	3.01	2.99	2.97	2.95	2.87	2.77	2.66	2.55	2.43
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.45	3.37	3.20	3.01	2.99	2.96	2.94	2.92	2.90	2.82	2.72	2.61	2.50	2.38
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.25	3.18	3.01	2.82	2.80	2.77	2.75	2.73	2.71	2.63	2.52	2.42	2.30	2.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	3.03	2.95	2.78	2.60	2.57	2.55	2.52	2.50	2.48	2.40	2.30	2.18	2.06	1.93
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.82	2.74	2.57	2.39	2.36	2.33	2.31	2.29	2.27	2.19	2.08	1.96	1.83	1.69
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.62	2.54	2.37	2.19	2.16	2.13	2.11	2.09	2.07	1.98	1.87	1.75	1.61	1.43
Infinity	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.43	2.36	2.19	2.00	1.97	1.95	1.92	1.90	1.88	1.79	1.67	1.53	1.36	1.00

With 1 degree of freedom (df) in the numerator and 3 df in the denominator, the critical *F*-value is 10.1 for a right-hand tail area equal to 0.05.

Quantitative Methods for Investment Analysis, Second Edition, by Richard A. DeFusco, CFA, Dennis W. McLeavey, CFA, Jerald E. Pinto, CFA, and David E. Runkle, CFA. Copyright © 2004 by CFA Institute.

Appendix E
Critical Values for the Durbin-Watson Statistic (α = .05)

	K = 1		K = 2		K = 3		K = 4		K = 5	
n	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Note: K = the number of slope parameters in the model.

Source: From J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression, II." *Biometrika* 38 (1951): 159–178.

Glossary

Abandonment option The ability to terminate a project at some future time if the financial results are disappointing.

Abnormal earnings See *residual income.*

Abnormal return The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

Absolute convergence The idea that developing countries, regardless of their particular characteristics, will eventually catch up with the developed countries and match them in per capita output.

Absolute valuation model A model that specifies an asset's intrinsic value.

Absolute version of PPP An extension of the law of one price whereby the prices of goods and services will not differ internationally once exchange rates are considered.

Accounting estimates Estimates used in calculating the value of assets or liabilities and in the amount of revenue and expense to allocate to a period. Examples of accounting estimates include, among others, the useful lives of depreciable assets, the salvage value of depreciable assets, product returns, warranty costs, and the amount of uncollectible receivables.

Accumulated benefit obligation The actuarial present value of benefits (whether vested or non-vested) attributed, generally by the pension benefit formula, to employee service rendered before a specified date and based on employee service and compensation (if applicable) before that date. The accumulated benefit obligation differs from the projected benefit obligation in that it includes no assumption about future compensation levels.

Accuracy The percentage of correctly predicted classes out of total predictions. It is an overall performance metric in classification problems.

Acquirer The company in a merger or acquisition that is acquiring the target.

Acquiring company See *acquirer*.

Acquisition The purchase of some portion of one company by another; the purchase may be for assets, a definable segment of another entity, or the entire company.

Activation function A functional part of a neural network's node that transforms the total net input received into the final output of the node. The activation function operates like a light dimmer switch that decreases or increases the strength of the input.

Active factor risk The contribution to active risk squared resulting from the portfolio's different-than-benchmark exposures relative to factors specified in the risk model.

Active return The return on a portfolio minus the return on the portfolio's benchmark.

Active risk The standard deviation of active returns.

Active risk squared The variance of active returns; active risk raised to the second power.

Active share A measure of how similar a portfolio is to its benchmark. A manager who precisely replicates the benchmark will have an active share of zero; a manager with no holdings in common with the benchmark will have an active share of one.

Active specific risk The contribution to active risk squared resulting from the portfolio's active weights on individual assets as those weights interact with assets' residual risk.

Adjusted funds from operations (AFFO) Funds from operations adjusted to remove any non-cash rent reported under straight-line rent accounting and to subtract maintenance-type capital expenditures and leasing costs, including leasing agents' commissions and tenants' improvement allowances.

Adjusted present value As an approach to valuing a company, the sum of the value of the company, assuming no use of debt, and the net present value of any effects of debt on company value.

Adjusted R^2 A measure of goodness-of-fit of a regression that is adjusted for degrees of freedom and hence does not automatically increase when another independent variable is added to a regression.

Administrative regulations or administrative law Rules issued by government agencies or other regulators.

Advanced set An arrangement in which the reference interest rate is set at the time the money is deposited.

Advanced settled An arrangement in which a forward rate agreement (FRA) expires and settles at the same time, at the FRA expiration date.

Agency costs Costs associated with the conflict of interest present when a company is managed by non-owners. Agency costs result from the inherent conflicts of interest between managers and equity owners.

Agency costs of equity The smaller the stake managers have in the company, the less their share in bearing the cost of excessive perquisite consumption—consequently, the less their desire to give their best efforts in running the company.

Agency issues Conflicts of interest that arise when the agent in an agency relationship has goals and incentives that differ from the principal to whom the agent owes a fiduciary duty. Also called *agency problems* or *principal–agent problems*.

Agglomerative clustering A bottom-up hierarchical clustering method that begins with each observation being treated as its own cluster. The algorithm finds the two closest clusters, based on some measure of distance (similarity), and combines them into one new larger cluster. This process is repeated iteratively until all observations are clumped into a single large cluster.

Allowance for loan losses A balance sheet account; it is a contra asset account to loans.

Alpha The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

American Depositary Receipt A negotiable certificate issued by a depositary bank that represents ownership in a non-US company's deposited equity (i.e., equity held in custody by the depositary bank in the company's home market).

Analysis of variance (ANOVA) The analysis that breaks the total variability of a dataset (such as observations on the dependent variable in a regression) into components representing different sources of variation. With reference to regression, ANOVA provides the inputs for an *F*-test of

the significance of the regression as a whole, as well as the inputs for the coefficient of determination and the standard error of the estimate.

Application programming interface (API) A set of well-defined methods of communication between various software components and typically used for accessing external data.

Arbitrage (1) The simultaneous purchase of an undervalued asset or portfolio and sale of an overvalued but equivalent asset or portfolio in order to obtain a riskless profit on the price differential. Taking advantage of a market inefficiency in a risk-free manner. (2) The condition in a financial market in which equivalent assets or combinations of assets sell for two different prices, creating an opportunity to profit at no risk with no commitment of money. In a well-functioning financial market, few arbitrage opportunities are possible. (3) A risk-free operation that earns an expected positive net profit but requires no net investment of money.

Arbitrage-free models Term structure models that project future interest rate paths that emanate from the existing term structure. Resulting prices are based on a no-arbitrage condition.

Arbitrage-free valuation An approach to valuation that determines security values consistent with the absence of any opportunity to earn riskless profits without any net investment of money.

Arbitrage opportunity An opportunity to conduct an arbitrage; an opportunity to earn an expected positive net profit without risk and with no net investment of money.

Arbitrage portfolio The portfolio that exploits an arbitrage opportunity.

Ask price The price at which a trader will sell a specified quantity of a security. Also called *ask*, *offer price*, or *offer*.

Asset-based approach Approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities.

Asset-based valuation An approach to valuing natural resource companies that estimates company value on the basis of the market value of the natural resources the company controls.

Asset beta The unlevered beta; reflects the business risk of the assets; the asset's systematic risk.

Asset purchase An acquisition in which the acquirer purchases the target company's assets and payment is made directly to the target company.

Asymmetric information The differential of information between corporate insiders and outsiders regarding the company's performance and prospects. Managers typically have more information about the company's performance and prospects than owners and creditors.

At market contract When a forward contract is established, the forward price is negotiated so that the market value of the forward contract on the initiation date is zero.

Authorized participants (APs) A special group of institutional investors who are authorized by the ETF issuer to participate in the creation/redemption process. APs are large broker/dealers, often market makers.

Autocorrelations The correlations of a time series with its own past values.

Autoregressive model (AR) A time series regressed on its own past values in which the independent variable is a lagged value of the dependent variable.

Backtesting The process that approximates the real-life investment process, using historical data, to assess whether an investment strategy would have produced desirable results.

Backward integration A merger involving the purchase of a target ahead of the acquirer in the value or production chain; for example, to acquire a supplier.

Backward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving backward through the network's layers.

Backwardation A condition in futures markets in which the spot price exceeds the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is higher than the longer-term futures contract price.

Bag-of-words (BOW) A collection of a distinct set of tokens from all the texts in a sample dataset. BOW does not capture the position or sequence of words present in the text.

Bankruptcy A declaration provided for by a country's laws that typically involves the establishment of a legal procedure that forces creditors to defer their claims.

Barbell portfolio Fixed-income portfolio that combines short and long maturities.

Base error Model error due to randomness in the data.

Basic earnings per share (EPS) Net earnings available to common shareholders (i.e., net income minus preferred dividends) divided by the weighted average number of common shares outstanding during the period.

Basis The difference between the spot price and the futures price. As the maturity date of the futures contract nears, the basis converges toward zero.

Basis trade A trade based on the pricing of credit in the bond market versus the price of the same credit in the CDS market. To execute a basis trade, go long the "underpriced" credit and short the "overpriced" credit. A profit is realized as the implied credit prices converge.

Bear hug A tactic used by acquirers to circumvent target management's objections to a proposed merger by submitting the proposal directly to the target company's board of directors.

Bearish flattening Term structure shift in which short-term bond yields rise more than long-term bond yields, resulting in a flatter yield curve.

Benchmark value of the multiple In using the method of comparables, the value of a price multiple for the comparison asset; when we have comparison assets (a group), the mean or median value of the multiple for the group of assets.

Best ask The offer to sell with the lowest ask price. Also called *best offer* or *inside ask*.

Best bid The offer to buy with the highest bid price. Also called the *inside bid*.

Best offer See *best ask*.

Bias error Describes the degree to which a model fits the training data. Algorithms with erroneous assumptions produce high bias error with poor approximation, causing underfitting and high in-sample error.

Bid–ask spread The ask price minus the bid price.

Bid price The price at which a trader will buy a specified quantity of a security. Also called *bid*.

Bill-and-hold basis Sales on a bill-and-hold basis involve selling products but not delivering those products until a later date.

Blockage factor An illiquidity discount that occurs when an investor sells a large amount of stock relative to its trading volume (assuming it is not large enough to constitute a controlling ownership).

Bond indenture A legal contract specifying the terms of a bond issue.

Bond risk premium The expected excess return of a default-free long-term bond less that of an equivalent short-term bond.

Bond yield plus risk premium method An estimate of the cost of common equity that is produced by summing the before-tax cost of debt and a risk premium that captures the additional yield on a company's stock relative to its bonds. The additional yield is often estimated using historical spreads between bond yields and stock yields.

Bonding costs Costs borne by management to assure owners that they are working in the owners' best interest (e.g., implicit cost of non-compete agreements).

Bonus issue of shares *See stock dividend.*

Book value Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value of equity Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value per share The amount of book value (also called carrying value) of common equity per share of common stock, calculated by dividing the book value of shareholders' equity by the number of shares of common stock outstanding.

Bootstrap aggregating (or bagging) A technique whereby the original training dataset is used to generate *n* new training datasets or bags of data. Each new bag of data is generated by random sampling with replacement from the initial training set.

Bootstrapping The use of a forward substitution process to determine zero-coupon rates by using the par yields and solving for the zero-coupon rates one by one, from the shortest to longest maturities.

Bottom-up approach With respect to forecasting, an approach that usually begins at the level of the individual company or a unit within the company.

Breakup value The value derived using a sum-of-the-parts valuation.

Breusch–Pagan test A test for conditional heteroskedasticity in the error term of a regression.

Bullet portfolio A fixed-income portfolio concentrated in a single maturity.

Bullish flattening Term structure change in which the yield curve flattens in response to a greater decline in long-term rates than short-term rates.

Bullish steepening Term structure change in which short-term rates fall by more than long-term yields, resulting in a steeper term structure.

Buy-side analysts Analysts who work for investment management firms, trusts, bank trust departments, and similar institutions.

Buyback *See share repurchase.*

Callable bond Bond that includes an embedded call option that gives the issuer the right to redeem the bond issue prior to maturity, typically when interest rates have fallen or when the issuer's credit quality has improved.

Canceled shares Shares that were issued, subsequently repurchased by the company, and then retired (cannot be reissued).

Cannibalization Cannibalization occurs when an investment takes customers and sales away from another part of the company.

Capital charge The company's total cost of capital in money terms.

Capital deepening An increase in the capital-to-labor ratio.

Capital rationing A capital rationing environment assumes that the company has a fixed amount of funds to invest.

Capital structure The mix of debt and equity that a company uses to finance its business; a company's specific mixture of long-term financing.

Capitalization of earnings method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capitalization rate The divisor in the expression for the value of perpetuity. In the context of real estate, it is the divisor in the direct capitalization method of estimating value. The cap rate equals net operating income divided by value.

Capitalized cash flow method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity. Also called *capitalized cash flow model*.

Capitalized income method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capped floater Floating-rate bond with a cap provision that prevents the coupon rate from increasing above a specified maximum rate. It protects the issuer against rising interest rates.

Carry arbitrage model A no-arbitrage approach in which the underlying instrument is either bought or sold along with an opposite position in a forward contract.

Carry benefits Benefits that arise from owning certain underlyings; for example, dividends, foreign interest, and bond coupon payments.

Carry costs Costs that arise from owning certain underlyings. They are generally a function of the physical characteristics of the underlying asset and also the interest forgone on the funds tied up in the asset.

Cash available for distribution See *adjusted funds from operations*.

Cash-generating unit The smallest identifiable group of assets that generates cash inflows that are largely independent of the cash inflows of other assets or groups of assets.

Cash offering A merger or acquisition that is to be paid for with cash; the cash for the merger might come from the acquiring company's existing assets or from a debt issue.

Cash settlement A procedure used in certain derivative transactions that specifies that the long and short parties settle the derivative's difference in value between them by making a cash payment.

Catalyst An event or piece of information that causes the marketplace to re-evaluate the prospects of a company.

Categorical dependent variables An alternative term for qualitative dependent variables.

CDS spread A periodic premium paid by the buyer to the seller that serves as a return over a market reference rate required to protect against credit risk.

Ceiling analysis A systematic process of evaluating different components in the pipeline of model building. It helps to understand what part of the pipeline can potentially improve in performance by further tuning.

Centroid The center of a cluster formed using the *k*-means clustering algorithm.

Chain rule of forecasting A forecasting process in which the next period's value as predicted by the forecasting equation is substituted into the right-hand side of the equation to give a predicted value two periods ahead.

Cheapest-to-deliver The debt instrument that can be purchased and delivered at the lowest cost yet has the same seniority as the reference obligation.

Classification and regression tree A supervised machine learning technique that can be applied to predict either a categorical target variable, producing a classification tree, or a continuous target variable, producing a regression tree. CART is commonly applied to binary classification or regression.

Clean surplus relation The relationship between earnings, dividends, and book value in which ending book value is equal to the beginning book value plus earnings less dividends, apart from ownership transactions.

Club convergence The idea that only rich and middle-income countries sharing a set of favorable attributes (i.e., are members of the "club") will converge to the income level of the richest countries.

Cluster A subset of observations from a dataset such that all the observations within the same cluster are deemed "similar."

Clustering The sorting of observations into groups (clusters) such that observations in the same cluster are more similar to each other than they are to observations in other clusters.

Cobb–Douglas production function A function of the form $Y = K^{\alpha} L^{1-\alpha}$ relating output (Y) to labor (L) and capital (K) inputs.

Coefficient of determination The percentage of the variation of the dependent variable that is explained by the independent variable. Also referred to as the "*R*-squared" or "R^2."

Cointegrated Describes two time series that have a long-term financial or economic relationship such that they do not diverge from each other without bound in the long run.

Collateral return The component of the total return on a commodity futures position attributable to the yield for the bonds or cash used to maintain the futures position. Also called *collateral yield*.

Collection frequency (CF) The number of times a given word appears in the whole corpus (i.e., collection of sentences) divided by the total number of words in the corpus.

Commercial real estate properties Income-producing real estate properties; properties purchased with the intent to let, lease, or rent (in other words, produce income).

Commodity swap A type of swap involving the exchange of payments over multiple dates as determined by specified reference prices or indexes relating to commodities.

Common size statements Financial statements in which all elements (accounts) are stated as a percentage of a key figure, such as revenue for an income statement or total assets for a balance sheet.

Company fundamental factors Factors related to the company's internal performance, such as factors relating to earnings growth, earnings variability, earnings momentum, and financial leverage.

Company share-related factors Valuation measures and other factors related to share price or the trading characteristics of the shares, such as earnings yield, dividend yield, and book-to-market value.

Comparables Assets used as benchmarks when applying the method of comparables to value an asset. Also called *comps*, *guideline assets*, or *guideline companies*.

Competition laws A law that promotes or maintains market competition by regulating anti-competitive conduct. Known as "antitrust law" in the United States, "anti-monopoly law" in China and Russia, and often referred to as "trade practices law" in the United Kingdom and Australia.

Compiled financial statements Financial statements that are not accompanied by an auditor's opinion letter.

Complexity A term referring to the number of features, parameters, or branches in a model and to whether the model is linear or non-linear (non-linear is more complex).

Composite variable A variable that combines two or more variables that are statistically strongly related to each other.

Comprehensive income All changes in equity other than contributions by, and distributions to, owners; income under clean surplus accounting; includes all changes in equity during a period except those resulting from investments by owners and distributions to owners. Comprehensive income equals net income plus other comprehensive income.

Comps Assets used as benchmarks when applying the method of comparables to value an asset.

Concentrated ownership Ownership structure consisting of an individual shareholder or a group (controlling shareholders) with the ability to exercise control over the corporation.

Conditional convergence The idea that convergence of per capita income is conditional on the countries having the same savings rate, population growth rate, and production function.

Conditional heteroskedasticity Heteroskedasticity in the error variance that is correlated with the values of the independent variable(s) in the regression.

Conditional VaR (CVaR) The weighted average of all loss outcomes in the statistical (i.e., return) distribution that exceed the VaR loss. Thus, CVaR is a more comprehensive measure of tail loss than VaR is. Sometimes referred to as the *expected tail loss* or *expected shortfall*.

Confusion matrix A grid used for error analysis in classification problems, it presents values for four evaluation metrics including true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Conglomerate discount The discount possibly applied by the market to the stock of a company operating in multiple, unrelated businesses.

Conglomerate merger A merger involving companies that are in unrelated businesses.

Consolidation The combining of the results of operations of subsidiaries with the parent company to present financial statements as if they were a single economic unit. The assets, liabilities, revenues, and expenses of the subsidiaries are combined with those of the parent company, eliminating intercompany transactions.

Constant dividend payout ratio policy A policy in which a constant percentage of net income is paid out in dividends.

Constant returns to scale The condition that if all inputs into the production process are increased by a given percentage, then output rises by that same percentage.

Contango A condition in futures markets in which the spot price is lower than the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is lower than the longer-term futures contract price.

Contingent consideration Potential future payments to the seller that are contingent on the achievement of certain agreed-on occurrences.

Continuing earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *persistent earnings*, or *underlying earnings*.

Continuing residual income Residual income after the forecast horizon.

Continuing value The analyst's estimate of a stock's value at a particular point in the future.

Control premium An increment or premium to value associated with a controlling ownership interest in a company.

Convergence The property by which as expiration approaches, the price of a newly created forward or futures contract will approach the price of a spot transaction. At expiration, a forward or futures contract is equivalent to a spot transaction in the underlying.

Conversion period For a convertible bond, the period during which bondholders have the right to convert their bonds into shares.

Conversion price For a convertible bond, the price per share at which the bond can be converted into shares.

Conversion rate (or ratio) For a convertible bond, the number of shares of common stock that a bondholder receives from converting the bond into shares.

Conversion value For a convertible bond, the value of the bond if it is converted at the market price of the shares. Also called *parity value*.

Convertible bond Bond with an embedded conversion option that gives bondholders the right to convert their bonds into the issuer's common stock during a pre-determined period at a pre-determined price.

Convexity A measure of how interest rate sensitivity changes with a change in interest rates.

Core earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *persistent earnings*, or *underlying earnings*.

Core real estate investment style Investing in high-quality, well-leased, core property types with low leverage (no more than 30% of asset value) in the largest markets with strong, diversified economies. It is a conservative strategy designed to avoid real estate–specific risks, including leasing, development, and speculation in favor of steady returns. Hotel properties are excluded from the core categories because of the higher cash flow volatility resulting from single-night leases and the greater importance of property operations, brand, and marketing.

Corpus A collection of text data in any form, including list, matrix, or data table forms.

Cost approach An approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities. In the context of real estate, this approach estimates the value of a property based on what it would cost to buy the land and construct a new property on the site that has the same utility or functionality as the property being appraised.

Cost of carry model A model that relates the forward price of an asset to the spot price by considering the cost of carry (also referred to as future-spot parity model).

Cost of debt The cost of debt financing to a company, such as when it issues a bond or takes out a bank loan.

Cost of equity The required rate of return on common stock.

Covariance stationary Describes a time series when its expected value and variance are constant and finite in all periods and when its covariance with itself for a fixed number of periods in the past or future is constant and finite in all periods.

Covered bonds A senior debt obligation of a financial institution that gives recourse to the originator/issuer and a predetermined underlying collateral pool.

Covered interest rate parity The relationship among the spot exchange rate, the forward exchange rate, and the interest rates in two currencies that ensures that the return on a hedged (i.e., covered) foreign risk-free investment is the same as the return on a domestic risk-free investment. Also called *interest rate parity*.

Cox-Ingersoll-Ross model A general equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is directly related to the level of interest rates.

Creation basket The list of securities (and share amounts) the authorized participant (AP) must deliver to the ETF manager in exchange for ETF shares. The creation basket is published each business day.

Creation/redemption The process in which ETF shares are created or redeemed by authorized participants transacting with the ETF issuer.

Creation units Large blocks of ETF shares transacted between the authorized participant (AP) and the ETF manager that are usually but not always equal to 50,000 shares of the ETF.

Credit correlation The correlation of credit (or default) risks of the underlying single-name CDS contained in an index CDS.

Credit curve The credit spreads for a range of maturities of a company's debt.

Credit default swap A derivative contract between two parties in which the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit derivative A derivative instrument in which the underlying is a measure of the credit quality of a borrower.

Credit event The event that triggers a payment from the credit protection seller to the credit protection buyer.

Credit protection buyer One party to a credit default swap; the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit protection seller One party to a credit default swap; the seller makes a promise to pay compensation for credit losses resulting from the default.

Credit risk The risk that the borrower will not repay principal and interest. Also called *default risk*.

Credit valuation adjustment The value of the credit risk of a bond in present value terms.

Cross-validation A technique for estimating out-of-sample error directly by determining the error in validation samples.

Current exchange rate For accounting purposes, the spot exchange rate on the balance sheet date.

Current rate method Approach to translating foreign currency financial statements for consolidation in which all assets and liabilities are translated at the current exchange rate. The current rate method is the prevalent method of translation.

Curvature One of the three factors (the other two are level and steepness) that empirically explain most of the changes in the shape of the yield curve. A shock to the curvature factor affects mid-maturity interest rates, resulting in the term structure becoming either more or less hump-shaped.

Curve trade Buying a CDS of one maturity and selling a CDS on the same reference entity with a different maturity.

Cyclical businesses Businesses with high sensitivity to business- or industry-cycle influences.

Data mining The practice of determining a model by extensive searching through a dataset for statistically significant patterns.

Data preparation (cleansing) The process of examining, identifying, and mitigating (i.e., cleansing) errors in raw data.

Data snooping The subconscious or conscious manipulation of data in a way that produces a statistically significant result (i.e., the *p*-value is sufficiently small or the *t*-statistic sufficiently large to indicate statistical significance), such as by running multiple simulations and naively accepting the best result. Also known as p-hacking.

Data wrangling (preprocessing) This task performs transformations and critical processing steps on cleansed data to make the data ready for ML model training (i.e., preprocessing), and includes dealing with outliers, extracting useful variables from existing data points, and scaling the data.

"Dead-hand" provision A poison pill provision that allows for the redemption or cancellation of a poison pill provision only by a vote of continuing directors (generally directors who were on the target company's board prior to the takeover attempt).

Debt rating An objective measure of the quality and safety of a company's debt based upon an analysis of the company's ability to pay the promised cash flows. It includes an analysis of any indentures.

Deep learning Algorithms based on deep neural networks, ones with many hidden layers (more than two), that address highly complex tasks, such as image classification, face recognition, speech recognition, and natural language processing.

Deep neural networks Neural networks with many hidden layers—at least 2 but potentially more than 20—that have proven successful across a wide range of artificial intelligence applications.

Default risk See *credit risk*.

Defined benefit pension plans Plan in which the company promises to pay a certain annual amount (defined benefit) to the employee after retirement. The company bears the investment risk of the plan assets.

Defined contribution pension plans Individual accounts to which an employee and typically the employer makes contributions, generally on a tax-advantaged basis. The amounts of contributions are defined at the outset, but the future value of the benefit is unknown. The employee bears the investment risk of the plan assets.

Definitive merger agreement A contract signed by both parties to a merger that clarifies the details of the transaction, including the terms, warranties, conditions, termination details, and the rights of all parties.

Delay costs Implicit trading costs that arise from the inability to complete desired trades immediately. Also called *slippage.*

Delta The relationship between the option price and the underlying price, which reflects the sensitivity of the price of the option to changes in the price of the underlying. Delta is a good approximation of how an option price will change for a small change in the stock.

Dendrogram A type of tree diagram used for visualizing a hierarchical cluster analysis; it highlights the hierarchical relationships among the clusters.

Dependent variable The variable whose variation about its mean is to be explained by the regression; the left-side variable in a regression equation. Also referred to as the *explained variable*.

Depository Trust and Clearinghouse Corporation A US-headquartered entity providing post-trade clearing, settlement, and information services.

Descriptive statistics The study of how data can be summarized effectively.

Diluted earnings per share (Diluted EPS)Net income, minus preferred dividends, divided by the weighted average number of common shares outstanding considering all dilutive securities (e.g., convertible debt and options); the EPS that would result if all dilutive securities were converted into common shares.

Dilution A reduction in proportional ownership interest as a result of the issuance of new shares.

Dimension reduction A set of techniques for reducing the number of features in a dataset while retaining variation across observations to preserve the information contained in that variation.

Diminishing marginal productivity When each additional unit of an input, keeping the other inputs unchanged, increases output by a smaller increment.

Direct capitalization method In the context of real estate, this method estimates the value of an income-producing property based on the level and quality of its net operating income.

Discount To reduce the value of a future payment in allowance for how far away it is in time; to calculate the present value of some future amount. Also, the amount by which an instrument is priced below its face value.

Discount factor The present value or price of a risk-free single-unit payment when discounted using the appropriate spot rate.

Discount for lack of control An amount or percentage deducted from the pro rata share of 100% of the value of an equity interest in a business to reflect the absence of some or all of the powers of control.

Discount for lack of marketability An amount of percentage deducted from the value of an ownership interest to reflect the relative absence of marketability.

Discount function Discount factors for the range of all possible maturities. The spot curve can be derived from the discount function and vice versa.

Discount rate Any rate used in finding the present value of a future cash flow.

Discounted abnormal earnings model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Discounted cash flow (DCF) analysis In the context of merger analysis, an estimate of a target company's value found by discounting the company's expected future free cash flows to the present.

Discounted cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows. In the context of real estate, this method estimates the value of an income-producing property based on discounting future projected cash flows.

Discounted cash flow model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Discriminant analysis A multivariate classification technique used to discriminate between groups, such as companies that either will or will not become bankrupt during some time frame.

Dispersed ownership Ownership structure consisting of many shareholders, none of which has the ability to individually exercise control over the corporation.

Divestiture The sale, liquidation, or spin-off of a division or subsidiary.

Dividend A distribution paid to shareholders based on the number of shares owned.

Dividend coverage ratio The ratio of net income to dividends.

Dividend discount model (DDM) A present value model of stock value that views the intrinsic value of a stock as present value of the stock's expected future dividends.

Dividend displacement of earnings The concept that dividends paid now displace earnings in all future periods.

Dividend imputation tax system A taxation system that effectively assures corporate profits distributed as dividends are taxed just once and at the shareholder's tax rate.

Dividend index point A measure of the quantity of dividends attributable to a particular index.

Dividend payout ratio The ratio of cash dividends paid to earnings for a period.

Dividend policy The strategy a company follows with regard to the amount and timing of dividend payments.

Dividend rate The annualized amount of the most recent dividend.

Dividend yield Annual dividends per share divided by share price.

Divisive clustering A top-down hierarchical clustering method that starts with all observations belonging to a single large cluster. The observations are then divided into two clusters based on some measure of distance (similarity). The algorithm then progressively partitions the intermediate clusters into smaller ones until each cluster contains only one observation.

Document frequency (DF) The number of documents (texts) that contain a particular token divided by the total number of documents. It is the simplest feature selection method and often performs well when many thousands of tokens are present.

Document term matrix (DTM) A matrix where each row belongs to a document (or text file), and each column represents a token (or term). The number of rows is equal to the number of documents (or text files) in a sample text dataset. The number of columns is equal to the number of tokens from the BOW built using all the documents in the sample dataset. The cells typically contain the counts of the number of times a token is present in each document.

Dominance An arbitrage opportunity when a financial asset with a risk-free payoff in the future must have a positive price today.

Double taxation system Corporate earnings are taxed twice when paid out as dividends. First, corporate pretax earnings are taxed regardless of whether they will be distributed as dividends or retained at the corporate level. Second, dividends are taxed again at the individual shareholder level.

Downstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary) such that the investor company records a profit on its income statement. An example is a sale of inventory by the investor company to the associate or by a parent to a subsidiary company.

Dual-class shares Shares that grant one share class superior or even sole voting rights, whereas the other share class has inferior or no voting rights.

Due diligence Investigation and analysis in support of a recommendation; the failure to exercise due diligence may sometimes result in liability according to various securities laws.

Dummy variable A type of qualitative variable that takes on a value of 1 if a particular condition is true and 0 if that condition is false.

Duration A measure of the approximate sensitivity of a security to a change in interest rates (i.e., a measure of interest rate risk).

Dutch disease A situation in which currency appreciation driven by strong export demand for resources makes other segments of the economy (particularly manufacturing) globally uncompetitive.

Earnings surprise The difference between reported EPS and expected EPS. Also referred to as *unexpected earnings.*

Earnings yield EPS divided by price; the reciprocal of the P/E.

Economic profit See *residual income.*

Economic sectors Large industry groupings.

Economic value added (EVA®) A commercial implementation of the residual income concept; the computation of EVA® is the net operating profit after taxes minus the cost of capital, where these inputs are adjusted for a number of items.

Economies of scale A situation in which average costs per unit of good or service produced fall as volume rises. In reference to mergers, the savings achieved through the consolidation of operations and elimination of duplicate resources.

Edwards–Bell–Ohlson model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Effective convexity Sensitivity of duration to changes in interest rates.

Effective duration Sensitivity of the bond's price to a 100 bps parallel shift of the benchmark yield curve, assuming no change in the bond's credit spread.

Effective spread Two times the difference between the execution price and the midpoint of the market quote at the time an order is entered.

Eigenvalue A measure that gives the proportion of total variance in the initial dataset that is explained by each eigenvector.

Eigenvector A vector that defines new mutually uncorrelated composite variables that are linear combinations of the original features.

Embedded options Contingency provisions found in a bond's indenture or offering circular representing rights that enable their holders to take advantage of interest rate movements. They can be exercised by the issuer, by the bondholder, or automatically depending on the course of interest rates.

Ensemble learning A technique of combining the predictions from a collection of models to achieve a more accurate prediction.

Ensemble method The method of combining multiple learning algorithms, as in ensemble learning.

Enterprise value Total company value (the market value of debt, common equity, and preferred equity) minus the value of cash and investments.

Enterprise value multiple A valuation multiple that relates the total market value of all sources of a company's capital (net of cash) to a measure of fundamental value for the entire company (such as a pre-interest earnings measure).

Equilibrium The condition in which supply equals demand.

Equity carve-out A form of restructuring that involves the creation of a new legal entity and the sale of equity in it to outsiders.

Equity charge The estimated cost of equity capital in money terms.

Equity REITs REITs that own, operate, and/or selectively develop income-producing real estate.

Equity swap A swap transaction in which at least one cash flow is tied to the return on an equity portfolio position, often an equity index.

Error autocorrelations The autocorrelations of the error term.

Error term The difference between an observation and its expected value, where the expected value is based on the true underlying population relation between the dependent and independent variables. Also known simply as the *error*.

ESG integration An ESG investment approach that focuses on systematic consideration of material ESG factors in asset allocation, security selection, and portfolio construction decisions for the purpose of achieving the product's stated investment objectives.

Estimated parameters With reference to a regression analysis, the estimated values of the population intercept and population slope coefficients in a regression.

***Ex ante* tracking error** A measure of the degree to which the performance of a given investment portfolio might be expected to deviate from its benchmark; also known as *relative VaR*.

***Ex ante* version of PPP** The hypothesis that expected changes in the spot exchange rate are equal to expected differences in national inflation rates. An extension of relative purchasing power parity to expected future changes in the exchange rate.

Ex-dividend Trading ex-dividend refers to shares that no longer carry the right to the next dividend payment.

Ex-dividend date The first date that a share trades without (i.e., "ex") the right to receive the declared dividend for the period.

Excess earnings method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Exchange ratio The number of shares that target stockholders are to receive in exchange for each of their shares in the target company.

Exercise date The date when employees actually exercise stock options and convert them to stock.

Exercise value The value of an option if it were exercised. Also sometimes called *intrinsic value*.

Expanded CAPM An adaptation of the CAPM that adds to the CAPM a premium for small size and company-specific risk.

Expectations approach A procedure for obtaining the value of an option derived from discounting at the risk-free rate its expected future payoff based on risk neutral probabilities.

Expected exposure The projected amount of money an investor could lose if an event of default occurs, before factoring in possible recovery.

Expected holding-period return The expected total return on an asset over a stated holding period; for stocks, the sum of the expected dividend yield and the expected price appreciation over the holding period.

Expected shortfall See *conditional VaR*.

Expected tail loss See *conditional VaR*.

Exploratory data analysis (EDA) The preliminary step in data exploration, where graphs, charts, and other visualizations (heat maps and word clouds) as well as quantitative methods (descriptive statistics and central tendency measures) are used to observe and summarize data.

Exposure to foreign exchange risk The risk of a change in value of an asset or liability denominated in a foreign currency due to a change in exchange rates.

Extendible bond Bond with an embedded option that gives the bondholder the right to keep the bond for a number of years after maturity, possibly with a different coupon.

External growth Company growth in output or sales that is achieved by buying the necessary resources externally (i.e., achieved through mergers and acquisitions).

Extra dividend *See special dividend*.

F1 score The harmonic mean of precision and recall. F1 score is a more appropriate overall performance metric (than accuracy) when there is unequal class distribution in the dataset and it is necessary to measure the equilibrium of precision and recall.

Factor A common or underlying element with which several variables are correlated.

Factor betas An asset's sensitivity to a particular factor; a measure of the response of return to each unit of increase in a factor, holding all other factors constant.

Factor portfolio See *pure factor portfolio*.

Factor price The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors.

Factor risk premium The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors. Also called *factor price*.

Factor sensitivity See *factor betas*.

Failure to pay When a borrower does not make a scheduled payment of principal or interest on any outstanding obligations after a grace period.

Fair market value The market price of an asset or liability that trades regularly.

Fair value The amount at which an asset (or liability) could be bought (or incurred) or sold (or settled) in a current transaction between willing parties, that is, other than in a forced or liquidation sale. As defined in IFRS and US GAAP, it is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date.

Feature engineering A process of creating new features by changing or transforming existing features.

Feature selection A process whereby only pertinent features from the dataset are selected for model training. Selecting fewer features decreases model complexity and training time.

Features The independent variables (X's) in a labeled dataset.

Financial contagion A situation in which financial shocks spread from their place of origin to other locales. In essence, a faltering economy infects other, healthier economies.

Financial distress Heightened uncertainty regarding a company's ability to meet its various obligations because of lower or negative earnings.

Financial transaction A purchase involving a buyer having essentially no material synergies with the target (e.g., the purchase of a private company by a company in an unrelated industry or by a private equity firm would typically be a financial transaction).

First-differencing A transformation that subtracts the value of the time series in period $t - 1$ from its value in period t.

First-order serial correlation Correlation between adjacent observations in a time series.

Fitting curve A curve which shows in- and out-of-sample error rates (E_{in} and E_{out}) on the y-axis plotted against model complexity on the x-axis.

Fixed price tender offer Offer made by a company to repurchase a specific number of shares at a fixed price that is typically at a premium to the current market price.

Fixed-rate perpetual preferred stock Non-convertible, non-callable preferred stock with a specified dividend rate that has a claim on earnings senior to the claim of common stock, and no maturity date.

Flight to quality During times of market stress, investors sell higher-risk asset classes such as stocks and commodities in favor of default-risk-free government bonds.

Flip-in pill A poison pill takeover defense that dilutes an acquirer's ownership in a target by giving other existing target company shareholders the right to buy additional target company shares at a discount.

Flip-over pill A poison pill takeover defense that gives target company shareholders the right to purchase shares of the acquirer at a significant discount to the market price, which has the effect of causing dilution to all existing acquiring company shareholders.

Float Amounts collected as premium and not yet paid out as benefits.

Floored floater Floating-rate bond with a floor provision that prevents the coupon rate from decreasing below a specified minimum rate. It protects the investor against declining interest rates.

Flotation cost Fees charged to companies by investment bankers and other costs associated with raising new capital.

Forced conversion For a convertible bond, when the issuer calls the bond and forces bondholders to convert their bonds into shares, which typically happens when the underlying share price increases above the conversion price.

Foreign currency transactions Transactions that are denominated in a currency other than a company's functional currency.

Forward curve The term structure of forward rates for loans made on a specific initiation date.

Forward dividend yield A dividend yield based on the anticipated dividend during the next 12 months.

Forward integration A merger involving the purchase of a target that is farther along the value or production chain; for example, to acquire a distributor.

Forward P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Forward price The fixed price or rate at which the transaction, scheduled to occur at the expiration of a forward contract, will take place. This price is agreed to at the initiation date of the forward contract.

Forward pricing model The model that describes the valuation of forward contracts.

Forward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving forward through the network's layers.

Forward rate An interest rate determined today for a loan that will be initiated in a future period.

Forward rate agreement An over-the-counter forward contract in which the underlying is an interest rate on a deposit. A forward rate agreement (FRA) calls for one party to make a fixed interest payment and the other to make an interest payment at a rate to be determined at contract expiration.

Forward rate model The forward pricing model expressed in terms of spot and forward interest rates.

Forward rate parity The proposition that the forward exchange rate is an unbiased predictor of the future spot exchange rate.

Forward value The monetary value of an existing forward contract.

Franking credit A tax credit received by shareholders for the taxes that a corporation paid on its distributed earnings.

Free cash flow The actual cash that would be available to the company's investors after making all investments necessary to maintain the company as an ongoing enterprise (also referred to as free cash flow to the firm); the internally generated funds that can be distributed to the company's investors (e.g., shareholders and bondholders) without impairing the value of the company.

Free cash flow hypothesis The hypothesis that higher debt levels discipline managers by forcing them to make fixed debt service payments and by reducing the company's free cash flow.

Free cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows.

Free cash flow to equity The cash flow available to a company's common shareholders after all operating expenses, interest, and principal payments have been made and necessary investments in working and fixed capital have been made.

Free cash flow to equity model A model of stock valuation that views a stock's intrinsic value as the present value of expected future free cash flows to equity.

Free cash flow to the firm The cash flow available to the company's suppliers of capital after all operating expenses (including taxes) have been paid and necessary investments in working and fixed capital have been made.

Free cash flow to the firm model A model of stock valuation that views the value of a firm as the present value of expected future free cash flows to the firm.

Frequency analysis The process of quantifying how important tokens are in a sentence and in the corpus as a whole. It helps in filtering unnecessary tokens (or features).

Friendly transaction A potential business combination that is endorsed by the managers of both companies.

Functional currency The currency of the primary economic environment in which an entity operates.

Fundamental factor models A multifactor model in which the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices.

Fundamentals Economic characteristics of a business, such as profitability, financial strength, and risk.

Funds available for distribution (FAD) See *adjusted funds from operations*.

Funds from operations (FFO) Net income (computed in accordance with generally accepted accounting principles) *plus* (1) gains and losses from sales of properties and (2) depreciation and amortization.

Futures price The price at which the parties to a futures contract agree to exchange the underlying (or cash). In commodity markets, the price agreed on to deliver or receive a defined quantity (and often quality) of a commodity at a future date.

Futures value The monetary value of an existing futures contract.

FX carry trade An investment strategy that involves taking long positions in high-yield currencies and short positions in low-yield currencies.

Gamma A measure of how sensitive an option's delta is to a change in the underlying. The change in a given instrument's delta for a given small change in the underlying's value, holding everything else constant.

Generalize When a model retains its explanatory power when predicting out-of-sample (i.e., using new data).

Generalized least squares A regression estimation technique that addresses heteroskedasticity of the error term.

Going-concern assumption The assumption that the business will maintain its business activities into the foreseeable future.

Going-concern value A business's value under a going-concern assumption.

Goodwill An intangible asset that represents the excess of the purchase price of an acquired company over the value of the net identifiable assets acquired.

Grant date The day that stock options are granted to employees.

Green bond Bonds in which the proceeds are designated by issuers to fund a specific project or portfolio of projects that have environmental or climate benefits.

Greenmail The purchase of the accumulated shares of a hostile investor by a company that is targeted for takeover by that investor, usually at a substantial premium over market price.

Greenwashing The risk that a green bond's proceeds are not actually used for a beneficial environmental or climate-related project.

Grid search A method of systematically training a model by using various combinations of hyperparameter values, cross validating each model, and determining which combination of hyperparameter values ensures the best model performance.

Gross domestic product A money measure of the goods and services produced within a country's borders over a stated period.

Gross lease A lease under which the tenant pays a gross rent to the landlord, who is responsible for all operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Ground truth The known outcome (i.e., target variable) of each observation in a labelled dataset.

Growth accounting equation The production function written in the form of growth rates. For the basic Cobb–Douglas production function, it states that the growth rate of output equals the rate of technological change plus α multiplied by the growth rate of capital plus $(1 - \alpha)$ multiplied by the growth rate of labor.

Growth capital expenditures Capital expenditures needed for expansion.

Growth option The ability to make additional investments in a project at some future time if the financial results are strong. Also called *expansion option*.

Guideline assets Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline companies Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline public companies Public-company comparables for the company being valued.

Guideline public company method A variation of the market approach; establishes a value estimate based on the observed multiples from trading activity in the shares of public companies viewed as reasonably comparable to the subject private company.

Guideline transactions method A variation of the market approach; establishes a value estimate based on pricing multiples derived from the acquisition of control of entire public or private companies that were acquired.

Harmonic mean A type of weighted mean computed by averaging the reciprocals of the observations and then taking the reciprocal of that average.

Hazard rate The probability that an event will occur, given that it has not already occurred.

Hedonic index Unlike a repeat-sales index, a hedonic index does not require repeat sales of the same property. It requires only one sale. The way it controls for the fact that different properties are selling each quarter is to include variables in the regression that control for differences in the characteristics of the property, such as size, age, quality of construction, and location.

Heteroskedastic With reference to the error term of regression, having a variance that differs across observations.

Heteroskedasticity The property of having a nonconstant variance; refers to an error term with the property that its variance differs across observations.

Heteroskedasticity-consistent standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Hierarchical clustering An iterative unsupervised learning procedure used for building a hierarchy of clusters.

Highest and best use The concept that the best use of a vacant site is the use that would result in the highest value for the land. Presumably, the developer that could earn the highest risk-adjusted profit based on time, effort, construction and development cost, leasing, and exit value would be the one to pay the highest price for the land.

Historical exchange rates For accounting purposes, the exchange rates that existed when the assets and liabilities were initially recorded.

Historical scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Historical simulation A simulation method that uses past return data and a random number generator that picks observations from the historical series to simulate an asset's future returns.

Historical simulation method The application of historical price changes to the current portfolio.

Historical stress testing The process that tests how investment strategies would perform under some of the most negative (i.e., adverse) combinations of events and scenarios.

Ho–Lee model The first arbitrage-free term structure model. The model is calibrated to market data and uses a binomial lattice approach to generate a distribution of possible future interest rates.

Holding period return The return that an investor earns during a specified holding period; a synonym for total return.

Holdout samples Data samples that are not used to train a model.

Homoskedasticity The property of having a constant variance; refers to an error term that is constant across observations.

Horizontal merger A merger involving companies in the same line of business, usually as competitors.

Horizontal ownership Companies with mutual business interests (e.g., key customers or suppliers) that have cross-holding share arrangements with each other.

Hostile transaction An attempt to acquire a company against the wishes of the target's managers.

Human capital The accumulated knowledge and skill that workers acquire from education, training, or life experience.

Hybrid approach With respect to forecasting, an approach that combines elements of both top-down and bottom-up analyses.

Hyperparameter A parameter whose value must be set by the researcher before learning begins.

I-spreads Shortened form of "interpolated spreads" and a reference to a linearly interpolated yield.

Illiquidity discount A reduction or discount to value that reflects the lack of depth of trading or liquidity in that asset's market.

Impairment Diminishment in value as a result of carrying (book) value exceeding fair value and/or recoverable value.

Impairment of capital rule A legal restriction that dividends cannot exceed retained earnings.

Implementation shortfall The difference between the money return (or value) on a notional or paper portfolio and the actual portfolio return (or value).

Implied volatility The standard deviation that causes an option pricing model to give the current option price.

In-sample forecast errors The residuals from a fitted time-series model within the sample period used to fit the model.

iNAVs "Indicated" net asset values are intraday "fair value" estimates of an ETF share based on its creation basket.

Income approach A valuation approach that values an asset as the present discounted value of the income expected from it. In the context of real estate, this approach estimates the value of a property based on an expected rate of return. The estimated value is the present value of the expected future income from the property, including proceeds from resale at the end of a typical investment holding period.

Incremental VaR (IVaR) A measure of the incremental effect of an asset on the VaR of a portfolio by measuring the difference between the portfolio's VaR while including a specified asset and the portfolio's VaR with that asset eliminated.

Indenture A written contract between a lender and borrower that specifies the terms of the loan, such as interest rate, interest payment schedule, or maturity.

Independent board directors Directors with no material relationship with the company with regard to employment, ownership, or remuneration.

Independent regulators Regulators recognized and granted authority by a government body or agency. They are not government agencies per se and typically do not rely on government funding.

Independent variable A variable used to explain the dependent variable in a regression; a right-side variable in a regression equation. Also referred to as the *explanatory variable*.

Index CDS A type of credit default swap that involves a combination of borrowers.

Indicator variable A variable that takes on only one of two values, 0 or 1, based on a condition. In simple linear regression, the slope is the difference in the dependent variable for the two conditions. Also referred to as a *dummy variable*.

Industry structure An industry's underlying economic and technical characteristics.

Information gain A metric which quantifies the amount of information that the feature holds about the response. Information gain can be regarded as a form of non-linear correlation between Y and X.

Information ratio (IR) Mean active return divided by active risk; or alpha divided by the standard deviation of diversifiable risk.

Informational frictions Forces that restrict availability, quality, and/or flow of information and its use.

Inside ask See *best ask*.

Inside bid See *best bid*.

Inside spread The spread between the best bid price and the best ask price. Also called the *market bid-ask spread*, *inside bid-ask spread*, or *market spread*.

Insiders Corporate managers and board directors who are also shareholders of a company.

Inter-temporal rate of substitution The ratio of the marginal utility of consumption s periods in the future (the numerator) to the marginal utility of consumption today (the denominator).

Intercept The expected value of the dependent variable when the independent variable in a simple linear regression is equal to zero.

Interest rate risk The risk that interest rates will rise and therefore the market value of current portfolio holdings will fall so that their current yields to maturity then match comparable instruments in the marketplace.

Interlocking directorates Corporate structure in which individuals serve on the board of directors of multiple corporations.

Internal rate of return Abbreviated as IRR. Rate of return that discounts future cash flows from an investment to the exact amount of the investment; the discount rate that makes the present value of an investment's costs (outflows) equal to the present value of the investment's benefits (inflows).

International Fisher effect The proposition that nominal interest rate differentials across currencies are determined by expected inflation differentials.

Intrinsic value The value of an asset given a hypothetically complete understanding of the asset's investment characteristics; the value obtained if an option is exercised based on current conditions. The difference between the spot exchange rate and the strike price of a currency.

Inverse price ratio The reciprocal of a price multiple—for example, in the case of a P/E, the "earnings yield" E/P (where P is share price and E is earnings per share).

Investment value The value to a specific buyer, taking account of potential synergies based on the investor's requirements and expectations.

ISDA Master Agreement A standard or "master" agreement published by the International Swaps and Derivatives Association. The master agreement establishes the terms for each party involved in the transaction.

Judicial law Interpretations of courts.

Justified (fundamental) P/E The price-to-earnings ratio that is fair, warranted, or justified on the basis of forecasted fundamentals.

Justified price multiple The estimated fair value of the price multiple, usually based on forecasted fundamentals or comparables.

***K*-fold cross-validation** A technique in which data (excluding test sample and fresh data) are shuffled randomly and then are divided into k equal sub-samples, with $k - 1$ samples used as training samples and one sample, the kth, used as a validation sample.

***K*-means** A clustering algorithm that repeatedly partitions observations into a fixed number, k, of non-overlapping clusters.

***K*-nearest neighbor** A supervised learning technique that classifies a new observation by finding similarities ("nearness") between this new observation and the existing data.

Kalotay–Williams–Fabozzi (KWF) model An arbitrage-free term structure model that describes the dynamics of the log of the short rate and assumes constant drift, no mean reversion, and constant volatility.

Key rate durations Sensitivity of a bond's price to changes in specific maturities on the benchmark yield curve. Also called *partial durations*.

***k*th-order autocorrelation** The correlation between observations in a time series separated by k periods.

Labeled dataset A dataset that contains matched sets of observed inputs or features (X's) and the associated output or target (Y).

Labor force Everyone of working age (ages 16 to 64) who either is employed or is available for work but not working.

Labor force participation rate The percentage of the working age population that is in the labor force.

Labor productivity The quantity of real GDP produced by an hour of labor. More generally, output per unit of labor input.

Labor productivity growth accounting equation States that potential GDP growth equals the growth rate of the labor input plus the growth rate of labor productivity.

Lack of marketability discount An extra return to investors to compensate for lack of a public market or lack of marketability.

LASSO Least absolute shrinkage and selection operator is a type of penalized regression which involves minimizing the sum of the absolute values of the regression coefficients. LASSO can also be used for regularization in neural networks.

Latency The elapsed time between the occurrence of an event and a subsequent action that depends on that event.

Law of one price A principle that states that if two investments have the same or equivalent future cash flows regardless of what will happen in the future, then these two investments should have the same current price.

Leading dividend yield Forecasted dividends per share over the next year divided by current stock price.

Leading P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Learning curve A curve that plots the accuracy rate (= 1 – error rate) in the validation or test samples (i.e., out-of-sample) against the amount of data in the training sample, which is thus useful for describing under- and overfitting as a function of bias and variance errors.

Learning rate A parameter that affects the magnitude of adjustments in the weights in a neural network.

Level One of the three factors (the other two are steepness and curvature) that empirically explain most yield curve shape changes. A shock to the level factor changes the yield for all maturities by an almost identical amount.

Leveraged buyout A transaction whereby the target company management team converts the target to a privately held company by using heavy borrowing to finance the purchase of the target company's outstanding shares.

Leveraged recapitalization A post-offer takeover defense mechanism that involves the assumption of a large amount of debt that is then used to finance share repurchases. The effect is to dramatically change the company's capital structure while attempting to deliver a value to target shareholders in excess of a hostile bid.

Libor–OIS spread The difference between Libor and the overnight indexed swap rate.

Limit order book The book or list of limit orders to buy and sell that pertains to a security.

Lin-log model A regression model in which the independent variable is in logarithmic form.

Linear classifier A binary classifier that makes its classification decision based on a linear combination of the features of each data point.

Linear regression Regression that models the straight-line relationship between the dependent and independent variables. Also known as *least squares regression* and *ordinary least squares regression*.

Linear trend A trend in which the dependent variable changes at a constant rate with time.

Liquidating dividend A dividend that is a return of capital rather than a distribution from earnings or retained earnings.

Liquidation To sell the assets of a company, division, or subsidiary piecemeal, typically because of bankruptcy; the form of bankruptcy that allows for the orderly satisfaction of creditors' claims after which the company ceases to exist.

Liquidation value The value of a company if the company were dissolved and its assets sold individually.

Liquidity preference theory A term structure theory that asserts liquidity premiums exist to compensate investors for the added interest rate risk they face when lending long term.

Liquidity premium The premium or incrementally higher yield that investors demand for lending long term.

Local currency The currency of the country where a company is located.

Local expectations theory A term structure theory that contends the return for all bonds over short periods is the risk-free rate.

Log-lin model A regression model in which the dependent variable is in logarithmic form.

Log-linear model With reference to time-series models, a model in which the growth rate of the time series as a function of time is constant.

Log-log model A regression model in which both the dependent and independent variables are in logarithmic form. Also known as the *double-log model*.

Log-log regression model A regression that expresses the dependent and independent variables as natural logarithms.

Logistic regression (logit model) A qualitative-dependent-variable multiple regression model based on the logistic probability distribution.

Long/short credit trade A credit protection seller with respect to one entity combined with a credit protection buyer with respect to another entity.

Look-ahead bias The bias created by using information that was unknown or unavailable in the time periods over which backtesting is conducted, such as company earnings and macroeconomic indicator values.

Lookback period The time period used to gather a historical data set.

Loss given default The amount that will be lost if a default occurs.

Macroeconomic factor model A multifactor model in which the factors are surprises in macroeconomic variables that significantly explain equity returns.

Macroeconomic factors Factors related to the economy, such as the inflation rate, industrial production, or economic sector membership.

Maintenance capital expenditures Capital expenditures needed to maintain operations at the current level.

Majority shareholders Shareholders that own more than 50% of a corporation's shares.

Majority-vote classifier A classifier that assigns to a new data point the predicted label with the most votes (i.e., occurrences).

Managerialism theories Theories that posit that corporate executives are motivated to engage in mergers to maximize the size of their company rather than shareholder value (a form of agency cost).

Marginal VaR (MVaR) A measure of the effect of a small change in a position size on portfolio VaR.

Market approach Valuation approach that values an asset based on pricing multiples from sales of assets viewed as similar to the subject asset.

Market conversion premium per share For a convertible bond, the difference between the market conversion price and the underlying share price, which allows investors to identify the premium or discount payable when buying a convertible bond rather than the underlying common stock.

Market conversion premium ratio For a convertible bond, the market conversion premium per share expressed as a percentage of the current market price of the shares.

Market efficiency A finance perspective on capital markets that deals with the relationship of price to intrinsic value. The **traditional efficient markets formulation** asserts that an asset's price is the best available estimate of its intrinsic value. The **rational efficient markets formulation** asserts that investors should expect to be rewarded for the costs of information gathering and analysis by higher gross returns.

Market fragmentation Trading the same instrument in multiple venues.

Market impact The effect of the trade on transaction prices. Also called *price impact.*

Market timing Asset allocation in which the investment in the market is increased if one forecasts that the market will outperform T-bills.

Market value of invested capital The market value of debt and equity.

Mature growth rate The earnings growth rate in a company's mature phase; an earnings growth rate that can be sustained long term.

Maximum drawdown The worst cumulative loss ever sustained by an asset or portfolio. More specifically, maximum drawdown is the difference between an asset's or a portfolio's maximum cumulative return and its subsequent lowest cumulative return.

Mean reversion The tendency of a time series to fall when its level is above its mean and rise when its level is below its mean; a mean-reverting time series tends to return to its long-term mean.

Mean square error (MSE) The sum of squares error divided by the degrees of freedom, $n - k - 1$; in a simple linear regression, $n - k - 1 = n - 2$.

Mean square regression (MSR) The sum of squares regression divided by the number of independent variables k; in a simple linear regression, $k = 1$.

Merger The absorption of one company by another; two companies become one entity and one or both of the pre-merger companies ceases to exist as a separate entity.

Metadata Data that describes and gives information about other data.

Method based on forecasted fundamentals An approach to using price multiples that relates a price multiple to forecasts of fundamentals through a discounted cash flow model.

Method of comparables An approach to valuation that involves using a price multiple to evaluate whether an asset is relatively fairly valued, relatively undervalued, or relatively overvalued when compared to a benchmark value of the multiple.

Midquote price The average, or midpoint, of the prevailing bid and ask prices.

Minority interest The proportion of the ownership of a subsidiary not held by the parent (controlling) company.

Minority shareholders Shareholders that own less than 50% of a corporation's shares.

Mispricing Any departure of the market price of an asset from the asset's estimated intrinsic value.

Mixed offering A merger or acquisition that is to be paid for with cash, securities, or some combination of the two.

Model specification With reference to regression, the set of variables included in the regression and the regression equation's functional form.

Molodovsky effect The observation that P/Es tend to be high on depressed EPS at the bottom of a business cycle and tend to be low on unusually high EPS at the top of a business cycle.

Momentum indicators Valuation indicators that relate either price or a fundamental (such as earnings) to the time series of their own past values (or in some cases to their expected value).

Monetary assets and liabilities Assets and liabilities with value equal to the amount of currency contracted for, a fixed amount of currency. Examples are cash, accounts receivable, accounts payable, bonds payable, and mortgages payable. Inventory is not a monetary asset. Most liabilities are monetary.

Monetary/non-monetary method Approach to translating foreign currency financial statements for consolidation in which monetary assets and liabilities are translated at the current exchange rate. Non-monetary assets and liabilities are translated at historical exchange rates (the exchange rates that existed when the assets and liabilities were acquired).

Monetizing Unwinding a position to either capture a gain or realize a loss.

Monitoring costs Costs borne by owners to monitor the management of the company (e.g., board of director expenses).

Monte Carlo simulation A technique that uses the inverse transformation method for converting a randomly generated uniformly distributed number into a simulated value of a random variable of a desired distribution. Each key decision variable in a Monte Carlo simulation requires an assumed statistical distribution; this assumption facilitates incorporating non-normality, fat tails, and tail dependence as well as solving high-dimensionality problems.

Mortgages Loans with real estate serving as collateral for the loans.

Multicollinearity A regression assumption violation that occurs when two or more independent variables (or combinations of independent variables) are highly but not perfectly correlated with each other.

Multiple linear regression Linear regression involving two or more independent variables.

Multiple linear regression model A linear regression model with two or more independent variables.

Mutual information Measures how much information is contributed by a token to a class of texts. MI will be 0 if the token's distribution in all text classes is the same. MI approaches 1 as the token in any one class tends to occur more often in only that particular class of text.

Mutually exclusive projects Mutually exclusive projects compete directly with each other. For example, if Projects A and B are mutually exclusive, you can choose A or B, but you cannot choose both.

N-grams A representation of word sequences. The length of a sequence varies from 1 to *n*. When one word is used, it is a unigram; a two-word sequence is a bigram; and a 3-word sequence is a trigram; and so on.

***n*-Period moving average** The average of the current and immediately prior $n - 1$ values of a time series.

Naked credit default swap A position where the owner of the CDS does not have a position in the underlying credit.

Name entity recognition An algorithm that analyzes individual tokens and their surrounding semantics while referring to its dictionary to tag an object class to the token.

Negative serial correlation Serial correlation in which a positive error for one observation increases the chance of a negative error for another observation, and vice versa.

Net asset balance sheet exposure When assets translated at the current exchange rate are greater in amount than liabilities translated at the current exchange rate. Assets exposed to translation gains or losses exceed the exposed liabilities.

Net asset value The difference between assets and liabilities, all taken at current market values instead of accounting book values.

Net asset value per share Net asset value divided by the number of shares outstanding.

Net lease A lease under which the tenant pays a net rent to the landlord and an additional amount based on the tenant's pro rata share of the operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Net liability balance sheet exposure When liabilities translated at the current exchange rate are greater assets translated at the current exchange rate. Liabilities exposed to translation gains or losses exceed the exposed assets.

Net operating income Gross rental revenue minus operating costs but before deducting depreciation, corporate overhead, and interest expense. In the context of real estate, a measure of the income from the property after deducting operating expenses for such items as property taxes, insurance, maintenance, utilities, repairs, and insurance but before deducting any costs associated with financing and before deducting federal income taxes. It is similar to EBITDA in a financial reporting context.

Net regulatory burden The private costs of regulation less the private benefits of regulation.

Network externalities The impact that users of a good, a service, or a technology have on other users of that product; it can be positive (e.g., a critical mass of users makes a product more useful) or negative (e.g., congestion makes the product less useful).

Neural networks Highly flexible machine learning algorithms that have been successfully applied to a variety of supervised and unsupervised tasks characterized by nonlinearities and interactions among features.

No-arbitrage approach A procedure for obtaining the value of an option based on the creation of a portfolio that replicates the payoffs of the option and deriving the option value from the value of the replicating portfolio.

No-growth company A company without positive expected net present value projects.

No-growth value per share The value per share of a no-growth company, equal to the expected level amount of earnings divided by the stock's required rate of return.

Non-cash rent An amount equal to the difference between the average contractual rent over a lease term (the straight-line rent) and the cash rent actually paid during a period. This figure is one of the deductions made from FFO to calculate AFFO.

Non-convergence trap A situation in which a country remains relatively poor, or even falls further behind, because it fails to implement necessary institutional reforms and/or adopt leading technologies.

Non-monetary assets and liabilities Assets and liabilities that are not monetary assets and liabilities. Non-monetary assets include inventory, fixed assets, and intangibles, and non-monetary liabilities include deferred revenue.

Non-renewable resources Finite resources that are depleted once they are consumed; oil and coal are examples.

Non-residential properties Commercial real estate properties other than multi-family properties, farmland, and timberland.

Nonearning assets Cash and investments (specifically cash, cash equivalents, and short-term investments).

Nonstationarity With reference to a random variable, the property of having characteristics, such as mean and variance, that are not constant through time.

Normal EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normalized EPS*.

Normalized earnings The expected level of mid-cycle earnings for a company in the absence of any unusual or temporary factors that affect profitability (either positively or negatively).

Normalized EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normal EPS*.

Normalized P/E P/E based on normalized EPS data.

Notional amount The amount of protection being purchased in a CDS.

NTM P/E Next 12-month P/E: current market price divided by an estimated next 12-month EPS.

Off-the-run A series of securities or indexes that were issued/created prior to the most recently issued/created series.

On-the-run The most recently issued/created series of securities or indexes.

One hot encoding The process by which categorical variables are converted into binary form (0 or 1) for machine reading. It is one of the most common methods for handling categorical features in text data.

One-sided durations Effective durations when interest rates go up or down, which are better at capturing the interest rate sensitivity of bonds with embedded options that do not react symmetrically to positive and negative changes in interest rates of the same magnitude.

One-tier board Board structure consisting of a single board of directors, composed of executive (internal) and non-executive (external) directors.

Opportunity cost The value that investors forgo by choosing a particular course of action; the value of something in its best alternative use.

Optimal capital structure The capital structure at which the value of the company is maximized.

Option-adjusted spread (OAS) Constant spread that, when added to all the one-period forward rates on the interest rate tree, makes the arbitrage-free value of the bond equal to its market price.

Orderly liquidation value The estimated gross amount of money that could be realized from the liquidation sale of an asset or assets, given a reasonable amount of time to find a purchaser or purchasers.

Organic growth Company growth in output or sales that is achieved by making investments internally (i.e., excludes growth achieved through mergers and acquisitions).

Other comprehensive income Changes to equity that bypass (are not reported in) the income statement; the difference between comprehensive income and net income.

Other post-employment benefits Promises by the company to pay benefits in the future, such as life insurance premiums and all or part of health care insurance for its retirees.

Out-of-sample forecast errors The differences between actual and predicted values of time series outside the sample period used to fit the model.

Overfitting When a model fits the training data too well and so does not generalize well to new data.

Overnight indexed swap (OIS) rate An interest rate swap in which the periodic floating rate of the swap equals the geometric average of a daily unsecured overnight rate (or overnight index rate).

Pairs trading An approach to trading that uses pairs of closely related stocks, buying the relatively undervalued stock and selling short the relatively overvalued stock.

Par curve A hypothetical yield curve for coupon-paying Treasury securities that assumes all securities are priced at par.

Par swap A swap in which the fixed rate is set so that no money is exchanged at contract initiation.

Parametric method A method of estimating VaR that uses the historical mean, standard deviation, and correlation of security price movements to estimate the portfolio VaR. Generally assumes a normal distribution but can be adapted to non-normal distributions with the addition of skewness and kurtosis. Sometimes called the *variance–covariance method* or the *analytical method*.

Partial regression coefficients The slope coefficients in a multiple regression. Also called *partial slope coefficients*.

Partial slope coefficients The slope coefficients in a multiple regression. Also called *partial regression coefficients*.

Parts of speech An algorithm that uses language structure and dictionaries to tag every token in the text with a corresponding part of speech (i.e., noun, verb, adjective, proper noun, etc.).

Payout amount The loss given default times the notional.

Payout policy The principles by which a company distributes cash to common shareholders by means of cash dividends and/or share repurchases.

Payouts Cash dividends and the value of shares repurchased in any given year.

Pecking order theory The theory that managers consider how their actions might be interpreted by outsiders and thus order their preferences for various forms of corporate financing. Forms of financing that are least visible to outsiders (e.g., internally generated funds) are most preferable to managers and those that are most visible (e.g., equity) are least preferable.

PEG ratio The P/E-to-growth ratio, calculated as the stock's P/E divided by the expected earnings growth rate.

Penalized regression A regression that includes a constraint such that the regression coefficients are chosen to minimize the sum of squared residuals *plus* a penalty term that increases in size with the number of included features.

Pension obligation The present value of future benefits earned by employees for service provided to date.

Perfect capital markets Markets in which, by assumption, there are no taxes, transaction costs, or bankruptcy costs and in which all investors have equal ("symmetric") information.

Perpetuity A perpetual annuity, or a set of never-ending level sequential cash flows, with the first cash flow occurring one period from now.

Persistent earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *continuing earnings*, or *underlying earnings*.

Pet projects Projects in which influential managers want the corporation to invest. Often, unfortunately, pet projects are selected without undergoing normal capital budgeting analysis.

Physical settlement Involves actual delivery of the debt instrument in exchange for a payment by the credit protection seller of the notional amount of the contract.

Point-in-time data Data consisting of the exact information available to market participants as of a given point in time. Point-in-time data is used to address look-ahead bias.

Poison pill A pre-offer takeover defense mechanism that makes it prohibitively costly for an acquirer to take control of a target without the prior approval of the target's board of directors.

Poison puts A pre-offer takeover defense mechanism that gives target company bondholders the right to sell their bonds back to the target at a pre-specified redemption price, typically at or above par value; this defense increases the need for cash and raises the cost of the acquisition.

Portfolio balance approach A theory of exchange rate determination that emphasizes the portfolio investment decisions of global investors and the requirement that global investors willingly hold all outstanding securities denominated in each currency at prevailing prices and exchange rates.

Positive serial correlation Serial correlation in which a positive error for one observation increases the chance of a positive error for another observation; a negative error for one observation increases the chance of a negative error for another observation.

Potential GDP The maximum amount of output an economy can sustainably produce without inducing an increase in the inflation rate. The output level that corresponds to full employment with consistent wage and price expectations.

Precision In error analysis for classification problems it is ratio of correctly predicted positive classes to all predicted positive classes. Precision is useful in situations where the cost of false positives (FP), or Type I error, is high.

Preferred habitat theory A term structure theory that contends that investors have maturity preferences and require yield incentives before they will buy bonds outside of their preferred maturities.

Premise of value The status of a company in the sense of whether it is assumed to be a going concern or not.

Premium leg The series of payments the credit protection buyer promises to make to the credit protection seller.

Premiums Amounts paid by the purchaser of insurance products.

Present value model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Present value of growth opportunities The difference between the actual value per share and the no-growth value per share. Also called *value of growth*.

Presentation currency The currency in which financial statement amounts are presented.

Price improvement When trade execution prices are better than quoted prices.

Price momentum A valuation indicator based on past price movement.

Price multiples The ratio of a stock's market price to some measure of value per share.

Price-setting option The operational flexibility to adjust prices when demand varies from what is forecast. For example, when demand exceeds capacity, the company could benefit from the excess demand by increasing prices.

Price-to-earnings ratio (P/E) The ratio of share price to earnings per share.

Priced risk Risk for which investors demand compensation for bearing (e.g., equity risk, company-specific factors, macroeconomic factors).

Principal components analysis (PCA) An unsupervised ML technique used to transform highly correlated features of data into a few main, uncorrelated composite variables.

Principle of no arbitrage In well-functioning markets, prices will adjust until there are no arbitrage opportunities.

Prior transaction method A variation of the market approach; considers actual transactions in the stock of the subject private company.

Private market value The value derived using a sum-of-the-parts valuation.

Probability of default The probability that a bond issuer will not meet its contractual obligations on schedule.

Probability of survival The probability that a bond issuer will meet its contractual obligations on schedule.

Procedural law The body of law that focuses on the protection and enforcement of the substantive laws.

Production-flexibility option The operational flexibility to alter production when demand varies from forecast. For example, if demand is strong, a company may profit from employees working overtime or from adding additional shifts.

Project sequencing To defer the decision to invest in a future project until the outcome of some or all of a current project is known. Projects are sequenced through time, so that investing in a project creates the option to invest in future projects.

Projection error The vertical (perpendicular) distance between a data point and a given principal component.

Prospective P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Protection leg The contingent payment that the credit protection seller may have to make to the credit protection buyer.

Protection period Period during which a bond's issuer cannot call the bond.

Provision for loan losses An income statement expense account that increases the amount of the allowance for loan losses.

Proxy fight An attempt to take control of a company through a shareholder vote.

Proxy statement A public document that provides the material facts concerning matters on which shareholders will vote.

Prudential supervision Regulation and monitoring of the safety and soundness of financial institutions to promote financial stability, reduce system-wide risks, and protect customers of financial institutions.

Pruning A regularization technique used in CART to reduce the size of the classification or regression tree—by pruning, or removing, sections of the tree that provide little classifying power.

Purchasing power gain A gain in value caused by changes in price levels. Monetary liabilities experience purchasing power gains during periods of inflation.

Purchasing power loss A loss in value caused by changes in price levels. Monetary assets experience purchasing power loss during periods of inflation.

Purchasing power parity (PPP) The idea that exchange rates move to equalize the purchasing power of different currencies.

Pure expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *unbiased expectations theory*.

Pure factor portfolio A portfolio with sensitivity of 1 to the factor in question and a sensitivity of 0 to all other factors.

Putable bond Bond that includes an embedded put option, which gives the bondholder the right to put back the bonds to the issuer prior to maturity, typically when interest rates have risen and higher-yielding bonds are available.

Qualitative dependent variables Dummy variables used as dependent variables rather than as independent variables.

Quality of earnings analysis The investigation of issues relating to the accuracy of reported accounting results as reflections of economic performance. Quality of earnings analysis is broadly understood to include not only earnings management but also balance sheet management.

Random forest classifier A collection of a large number of decision trees trained via a bagging method.

Random walk A time series in which the value of the series in one period is the value of the series in the previous period plus an unpredictable random error.

Rational efficient markets formulation See *market efficiency.*

Readme files Text files provided with raw data that contain information related to a data file. They are useful for understanding the data and how they can be interpreted correctly.

Real estate investment trusts Tax-advantaged entities (companies or trusts) that own, operate, and—to a limited extent—develop income-producing real estate property.

Real estate operating companies Regular taxable real estate ownership companies that operate in the real estate industry in countries that do not have a tax-advantaged REIT regime in place or that are engage in real estate activities of a kind and to an extent that do not fit in their country's REIT framework.

Real interest rate parity The proposition that real interest rates will converge to the same level across different markets.

Real options Options that relate to investment decisions such as the option to time the start of a project, the option to adjust its scale, or the option to abandon a project that has begun.

Rebalance return A return from rebalancing the component weights of an index.

Recall Also known as *sensitivity*, in error analysis for classification problems it is the ratio of correctly predicted positive classes to all actual positive classes. Recall is useful in situations where the cost of false negatives (FN), or Type II error, is high.

Reconstitution When dealers recombine appropriate individual zero-coupon securities and reproduce an underlying coupon Treasury.

Recovery rate The percentage of the loss recovered.

Redemption basket The list of securities (and share amounts) the authorized participant (AP) receives when it redeems ETF shares back to the ETF manager. The redemption basket is published each business day.

Reference entity The borrower (debt issuer) covered by a single-name CDS.

Reference obligation A particular debt instrument issued by the borrower that is the designated instrument being covered.

Regime With reference to a time series, the underlying model generating the times series.

Regression analysis A tool for examining whether a variable is useful for explaining another variable.

Regression coefficients The intercept and slope coefficient(s) of a regression.

Regular expression (regex) A series of texts that contains characters in a particular order. Regex is used to search for patterns of interest in a given text.

Regularization A term that describes methods for reducing statistical variability in high-dimensional data estimation problems.

Regulatory arbitrage Entities identify and use some aspect of regulations that allows them to exploit differences in economic substance and regulatory interpretation or in foreign and domestic regulatory regimes to their (the entities') advantage.

Regulatory burden The costs of regulation for the regulated entity.

Regulatory capture Theory that regulation often arises to enhance the interests of the regulated.

Regulatory competition Regulators may compete to provide a regulatory environment designed to attract certain entities.

Reinforcement learning Machine learning in which a computer learns from interacting with itself or data generated by the same algorithm.

Relative-strength indicators Valuation indicators that compare a stock's performance during a period either to its own past performance or to the performance of some group of stocks.

Relative valuation models A model that specifies an asset's value relative to the value of another asset.

Relative VaR See *ex ante tracking error*.

Relative version of PPP The hypothesis that changes in (nominal) exchange rates over time are equal to national inflation rate differentials.

Renewable resources Resources that can be replenished, such as a forest.

Rental price of capital The cost per unit of time to rent a unit of capital.

Repeat sales index As the name implies, this type of index relies on repeat sales of the same property. In general, the idea supporting this type of index is that because it is the same property that sold twice, the change in value between the two sale dates indicates how market conditions have changed over time.

Replacement cost In the context of real estate, the value of a building assuming it was built today using current construction costs and standards.

Reporting unit For financial reporting under US GAAP, an operating segment or one level below an operating segment (referred to as a component).

Required rate of return The minimum rate of return required by an investor to invest in an asset, given the asset's riskiness.

Residential properties Properties that provide housing for individuals or families. Single-family properties may be owner-occupied or rental properties, whereas multi-family properties are rental properties even if the owner or manager occupies one of the units.

Residual The difference between an observation and its predicted value, where the predicted value is based on the estimated linear relation between the dependent and independent variables using sample data.

Residual autocorrelations The sample autocorrelations of the residuals.

Residual income Earnings for a given period, minus a deduction for common shareholders' opportunity cost in generating the earnings. Also called *economic profit* or *abnormal earnings*.

Residual income method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Residual income model (RIM) A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share. Also called *discounted abnormal earnings model* or *Edwards–Bell–Ohlson model*.

Residual loss Agency costs that are incurred despite adequate monitoring and bonding of management.

Restructuring Reorganizing the capital structure of a firm.

Return on capital employed Operating profit divided by capital employed (debt and equity capital).

Return on invested capital A measure of the after-tax profitability of the capital invested by the company's shareholders and debtholders.

Reverse carry arbitrage A strategy involving the short sale of the underlying and an offsetting opposite position in the derivative.

Reverse stock split A reduction in the number of shares outstanding with a corresponding increase in share price but no change to the company's underlying fundamentals.

Reverse stress testing A risk management approach in which the user identifies key risk exposures in the portfolio and subjects those exposures to extreme market movements.

Reviewed financial statements A type of non-audited financial statements; typically provide an opinion letter with representations and assurances by the reviewing accountant that are less than those in audited financial statements.

Rho The change in a given derivative instrument for a given small change in the risk-free interest rate, holding everything else constant. Rho measures the sensitivity of the option to the risk-free interest rate.

Risk budgeting The allocation of an asset owner's total risk appetite among groups or divisions (in the case of a trading organization) or among strategies and managers (in the case of an institutional or individual investor).

Risk decomposition The process of converting a set of holdings in a portfolio into a set of exposures to risk factors.

Risk factors Variables or characteristics with which individual asset returns are correlated. Sometimes referred to simply as *factors*.

Risk parity A portfolio allocation scheme that weights stocks or factors based on an equal risk contribution.

Robust standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Roll When an investor moves its investment position from an older series to the most current series.

Roll return The component of the return on a commodity futures contract attributable to rolling long futures positions forward through time. Also called *roll yield*.

Rolling down the yield curve A maturity trading strategy that involves buying bonds with a maturity longer than the intended investment horizon. Also called *riding the yield curve*.

Rolling windows A backtesting method that uses a rolling-window (or walk-forward) framework, rebalances the portfolio after each period, and then tracks performance over time. As new information arrives each period, the investment manager optimizes (revises and tunes) the model and readjusts stock positions.

Root mean squared error (RMSE) The square root of the average squared forecast error; used to compare the out-of-sample forecasting performance of forecasting models.

Sale-leaseback A situation in which a company sells the building it owns and occupies to a real estate investor and the company then signs a long-term lease with the buyer to continue to occupy the building. At the end of the lease, use of the property reverts to the landlord.

Sales comparison approach In the context of real estate, this approach estimates value based on what similar or comparable properties (comparables) transacted for in the current market.

Scaled earnings surprise Unexpected earnings divided by the standard deviation of analysts' earnings forecasts.

Scaling The process of adjusting the range of a feature by shifting and changing the scale of the data. Two of the most common ways of scaling are normalization and standardization.

Scatter plot A chart in which two variables are plotted along the axis and points on the chart represent pairs of the two variables. In regression, the dependent variable is plotted on the vertical axis and the independent variable is plotted along the horizontal axis. Also known as a scattergram and a *scatter diagram*.

Scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Scree plots A plot that shows the proportion of total variance in the data explained by each principal component.

Screening The application of a set of criteria to reduce a set of potential investments to a smaller set having certain desired characteristics.

Seasonality A characteristic of a time series in which the data experience regular and predictable periodic changes; for example, fan sales are highest during the summer months.

Secured overnight financing rate (SOFR) A daily volume-weighted index of rates on qualified cash borrowings collateralized by US Treasuries that is expected to replace Libor as a floating reference rate for swaps.

Securities offering A merger or acquisition in which target shareholders are to receive shares of the acquirer's common stock as compensation.

Security selection risk See *active specific risk*.

Segmented markets theory A term structure theory that contends yields are solely a function of the supply and demand for funds of a particular maturity.

Self-regulating organizations (SROs) Self-regulating bodies that are given recognition and authority, including enforcement power, by a government body or agency.

Self-regulatory bodies Private, non-governmental organizations that both represent and regulate their members. Some self-regulating organizations are also independent regulators.

Sell-side analysts Analysts who work at brokerages.

Sensitivity analysis A technique for exploring how a target variable (e.g., portfolio returns) and risk profiles are affected by changes in input variables (e.g., the distribution of asset or factor returns).

Sentence length The number of characters, including spaces, in a sentence.

Serially correlated With reference to regression errors, errors that are correlated across observations.

Service period For employee stock options, usually the period between the grant date and the vesting date.

Settled in arrears An arrangement in which the interest payment is made (i.e., settlement occurs) at the maturity of the underlying instrument.

Settlement In the case of a credit event, the process by which the two parties to a CDS contract satisfy their respective obligations.

Shaping risk The sensitivity of a bond's price to the changing shape of the yield curve.

Share repurchase A transaction in which a company buys back its own shares. Unlike stock dividends and stock splits, share repurchases use corporate cash.

Shareholder activism Strategies used by shareholders to attempt to compel a company to act in a desired manner.

Shareholders' equity Total assets minus total liabilities.

Shark repellents A pre-offer takeover defense mechanism involving the corporate charter (e.g., staggered boards of directors and supermajority provisions).

Simple linear regression (SLR) A regression that summarizes the relation between the dependent variable and a single independent variable.

Simulation A technique for exploring how a target variable (e.g. portfolio returns) would perform in a hypothetical environment specified by the user, rather than a historical setting.

Single-name CDS Credit default swap on one specific borrower.

Sinking fund bond A bond that requires the issuer to set aside funds over time to retire the bond issue, thus reducing credit risk.

Slope coefficient The coefficient of an independent variable that represents the average change in the dependent variable for a one-unit change in the independent variable.

Soft margin classification An adaptation in the support vector machine algorithm which adds a penalty to the objective function for observations in the training set that are misclassified.

Special dividend A dividend paid by a company that does not pay dividends on a regular schedule or a dividend that supplements regular cash dividends with an extra payment.

Spin-off A form of restructuring in which shareholders of a parent company receive a proportional number of shares in a new, separate entity; shareholders end up owning stock in two different companies where there used to be one.

Split-off A form of restructuring in which shareholders of the parent company are given shares in a newly created entity in exchange for their shares of the parent company.

Split-rate tax system In reference to corporate taxes, a split-rate system taxes earnings to be distributed as dividends at a different rate than earnings to be retained. Corporate profits distributed as dividends are taxed at a lower rate than those retained in the business.

Spot curve The term structure of spot rates for loans made today.

Spot price The current price of an asset or security. For commodities, the current price to deliver a physical commodity to a specific location or purchase and transport it away from a designated location.

Spot rate The interest rate that is determined today for a risk-free, single-unit payment at a specified future date.

Spot yield curve The term structure of spot rates for loans made today.

Stabilized NOI In the context of real estate, the expected NOI when a renovation is complete.

Stable dividend policy A policy in which regular dividends are paid that reflect long-run expected earnings. In contrast to a constant dividend payout ratio policy, a stable dividend policy does not reflect short-term volatility in earnings.

Standard error of the estimate A measure of the fit of a regression line, calculated as the square root of the mean square error. Also known as the *standard error of the regression* and the *root mean square error*.

Standard error of the forecast A measure of the uncertainty associated with a forecasted value of the dependent variable that depends on the standard error of the estimate, the variability of the independent variable, the deviation of the forecasted independent variable from the mean in the regression, and the number of observations.

Standard error of the slope coefficient The standard error of the slope, which in a simple linear regression is the ratio of the model's standard error of the estimate (s_e) to the square root of the variation of the independent variable.

Standardized beta With reference to fundamental factor models, the value of the attribute for an asset minus the average value of the attribute across all stocks, divided by the standard deviation of the attribute across all stocks.

Standardized unexpected earnings Unexpected earnings per share divided by the standard deviation of unexpected earnings per share over a specified prior time period.

Static trade-off theory of capital structure A theory pertaining to a company's optimal capital structure. The optimal level of debt is found at the point where additional debt would cause the costs of financial distress to increase by a greater amount than the benefit of the additional tax shield.

Statistical factor model A multifactor model in which statistical methods are applied to a set of historical returns to determine portfolios that best explain either historical return covariances or variances.

Statutes Laws enacted by legislative bodies.

Statutory merger A merger in which one company ceases to exist as an identifiable entity and all its assets and liabilities become part of a purchasing company.

Steady-state rate of growth The constant growth rate of output (or output per capita) that can or will be sustained indefinitely once it is reached. Key ratios, such as the capital–output ratio, are constant on the steady-state growth path.

Steepness The difference between long-term and short-term yields that constitutes one of the three factors (the other two are level and curvature) that empirically explain most of the changes in the shape of the yield curve.

Stock dividend A type of dividend in which a company distributes additional shares of its common stock to shareholders instead of cash.

Stock purchase An acquisition in which the acquirer gives the target company's shareholders some combination of cash and securities in exchange for shares of the target company's stock.

Stop-loss limit Constraint used in risk management that requires a reduction in the size of a portfolio, or its complete liquidation, when a loss of a particular size occurs in a specified period.

Straight bond An underlying option-free bond with a specified issuer, issue date, maturity date, principal amount and repayment structure, coupon rate and payment structure, and currency denomination.

Straight-line rent The average annual rent under a multi-year lease agreement that contains contractual increases in rent during the life of the lease.

Straight-line rent adjustment See *non-cash rent.*

Straight voting Voting structure in which shareholders are granted the right of one vote for each share owned.

Stranded assets Assets that are obsolete or not economically viable.

Strategic transaction A purchase involving a buyer that would benefit from certain synergies associated with owning the target firm.

Stress tests A risk management technique that assesses the portfolio's response to extreme market movements.

Stripping A dealer's ability to separate a bond's individual cash flows and trade them as zero-coupon securities.

Subsidiary merger A merger in which the company being purchased becomes a subsidiary of the purchaser.

Substantive law The body of law that focuses on the rights and responsibilities of entities and relationships among entities.

Succession event A change of corporate structure of the reference entity, such as through a merger, a divestiture, a spinoff, or any similar action, in which ultimate responsibility for the debt in question is unclear.

Sum of squares error (SSE) The sum of the squared deviations of (1) the value of the dependent variable and (2) the value of the dependent variable based on the estimated regression line. Also referred to as the *residual sum of squares.*

Sum of squares regression (SSR) The sum of the squared deviations of (1) the value of the dependent variable based on the estimated regression line and (2) the mean of the dependent variable.

Sum of squares total (SST) The sum of the squared deviations of the dependent variable from its mean; the variation of the dependent variable. Also referred to as the *total sum of squares.*

Sum-of-the-parts valuation A valuation that sums the estimated values of each of a company's businesses as if each business were an independent going concern.

Summation operator A functional part of a neural network's node that multiplies each input value received by a weight and sums the weighted values to form the total net input, which is then passed to the activation function.

Supernormal growth Above-average or abnormally high growth rate in earnings per share.

Supervised learning Machine learning where algorithms infer patterns between a set of inputs (the *X*'s) and the desired output (*Y*). The inferred pattern is then used to map a given input set into a predicted output.

Support vector machine A linear classifier that determines the hyperplane that optimally separates the observations into two sets of data points.

Survivorship bias The bias that results when data as of a given date reflects only those entities that have survived to that date. Entities can include any element of an index or list that is constituted through time: stocks, investment funds, etc. Survivorship bias is a form of look-ahead bias.

Sustainable growth rate The rate of dividend (and earnings) growth that can be sustained over time for a given level of return on equity, keeping the capital structure constant and without issuing additional common stock.

Swap curve The term structure of swap rates.

Swap rate The "price" that swap traders quote among one another. It is the rate at which the present value of all the expected floating-rate payments received over the life of the floating-rate bond equal the present value of all the expected fixed-rate payments made over the life of the fixed-rate bond.

Swap rate curve The term structure of swap rates.

Swap spread The difference between the fixed rate on an interest rate swap and the rate on a Treasury note with equivalent maturity; it reflects the general level of credit risk in the market.

Systematic risk Risk that affects the entire market or economy; it cannot be avoided and is inherent in the overall market. Systematic risk is also known as non-diversifiable or market risk.

Systemic risk The risk of failure of the financial system.

Tail risk The risk that losses in extreme events could be greater than would be expected for a portfolio of assets with a normal distribution.

Takeover A merger; the term may be applied to any transaction but is often used in reference to hostile transactions.

Takeover premium The amount by which the takeover price for each share of stock must exceed the current stock price in order to entice shareholders to relinquish control of the company to an acquirer.

Tangible book value per share Common shareholders' equity minus intangible assets reported on the balance sheet, divided by the number of shares outstanding.

Target In machine learning, the dependent variable (*Y*) in a labeled dataset; the company in a merger or acquisition that is being acquired.

Target capital structure A company's chosen proportions of debt and equity.

Target company See *target.*

Target payout ratio A strategic corporate goal representing the long-term proportion of earnings that the company intends to distribute to shareholders as dividends.

Taxable REIT subsidiaries Subsidiaries that pay income taxes on earnings from non-REIT-qualifying activities like merchant development or third-party property management.

Technical indicators Momentum indicators based on price.

TED spread A measure of perceived credit risk determined as the difference between Libor and the T-bill yield of matching maturity.

Temporal method A variation of the monetary/non-monetary translation method that requires not only monetary assets and liabilities, but also non-monetary assets and liabilities that are measured at their current value on the balance sheet date to be translated at the current exchange rate. Assets and liabilities are translated at rates consistent with the timing of their measurement value. This method is typically used when the functional currency is other than the local currency.

Tender offer A public offer whereby the acquirer invites target shareholders to submit ("tender") their shares in return for the proposed payment.

Term frequency (TF) Ratio of the number of times a given token occurs in all the texts in the dataset to the total number of tokens in the dataset.

Term premium The additional return required by lenders to invest in a bond to maturity net of the expected return from continually reinvesting at the short-term rate over that same time horizon.

Terminal price multiples The price multiple for a stock assumed to hold at a stated future time.

Terminal share price The share price at a particular point in the future.

Terminal value of the stock The analyst's estimate of a stock's value at a particular point in the future. Also called *continuing value of the stock*.

Test sample A data sample that is used to test a model's ability to predict well on new data.

Theta The change in a derivative instrument for a given small change in calendar time, holding everything else constant. Specifically, the theta calculation assumes nothing changes except calendar time. Theta also reflects the rate at which an option's time value decays.

Time series A set of observations on a variable's outcomes in different time periods.

Tobin's *q* The ratio of the market value of debt and equity to the replacement cost of total assets.

Token The equivalent of a word (or sometimes a character).

Tokenization The process of splitting a given text into separate tokens. Tokenization can be performed at the word or character level but is most commonly performed at word level.

Top-down approach With respect to forecasting, an approach that usually begins at the level of the overall economy. Forecasts are then made at more narrowly defined levels, such as sector, industry, and market for a specific product.

Total factor productivity (TFP) A multiplicative scale factor that reflects the general level of productivity or technology in the economy. Changes in total factor productivity generate proportional changes in output for any input combination.

Total invested capital The sum of market value of common equity, book value of preferred equity, and face value of debt.

Tracking error The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking risk*.

Tracking risk The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking error*.

Trailing dividend yield The reciprocal of current market price divided by the most recent annualized dividend.

Trailing P/E A stock's current market price divided by the most recent four quarters of EPS (or the most recent two semi-annual periods for companies that report interim data semi-annually). Also called *current P/E*.

Training sample A data sample that is used to train a model.

Tranche CDS A type of credit default swap that covers a combination of borrowers but only up to pre-specified levels of losses.

Transaction exposure The risk of a change in value between the transaction date and the settlement date of an asset of liability denominated in a foreign currency.

Treasury shares/stock Shares that were issued and subsequently repurchased by the company.

Trend A long-term pattern of movement in a particular direction.

Triangular arbitrage An arbitrage transaction involving three currencies that attempts to exploit inconsistencies among pairwise exchange rates.

Trimming Also called truncation, it is the process of removing extreme values and outliers from a dataset.

Triple-net leases Common leases in the United States and Canada that require each tenant to pay its share of the following three operating expenses: common area maintenance and repair expenses; property taxes; and building insurance costs. Also known as *NNN leases*.

Two-tier board Board structure consisting of a supervisory board that oversees a management board.

Unbiased expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *pure expectations theory*.

Unconditional heteroskedasticity Heteroskedasticity of the error term that is not correlated with the values of the independent variable(s) in the regression.

Uncovered interest rate parity The proposition that the expected return on an uncovered (i.e., unhedged) foreign currency (risk-free) investment should equal the return on a comparable domestic currency investment.

Underlying earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *core earnings*, or *persistent earnings*.

Unexpected earnings The difference between reported EPS and expected EPS. Also referred to as an *earnings surprise*.

Unit root A time series that is not covariance stationary is said to have a unit root.

Unsupervised learning Machine learning that does not make use of labeled data.

Upfront payment The difference between the credit spread and the standard rate paid by the protection buyer if the standard rate is insufficient to compensate the protection seller. Also called *upfront premium*.

Upfront premium See *upfront payment*.

Upstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary company) such that the associate company records a profit on its income statement. An example is a sale of inventory by the associate to the investor company or by a subsidiary to a parent company.

Validation sample A data sample that is used to validate and tune a model.

Valuation The process of determining the value of an asset or service either on the basis of variables perceived to be related to future investment returns or on the basis of comparisons with closely similar assets.

Value additivity An arbitrage opportunity when the value of the whole equals the sum of the values of the parts.

Value at risk (VaR) The minimum loss that would be expected a certain percentage of the time over a certain period of time given the assumed market conditions.

Value of growth The difference between the actual value per share and the no-growth value per share.

Variance error Describes how much a model's results change in response to new data from validation and test samples. Unstable models pick up noise and produce high variance error, causing overfitting and high out-of-sample error.

Vasicek model A partial equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is constant.

Vega The change in a given derivative instrument for a given small change in volatility, holding everything else constant. A sensitivity measure for options that reflects the effect of volatility.

Venture capital investors Private equity investors in development-stage companies.

Vertical merger A merger involving companies at different positions of the same production chain; for example, a supplier or a distributor.

Vertical ownership Ownership structure in which a company or group that has a controlling interest in two or more holding companies, which in turn have controlling interests in various operating companies.

Vested benefit obligation The actuarial present value of vested benefits.

Vesting date The date that employees can first exercise stock options.

Visibility The extent to which a company's operations are predictable with substantial confidence.

Voting caps Legal restrictions on the voting rights of large share positions.

Web spidering (scraping or crawling) programs Programs that extract raw content from a source, typically web pages.

Weighted average cost of capital (WACC) A weighted average of the after-tax required rates of return on a company's common stock, preferred stock, and long-term debt, where the weights are the fraction of each source of financing in the company's target capital structure.

Weighted harmonic mean See *harmonic mean*.

White-corrected standard errors A synonym for robust standard errors.

White knight A third party that is sought out by the target company's board to purchase the target in lieu of a hostile bidder.

White squire A third party that is sought out by the target company's board to purchase a substantial minority stake in the target—enough to block a hostile takeover without selling the entire company.

Winner's curse The tendency for the winner in certain competitive bidding situations to overpay, whether because of overestimation of intrinsic value, emotion, or information asymmetries.

Winsorization The process of replacing extreme values and outliers in a dataset with the maximum (for large value outliers) and minimum (for small value outliers) values of data points that are not outliers.

Write-down A reduction in the value of an asset as stated in the balance sheet.

Yield curve factor model A model or a description of yield curve movements that can be considered realistic when compared with historical data.

Zero A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

Zero-coupon bond A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

FINANCIAL STATEMENT ANALYSIS

CFA® Program Curriculum

2022 • LEVEL II • VOLUME 2

WILEY

CONTENTS

How to Use the CFA Program Curriculum

Congratulations on your decision to enter the Chartered Financial Analyst (CFA®) Program. This exciting and rewarding program of study reflects your desire to become a serious investment professional. You are embarking on a program noted for its high ethical standards and the breadth of knowledge, skills, and abilities (competencies) it develops. Your commitment should be educationally and professionally rewarding.

The credential you seek is respected around the world as a mark of accomplishment and dedication. Each level of the program represents a distinct achievement in professional development. Successful completion of the program is rewarded with membership in a prestigious global community of investment professionals. CFA charterholders are dedicated to life-long learning and maintaining currency with the ever-changing dynamics of a challenging profession. CFA Program enrollment represents the first step toward a career-long commitment to professional education.

The CFA exam measures your mastery of the core knowledge, skills, and abilities required to succeed as an investment professional. These core competencies are the basis for the Candidate Body of Knowledge (CBOK™). The CBOK consists of four components:

- A broad outline that lists the major CFA Program topic areas (www.cfainstitute.org/programs/cfa/curriculum/cbok);
- Topic area weights that indicate the relative exam weightings of the top-level topic areas (www.cfainstitute.org/programs/cfa/curriculum);
- Learning outcome statements (LOS) that advise candidates about the specific knowledge, skills, and abilities they should acquire from readings covering a topic area (LOS are provided in candidate study sessions and at the beginning of each reading); and
- CFA Program curriculum that candidates receive upon exam registration.

Therefore, the key to your success on the CFA exams is studying and understanding the CBOK. The following sections provide background on the CBOK, the organization of the curriculum, features of the curriculum, and tips for designing an effective personal study program.

BACKGROUND ON THE CBOK

CFA Program is grounded in the practice of the investment profession. CFA Institute performs a continuous practice analysis with investment professionals around the world to determine the competencies that are relevant to the profession, beginning with the Global Body of Investment Knowledge (GBIK®). Regional expert panels and targeted surveys are conducted annually to verify and reinforce the continuous feedback about the GBIK. The practice analysis process ultimately defines the CBOK. The CBOK reflects the competencies that are generally accepted and applied by investment professionals. These competencies are used in practice in a generalist context and are expected to be demonstrated by a recently qualified CFA charterholder.

© 2021 CFA Institute. All rights reserved.

The CFA Institute staff—in conjunction with the Education Advisory Committee and Curriculum Level Advisors, who consist of practicing CFA charterholders—designs the CFA Program curriculum in order to deliver the CBOK to candidates. The exams, also written by CFA charterholders, are designed to allow you to demonstrate your mastery of the CBOK as set forth in the CFA Program curriculum. As you structure your personal study program, you should emphasize mastery of the CBOK and the practical application of that knowledge. For more information on the practice analysis, CBOK, and development of the CFA Program curriculum, please visit www.cfainstitute.org.

ORGANIZATION OF THE CURRICULUM

The Level II CFA Program curriculum is organized into 10 topic areas. Each topic area begins with a brief statement of the material and the depth of knowledge expected. It is then divided into one or more study sessions. These study sessions should form the basic structure of your reading and preparation. Each study session includes a statement of its structure and objective and is further divided into assigned readings. An outline illustrating the organization of these study sessions can be found at the front of each volume of the curriculum.

The readings are commissioned by CFA Institute and written by content experts, including investment professionals and university professors. Each reading includes LOS and the core material to be studied, often a combination of text, exhibits, and in-text examples and questions. End of Reading Questions (EORQs) followed by solutions help you understand and master the material. The LOS indicate what you should be able to accomplish after studying the material. The LOS, the core material, and the EORQs are dependent on each other, with the core material and EORQs providing context for understanding the scope of the LOS and enabling you to apply a principle or concept in a variety of scenarios.

The entire readings, including the EORQs, are the basis for all exam questions and are selected or developed specifically to teach the knowledge, skills, and abilities reflected in the CBOK.

You should use the LOS to guide and focus your study because each exam question is based on one or more LOS and the core material and practice problems associated with the LOS. As a candidate, you are responsible for the entirety of the required material in a study session.

We encourage you to review the information about the LOS on our website (www.cfainstitute.org/programs/cfa/curriculum/study-sessions), including the descriptions of LOS "command words" on the candidate resources page at www.cfainstitute.org.

FEATURES OF THE CURRICULUM

End of Reading Questions/Solutions *All End of Reading Questions (EORQs) as well as their solutions are part of the curriculum and are required material for the exam.* In addition to the in-text examples and questions, these EORQs help demonstrate practical applications and reinforce your understanding of the concepts presented. Some of these EORQs are adapted from past CFA exams and/or may serve as a basis for exam questions.

Glossary For your convenience, each volume includes a comprehensive Glossary. Throughout the curriculum, a **bolded** word in a reading denotes a term defined in the Glossary.

Note that the digital curriculum that is included in your exam registration fee is searchable for key words, including Glossary terms.

LOS Self-Check We have inserted checkboxes next to each LOS that you can use to track your progress in mastering the concepts in each reading.

Source Material The CFA Institute curriculum cites textbooks, journal articles, and other publications that provide additional context or information about topics covered in the readings. As a candidate, you are not responsible for familiarity with the original source materials cited in the curriculum.

Note that some readings may contain a web address or URL. The referenced sites were live at the time the reading was written or updated but may have been deactivated since then.

Some readings in the curriculum cite articles published in the *Financial Analysts Journal*®, which is the flagship publication of CFA Institute. Since its launch in 1945, the *Financial Analysts Journal* has established itself as the leading practitioner-oriented journal in the investment management community. Over the years, it has advanced the knowledge and understanding of the practice of investment management through the publication of peer-reviewed practitioner-relevant research from leading academics and practitioners. It has also featured thought-provoking opinion pieces that advance the common level of discourse within the investment management profession. Some of the most influential research in the area of investment management has appeared in the pages of the *Financial Analysts Journal*, and several Nobel laureates have contributed articles.

Candidates are not responsible for familiarity with *Financial Analysts Journal* articles that are cited in the curriculum. But, as your time and studies allow, we strongly encourage you to begin supplementing your understanding of key investment management issues by reading this, and other, CFA Institute practice-oriented publications through the Research & Analysis webpage (www.cfainstitute.org/en/research).

Errata The curriculum development process is rigorous and includes multiple rounds of reviews by content experts. Despite our efforts to produce a curriculum that is free of errors, there are times when we must make corrections. Curriculum errata are periodically updated and posted by exam level and test date online (www.cfainstitute.org/en/programs/submit-errata). If you believe you have found an error in the curriculum, you can submit your concerns through our curriculum errata reporting process found at the bottom of the Curriculum Errata webpage.

DESIGNING YOUR PERSONAL STUDY PROGRAM

Create a Schedule An orderly, systematic approach to exam preparation is critical. You should dedicate a consistent block of time every week to reading and studying. Complete all assigned readings and the associated problems and solutions in each study session. Review the LOS both before and after you study each reading to ensure that

you have mastered the applicable content and can demonstrate the knowledge, skills, and abilities described by the LOS and the assigned reading. Use the LOS self-check to track your progress and highlight areas of weakness for later review.

Successful candidates report an average of more than 300 hours preparing for each exam. Your preparation time will vary based on your prior education and experience, and you will probably spend more time on some study sessions than on others.

You should allow ample time for both in-depth study of all topic areas and additional concentration on those topic areas for which you feel the least prepared.

CFA INSTITUTE LEARNING ECOSYSTEM (LES)

As you prepare for your exam, we will email you important exam updates, testing policies, and study tips. Be sure to read these carefully.

Your exam registration fee includes access to the CFA Program Learning Ecosystem (LES). This digital learning platform provides access, even offline, to all of the readings and End of Reading Questions found in the print curriculum organized as a series of shorter online lessons with associated EORQs. This tool is your one-stop location for all study materials, including practice questions and mock exams.

The LES provides the following supplemental study tools:

Structured and Adaptive Study Plans The LES offers two ways to plan your study through the curriculum. The first is a structured plan that allows you to move through the material in the way that you feel best suits your learning. The second is an adaptive study plan based on the results of an assessment test that uses actual practice questions.

Regardless of your chosen study path, the LES tracks your level of proficiency in each topic area and presents you with a dashboard of where you stand in terms of proficiency so that you can allocate your study time efficiently.

Flashcards and Game Center The LES offers all the Glossary terms as Flashcards and tracks correct and incorrect answers. Flashcards can be filtered both by curriculum topic area and by action taken—for example, answered correctly, unanswered, and so on. These Flashcards provide a flexible way to study Glossary item definitions.

The Game Center provides several engaging ways to interact with the Flashcards in a game context. Each game tests your knowledge of the Glossary terms a in different way. Your results are scored and presented, along with a summary of candidates with high scores on the game, on your Dashboard.

Discussion Board The Discussion Board within the LES provides a way for you to interact with other candidates as you pursue your study plan. Discussions can happen at the level of individual lessons to raise questions about material in those lessons that you or other candidates can clarify or comment on. Discussions can also be posted at the level of topics or in the initial Welcome section to connect with other candidates in your area.

Practice Question Bank The LES offers access to a question bank of hundreds of practice questions that are in addition to the End of Reading Questions. These practice questions, only available on the LES, are intended to help you assess your mastery of individual topic areas as you progress through your studies. After each practice question, you will receive immediate feedback noting the correct response and indicating the relevant assigned reading so you can identify areas of weakness for further study.

Mock Exams The LES also includes access to three-hour Mock Exams that simulate the morning and afternoon sessions of the actual CFA exam. These Mock Exams are intended to be taken after you complete your study of the full curriculum and take practice questions so you can test your understanding of the curriculum and your readiness for the exam. If you take these Mock Exams within the LES, you will receive feedback afterward that notes the correct responses and indicates the relevant assigned readings so you can assess areas of weakness for further study. We recommend that you take Mock Exams during the final stages of your preparation for the actual CFA exam. For more information on the Mock Exams, please visit www.cfainstitute.org.

PREP PROVIDERS

You may choose to seek study support outside CFA Institute in the form of exam prep providers. After your CFA Program enrollment, you may receive numerous solicitations for exam prep courses and review materials. When considering a prep course, make sure the provider is committed to following the CFA Institute guidelines and high standards in its offerings.

Remember, however, that there are no shortcuts to success on the CFA exams; reading and studying the CFA Program curriculum *is* the key to success on the exam. The CFA Program exams reference only the CFA Institute assigned curriculum; no prep course or review course materials are consulted or referenced.

SUMMARY

Every question on the CFA exam is based on the content contained in the required readings and on one or more LOS. Frequently, an exam question is based on a specific example highlighted within a reading or on a specific practice problem and its solution. To make effective use of the CFA Program curriculum, please remember these key points:

1. All pages of the curriculum are required reading for the exam.
2. All questions, problems, and their solutions are part of the curriculum and are required study material for the exam. These questions are found at the end of the readings in the print versions of the curriculum. In the LES, these questions appear directly after the lesson with which they are associated. The LES provides immediate feedback on your answers and tracks your performance on these questions throughout your study.
3. We strongly encourage you to use the CFA Program Learning Ecosystem. In addition to providing access to all the curriculum material, including EORQs, in the form of shorter, focused lessons, the LES offers structured and adaptive study planning, a Discussion Board to communicate with other candidates, Flashcards, a Game Center for study activities, a test bank of practice questions, and online Mock Exams. Other supplemental study tools, such as eBook and PDF versions of the print curriculum, and additional candidate resources are available at www.cfainstitute.org.
4. Using the study planner, create a schedule and commit sufficient study time to cover the study sessions. You should also plan to review the materials, answer practice questions, and take Mock Exams.
5. Some of the concepts in the study sessions may be superseded by updated rulings and/or pronouncements issued after a reading was published. Candidates are expected to be familiar with the overall analytical framework contained in the assigned readings. Candidates are not responsible for changes that occur after the material was written.

FEEDBACK

At CFA Institute, we are committed to delivering a comprehensive and rigorous curriculum for the development of competent, ethically grounded investment professionals. We rely on candidate and investment professional comments and feedback as we work to improve the curriculum, supplemental study tools, and candidate resources.

Please send any comments or feedback to info@cfainstitute.org. You can be assured that we will review your suggestions carefully. Ongoing improvements in the curriculum will help you prepare for success on the upcoming exams and for a lifetime of learning as a serious investment professional.

Financial Statement Analysis

STUDY SESSIONS

Study Session 4	Financial Statement Analysis (1)
Study Session 5	Financial Statement Analysis (2)

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to analyze the effects of financial reporting choices on financial statements and ratios. The candidate also should be able to analyze and interpret financial statements and accompanying disclosures and to evaluate financial reporting quality.

Investments in other entities, post-employment (retirement) benefits, and cross-border transactions introduce complexity for the financial analyst. Based on their jurisdiction and depending on whether IFRS or GAAP accounting standards are applied, companies may look and report differently with respect to these items. By identifying and reconciling these reporting differences, analysts can more accurately assess a company's financial condition and position compared with its peers.

Note: Changes in accounting standards as well as new rulings and/or pronouncements issued after the publication of the readings on financial reporting and analysis may cause some of the information in these readings to become dated. Candidates are *not* responsible for anything that occurs after the readings were published. In addition, candidates are expected to be familiar with the analytical frameworks contained in the readings, as well as the implications of alternative accounting methods for financial analysis and valuation discussed in the readings. Candidates are also responsible for the content of accounting standards, but not for the actual reference numbers. Finally, candidates should be aware that certain ratios may be defined and calculated differently. When alternative ratio definitions exist and no specific definition is given, candidates should use the ratio definitions emphasized in the readings.

© 2021 CFA Institute. All rights reserved.

FINANCIAL RATIO LIST

Candidates should be aware that certain ratios may be defined differently. Such differences are part of the nature of practical financial analysis. For examination purposes, when alternative ratio definitions exist and no specific definition is given in the question, candidates should use the definition provided in this list of ratios.

1. Current ratio = Current assets ÷ Current liabilities
2. Quick ratio = (Cash + Short-term marketable investments + Receivables) ÷ Current liabilities
3. Cash ratio = (Cash + Short-term marketable investments) ÷ Current liabilities
4. Defensive interval ratio = (Cash + Short-term marketable investments + Receivables) ÷ Daily cash expenditures
5. Receivables turnover ratio = Total revenue ÷ Average receivables
6. Days of sales outstanding (DSO) = Number of days in period ÷ Receivables turnover ratio
7. Inventory turnover ratio = Cost of goods sold ÷ Average inventory
8. Days of inventory on hand (DOH) = Number of days in period ÷ Inventory turnover ratio
9. Payables turnover ratio = Purchases ÷ Average trade payables
10. Number of days of payables = Number of days in period ÷ Payables turnover ratio
11. Cash conversion cycle (net operating cycle) = DOH + DSO − Number of days of payables
12. Working capital turnover ratio = Total revenue ÷ Average working capital
13. Fixed asset turnover ratio = Total revenue ÷ Average net fixed assets
14. Total asset turnover ratio = Total revenue ÷ Average total assets
15. Gross profit margin = Gross profit ÷ Total revenue
16. Operating profit margin = Operating profit ÷ Total revenue
17. Pretax margin = Earnings before tax but after interest ÷ Total revenue
18. Net profit margin = Net income ÷ Total revenue
19. Operating return on assets = Operating income ÷ Average total assets
20. Return on assets = Net income ÷ Average total assets
21. Return on equity = Net income ÷ Average shareholders' equity
22. Return on total capital = Earnings before interest and taxes ÷ (Interest bearing debt + Shareholders' equity)
23. Return on common equity = (Net income − Preferred dividends) ÷ Average common shareholders' equity
24. Tax burden = Net income ÷ Earnings before taxes
25. Interest burden = Earnings before taxes ÷ Earnings before interest and taxes
26. EBIT margin = Earnings before interest and taxes ÷ Total revenue
27. Financial leverage ratio (equity multiplier) = Average total assets ÷ Average shareholders' equity
28. Total debt = The total of interest-bearing short-term and long-term debt, excluding liabilities such as accrued expenses and accounts payable
29. Debt-to-assets ratio = Total debt ÷ Total assets
30. Debt-to-equity ratio = Total debt ÷ Total shareholders' equity

31 Debt-to-capital ratio = Total debt ÷ (Total debt + Total shareholders' equity)

32 Interest coverage ratio = Earnings before interest and taxes ÷ Interest payments

33 Fixed charge coverage ratio = (Earnings before interest and taxes + Lease payments) ÷ (Interest payments + Lease payments)

34 Dividend payout ratio = Common share dividends ÷ Net income attributable to common shares

35 Retention rate = (Net income attributable to common shares – Common share dividends) ÷ Net income attributable to common shares = 1 – Payout ratio

36 Sustainable growth rate = Retention rate × Return on equity

37 Earnings per share = (Net income – Preferred dividends) ÷ Weighted average number of ordinary shares outstanding

38 Book value per share = Common stockholders' equity ÷ Total number of common shares outstanding

39 Free cash flow to equity (FCFE) = Cash flow from operating activities – Investment in fixed capital + Net borrowing

40 Free cash flow to the firm (FCFF) = Cash flow from operating activities + Interest expense × (1 – Tax rate) – Investment in fixed capital *(Interest expense should be added back only if it was subtracted in determining cash flow from operating activities. This may not be the case for companies electing an alternative treatment under IFRS.)*

FINANCIAL STATEMENT ANALYSIS

STUDY SESSION

4

Financial Statement Analysis (1)

This study session covers investments in other entities (intercorporate investments), post-employment benefits, and foreign currency transactions. Intercorporate investments take the form of investments in 1) financial assets, 2) associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities. Current and new reporting standards for these investments are examined. The valuation and treatment of post-employment benefits follows, including share-based compensation (grants, options). Differences in valuation methods between defined-contribution and defined-benefit plans are described. The effect of foreign currency on a business's financials and methods to translate foreign currency from operations for consolidated financial statement reporting is examined. Analysis of financial institutions, including factors for consideration and an analysis approach (CAMELS), concludes the session.

Note: Changes in accounting standards as well as new rulings and/or pronouncements issued after the publication of the readings on financial reporting and analysis may cause some of the information in these readings to become dated. Candidates are *not* responsible for anything that occurs after the readings were published. In addition, candidates are expected to be familiar with the analytical frameworks contained in the readings, as well as the implications of alternative accounting methods for financial analysis and valuation discussed in the readings. Candidates are also responsible for the content of accounting standards, but not for the actual reference numbers. Finally, candidates should be aware that certain ratios may be defined and calculated differently. When alternative ratio definitions exist and no specific definition is given, candidates should use the ratio definitions emphasized in the readings.

READING ASSIGNMENTS

© 2021 CFA Institute. All rights reserved.

READING

9

Intercorporate Investments

by Susan Perry Williams, CPA, CMA, PhD

Susan Perry Williams, CPA, CMA, PhD, is Professor Emeritus at the McIntire School of Commerce, University of Virginia (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;
☐	**b.** compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;
☐	**c.** analyze how different methods used to account for intercorporate investments affect financial statements and ratios.

Note: New rulings and/or pronouncements issued after the publication of the readings in financial reporting and analysis may cause some of the information in these readings to become dated. Candidates are expected to be familiar with the overall analytical framework contained in the study session readings, as well as the implications of alternative accounting methods for financial analysis and valuation, as provided in the assigned readings. Candidates are not responsible for changes that occur after the material was written.

© 2019 CFA Institute. All rights reserved.

1

INTRODUCTION

Intercorporate investments (investments in other companies) can have a significant impact on an investing company's financial performance and position. Companies invest in the debt and equity securities of other companies to diversify their asset base, enter new markets, obtain competitive advantages, deploy excess cash, and achieve additional profitability. Debt securities include commercial paper, corporate and government bonds and notes, redeemable preferred stock, and asset-backed securities. Equity securities include common stock and non-redeemable preferred stock. The percentage of equity ownership a company acquires in an investee depends on the resources available, the ability to acquire the shares, and the desired level of influence or control.

The International Accounting Standards Board (IASB) and the US Financial Accounting Standards Board (FASB) worked to reduce differences in accounting standards that apply to the classification, measurement, and disclosure of intercorporate investments. The resulting standards have improved the relevance, transparency, and comparability of information provided in financial statements.

Complete convergence between IFRS accounting standards and US GAAP did not occur for accounting for financial instruments, and some differences still exist. The terminology used in this reading is IFRS-oriented. US GAAP may not use identical terminology, but in most cases the terminology is similar.

This reading is organized as follows: Section 2 explains the basic categorization of corporate investments. Section 3 describes reporting under IFRS 9, the IASB standard for financial instruments. Sections 4–6 describe equity method reporting for investments in associates where significant influence can exist including the reporting for joint ventures, a type of investment where control is shared. Sections 7–11 describe reporting for business combinations, the parent/subsidiary relationship, and variable interest and special purpose entities. A summary concludes the reading.

2

BASIC CORPORATE INVESTMENT CATEGORIES

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

In general, investments in marketable debt and equity securities can be categorized as 1) investments in financial assets in which the investor has no significant influence or control over the operations of the investee, 2) investments in associates in which the investor can exert significant influence (but not control) over the investee, 3) joint ventures where control is shared by two or more entities, and 4) business combinations, including investments in subsidiaries, in which the investor obtains a controlling interest over the investee. The distinction between investments in financial assets, investments in associates, and business combinations is based on the degree of influence or control rather than purely on the percent holding. However, lack of influence is generally presumed when the investor holds less than a 20% equity interest, significant influence is generally presumed between 20% and 50%, and control is presumed when the percentage of ownership exceeds 50%.

The following excerpt from Note 2 to the Financial Statements in the 2017 Annual Report of GlaxoSmithKline, a British pharmaceutical and healthcare company, illustrates the categorization and disclosure in practice:

> Entities over which the Group has the power to direct the relevant activities so as to affect the returns to the Group, generally through control over the financial and operating policies, are accounted for as subsidiaries.
>
> Where the Group has the ability to exercise joint control over, and rights to the net assets of, entities, the entities are accounted for as joint ventures. Where the Group has the ability to exercise joint control over an arrangement, but has rights to specified assets and obligations for specified liabilities of the arrangement, the arrangement is accounted for as a joint operation. Where the Group has the ability to exercise significant influence over entities, they are accounted for as associates. The results and assets and liabilities of associates and joint ventures are incorporated into the consolidated financial statements using the equity method of accounting. The Group's rights to assets, liabilities, revenue and expenses of joint operations are included in the consolidated financial statements in accordance with those rights and obligations.

A summary of the financial reporting and relevant standards for various types of corporate investment is presented in Exhibit 1 (the headings in Exhibit 1 use the terminology of IFRS; US GAAP categorizes intercorporate investments similarly but not identically). The reader should be alert to the fact that value measurement and/or the treatment of changes in value can vary depending on the classification and whether IFRS or US GAAP is used. The alternative treatments are discussed in greater depth later in this reading.

Exhibit 1 Summary of Accounting Treatments for Investments

	In Financial Assets	In Associates	Business Combinations	In Joint Ventures
Influence	Not significant	Significant	Controlling	Shared control
Typical percentage interest	Usually < 20%	Usually 20% to 50%	Usually > 50% or other indications of control	
US GAAP [b]	FASB ASC Topic 320	FASB ASC Topic 323	FASB ASC Topics 805 and 810	FASB ASC Topic 323
Financial Reporting	Classified as: ■ Fair value through profit or loss ■ Fair value through other comprehensive income ■ Amortized cost	Equity method	Consolidation	IFRS: Equity method

(continued)

Exhibit 1 (Continued)

	In Financial Assets	In Associates	Business Combinations	In Joint Ventures
Applicable IFRS [a]	IFRS 9	IAS 28	IAS 27 IFRS 3 IFRS 10	IFRS 11 IFRS 12 IAS 28
US GAAP [b]	FASB ASC Topic 320	FASB ASC Topic 323	FASB ASC Topics 805 and 810	FASB ASC Topic 323

[a] IFRS 9 Financial Instruments; IAS 28 Investments in Associates; IAS 27 Separate Financial Statements; IFRS 3 Business Combinations; IFRS 10 Consolidated Financial Statements; IFRS 11 Joint Arrangements; IFRS 12, Disclosure of Interests in Other Entities.
[b] FASB ASC Topic 320 [Investments–Debt and Equity Securities]; FASB ASC Topic 323 [Investments– Equity Method and Joint Ventures]; FASB ASC Topics 805 [Business Combinations] and 810 [Consolidations].

3 INVESTMENTS IN FINANCIAL ASSETS: IFRS 9

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

Both IASB and FASB developed revised standards for financial investments. The IASB issued the first phase of their project dealing with classification and measurement of financial instruments by including relevant chapters in IFRS 9, *Financial Instruments*. IFRS 9, which replaces IAS 39, became effective for annual periods on 1 January 2018. The FASB's guidance relating to the accounting for investments in financial instruments is contained in ASC 825, *Financial Instruments*, which has been updated several times, with the standard being effective for periods after 15 December 2017. The resulting US GAAP guidance has many consistencies with IFRS requirements, but there are also some differences.

IFRS 9 is based on an approach that considers the contractual characteristics of cash flows as well as the management of the financial assets. The portfolio approach of the previous standard (i.e., designation of held for trading, available-for-sale, and held-to-maturity) is no longer appropriate, and the terms *available-for-sale* and *held-to-maturity* no longer appear in IFRS 9. Another key change in IFRS 9, compared with IAS 39, relates to the approach to loan impairment. In particular, companies are required to migrate from an incurred loss model to an expected credit loss model. This results in companies evaluating not only historical and current information about loan performance, but also forward-looking information.[1]

1 Under US GAAP, requirements for assessing credit impairment are included in ASC 326, which is effective for most public companies beginning January 1, 2020.

The criteria for using amortized cost are similar to those of the IAS 39 "management intent to hold-to-maturity" classification. Specifically, to be measured at amortized cost, financial assets must meet two criteria:[2]

1 A business model test:[3] The financial assets are being held to collect contractual cash flows; and
2 A cash flow characteristic test: The contractual cash flows are solely payments of principal and interest on principal.

3.1 Classification and Measurement

IFRS 9 divides all financial assets into two classifications—those measured at amortized cost and those measured at fair value. Under this approach, there are three different categories of measurement:

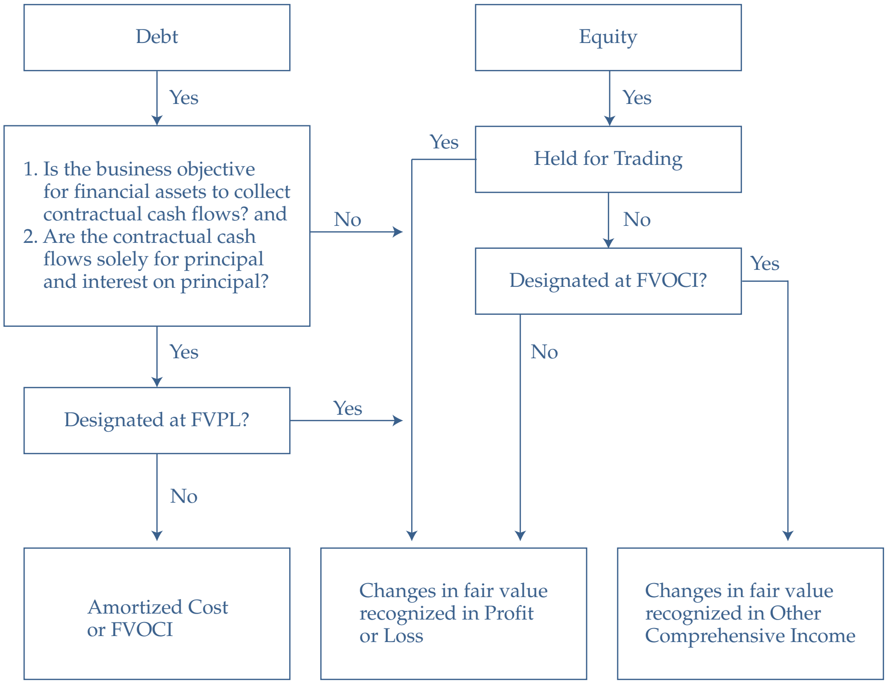
- Amortised cost
- Fair value through profit or loss (FVPL) or
- Fair Value through Other comprehensive income (FVOCI).

All financial assets are measured at fair value when initially acquired (which will generally be equal to the cost basis on the date of acquisition). Subsequently, financial assets are measured at either fair value or amortized cost. Financial assets that meet the two criteria above are generally measured at amortized cost. If the financial asset meets the criteria above but may be sold, a "hold-to-collect and sell" business model, it may be measured at fair value through other comprehensive income (FVOCI). However, management may choose the "fair value through profit or loss" (FVPL) option to avoid an accounting mismatch.[4] An "accounting mismatch" refers to an inconsistency resulting from different measurement bases for assets and liabilities, i.e., some are measured at amortized cost and some at fair value. Debt instruments are measured at amortized cost, fair value through other comprehensive income (FVOCI), or fair value through profit or loss (FVPL) depending upon the business model.

Equity instruments are measured at FVPL or at FVOCI; they are not eligible for measurement at amortized cost. Equity investments held-for-trading must be measured at FVPL. Other equity investments can be measured at FVPL or FVOCI; however, the choice is irrevocable. If the entity uses the FVOCI option, only the dividend income is recognized in profit or loss. Furthermore, the requirements for reclassifying gains or losses recognized in other comprehensive income are different for debt and equity instruments.

2 IFRS 9, paragraph 4.1.2.
3 A business model refers to how an entity manages its financial assets in order to generate cash flows – by collecting contractual cash flows, selling financial assets or both. (IFRS 9 Financial Instruments, Project Summary, July 2014)
4 IFRS 9, paragraph 4.1.5.

Exhibit 2 Financial Assets Classification and Measurement Model, IFRS 9



Financial assets that are derivatives are measured at fair value through profit or loss (except for hedging instruments). Embedded derivatives are not separated from the hybrid contract if the asset falls within the scope of this standard and the asset as a whole is measured at FVPL.

Exhibit 3 contains an excerpt from the 2017 Deutsche Bank financial statements that describes how financial assets and financial liabilities are determined, measured, and recognized on its financial statements.

Exhibit 3 Excerpt from Deutsche Bank's 2017 Financial Statements

Financial Assets

IFRS 9 requires that an entity's business model and a financial instrument's contractual cash flows will determine its classification and measurement in the financial statements. Upon initial recognition each financial asset will be classified as either fair value through profit or loss ('FVTPL'), amortized cost, or fair value through Other Comprehensive Income ('FVOCI'). As the requirements under IFRS 9 are different than the assessments under the existing IAS 39 rules, there will be some differences from the classification and measurement of financial assets under IAS 39, including whether to elect the fair value option on certain assets. The classification and measurement of financial liabilities remain largely unchanged under IFRS 9 from current requirements.

In 2015, the Group made an initial determination of business models and assessed the contractual cash flow characteristics of the financial assets within such business models to determine the potential classification and measurement changes as a result of IFRS 9. As a result of the initial analysis performed, in 2016 the Group identified a population of financial assets which are to be measured at either amortized cost or fair value through other comprehensive income, which will be subject to the IFRS 9 impairment rules. In 2017, the Group updated its

Exhibit 3 (Continued)

business model assessments and completed outstanding classification decisions. On initial recognition of an equity investment not held for trading, the Group may on an investment-by-investment basis, irrevocably elect to present subsequent fair value changes in OCI. The Group has not made any such elections. Where issued debt liabilities are designated at fair value, the fair value movements attributable to an entity's own credit risk will be recognized in Other Comprehensive Income rather than in the Statement of Income. The standard also allows the Group the option to elect to apply early the presentation of fair value movements of an entity's credit risk in Other Comprehensive Income prior to adopting IFRS 9 in full. The Group did not early adopt this requirement

3.2 Reclassification of Investments

Under IFRS 9, the reclassification of equity instruments is not permitted because an entity's initial classification of FVPL and FVOCI is irrevocable. Reclassification of debt instruments is only permitted if the business model for the financial assets (objective for holding the financial assets) has changed in a way that significantly affects operations. Changes to the business model will require judgment and are expected to be very infrequent.

When reclassification is deemed appropriate, there is no restatement of prior periods at the reclassification date. For example, if the financial asset is reclassified from amortized cost to FVPL, the asset is then measured at fair value with any gain or loss immediately recognized in profit or loss. If the financial asset is reclassified from FVPL to amortized cost, the fair value at the reclassification date becomes the carrying amount.

In summary, the major changes made by IFRS 9 are:

- A business model approach to classification of debt instruments.
- Three classifications for financial assets:
  - Fair value through profit or loss (FVPL),
  - fair value through other comprehensive income (FVOCI), and
  - amortized cost.
- Reclassifications of debt instruments are permitted only when the business model changes. The choice to measure equity investments at FVOCI or FVPL is irrevocable.
- A redesign of the provisioning models for financial assets, financial guarantees, loan commitments, and lease receivables. The new standard moves the recognition criteria from an "incurred loss" model to an "expected loss" model. Under the new criteria, there is an earlier recognition of impairment—12 month expected losses for performing assets and lifetime expected losses for non-performing assets, to be captured upfront.[5]

Analysts typically evaluate performance separately for operating and investing activities. Analysis of operating performance should exclude items related to investing activities such as interest income, dividends, and realized and unrealized gains and losses. For comparative purposes, analysts should exclude non-operating assets in the determination of return on net operating assets. IFRS and US GAAP[6] require

5 IFRS 9, paragraphs 5.5.4, 5.5.5, 5.5.15, 5.5.16.

6 IFRS 7 Financial Instruments: Disclosures and FASB ASC Section 320-10-50 [Investments–Debt and Equity Securities–Overall–Disclosure].

disclosure of fair value of each class of investment in financial assets. Using market values and adjusting pro forma financial statements for consistency improves assessments of performance ratios across companies.

4 INVESTMENTS IN ASSOCIATES AND JOINT VENTURES: EQUITY METHOD OF ACCOUNTING, BASIS PRINCIPLES

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

c analyze how different methods used to account for intercorporate investments affect financial statements and ratios.

Under both IFRS and US GAAP, when a company (investor) holds 20 to 50% of the voting rights of an associate (investee), either directly or indirectly (i.e., through subsidiaries), it is presumed that the company has (or can exercise) significant influence, but not control, over the investee's business activities.[7] Conversely, if the investor holds, directly or indirectly, less than 20% of the voting power of the associate (investee), it is presumed that the investor cannot exercise significant influence, unless such influence can be demonstrated. IAS 28 (IFRS) and FASB ASC Topic 323 (US GAAP) apply to most investments in which an investor has significant influence; they also provide guidance on accounting for investments in associates using the equity method.[8] These standards note that significant influence may be evidenced by

- representation on the board of directors;
- participation in the policy-making process;
- material transactions between the investor and the investee;
- interchange of managerial personnel; or
- technological dependency.

The ability to exert significant influence means that the financial and operating performance of the investee is partly influenced by management decisions and operational skills of the investor. The equity method of accounting for the investment reflects the economic reality of this relationship and provides a more objective basis for reporting investment income.

---

7 The determination of significant influence under IFRS also includes currently exercisable or convertible warrants, call options, or convertible securities that the investor owns, which give it additional voting power or reduce another party's voting power over the financial and operating policies of the investee. Under US GAAP, the determination of an investor's voting stock interest is based only on the voting shares outstanding at the time of the purchase. The existence and effect of securities with potential voting rights are not considered.

8 IAS 28 Investments in Associates and Joint Ventures and FASB ASC Topic 323 [Investments–Equity Method and Joint Ventures].

Joint ventures—ventures undertaken and controlled by two or more parties—can be a convenient way to enter foreign markets, conduct specialized activities, and engage in risky projects. They can be organized in a variety of different forms and structures. Some joint ventures are primarily contractual relationships, whereas others have common ownership of assets. They can be partnerships, limited liability companies (corporations), or other legal forms (unincorporated associations, for example). IFRS identify the following common characteristics of joint ventures: 1) A contractual arrangement exists between two or more venturers, and 2) the contractual arrangement establishes joint control. Both IFRS and US GAAP[9] require the equity method of accounting for joint ventures.[10]

Only under rare circumstances will joint ventures be allowed to use proportionate consolidation under IFRS and US GAAP. On the venturer's financial statements, proportionate consolidation requires the venturer's share of the assets, liabilities, income, and expenses of the joint venture to be combined or shown on a line-by-line basis with similar items under its sole control. In contrast, the equity method results in a single line item (equity in income of the joint venture) on the income statement and a single line item (investment in joint venture) on the balance sheet.

Because the single line item on the income statement under the equity method reflects the net effect of the sales and expenses of the joint venture, the total income recognized is identical under the two methods. In addition, because the single line item on the balance sheet item (investment in joint venture) under the equity method reflects the investors' share of the net assets of the joint venture, the total net assets of the investor is identical under both methods. There can be significant differences, however, in ratio analysis between the two methods because of the differential effects on values for total assets, liabilities, sales, expenses, etc.

4.1 Equity Method of Accounting: Basic Principles

Under the equity method of accounting, the equity investment is initially recorded on the investor's balance sheet at cost. In subsequent periods, the carrying amount of the investment is adjusted to recognize the investor's proportionate share of the investee's earnings or losses, and these earnings or losses are reported in income. Dividends or other distributions received from the investee are treated as a return of capital and reduce the carrying amount of the investment and are not reported in the investor's profit or loss. The equity method is often referred to as "one-line consolidation" because the investor's proportionate ownership interest in the assets and liabilities of the investee is disclosed as a single line item (net assets) on its balance sheet, and the investor's share of the revenues and expenses of the investee is disclosed as a single line item on its income statement. (Contrast these disclosures with the disclosures on consolidated statements in Section 9.) Equity method investments are classified as non-current assets on the balance sheet. The investor's share of the profit or loss of equity method investments, and the carrying amount of those investments, must be separately disclosed on the income statement and balance sheet.

9 Under US GAAP, ASC 323-10 provides guidance on the application of the equity method of accounting.
10 IFRS 11, Joint Arrangements classifies joint arrangements as either a joint operation or a joint venture. Joint ventures are arrangements wherein parties with joint control have rights to the net assets of the arrangement. Joint ventures are required to use equity method under IAS 28.

EXAMPLE 1

Equity Method: Balance in Investment Account

Branch (a fictitious company) purchases a 20% interest in Williams (a fictitious company) for €200,000 on 1 January 2016. Williams reports income and dividends as follows:

	Income	Dividends
2016	€200,000	€50,000
2017	300,000	100,000
2018	400,000	200,000
	€900,000	€350,000

Calculate the investment in Williams that appears on Branch's balance sheet as of the end of 2018.

Solution:

Investment in Williams at 31 December 2018:

Initial cost	€200,000	
Equity income 2016	€40,000	= (20% of €200,000 Income)
Dividends received 2016	(€10,000)	= (20% of €50,000 Dividends)
Equity income 2017	€60,000	= (20% of €300,000 Income)
Dividends received 2017	(€20,000)	= (20% of €100,000 Dividends)
Equity income 2018	€80,000	= (20% of €400,000 Income)
Dividends received 2018	(€40,000)	= (20% of €200,000 Dividends)
Balance-Equity Investment	€310,000	= [€200,000 + 20% × (€900,000 − €350,000)]

This simple example implicitly assumes that the purchase price equals the purchased equity (20%) in the book value of Williams' net assets. Sections 6 and 7 will cover the more typical case in which the purchase price does not equal the proportionate share of the book value of the investee's net assets.

Using the equity method, the investor includes its share of the investee's profit and losses on the income statement. The equity investment is carried at cost, plus its share of post-acquisition income, less dividends received. The recorded investment value can decline as a result of investee losses or a permanent decline in the investee's market value (see Section 9 for treatment of impairments). If the investment value is reduced to zero, the investor usually discontinues the equity method and does not record further losses. If the investee subsequently reports profits, the equity method is resumed after the investor's share of the profits equals the share of losses not recognized during the suspension of the equity method. Exhibit 4 contains excerpts from Deutsche Bank's 2017 annual report that describes its accounting treatment for investments in associates.

Exhibit 4 Excerpt from Deutsche Bank 2017 Annual Report

[From Note 01] ASSOCIATES

An associate is an entity in which the Group has significant influence, but not a controlling interest, over the operating and financial management policy decisions of the entity. Significant influence is generally presumed when the Group holds between 20 % and 50 % of the voting rights. The existence and effect of potential voting rights that are currently exercisable or convertible are considered in assessing whether the Group has significant influence. Among the other factors that are considered in determining whether the Group has significant influence are representation on the board of directors (supervisory board in the case of German stock corporations) and material intercompany transactions. The existence of these factors could require the application of the equity method of accounting for a particular investment even though the Group's investment is less than 20 % of the voting stock.

Investments in associates are accounted for under the equity method of accounting. The Group's share of the results of associates is adjusted to conform to the accounting policies of the Group and is reported in the Consolidated Statement of Income as Net income (loss) from equity method investments. The Group's share in the associate's profits and losses resulting from intercompany sales is eliminated on consolidation.

If the Group previously held an equity interest in an entity (for example, as available for sale) and subsequently gained significant influence, the previously held equity interest is remeasured to fair value and any gain or loss is recognized in the Consolidated Statement of Income. Any amounts previously recognized in other comprehensive income associated with the equity interest would be reclassified to the Consolidated Statement of Income at the date the Group gains significant influence, as if the Group had disposed of the previously held equity interest.

Under the equity method of accounting, the Group's investments in associates and jointly controlled entities are initially recorded at cost including any directly related transaction costs incurred in acquiring the associate, and subsequently increased (or decreased) to reflect both the Group's pro-rata share of the post-acquisition net income (or loss) of the associate or jointly controlled entity and other movements included directly in the equity of the associate or jointly controlled entity. Goodwill arising on the acquisition of an associate or a jointly controlled entity is included in the carrying value of the investment (net of any accumulated impairment loss). As goodwill is not reported separately it is not specifically tested for impairment. Rather, the entire equity method investment is tested for impairment at each balance sheet date.

If there is objective evidence of impairment, an impairment test is performed by comparing the investment's recoverable amount, which is the higher of its value in use and fair value less costs to sell, with its carrying amount. An impairment loss recognized in prior periods is only reversed if there has been a change in the estimates used to determine the in-vestment's recoverable amount since the last impairment loss was recognized. If this is the case the carrying amount of the investment is increased to its higher recoverable amount. The increased carrying amount of the investment in associate attributable to a reversal of an impairment loss shall not exceed the carrying amount that would have been determined had no impairment loss been recognized for the investment in prior years.

At the date that the Group ceases to have significant influence over the associate or jointly controlled entity the Group recognizes a gain or loss on the disposal of the equity method investment equal to the difference between the sum

(continued)

Exhibit 4 (Continued)

of the fair value of any retained investment and the proceeds from disposing of the associate and the carrying amount of the investment. Amounts recognized in prior periods in other comprehensive income in relation to the associate are accounted for on the same basis as would have been required if the investee had directly disposed of the related assets or liabilities.

[From Note 17] EQUITY METHOD INVESTMENTS

Investments in associates and jointly controlled entities are accounted for using the equity method of accounting.

The Group holds interests in 77 (2016: 92) associates and 13 (2016: 14) jointly controlled entities. There are no individually material investments in associates and joint ventures.

Aggregated financial information on the Group's share in associates and joint ventures that are individually immaterial (in €m)	Dec 31, 2017	Dec 31, 2016
Carrying amount of all associated that are individually immaterial to the Group	866	1,027
Aggregated amount of the Group's share of profit (loss) from continuing operations	141	183
Aggregated amount of the Group's share of post-tax profit (loss) from discontinued operations	0	0
Aggregated amount of the Group's share of other comprehensive income	(36)	11
Aggregated amount of the Group's share of total comprehensive income	105	194

It is interesting to note the explanations for the treatment of associates when the ownership percentage is less than 20% or is greater than 50%. The equity method reflects the strength of the relationship between the investor and its associates. In the instances where the percentage ownership is less than 20%, Deutsche Bank uses the equity method because it has significant influence over these associates' operating and financial policies either through its representation on their boards of directors and/or other measures. The equity method provides a more objective basis for reporting investment income than the accounting treatment for investments in financial assets because the investor can potentially influence the timing of dividend distributions.

5 INVESTMENT COSTS THAT EXCEED THE BOOK VALUE OF THE INVESTEE, AMORTIZATION OF EXCESS PURCHASE PRICE, FAIR VALUE OPTION AND IMPAIRMENT

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

The cost (purchase price) to acquire shares of an investee is often greater than the book value of those shares. This is because, among other things, many of the investee's assets and liabilities reflect historical cost rather than fair value. IFRS allow a company to measure its property, plant, and equipment using either historical cost or fair value (less accumulated depreciation).[11] US GAAP, however, require the use of historical cost (less accumulated depreciation) to measure property, plant, and equipment.[12]

When the cost of the investment exceeds the investor's proportionate share of the book value of the investee's (associate's) net identifiable tangible and intangible assets (e.g., inventory, property, plant and equipment, trademarks, patents), the difference is first allocated to specific assets (or categories of assets) using fair values. These differences are then amortized to the investor's proportionate share of the investee's profit or loss over the economic lives of the assets whose fair values exceeded book values. It should be noted that the allocation is not recorded formally; what appears initially in the investment account on the balance sheet of the investor is the cost. Over time, as the differences are amortized, the balance in the investment account will come closer to representing the ownership percentage of the book value of the net assets of the associate.

IFRS and US GAAP both treat the difference between the cost of the acquisition and investor's share of the fair value of the net identifiable assets as goodwill. Therefore, any remaining difference between the acquisition cost and the fair value of net identifiable assets that cannot be allocated to specific assets is treated as goodwill and is not amortized. Instead, it is reviewed for impairment on a regular basis, and written down for any identified impairment. Goodwill, however, is included in the carrying amount of the investment, because investment is reported as a single line item on the investor's balance sheet.[13]

11 After initial recognition, an entity can choose to use either a cost model or a revaluation model to measure its property, plant, and equipment. Under the revaluation model, property, plant, and equipment whose fair value can be measured reliably can be carried at a revalued amount. This revalued amount is its fair value at the date of the revaluation less any subsequent accumulated depreciation

12 Successful companies should be able to generate, through the productive use of assets, economic value in excess of the resale value of the assets themselves. Therefore, investors may be willing to pay a premium in anticipation of future benefits. These benefits could be a result of general market conditions, the investor's ability to exert significant influence on the investee, or other synergies.

13 If the investor's share of the fair value of the associate's net assets (identifiable assets, liabilities, and contingent liabilities) is greater than the cost of the investment, the difference is excluded from the carrying amount of the investment and instead included as income in the determination of the investor's share of the associate's profit or loss in the period in which the investment is acquired.

EXAMPLE 2

Equity Method Investment in Excess of Book Value

Blake Co. and Brown Co. are two hypothetical companies. Assume that Blake Co. acquires 30% of the outstanding shares of Brown Co. At the acquisition date, book values and fair values of Brown's recorded assets and liabilities are as follows:

	Book Value	Fair Value
Current assets	€10,000	€10,000
Plant and equipment	190,000	220,000
Land	120,000	140,000
	€320,000	€370,000
Liabilities	100,000	100,000
Net assets	€220,000	€270,000

Blake Co. believes the value of Brown Co. is higher than the fair value of its identifiable net assets. They offer €100,000 for a 30% interest in Brown, which represents a €34,000 excess purchase price. The difference between the fair value and book value of the net identifiable assets is €50,000 (€270,000 – 220,000). Based on Blake Co.'s 30% ownership, €15,000 of the excess purchase price is attributable to the net identifiable assets, and the residual is attributable to goodwill. Calculate goodwill.

Solution:

Purchase price	€100,000
30% of book value of Brown (30% × €220,000)	66,000
Excess purchase price	€34,000
Attributable to net assets	
Plant and equipment (30% × €30,000)	€9,000
Land (30% × €20,000)	6,000
Goodwill (residual)	19,000
	€34,000

As illustrated above, goodwill is the residual excess not allocated to identifiable assets or liabilities. The investment is carried as a non-current asset on the Blake's book as a single line item (Investment in Brown, €100,000) on the acquisition date.

5.1 Amortization of Excess Purchase Price

The excess purchase price allocated to the assets and liabilities is accounted for in a manner that is consistent with the accounting treatment for the specific asset or liability to which it is assigned. Amounts allocated to assets and liabilities that are expensed (such as inventory) or periodically depreciated or amortized (plant, property, and intangible assets) must be treated in a similar manner. These allocated amounts are not reflected on the financial statements of the investee (associate), and the investee's income statement will not reflect the necessary periodic adjustments. Therefore, the investor must directly record these adjustment effects by reducing the carrying amount of the investment on its balance sheet and by reducing the investee's profit recognized on its income statement. Amounts allocated to assets or liabilities that are not systematically amortized (e.g., land) will continue to be reported at their fair value as of the date the investment was acquired. As stated above, goodwill is included in the carrying amount of the investment instead of being separately recognized. It is not amortized because it is considered to have an indefinite life.

Using the example above and assuming a 10-year useful life for plant, property, and equipment and using straight-line depreciation, the annual amortization is as follows:

Account	Excess Price (€)	Useful Life	Amortization/Year (€)
Plant and equipment	9,000	10 years	900
Land	6,000	Indefinite	0
Goodwill	19,000	Indefinite	0

Annual amortization would reduce the investor's share of the investee's reported income (equity income) and the balance in the investment account by €900 for each year over the 10-year period.

EXAMPLE 3

Equity Method Investments with Goodwill

On 1 January 2018, Parker Company acquired 30% of Prince Inc. common shares for the cash price of €500,000 (both companies are fictitious). It is determined that Parker has the ability to exert significant influence on Prince's financial and operating decisions. The following information concerning Prince's assets and liabilities on 1 January 2018 is provided:

Prince, Inc.			
	Book Value	Fair Value	Difference
Current assets	€100,000	€100,000	€0
Plant and equipment	1,900,000	2,200,000	300,000
	€2,000,000	€2,300,000	€300,000
Liabilities	800,000	800,000	0
Net assets	€1,200,000	€1,500,000	€300,000

The plant and equipment are depreciated on a straight-line basis and have 10 years of remaining life. Prince reports net income for 2018 of €100,000 and pays dividends of €50,000. Calculate the following:

1 Goodwill included in the purchase price.

2 Investment in associate (Prince) at the end of 2018.

Solution to 1:

Purchase price	€500,000
Acquired equity in book value of Prince's net assets (30% × €1,200,000)	360,000
Excess purchase price	€140,000
Attributable to plant and equipment (30% × €300,000)	(90,000)
Goodwill (residual)	€50,000

Solution to 2:

Investment in associate

Purchase price	€500,000
Parker's share of Prince's net income (30% × €100,000)	30,000
Dividends received (30% of €50,000)	(15,000)
Amortization of excess purchase price attributable to plant and equipment (€90,000 ÷ 10 years)	(9,000)
31 December 2018 balance in investment in Prince	€506,000

An alternate way to look at the balance in the investment account is that it reflects the basic valuation principle of the equity method. At any point in time, the investment account balance equals the investor's (Parker) proportionate share of the net equity (net assets at book value) of the investee (Prince) plus the unamortized balance of the original excess purchase price. Applying this principle to this example:

2018 Beginning net assets =	€1,200,000
Plus: Net income	100,000
Less: Dividends	(50,000)
2018 Ending net assets	€1,250,000
Parker's proportionate share of Prince's recorded net assets (30% × €1,250,000)	€375,000
Unamortized excess purchase price (€140,000 − 9,000)	131,000
Investment in Prince	€506,000

Note that the unamortized excess purchase price is a cost incurred by Parker, not Prince. Therefore, the total amount is included in the investment account balance.

5.2 Fair Value Option

Both IFRS and US GAAP give the investor the option to account for their equity method investment at fair value.[14] Under US GAAP, this option is available to all entities; however, under IFRS, its use is restricted to venture capital organizations, mutual funds, unit trusts, and similar entities, including investment-linked insurance funds.

Both standards require that the election to use the fair value option occur at the time of initial recognition and is irrevocable. Subsequent to initial recognition, the investment is reported at fair value with unrealized gains and losses arising from changes in fair value as well as any interest and dividends received included in the investor's profit or loss (income). Under the fair value method, the investment account on the investor's balance sheet does not reflect the investor's proportionate share of the investee's profit or loss, dividends, or other distributions. In addition, the excess of cost over the fair value of the investee's identifiable net assets is not amortized, nor is goodwill created.

14 IFRS 9 Financial Instruments. FASB ASC Section 825-10-25 [Financial Instruments–Overall–Recognition].

5.3 Impairment

Both IFRS and US GAAP require periodic reviews of equity method investments for impairment. If the fair value of the investment is below its carrying value and this decline is deemed to be other than temporary, an impairment loss must be recognized.

Under IFRS, there must be objective evidence of impairment as a result of one or more (loss) events that occurred after the initial recognition of the investment, and that loss event has an impact on the investment's future cash flows, which can be reliably estimated. Because goodwill is included in the carrying amount of the investment and is not separately recognized, it is not separately tested for impairment. Instead, the entire carrying amount of the investment is tested for impairment by comparing its recoverable amount with its carrying amount.[15] The impairment loss is recognized on the income statement, and the carrying amount of the investment on the balance sheet is either reduced directly or through the use of an allowance account.

US GAAP takes a different approach. If the fair value of the investment declines below its carrying value *and* the decline is determined to be permanent, US GAAP[16] requires an impairment loss to be recognized on the income statement and the carrying value of the investment on the balance sheet is reduced to its fair value.

Both IFRS and US GAAP prohibit the reversal of impairment losses even if the fair value later increases.

Section 8.4 of this reading discusses impairment tests for the goodwill attributed to a controlling investment (consolidated subsidiary). Note the distinction between the disaggregated goodwill impairment test for consolidated statements and the impairment test of the total fair value of equity method investments.

TRANSACTIONS WITH ASSOCIATES AND DISCLOSURE

6

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

Because an investor company can influence the terms and timing of transactions with its associates, profits from such transactions cannot be realized until confirmed through use or sale to third parties. Accordingly, the investor company's share of any unrealized profit must be deferred by reducing the amount recorded under the equity method. In the subsequent period(s) when this deferred profit is considered confirmed, it is added to the equity income. At that time, the equity income is again based on the recorded values in the associate's accounts.

Transactions between the two affiliates may be **upstream** (associate to investor) or **downstream** (investor to associate). In an upstream sale, the profit on the intercompany transaction is recorded on the associate's income (profit or loss) statement.

---

15 Recoverable amount is the higher of "value in use" or net selling price. Value in use is equal to the present value of estimated future cash flows expected to arise from the continuing use of an asset and from its disposal at the end of its useful life. Net selling price is equal to fair value less cost to sell.

16 FASB ASC Section 323-10-35 [Investments–Equity Method and Joint Ventures–Overall–Subsequent Measurement].

The investor's share of the unrealized profit is thus included in equity income on the investor's income statement. In a downstream sale, the profit is recorded on the investor's income statement. Both IFRS and US GAAP require that the unearned profits be eliminated to the extent of the investor's interest in the associate.[17] The result is an adjustment to equity income on the investor's income statement.

EXAMPLE 4

Equity Method with Sale of Inventory: Upstream Sale

On 1 January 2018, Wicker Company acquired a 25% interest in Foxworth Company (both companies are fictitious) for €1,000,000 and used the equity method to account for its investment. The book value of Foxworth's net assets on that date was €3,800,000. An analysis of fair values revealed that all fair values of assets and liabilities were equal to book values except for a building. The building was undervalued by €40,000 and has a 20-year remaining life. The company used straight-line depreciation for the building. Foxworth paid €3,200 in dividends in 2018. During 2018, Foxworth reported net income of €20,000. During the year, Foxworth sold inventory to Wicker. At the end of the year, there was €8,000 profit from the upstream sale in Foxworth's net income. The inventory sold to Wicker by Foxworth had not been sold to an outside party.

1 Calculate the equity income to be reported as a line item on Wicker's 2018 income statement.
2 Calculate the balance in the investment in Foxworth to be reported on the 31 December 2018 balance sheet.

Purchase price	€1,000,000
Acquired equity in book value of Foxworth's net assets (25% × €3,800,000)	950,000
Excess purchase price	€50,000
Attributable to:	
Building (25% × €40,000)	€10,000
Goodwill (residual)	40,000
	€50,000

Solution to 1:

Equity Income

Wicker's share of Foxworth's reported income (25% × €20,000)	€5,000
Amortization of excess purchase price attributable to building, (€10,000 ÷ 20)	(500)
Unrealized profit (25% × €8,000)	(2,000)
Equity income 2018	€2,500

Solution to 2:

Investment in Foxworth:

17 IAS 28 Investments in Associates and Joint Ventures; FASB ASC Topic 323 [Investments–Equity Method and Joint Ventures].

Purchase price	€1,000,000
Equity income 2018	2,500
Dividends received (25% × €3,200)	(800)
Investment in Foxworth, 31 Dec 2018	€1,001,700
Composition of investment account:	
Wicker's proportionate share of Foxworth's net equity (net assets at book value) [25% × (€3,800,000 + (20,000 − 8,000) − 3,200)]	€952,200
Unamortized excess purchase price (€50,000 − 500)	49,500
	€1,001,700

EXAMPLE 5

Equity Method with Sale of Inventory: Downstream Sale

Jones Company owns 25% of Jason Company (both fictitious companies) and appropriately applies the equity method of accounting. Amortization of excess purchase price, related to undervalued assets at the time of the investment, is €8,000 per year. During 2017 Jones sold €96,000 of inventory to Jason for €160,000. Jason resold €120,000 of this inventory during 2017. The remainder was sold in 2018. Jason reports income from its operations of €800,000 in 2017 and €820,000 in 2018.

1 Calculate the equity income to be reported as a line item on Jones's 2017 income statement.

2 Calculate the equity income to be reported as a line item on Jones's 2018 income statement.

Solution to 1:

Equity Income 2017

Jones's share of Jason's reported income (25% × €800,000)	€200,000
Amortization of excess purchase price	(8,000)
Unrealized profit (25% × €16,000)	(4,000)
Equity income 2017	€188,000

Jones's profit on the sale to Jason = €160,000 − 96,000 = €64,000

Jason sells 75% (€120,000/160,000) of the goods purchased from Jones; 25% is unsold.

Total unrealized profit = €64,000 × 25% = €16,000

Jones's share of the unrealized profit = €16,000 × 25% = €4,000

Alternative approach:

Jones's profit margin on sale to Jason: 40% (€64,000/€160,000)

Jason's inventory of Jones's goods at 31 Dec 2017: €40,000

Jones's profit margin on this was 40% × 40,000 = €16,000

Jones's share of profit on unsold goods = €16,000 × 25% = €4,000

Solution to 2:

Equity Income 2018

Jones's share of Jason's reported income (25% × €820,000)	€205,000
Amortization of excess purchase price	(8,000)
Realized profit (25% × €16,000)	4,000
Equity income 2018	€201,000

Jason sells the remaining 25% of the goods purchased from Jones.

6.1 Disclosure

The notes to the financial statements are an integral part of the information necessary for investors. Both IFRS and US GAAP require disclosure about the assets, liabilities, and results of equity method investments. For example, in their 2017 annual report, within its note titled "Principles of Consolidation," Deutsche Bank reports that:

> Investments in associates are accounted for under the equity method of accounting. The Group's share of the results of associates is adjusted to conform to the accounting policies of the Group and is reported in the Consolidated Statement of Income as Net income (loss) from equity method investments. The Group's share in the associate's profits and losses resulting from intercompany sales is eliminated on consolidation.
>
> If the Group previously held an equity interest in an entity (for example, as available for sale) and subsequently gained significant influence, the previously held equity interest is remeasured to fair value and any gain or loss is recognized in the Consolidated Statement of Income. Any amounts previously recognized in other comprehensive income associated with the equity interest would be reclassified to the Consolidated Statement of Income at the date the Group gains significant influence, as if the Group had disposed of the previously held equity interest.
>
> Under the equity method of accounting, the Group's investments in associates and jointly controlled entities are initially recorded at cost including any directly related transaction costs incurred in acquiring the associate, and subsequently increased (or decreased) to reflect both the Group's pro-rata share of the post-acquisition net income (or loss) of the associate or jointly controlled entity and other movements included directly in the equity of the associate or jointly controlled entity. Goodwill arising on the acquisition of an associate or a jointly controlled entity is included in the carrying value of the investment (net of any accumulated impairment loss). As goodwill is not reported separately it is not specifically tested for impairment. Rather, the entire equity method investment is tested for impairment at each balance sheet date.

For practical reasons, associated companies' results are sometimes included in the investor's accounts with a certain time lag, normally not more than one quarter. Dividends from associated companies are not included in investor income because it would be a double counting. Applying the equity method recognizes the investor's full share of the associate's income. Dividends received involve exchanging a portion of equity interest for cash. In the consolidated balance sheet, the book value of shareholdings in associated companies is increased by the investor's share of the company's net income and reduced by amortization of surplus values and the amount of dividends received.

6.2 Issues for Analysts

Equity method accounting presents several challenges for analysis. First, analysts should question whether the equity method is appropriate. For example, an investor holding 19% of an associate may in fact exert significant influence but may attempt to avoid using the equity method to avoid reporting associate losses. On the other hand, an investor holding 25% of an associate may be unable to exert significant influence and may be unable to access cash flows, and yet may prefer the equity method to capture associate income.

Second, the investment account represents the investor's percentage ownership in the net assets of the investee company through "one-line consolidation." There can be significant assets and liabilities of the investee that are not reflected on the investor's balance sheet, which will significantly affect debt ratios. Net margin ratios could be overstated because income for the associate is included in investor net income but is not specifically included in sales. An investor may actually control the investee with less than 50% ownership but prefer the financial results using the equity method. Careful analysis can reveal financial performance driven by accounting structure.

Finally, the analyst must consider the quality of the equity method earnings. The equity method assumes that a percentage of each dollar earned by the investee company is earned by the investor (i.e., a fraction of the dollar equal to the fraction of the company owned), even if cash is not received. Analysts should, therefore, consider potential restrictions on dividend cash flows (the statement of cash flows).

BUSINESS COMBINATIONS: ACQUISITION METHOD AND IMPACT OF THE ACQUISITION METHOD ON FINANCIAL STATEMENTS POST-ACQUISITION

7

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

c analyze how different methods used to account for intercorporate investments affect financial statements and ratios.

Business combinations (controlling interest investments) involve the combination of two or more entities into a larger economic entity. Business combinations are typically motivated by expectations of added value through synergies, including potential for increased revenues, elimination of duplicate costs, tax advantages, coordination of the production process, and efficiency gains in the management of assets.[18]

18 IFRS 3, *Business Combinations*, revised in 2008 and FASB ASC Topic 805 [*Business Combinations*] provide guidance on business combinations.

Under IFRS, there is no distinction among business combinations based on the resulting structure of the larger economic entity. For all business combinations, one of the parties to the business combination is identified as the acquirer. Under US GAAP, an acquirer is identified, but the business combinations are categorized as merger, acquisition, or consolidation based on the legal structure after the combination. Each of these types of business combinations has distinctive characteristics that are described in Exhibit 5. Features of variable interest and special purpose entities are also described in Exhibit 5 because these are additional instances where control is exerted by another entity. Under both IFRS and US GAAP, business combinations are accounted for using the *acquisition method*.

Exhibit 5 Types of Business Combinations

Merger

The distinctive feature of a merger is that only one of the entities remains in existence. One hundred percent of the target is absorbed into the acquiring company. Company A may issue common stock, preferred stock, bonds, or pay cash to acquire the net assets. The net assets of Company B are transferred to Company A. Company B ceases to exist and Company A is the only entity that remains.

Company A + Company B = Company A

Acquisition

The distinctive feature of an acquisition is the legal continuity of the entities. Each entity continues operations but is connected through a parent–subsidiary relationship. Each entity is an individual that maintains separate financial records, but the parent (the acquirer) provides consolidated financial statements in each reporting period. Unlike a merger or consolidation, the acquiring company does not need to acquire 100% of the target. In fact, in some cases, it may acquire less than 50% and still exert control. If the acquiring company acquires less than 100%, non-controlling (minority) shareholders' interests are reported on the consolidated financial statements.

Company A + Company B = (Company A + Company B)

Consolidation

The distinctive feature of a consolidation is that a new legal entity is formed and none of the predecessor entities remain in existence. A new entity is created to take over the net assets of Company A and Company B. Company A and Company B cease to exist and Company C is the only entity that remains.

Company A + Company B = Company C

Special Purpose or Variable Interest Entities

The distinctive feature of a special purpose (variable interest) entity is that control is not usually based on voting control, because equity investors do not have a sufficient amount at risk for the entity to finance its activities without additional subordinated financial support. Furthermore, the equity investors may lack a controlling financial interest. The sponsoring company usually creates a special purpose entity (SPE) for a narrowly defined purpose. IFRS require consolidation if the substance of the relationship indicates control by the sponsor.

Under IFRS 10, *Consolidated Financial Statements* and SIC-12, *Consolidation-Special Purpose Entities*, the definition of control extends to a broad range of activities. The control concept requires judgment and evaluation of relevant factors to determine whether control exists. Control is present when 1) the investor has the ability to exert influence on the financial and operating policy of the entity; and 2) is exposed, or has rights, to variable returns from its involvement with the investee. Consolidation criteria apply to all entities that meet the definition of control.

US GAAP uses a two-component consolidation model that includes both a variable interest component and a voting interest (control) component. Under the variable interest component, US GAAP[19] requires the primary beneficiary of a variable interest entity (VIE) to consolidate the VIE regardless of its voting interests (if any) in the VIE or its decision-making authority. The primary beneficiary is defined as the party that will absorb the majority of the VIE's expected losses, receive the majority of the VIE's expected residual returns, or both.

In the past, business combinations could be accounted for either as a purchase transaction or as a uniting (or pooling) of interests. However, the use of the pooling accounting method for acquisitions is no longer permitted, and IFRS and US GAAP now require that all business combinations be accounted for in a similar manner. The *acquisition method* developed by the IASB and the FASB replaces the purchase method, and substantially reduces any differences between IFRS and US GAAP for business combinations.[20]

7.1 Acquisition Method

IFRS and US GAAP require the acquisition method of accounting for business combinations, although both have a few specific exemptions.

Under this approach, the fair value of the consideration given by the acquiring company is the appropriate measurement for acquisitions and also includes the acquisition-date fair value of any contingent consideration. Direct costs of the business combination, such as professional and legal fees, valuation experts, and consultants, are expensed as incurred.

The acquisition method (which replaced the purchase method) addresses three major accounting issues that often arise in business combinations and the preparation of consolidated (combined) financial statements:

- The recognition and measurement of the assets and liabilities of the combined entity;
- The initial recognition and subsequent accounting for goodwill; and
- The recognition and measurement of any non-controlling interest.

7.1.1 *Recognition and Measurement of Identifiable Assets and Liabilities*

IFRS and US GAAP require that the acquirer measure the identifiable tangible and intangible assets and liabilities of the acquiree (acquired entity) at fair value as of the date of the acquisition. The acquirer must also recognize any assets and liabilities that the acquiree had not previously recognized as assets and liabilities in its financial statements. For example, identifiable intangible assets (for example, brand names, patents, technology) that the acquiree developed internally would be recognized by the acquirer.

19 FASB ASC Topic 810 [Consolidation].
20 IFRS 10, Consolidated Financial Statements; IFRS 3, Business Combinations; FASB ASC Topic 805 [Business Combinations]; FASB ASC Topic 810 [Consolidations].

7.1.2 *Recognition and Measurement of Contingent Liabilities*[21]

On the acquisition date, the acquirer must recognize any contingent liability assumed in the acquisition if 1) it is a present obligation that arises from past events, and 2) it can be measured reliably. Costs that the acquirer expects (but is not obliged) to incur, however, are not recognized as liabilities as of the acquisition date. Instead, the acquirer recognizes these costs in future periods as they are incurred. For example, expected restructuring costs arising from exiting an acquiree's business will be recognized in the period in which they are incurred.

There is a difference between IFRS and US GAAP with regard to treatment of contingent liabilities. IFRS include contingent liabilities if their fair values can be reliably measured. US GAAP includes only those contingent liabilities that are probable and can be reasonably estimated.

7.1.3 *Recognition and Measurement of Indemnification Assets*

On the acquisition date, the acquirer must recognize an indemnification asset if the seller (acquiree) contractually indemnifies the acquirer for the outcome of a contingency or an uncertainty related to all or part of a specific asset or liability of the acquiree. The seller may also indemnify the acquirer against losses above a specified amount on a liability arising from a particular contingency. For example, the seller guarantees that an acquired contingent liability will not exceed a specified amount. In this situation, the acquirer recognizes an indemnification asset at the same time it recognizes the indemnified liability, with both measured on the same basis. If the indemnification relates to an asset or a liability that is recognized at the acquisition date and measured at its acquisition date fair value, the acquirer will also recognize the indemnification asset at the acquisition date at its acquisition date fair value.

7.1.4 *Recognition and Measurement of Financial Assets and Liabilities*

At the acquisition date, identifiable assets and liabilities acquired are classified in accordance with IFRS (or US GAAP) standards. The acquirer reclassifies the financial assets and liabilities of the acquiree based on the contractual terms, economic conditions, and the acquirer's operating or accounting policies, as they exist at the acquisition date.

7.1.5 *Recognition and Measurement of Goodwill*

IFRS allows two options for recognizing goodwill at the transaction date. The goodwill option is on a transaction-by-transaction basis. "Partial goodwill" is measured as the fair value of the acquisition (fair value of consideration given) less the acquirer's share of the fair value of all identifiable tangible and intangible assets, liabilities, and contingent liabilities acquired. "Full goodwill" is measured as the fair value of the entity as a whole less the fair value of all identifiable tangible and intangible assets, liabilities, and contingent liabilities. US GAAP views the entity as a whole and requires full goodwill.[22]

Because goodwill is considered to have an indefinite life, it is not amortized. Instead, it is tested for impairment annually or more frequently if events or circumstances indicate that goodwill might be impaired.

---

21 A contingent liability must be recognized even if it is not probable that an outflow of resources or economic benefits will be used to settle the obligation.

22 FASB ASC Topic 805 [Business Combinations].

EXAMPLE 6

Recognition and Measurement of Goodwill

Acquirer contributes $800,000 for an 80% interest in Acquiree. The identifiable net assets have a fair value of $900,000. The fair value of the entire entity is determined to be $1 million.

	IFRS Partial Goodwill
Fair value of consideration	$800,000
80% of Fair value of identifiable net assets	720,000
Goodwill recognized	$80,000

	IFRS and US GAAP Full Goodwill
Fair value of entity	$1,000,000
Fair value of identifiable assets	900,000
Goodwill recognized	$100,000

7.1.6 *Recognition and Measurement when Acquisition Price Is Less than Fair Value*

Occasionally, a company faces adverse circumstances such that its market value drops below the fair value of its net assets. In an acquisition of such a company, where the purchase price is less than the fair value of the target's (acquiree's) net assets, the acquisition is considered to be a "bargain purchase" acquisition. IFRS and US GAAP require the difference between the fair value of the acquired net assets and the purchase price to be recognized immediately as a gain in profit or loss. Any contingent consideration must be measured and recognized at fair value at the time of the business combination. Any subsequent changes in value of the contingent consideration are recognized in profit or loss.

7.2 Impact of the Acquisition Method on Financial Statements, Post-Acquisition

Example 7 shows the consolidated balance sheet of an acquiring company after the acquisition.

EXAMPLE 7

Acquisition Method Post-Combination Balance Sheet

Franklin Company, a hypothetical company, acquired 100% of the outstanding shares of Jefferson, Inc. (another fictitious company) by issuing 1,000,000 shares of its €1 par common stock (€15 market value). Immediately before the transaction, the two companies compiled the following information:

	Franklin Book Value (000)	Jefferson Book Value (000)	Jefferson Fair Value (000)
Cash and receivables	€10,000	€300	€300
Inventory	12,000	1,700	3,000
PP&E (net)	27,000	2,500	4,500
	€49,000	€4,500	€7,800
Current payables	8,000	600	600
Long-term debt	16,000	2,000	1,800
	24,000	2,600	2,400
Net assets	€25,000	€1,900	€5,400
Shareholders' equity:			
Capital stock (€1 par)	€5,000	€400	
Additional paid in capital	6,000	700	
Retained earnings	€14,000	€800	

Jefferson has no identifiable intangible assets. Show the balances in the post-combination balance sheet using the acquisition method.

Solution:

Under the acquisition method, the purchase price allocation would be as follows:

Fair value of the stock issued	
(1,000,000 shares at market value of €15)	€15,000,000
Book value of Jefferson's net assets	1,900,000
Excess purchase price	€13,100,000
Fair value of the stock issued	€15,000,000
Fair value allocated to identifiable net assets	5,400,000
Goodwill	€9,600,000

Allocation of excess purchase price (based on the differences between fair values and book values):

Inventory	€1,300,000
PP&E (net)	2,000,000
Long-term debt	200,000
Goodwill	9,600,000
	€13,100,000

Both IFRS and US GAAP record the fair value of the acquisition at the market value of the stock issued, or €15,000,000. In this case, the purchase price exceeds the book value of Jefferson's net assets by €13,100,000. Inventory, PP&E (net), and long-term debt are adjusted to fair values. The excess of the purchase price over the fair value of identifiable net assets results in goodwill recognition of €9,600,000.

The post-combination balance sheet of the combined entity would appear as follows:

Franklin Consolidated Balance Sheet (Acquisition Method) (000)	
Cash and receivables	€10,300
Inventory	15,000
PP&E (net)	31,500
Goodwill	9,600
Total assets	€66,400
Current payables	€8,600
Long-term debt	17,800
Total liabilities	€26,400
Capital stock (€1 par)	€6,000
Additional paid in capital	20,000
Retained earnings	14,000
Total stockholders' equity	€40,000
Total liabilities and stockholders' equity	€66,400

Assets and liabilities are combined using book values of Franklin plus fair values for the assets and liabilities acquired from Jefferson. For example, the book value of Franklin's inventory (€12,000,000) is added to the fair value of inventory acquired from Jefferson (€3,000,000) for a combined inventory of €15,000,000. Long-term debt has a book value of €16,000,000 on Franklin's pre-acquisition statements, and Jefferson's fair value of debt is €1,800,000. The combined long-term debt is recorded as €17,800,000.

Franklin's post-merger financial statement reflects in stockholders' equity the stock issued by Franklin to acquire Jefferson. Franklin issues stock with a par value of €1,000,000; however, the stock is measured at fair value under both IFRS and US GAAP. Therefore, the consideration exchanged is 1,000,000 shares at market value of €15, or €15,000,000. Prior to the transaction, Franklin had 5,000,000 shares of €1 par stock outstanding (€5,000,000). The combined entity reflects the Franklin capital stock outstanding of €6,000,000 (€5,000,000 plus the additional 1,000,000 shares of €1 par stock issued to effect the transaction). Franklin's additional paid in capital of €6,000,000 is increased by the €14,000,000 additional paid in capital from the issuance of the 1,000,000 shares (€15,000,000 less par value of €1,000,000) for a total of €20,000,000. At the acquisition date, only the acquirer's retained earnings are carried to the combined entity. Earnings of the target are included on the consolidated income statement and retained earnings only in post-acquisition periods.

In the periods subsequent to the business combination, the financial statements continue to be affected by the acquisition method. Net income reflects the performance of the combined entity. Under the acquisition method, amortization/depreciation is based on historical cost of Franklin's assets and the fair value of Jefferson's assets. Using Example 7, as Jefferson's acquired inventory is sold, the cost of goods sold would be €1,300,000 higher and depreciation on PP&E would be €2,000,000 higher over the life of the asset than if the companies had not combined.

8 THE CONSOLIDATION PROCESS

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

Consolidated financial statements combine the separate financial statements for distinct legal entities, the parent and its subsidiaries, as if they were one economic unit. Consolidation combines the assets, liabilities, revenues, and expenses of subsidiaries with the parent company. Transactions between the parent and subsidiary (intercompany transactions) are eliminated to avoid double counting and premature income recognition. Consolidated statements are presumed to be more meaningful in terms of representational faithfulness. It is important for the analyst to consider the differences in IFRS and US GAAP, valuation bases, and other factors that could impair the validity of comparative analyses.

8.1 Business Combination with Less than 100% Acquisition

The acquirer purchases 100% of the equity of the target company in a transaction structured as a merger or consolidation. For a transaction structured as an acquisition, however, the acquirer does not have to purchase 100% of the equity of the target in order to achieve control. The acquiring company may purchase less than 100% of the target because it may be constrained by resources or it may be unable to acquire all the outstanding shares. As a result, both the acquirer and the target remain separate legal entities. Both IFRS and US GAAP presume a company has control if it owns more than 50% of the voting shares of an entity. In this case, the acquiring company is viewed as the parent, and the target company is viewed as the subsidiary. Both the parent and the subsidiary typically prepare their own financial records, but the parent also prepares consolidated financial statements at each reporting period. The consolidated financial statements are the primary source of information for investors and analysts.

8.2 Non-controlling (Minority) Interests: Balance Sheet

A non-controlling (minority) interest is the portion of the subsidiary's equity (residual interest) that is held by third parties (i.e., not owned by the parent). Non-controlling interests are created when the parent acquires less than a 100% controlling interest in a subsidiary. IFRS and US GAAP have similar treatment for how non-controlling interests are classified.[23] Non-controlling interests in consolidated subsidiaries are presented on the consolidated balance sheet as a separate component of stockholders' equity. IFRS and US GAAP differ, however, on the measurement of non-controlling interests. Under IFRS, the parent can measure the non-controlling interest at either its fair value (full goodwill method) or at the non-controlling interest's proportionate

23 IFRS 10, Consolidated Financial Statements and FASB ASC Topic 810 [Consolidation].

share of the acquiree's identifiable net assets (partial goodwill method). Under US GAAP, the parent must use the full goodwill method and measure the non-controlling interest at fair value.

Example 8 illustrates the differences in reporting requirements.

EXAMPLE 8

Non-controlling Asset Valuation

On 1 January 2018, the hypothetical Parent Co. acquired 90% of the outstanding shares of the hypothetical Subsidiary Co. in exchange for shares of Parent Co.'s no par common stock with a fair value of €180,000. The fair market value of the subsidiary's shares on the date of the exchange was €200,000. Below is selected financial information from the two companies immediately prior to the exchange of shares (before the parent recorded the acquisition):

	Parent Book	Subsidiary	
	Value	Book Value	Fair Value
Cash and receivables	€40,000	€15,000	€15,000
Inventory	125,000	80,000	80,000
PP&E (net)	235,000	95,000	155,000
	€400,000	€190,000	€250,000
Payables	55,000	20,000	20,000
Long-term debt	120,000	70,000	70,000
	175,000	90,000	90,000
Net assets	€225,000	€100,000	€160,000
Shareholders' equity:			
Capital stock (no par)	€87,000	€34,000	
Retained earnings	€138,000	€66,000	

1. Calculate the value of PP&E (net) on the consolidated balance sheet under both IFRS and US GAAP.
2. Calculate the value of goodwill and the value of the non-controlling interest at the acquisition date under the full goodwill method.
3. Calculate the value of goodwill and the value of the non-controlling interest at the acquisition date under the partial goodwill method.

Solution to 1:

Relative to fair value, the PP&E of the subsidiary is understated by €60,000. Under the acquisition method (IFRS and US GAAP), as long as the parent has control over the subsidiary (i.e., regardless of whether the parent had purchased 51% or 100% of the subsidiary's stock), it would include 100% of the subsidiary's assets and liabilities at fair value on the consolidated balance sheet. Therefore, PP&E on the consolidated balance sheet would be valued at €390,000.

Solution to 2:

Under the full goodwill method (mandatory under US GAAP and optional under IFRS), goodwill on the consolidated balance sheet would be the difference between the total fair value of the subsidiary and the fair value of the subsidiary's identifiable net assets.

Fair value of the subsidiary	€200,000
Fair value of subsidiary's identifiable net assets	160,000
Goodwill	€40,000

The value of the non-controlling interest is equal to the non-controlling interest's proportionate share of the subsidiary's fair value. The non-controlling interest's proportionate share of the subsidiary is 10% and the fair value of the subsidiary is €200,000 on the acquisition date. Under the full goodwill method, the value of the non-controlling interest would be €20,000 (10% × €200,000).

Solution to 3:

Under the partial goodwill method (IFRS only), goodwill on the parent's consolidated balance sheet would be €36,000, the difference between the purchase price and the parent's proportionate share of the subsidiary's identifiable assets.

Acquisition price	€180,000
90% of fair value	144,000
Goodwill	€36,000

The value of the non-controlling interest is equal to the non-controlling interest's proportionate share of the fair value of the subsidiary's identifiable net assets. The non-controlling interest's proportionate share is 10%, and the fair value of the subsidiary's identifiable net assets on the acquisition date is €160,000. Under the partial goodwill method, the value of the non-controlling interest would be €16,000 (10% × €160,000).

Regardless of which method is used, goodwill is not amortized under either IFRS or US GAAP but it is tested for impairment at least annually.

For comparative purposes, below is the balance sheet at the acquisition date under the full goodwill and partial goodwill methods.

Comparative Consolidated Balance Sheet at Acquisition Date: Acquisition Method

	Full Goodwill	Partial Goodwill
Cash and receivables	€55,000	€55,000
Inventory	205,000	205,000
PP&E (net)	390,000	390,000
Goodwill	40,000	36,000
Total assets	€690,000	€686,000
Payables	€75,000	€75,000
Long-term debt	190,000	190,000
Total liabilities	€265,000	€265,000
Shareholders' equity:		
Noncontrolling interests	€20,000	€16,000
Capital stock (no par)	€267,000	€267,000
Retained earnings	138,000	138,000
Total equity	€425,000	€421,000

(Continued)

	Full Goodwill	Partial Goodwill
Total liabilities and shareholders' equity	€690,000	€686,000

8.3 Non-controlling (Minority) Interests: Income Statement

On the income statement, non-controlling (minority) interests are presented as a line item reflecting the allocation of profit or loss for the period. Intercompany transactions, if any, are eliminated in full.

Using assumed data consistent with the facts in Example 8, the amounts included for the subsidiary in the consolidated income statements under IFRS and US GAAP are presented below:

	Full Goodwill	Partial Goodwill
Sales	€250,000	€250,000
Cost of goods sold	137,500	137,500
Interest expense	10,000	10,000
Depreciation expense	39,000	39,000
Income from continuing operations	€63,500	€63,500
Non-controlling interest (10%)	(6,350)	(6,350)
Consolidated net income to parent's shareholders	€57,150	€57,150

Income to the parent's shareholders is €57,150 using either method. This is because the fair value of the PP&E is allocated to non-controlling shareholders as well as to the controlling shareholders under the full goodwill and the partial goodwill methods. Therefore, the non-controlling interests will share in the adjustment for excess depreciation resulting from the €60,000 increase in PP&E. Because depreciation expense is the same under both methods, it results in identical net income to all shareholders, whichever method is used to recognize goodwill and to measure the non-controlling interest.

Although net income to parent's shareholders is the same, the impact on ratios would be different because total assets and stockholders' equity would differ.

Impact on Ratios

	Full Goodwill (%)	Partial Goodwill (%)
Return on assets	8.28	8.33
Return on equity	13.45	13.57

Over time, the value of the subsidiary will change as a result of net income and changes in equity. As a result, the value of the non-controlling interest on the parent's consolidated balance sheet will also change.

8.4 Goodwill Impairment

Although goodwill is not amortized, it must be tested for impairment at least annually or more frequently if events or changes in circumstances indicate that it might be impaired. If it is probable that some or all of the goodwill will not be recovered through the profitable operations of the combined entity, it should be partially or fully written off by charging it to an expense. Once written down, goodwill cannot be later restored.

IFRS and US GAAP differ on the definition of the levels at which goodwill is assigned and how goodwill is tested for impairment.

Under IFRS, at the time of acquisition, the total amount of goodwill recognized is allocated to each of the acquirer's cash-generating units that will benefit from the expected synergies resulting from the combination with the target. A cash-generating unit represents the lowest level within the combined entity at which goodwill is monitored for impairment purposes.[24] Goodwill impairment testing is then conducted under a one-step approach. The recoverable amount of a cash-generating unit is calculated and compared with the carrying value of the cash-generating unit.[25] An impairment loss is recognized if the recoverable amount of the cash-generating unit is less than its carrying value. The impairment loss (the difference between these two amounts) is first applied to the goodwill that has been allocated to the cash-generating unit. Once this has been reduced to zero, the remaining amount of the loss is then allocated to all of the other non-cash assets in the unit on a pro rata basis.

Under US GAAP, at the time of acquisition, the total amount of goodwill recognized is allocated to each of the acquirer's reporting units. A reporting unit is an operating segment or component of an operating segment that is one level below the operating segment as a whole. Goodwill impairment testing is then conducted under a two-step approach: identification of impairment and then measurement of the loss. First, the carrying amount of the reporting unit (including goodwill) is compared to its fair value. If the carrying value of the reporting unit exceeds its fair value, potential impairment has been identified. The second step is then performed to measure the amount of the impairment loss. The amount of the impairment loss is the difference between the implied fair value of the reporting unit's goodwill and its carrying amount. The implied fair value of goodwill is determined in the same manner as in a business combination (it is the difference between the fair value of the reporting unit and the fair value of the reporting unit's assets and liabilities). The impairment loss is applied to the goodwill that has been allocated to the reporting unit. After the goodwill of the reporting unit has been eliminated, no other adjustments are made automatically to the carrying values of any of the reporting unit's other assets or liabilities. However, it may be prudent to test other asset values for recoverability and possible impairment.

Under both IFRS and US GAAP, the impairment loss is recorded as a separate line item in the consolidated income statement.

---

24 A cash-generating unit is the smallest identifiable group of assets that generates cash inflows that are largely independent of the cash inflows from other assets or groups of assets.

25 The recoverable amount of a cash-generating unit is the higher of net selling price (i.e., fair value less costs to sell) and its value in use. Value in use is the present value of the future cash flows expected to be derived from the cash-generating unit. The carrying value of a cash-generating unit is equal to the carrying value of the unit's assets and liabilities including the goodwill that has been allocated to that unit.

EXAMPLE 9

Goodwill Impairment: IFRS

The cash-generating unit of a French company has a carrying value of €1,400,000, which includes €300,000 of allocated goodwill. The recoverable amount of the cash-generating unit is determined to be €1,300,000, and the estimated fair value of its identifiable net assets is €1,200,000. Calculate the impairment loss.

Solution:

Recoverable amount of unit	€1,300,000
Carrying amount of unit	1,400,000
Impairment loss	€100,000

The impairment loss of €100,000 is reported on the income statement, and the goodwill allocated to the cash-generating unit would be reduced by €100,000 to €200,000.

If the recoverable amount of the cash-generating unit had been €800,000 instead of €1,300,000, the impairment loss recognized would be €600,000. This would first be absorbed by the goodwill allocated to the unit (€300,000). Once this has been reduced to zero, the remaining amount of the impairment loss (€300,000) would then be allocated on a pro rata basis to the other non-cash assets within the unit.

EXAMPLE 10

Goodwill Impairment: US GAAP

A reporting unit of a US corporation (e.g., a division) has a fair value of $1,300,000 and a carrying value of $1,400,000 that includes recorded goodwill of $300,000. The estimated fair value of the identifiable net assets of the reporting unit at the impairment test date is $1,200,000. Calculate the impairment loss.

Solution:

Step 1 – Determination of an Impairment Loss

Because the fair value of the reporting unit is less than its carrying book value, a potential impairment loss has been identified.

Fair value of unit: $1,300,000 < $1,400,000

Step 2 – Measurement of the Impairment Loss

Fair value of reporting unit	$1,300,000
Less: net assets	1,200,000
Implied goodwill	$100,000

Current carrying value of goodwill	$300,000
Less: implied goodwill	100,000
Impairment loss	$200,000

The impairment loss of \$200,000 is reported on the income statement, and the goodwill allocated to the reporting unit would be reduced by \$200,000 to \$100,000.

If the fair value of the reporting unit was \$800,000 (instead of \$1,300,000), the implied goodwill would be a negative \$400,000. In this case, the maximum amount of the impairment loss recognized would be \$300,000, the carrying amount of goodwill.

9 FINANCIAL STATEMENT PRESENTATION SUBSEQUENT TO THE BUSINESS COMBINATION

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

The presentation of consolidated financial statements is similar under IFRS and US GAAP. For example, selected financial statements for GlaxoSmithKline are shown in Exhibits 6 and 7. GlaxoSmithKline is a leading pharmaceutical company headquartered in the United Kingdom.

The consolidated balance sheet in Exhibit 6 combines the operations of GlaxoSmithKline and its subsidiaries. The analyst can observe that in 2017 GlaxoSmithKline had investments in financial assets (other investments of £918,000,000 and liquid investments of £78,000,000), and investments in associates and joint ventures of £183,000,000. In 2017 GlaxoSmithKline did not acquire any additional companies, however, it made a number of small business disposals during the year for a net cash consideration of £342,000,000, including contingent consideration receivable of £86,000,000. In addition, during 2017 GlaxoSmithKline made cash investment of £15,000,000 in Associates and disposed of two associated for a cash consideration of £198,000,000.[26] The decrease in goodwill on the balance sheet reflects exchange adjustments recognized by GlaxoSmithKline due to the weakness of the functional currency of the parent (Pound Sterling). Note that GlaxoSmithKline has £6,172,000 in contingent consideration liabilities, which relate to future events such as development milestones or sales performance for acquired companies. Of the £6 billion total contingent liability, £1,076,000 is expected to be paid within one year in respect of the Novartis Vaccines business, which reached its sales milestone. The remaining contingent consideration relates to the acquisition of the Shionogi-ViiV Healthcare joint venture and Novartis Vaccines are expected to be paid over a number of years.[27] The analyst can also note that GlaxoSmithKline is the parent company in a less than 100% acquisition. The minority interest of £3,557,000,000 in the equity section is the portion of the combined entity that accrues to non-controlling shareholders.

26 Note 38: Acquisitions and Disposals, GlaxoSmithKline financial statements 2017

27 The notes state that the amount included in the balance sheet is the present value of the expected contingent consideration payments, which have been discounted using a rate of 8.5%.

Exhibit 6 GlaxoSmithKline Consolidated Balance Sheet at 31 December 2017

	Notes	2017 £m	2016 £m
Non-current assets			
Property, plant and equipment	17	**10,860**	10,808
Goodwill	18	**5,734**	5,965
Other intangible assets	19	**17,562**	18,776
Investments in associates and joint ventures	20	**183**	263
Other investments	21	**918**	985
Deferred tax assets	14	**3,796**	4,374
Derivative financial instruments	42	**8**	—
Other non-current assets	22	**1,413**	1,199
Total non-current assets		**40,474**	42,370
Current assets			
Inventories	23	**5,557**	5,102
Current tax recoverable	14	**258**	226
Trade and other receivables	24	**6,000**	6,026
Derivative financial instruments	42	**68**	156
Liquid investments	31	**78**	89
Cash and cash equivalents	25	**3,833**	4,897
Assets held for sale	26	**113**	215
Total current assets		**15,907**	16,711
Total assets		**56,381**	59,081
Current liabilities			
Short-term borrowings	31	**(2,825)**	(4,129)
Contingent consideration liabilities	39	**(1,076)**	(561)
Trade and other payables	27	**(20,970)**	(11,964)
Derivative financial instruments	42	**(74)**	(194)
Current tax payable	14	**(995)**	(1,305)
Short-term provisions	29	**(629)**	(848)
Total current liabilities		**(26,569)**	(19,001)
Non-current liabilities			
Long-term borrowings	31	**(14,264)**	(14,661)
Corporation tax payable	14	**(411)**	—
Deferred tax liabilities	14	**(1,396)**	(1,934)
Pensions and other post-employment benefits	28	**(3,539)**	(4,090)
Other provisions	29	**(636)**	(652)
Contingent consideration liabilities	39	**(5,096)**	(5,335)
Other non-current liabilities	30	**(981)**	(8,445)
Total non-current liabilities		**(26,323)**	(35,117)
Total liabilities		**(52,892)**	(54,118)
Net assets		**3,489**	4,963
Equity			

(continued)

Exhibit 6 (Continued)

	Notes	2017 £m	2016 £m
Share capital	33	**1,343**	1,342
Share premium account	33	**3,019**	2,954
Retained earnings	34	**(6,477)**	(5,392)
Other reserves	34	**2,047**	2,220
Shareholders' equity		**(68)**	1,124
Non-controlling interests		**3,557**	3,839
Total equity		**3,489**	4,963

The consolidated income statement for GlaxoSmithKline is presented in Exhibit 7. IFRS and US GAAP have similar formats for consolidated income statements. Each line item (e.g., turnover [sales], cost of sales, etc.) includes 100% of the parent and the subsidiary transactions after eliminating any **upstream** (subsidiary sells to parent) or **downstream** (parent sells to subsidiary) intercompany transactions. The portion of income accruing to non-controlling shareholders is presented as a separate line item on the consolidated income statement. Note that net income would be the same under IFRS and US GAAP.[28] The analyst will need to make adjustments for any analysis comparing specific line items that might differ between IFRS and US GAAP.

Exhibit 7 GlaxoSmithKline Consolidated Income Statement for the Year Ended 31 December 2017

		2017	2016	2015
	Notes	Total £m	£m	£m
Turnover	6	**30,186**	27,889	23,923
Cost of sales		**(10,342)**	(9,290)	(8,853)
Gross profit		**19,844**	18,599	15,070
Selling, general and administration		**(9,672)**	(9,366)	(9,232)
Research and development		**(4,476)**	(3,628)	(3,560)
Royalty income		**356**	398	329
Other operating income	7	**(1,965)**	(3,405)	7,715
Operating profit	8	**4,087**	2,598	10,322
Finance income	11	**65**	72	104
Finance costs	12	**(734)**	(736)	(757)
Profit on disposal of interests in Associates		**95**	—	843
Share of after tax profits of associates and joint ventures	13	**13**	5	14
Profit before taxation		**3,525**	1,939	10,526

28 It is possible, however, for differences to arise through the application of different accounting rules (e.g., valuation of fixed assets).

Exhibit 7 (Continued)

	Notes	2017 Total £m	2016 £m	2015 £m
Taxation	14	**(1,356)**	(877)	(2,154)
Profit after taxation for the year		**2,169**	1,062	8,372
Profit/(loss) attributable to non-controlling interests		**637**	150	(50)
Profit attributable to shareholders		**1,532**	912	8,472
		2,169	1,062	8,372
Basic earnings per share (pence)	15	**31.4p**	18.8p	174.3p
Diluted earnings per share (pence)	15	**31.0p**	18.6p	172.3p

VARIABLE INTEREST AND SPECIAL PURPOSE ENTITIES 10

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

Special purpose entities (SPEs) are enterprises that are created to accommodate specific needs of the sponsoring entity.[29] The sponsoring entity (on whose behalf the SPE is created) frequently transfers assets to the SPE, obtains the right to use assets held by the SPE, or performs services for the SPE, while other parties (capital providers) provide funding to the SPE. SPEs can be a legitimate financing mechanism for a company to segregate certain activities and thereby reduce risk. SPEs may take the form of a limited liability company (corporation), trust, partnership, or unincorporated entity. They are often created with legal arrangements that impose strict and sometimes permanent limits on the decision-making powers of their governing board or management.

Beneficial interest in an SPE may take the form of a debt instrument, an equity instrument, a participation right, or a residual interest in a lease. Some beneficial interests may simply provide the holder with a fixed or stated rate of return, while beneficial interests give the holder the rights or the access to future economic benefits of the SPE's activities. In most cases, the creator/sponsor of the entity retains a significant beneficial interest in the SPE even though it may own little or none of the SPE's voting equity.

29 The term "special purpose entity" is used by IFRS and "variable interest entity" and "special purpose entity" is used by US GAAP.

In the past, sponsors were able to avoid consolidating SPEs on their financial statements because they did not have "control" (i.e., own a majority of the voting interest) of the SPE. SPEs were structured so that the sponsoring company had financial control over their assets or operating activities, while third parties held the majority of the voting interest in the SPE.

These outside equity participants often funded their investments in the SPE with debt that was either directly or indirectly guaranteed by the sponsoring companies. The sponsoring companies, in turn, were able to avoid the disclosure of many of these guarantees as well as their economic significance. In addition, many sponsoring companies created SPEs to facilitate the transfer of assets and liabilities from their own balance sheets. As a result, they were able to recognize large amounts of revenue and gains, because these transactions were accounted for as sales. By avoiding consolidation, sponsoring companies did not have to report the assets and the liabilities of the SPE; financial performance as measured by the unconsolidated financial statements was potentially misleading. The benefit to the sponsoring company was improved asset turnover, lower operating and financial leverage metrics, and higher profitability.

Enron, for example, used SPEs to obtain off-balance sheet financing and artificially improve its financial performance. Its subsequent collapse was partly attributable to its guarantee of the debt of the SPEs it had created.

To address the accounting issues arising from the misuse and abuse of SPEs, the IASB and the FASB worked to improve the consolidation models to take into account financial arrangements where parties other than the holders of the majority of the voting interests exercise financial control over another entity. IFRS 10, *Consolidated Financial Statements*, revised the definition of control to encompass many special purpose entities. Special purpose entities involved in a structured financial transaction will require an evaluation of the purpose, design, and risks.

In developing new accounting standards to address this consolidation issue, the FASB used the more general term variable interest entity (VIE) to more broadly define an entity that is financially controlled by one or more parties that do not hold a majority voting interest. Therefore, under US GAAP, a VIE includes other entities besides SPEs. FASB ASC Topic 810 [*Consolidation*] provides guidance for US GAAP, which classifies special purpose entities as variable interest entities if:

1 total equity at risk is insufficient to finance activities without financial support from other parties, or

2 equity investors lack any one of the following:

   a the ability to make decisions;

   b the obligation to absorb losses; or

   c the right to receive returns.

Common examples of variable interests are entities created to lease real estate or other property, entities created for the securitization of financial assets, or entities created for research and development activity.

Under FASB ASC Topic 810 [*Consolidation*], the primary beneficiary of a VIE must consolidate it as a subsidiary regardless of how much of an equity investment the beneficiary has in the VIE. The primary beneficiary (which is often the sponsor) is the entity that is expected to absorb the majority of the VIE's expected losses, receive the majority of the VIE's residual returns, or both. If one entity will absorb a majority of the VIE's expected losses and another unrelated entity will receive a majority of the VIE's expected residual returns, the entity absorbing a majority of the losses must consolidate the VIE. If there are non-controlling interests in the VIE, these would also be shown in the consolidated balance sheet and consolidated income statement of the primary beneficiary. ASC Topic 810 also requires entities to disclose information about their relationships with VIEs, even if they are not considered the primary beneficiary.

10.1 Securitization of Assets

Example 11 shows the effects of securitizing assets on companies' balance sheets.

EXAMPLE 11

Receivables Securitization

Odena, a (fictional) Italian auto manufacturer, wants to raise €55M in capital by borrowing against its financial receivables. To accomplish this objective, Odena can choose between two alternatives:

Alternative 1 Borrow directly against the receivables; or

Alternative 2 Create a special purpose entity, invest €5M in the SPE, have the SPE borrow €55M, and then use the funds to purchase €60M of receivables from Odena.

Using the financial statement information provided below, describe the effect of each Alternative on Odena, assuming that Odena meets the definition of control and will consolidate the SPE.

Odena Balance Sheet	
Cash	€30,000,000
Accounts receivable	60,000,000
Other assets	40,000,000
Total assets	€130,000,000
Current liabilities	€27,000,000
Noncurrent liabilities	20,000,000
Total liabilities	€47,000,000
Shareholder equity	€83,000,000
Total liabilities and equity	€130,000,000

Alternative 1:

Odena's cash will increase by €55M (to €85M) and its debt will increase by €55M (to €75M). Its sales and net income will not change.

Odena: Alternative 1 Balance Sheet	
Cash	€85,000,000
Accounts receivable	60,000,000
Other assets	40,000,000
Total assets	€185,000,000
Current liabilities	€27,000,000
Noncurrent liabilities	75,000,000
Total liabilities	€102,000,000

(continued)

(Continued)	
Shareholder equity	€83,000,000
Total liabilities and equity	€185,000,000

Alternative 2:

Odena's accounts receivable will decrease by €60M and its cash will increase by €55 (it invests €5M in cash in the SPE). However, if Odena is able to sell the receivables to the SPE for more than their carrying value (for example, €65), it would also report a gain on the sale in its profit and loss. Equally important, the SPE may be able to borrow the funds at a lower rate than Odena, since they are bankruptcy remote from Odena (i.e., out of reach of Odena's creditors), and the lenders to the SPE are the claimants on its assets (i.e., the purchased receivables).

SPE Balance Sheet	
Accounts receivable	€60,000,000
Total assets	€60,000,000
Long-term debt	€55,000,000
Equity	5,000,000
Total liabilities and equity	€60,000,000

Because Odena consolidates the SPE, its financial balance sheet would look like the following:

Odena: Alternative 2 Consolidated Balance Sheet	
Cash	€85,000,000
Accounts receivable	60,000,000
Other assets	40,000,000
Total assets	€185,000,000
Current liabilities	€27,000,000
Noncurrent liabilities	75,000,000
Total liabilities	€102,000,000
Shareholder equity	€83,000,000
Total liabilities and equity	€185,000,000

Therefore, the consolidated balance sheet of Odena would look exactly the same as if it borrowed directly against the receivables. In addition, as a result of the consolidation, the transfer (sale) of the receivables to the SPE would be reversed along with any gain Odena recognized on the sale.

ADDITIONAL ISSUES IN BUSINESS COMBINATIONS THAT IMPAIR COMPARABILITY

11

a describe the classification, measurement, and disclosure under International Financial Reporting Standards (IFRS) for 1) investments in financial assets, 2) investments in associates, 3) joint ventures, 4) business combinations, and 5) special purpose and variable interest entities;

b compare and contrast IFRS and US GAAP in their classification, measurement, and disclosure of investments in financial assets, investments in associates, joint ventures, business combinations, and special purpose and variable interest entities;

Accounting for business combinations is a complex topic. In addition to the basics covered so far in this reading, we briefly mention some of the more common issues that impair comparability between IFRS and US GAAP.

11.1 Contingent Assets and Liabilities

Under IFRS, the cost of an acquisition is allocated to the fair value of assets, liabilities, and contingent liabilities. Contingent liabilities are recorded separately as part of the cost allocation process, provided that their fair values can be measured reliably. Subsequently, the contingent liability is measured at the higher of the amount initially recognized or the best estimate of the amount required to settle. As mentioned previously, GlaxoSmithKline had approximately £6 billion in contingent liabilities in relation to a number of purchases for the year ended 31 December 2017, with the notes to the financial statements further stating that the £6 billion was the expected value of the contingent consideration payments, discounted at an appropriate discount rate. Contingent assets are not recognized under IFRS.

Under US GAAP, contractual contingent assets and liabilities are recognized and recorded at their fair values at the time of acquisition. Non-contractual contingent assets and liabilities must also be recognized and recorded only if it is "more likely than not" they meet the definition of an asset or a liability at the acquisition date. Subsequently, a contingent liability is measured at the higher of the amount initially recognized or the best estimate of the amount of the loss. A contingent asset, however, is measured at the lower of the acquisition date fair value or the best estimate of the future settlement amount.

11.2 Contingent Consideration

Contingent consideration may be negotiated as part of the acquisition price. For example, the acquiring company (parent) may agree to pay additional money to the acquiree's (subsidiary's) former shareholders if certain agreed upon events occur. These can include achieving specified sales or profit levels for the acquiree and/or the combined entity. Under both IFRS and US GAAP, contingent consideration is initially measured at fair value. IFRS and US GAAP classify contingent consideration as an asset, liability or equity. In subsequent periods, changes in the fair value of liabilities (and assets, in the case of US GAAP) are recognized in the consolidated income statement. Both IFRS and US GAAP do not remeasure equity classified contingent consideration; instead, settlement is accounted for within equity.

11.3 In-Process R&D

IFRS and US GAAP recognize in-process research and development acquired in a business combination as a separate intangible asset and measure it at fair value (if it can be measured reliably). In subsequent periods, this research and development is subject to amortization if successfully completed (a marketable product results) or to impairment if no product results or if the product is not technically and/or financially viable.

11.4 Restructuring Costs

IFRS and US GAAP do not recognize restructuring costs that are associated with the business combination as part of the cost of the acquisition. Instead, they are recognized as an expense in the periods the restructuring costs are incurred.

SUMMARY

Intercompany investments play a significant role in business activities and create significant challenges for the analyst in assessing company performance. Investments in other companies can take five basic forms: investments in financial assets, investments in associates, joint ventures, business combinations, and investments in special purpose and variable interest entities. Key concepts are as follows:

- Investments in financial assets are those in which the investor has no significant influence. They can be measured and reported as
  - Fair value through profit or loss.
  - Fair value through other comprehensive income.
  - Amortized cost.

  IFRS and US GAAP treat investments in financial assets in a similar manner.
- Investments in associates and joint ventures are those in which the investor has significant influence, but not control, over the investee's business activities. Because the investor can exert significant influence over financial and operating policy decisions, IFRS and US GAAP require the equity method of accounting because it provides a more objective basis for reporting investment income.
  - The equity method requires the investor to recognize income as earned rather than when dividends are received.
  - The equity investment is carried at cost, plus its share of post-acquisition income (after adjustments) less dividends received.
  - The equity investment is reported as a single line item on the balance sheet and on the income statement.
- IFRS and US GAAP accounting standards require the use of the acquisition method to account for business combinations. Fair value of the consideration given is the appropriate measurement for identifiable assets and liabilities acquired in the business combination.
- Goodwill is the difference between the acquisition value and the fair value of the target's identifiable net tangible and intangible assets. Because it is considered to have an indefinite life, it is not amortized. Instead, it is evaluated at least annually for impairment. Impairment losses are reported on the income statement. IFRS use a one-step approach to determine and measure the impairment loss, whereas US GAAP uses a two-step approach.

- If the acquiring company acquires less than 100%, non-controlling (minority) shareholders' interests are reported on the consolidated financial statements. IFRS allows the non-controlling interest to be measured at either its fair value (full goodwill) or at the non-controlling interest's proportionate share of the acquiree's identifiable net assets (partial goodwill). US GAAP requires the non-controlling interest to be measured at fair value (full goodwill).
- Consolidated financial statements are prepared in each reporting period.
- Special purpose (SPEs) and variable interest entities (VIEs) are required to be consolidated by the entity which is expected to absorb the majority of the expected losses or receive the majority of expected residual benefits.

PRACTICE PROBLEMS

The following information relates to Questions 1–5

Cinnamon, Inc. is a diversified manufacturing company headquartered in the United Kingdom. It complies with IFRS. In 2017, Cinnamon held a 19 percent passive equity ownership interest in Cambridge Processing. In December 2017, Cinnamon announced that it would be increasing its ownership interest to 50 percent effective 1 January 2018 through a cash purchase. Cinnamon and Cambridge have no intercompany transactions.

Peter Lubbock, an analyst following both Cinnamon and Cambridge, is curious how the increased stake will affect Cinnamon's consolidated financial statements. He asks Cinnamon's CFO how the company will account for the investment, and is told that the decision has not yet been made. Lubbock decides to use his existing forecasts for both companies' financial statements to compare the outcomes of alternative accounting treatments.

Lubbock assembles abbreviated financial statement data for Cinnamon (Exhibit 1) and Cambridge (Exhibit 2) for this purpose.

Exhibit 1 Selected Financial Statement Information for Cinnamon, Inc. (£ Millions)

Year ending 31 December	2017	2018*
Revenue	1,400	1,575
Operating income	126	142
Net income	62	69
31 December	**2017**	**2018***
Total assets	1,170	1,317
Shareholders' equity	616	685

* Estimates made prior to announcement of increased stake in Cambridge.

Exhibit 2 Selected Financial Statement Information for Cambridge Processing (£ Millions)

Year ending 31 December	2017	2018*
Revenue	1,000	1,100
Operating income	80	88
Net income	40	44
Dividends paid	20	22
31 December	**2017**	**2018***

© 2019 CFA Institute. All rights reserved.

Exhibit 2 (Continued)

Year ending 31 December	2017	2018*
Total assets	800	836
Shareholders' equity	440	462

* Estimates made prior to announcement of increased stake by Cinnamon.

1 In 2018, if Cinnamon is deemed to have control over Cambridge, it will *most likely* account for its investment in Cambridge using:

A the equity method.

B the acquisition method.

C proportionate consolidation.

2 At 31 December 2018, Cinnamon's total shareholders' equity on its balance sheet would *most likely* be:

A highest if Cinnamon is deemed to have control of Cambridge.

B independent of the accounting method used for the investment in Cambridge.

C highest if Cinnamon is deemed to have significant influence over Cambridge.

3 In 2018, Cinnamon's net profit margin would be *highest* if:

A it is deemed to have control of Cambridge.

B it had not increased its stake in Cambridge.

C it is deemed to have significant influence over Cambridge.

4 At 31 December 2018, assuming control and recognition of goodwill, Cinnamon's reported debt to equity ratio will *most likely* be highest if it accounts for its investment in Cambridge using the:

A equity method.

B full goodwill method.

C partial goodwill method.

5 Compared to Cinnamon's operating margin in 2017, if it is deemed to have control of Cambridge, its operating margin in 2018 will *most likely* be:

A lower.

B higher.

C the same.

The following information relates to Questions 6–10

Zimt, AG is a consumer products manufacturer headquartered in Austria. It complies with IFRS. In 2017, Zimt held a 10 percent passive stake in Oxbow Limited. In December 2017, Zimt announced that it would be increasing its ownership to 50 percent effective 1 January 2018.

Franz Gelblum, an analyst following both Zimt and Oxbow, is curious how the increased stake will affect Zimt's consolidated financial statements. Because Gelblum is uncertain how the company will account for the increased stake, he uses his existing forecasts for both companies' financial statements to compare various alternative outcomes.

Gelblum gathers abbreviated financial statement data for Zimt (Exhibit 1) and Oxbow (Exhibit 2) for this purpose.

Exhibit 1 Selected Financial Statement Estimates for Zimt AG (€ Millions)

Year ending 31 December	2017	2018*
Revenue	1,500	1,700
Operating income	135	153
Net income	66	75
31 December	**2017**	**2018***
Total assets	1,254	1,421
Shareholders' equity	660	735

* Estimates made prior to announcement of increased stake in Oxbow.

Exhibit 2 Selected Financial Statement Estimates for Oxbow Limited (€ Millions)

Year ending 31 December	2017	2018*
Revenue	1,200	1,350
Operating income	120	135
Net income	60	68
Dividends paid	20	22
31 December	**2017**	**2018***
Total assets	1,200	1,283
Shareholders' equity	660	706

* Estimates made prior to announcement of increased stake by Zimt.

6 At 31 December 2018, Zimt's total assets balance would *most likely* be:
   - A highest if Zimt is deemed to have control of Oxbow.
   - B highest if Zimt is deemed to have significant influence over Oxbow.
   - C unaffected by the accounting method used for the investment in Oxbow.

7 Based on Gelblum's estimates, if Zimt is deemed to have significant influence over Oxbow, its 2018 net income (in € millions) would be *closest* to:
   - A €75.
   - B €109.
   - C €143.

8 Based on Gelblum's estimates, if Zimt is deemed to have joint control of Oxbow, and Zimt uses the proportionate consolidation method, its 31 December 2018 total liabilities (in € millions) will *most likely* be *closest* to:

A €686.

B €975.

C €1,263.

9 Based on Gelblum's estimates, if Zimt is deemed to have control over Oxbow, its 2018 consolidated sales (in € millions) will be *closest* to:

A €1,700.

B €2,375.

C €3,050.

10 Based on Gelblum's estimates, and holding the size of Zimt's ownership stake in Oxbow constant, Zimt's net income in 2018 will *most likely* be:

A highest if Zimt is deemed to have control of Oxbow.

B highest if Zimt is deemed to have significant influence over Oxbow.

C independent of the accounting method used for the investment in Oxbow.

The following information relates to Questions 11–16

Burton Howard, CFA, is an equity analyst with Maplewood Securities. Howard is preparing a research report on Confabulated Materials, SA, a publicly traded company based in France that complies with IFRS 9. As part of his analysis, Howard has assembled data gathered from the financial statement footnotes of Confabulated's 2018 Annual Report and from discussions with company management. Howard is concerned about the effect of this information on Confabulated's future earnings.

Information about Confabulated's investment portfolio for the years ended 31 December 2017 and 2018 is presented in Exhibit 1. As part of his research, Howard is considering the possible effect on reported income of Confabulated's accounting classification for fixed income investments.

Exhibit 1 Confabulated's Investment Portfolio (€ Thousands)

Characteristic	Bugle AG	Cathay Corp	Dumas SA
Classification	FVPL	FVOCI	Amortized cost
Cost*	€25,000	€40,000	€50,000
Market value, 31 December 2017	29,000	38,000	54,000
Market value, 31 December 2018	28,000	37,000	55,000

* All securities were acquired at par value.

In addition, Confabulated's annual report discusses a transaction under which receivables were securitized through a special purpose entity (SPE) for Confabulated's benefit.

11 The balance sheet carrying value of Confabulated's investment portfolio (in € thousands) at 31 December 2018 is *closest* to:

A 112,000.

B 115,000.

C 118,000.

12 The balance sheet carrying value of Confabulated's investment portfolio at 31 December 2018 would have been higher if which of the securities had been reclassified as FVPL security?

A Bugle.

B Cathay.

C Dumas.

13 Compared to Confabulated's reported interest income in 2018, if Dumas had been classified as FVPL, the interest income would have been:

A lower.

B the same.

C higher.

14 Compared to Confabulated's reported earnings before taxes in 2018, if Dumas had been classified as a FVPL security, the earnings before taxes (in € thousands) would have been:

A the same.

B €1,000 lower.

C €1,000 higher.

15 Confabulated's reported interest income would be lower if the cost was the same but the par value (in € thousands) of:

A Bugle was €28,000.

B Cathay was €37,000.

C Dumas was €55,000.

16 Confabulated's special purpose entity is *most likely* to be:

A held off-balance sheet.

B consolidated on Confabulated's financial statements.

C consolidated on Confabulated's financial statements only if it is a "qualifying SPE."

The following information relates to Questions 17–22

BetterCare Hospitals, Inc. operates a chain of hospitals throughout the United States. The company has been expanding by acquiring local hospitals. Its largest acquisition, that of Statewide Medical, was made in 2001 under the pooling of interests method. BetterCare complies with US GAAP.

BetterCare is currently forming a 50/50 joint venture with Supreme Healthcare under which the companies will share control of several hospitals. BetterCare plans to use the equity method to account for the joint venture. Supreme Healthcare complies with IFRS and will use the proportionate consolidation method to account for the joint venture.

Erik Ohalin is an equity analyst who covers both companies. He has estimated the joint venture's financial information for 2018 in order to prepare his estimates of each company's earnings and financial performance. This information is presented in Exhibit 1.

Exhibit 1 Selected Financial Statement Forecasts for Joint Venture ($ Millions)

Year ending 31 December	2018
Revenue	1,430
Operating income	128
Net income	62
31 December	**2018**
Total assets	1,500
Shareholders' equity	740

Supreme Healthcare recently announced it had formed a special purpose entity through which it plans to sell up to $100 million of its accounts receivable. Supreme Healthcare has no voting interest in the SPE, but it is expected to absorb any losses that it may incur. Ohalin wants to estimate the impact this will have on Supreme Healthcare's consolidated financial statements.

17 Compared to accounting principles currently in use, the pooling method BetterCare used for its Statewide Medical acquisition has *most likely* caused its reported:

- **A** revenue to be higher.
- **B** total equity to be lower.
- **C** total assets to be higher.

18 Based on Ohalin's estimates, the amount of joint venture revenue (in $ millions) included on BetterCare's consolidated 2018 financial statements should be *closest* to:

- **A** $0.
- **B** $715.
- **C** $1,430.

19 Based on Ohalin's estimates, the amount of joint venture net income included on the consolidated financial statements of each venturer will *most likely* be:

- **A** higher for BetterCare.
- **B** higher for Supreme Healthcare.
- **C** the same for both BetterCare and Supreme Healthcare.

20 Based on Ohalin's estimates, the amount of the joint venture's 31 December 2018 total assets (in $ millions) that will be included on Supreme Healthcare's consolidated financial statements will be *closest* to:

- **A** $0.
- **B** $750.
- **C** $1,500.

21 Based on Ohalin's estimates, the amount of joint venture shareholders' equity at 31 December 2018 included on the consolidated financial statements of each venturer will *most likely* be:

- **A** higher for BetterCare.
- **B** higher for Supreme Healthcare.
- **C** the same for both BetterCare and Supreme Healthcare.

22 If Supreme Healthcare sells its receivables to the SPE, its consolidated financial results will *least likely* show:

A a higher revenue for 2018.

B the same cash balance at 31 December 2018.

C the same accounts receivable balance at 31 December 2018.

The following information relates to Questions 23–28

Percy Byron, CFA, is an equity analyst with a UK-based investment firm. One firm Byron follows is NinMount PLC, a UK-based company. On 31 December 2018, NinMount paid £320 million to purchase a 50 percent stake in Boswell Company. The excess of the purchase price over the fair value of Boswell's net assets was attributable to previously unrecorded licenses. These licenses were estimated to have an economic life of six years. The fair value of Boswell's assets and liabilities other than licenses was equal to their recorded book values. NinMount and Boswell both use the pound sterling as their reporting currency and prepare their financial statements in accordance with IFRS.

Byron is concerned whether the investment should affect his "buy" rating on NinMount common stock. He knows NinMount could choose one of several accounting methods to report the results of its investment, but NinMount has not announced which method it will use. Byron forecasts that both companies' 2019 financial results (excluding any merger accounting adjustments) will be identical to those of 2018.

NinMount's and Boswell's condensed income statements for the year ended 31 December 2018, and condensed balance sheets at 31 December 2018, are presented in Exhibits 1 and 2, respectively.

Exhibit 1 NinMount PLC and Boswell Company Income Statements for the Year Ended 31 December 2018 (£ millions)

	NinMount	Boswell
Net sales	950	510
Cost of goods sold	(495)	(305)
Selling expenses	(50)	(15)
Administrative expenses	(136)	(49)
Depreciation & amortization expense	(102)	(92)
Interest expense	(42)	(32)
Income before taxes	125	17
Income tax expense	(50)	(7)
Net income	75	10

Exhibit 2 NinMount PLC and Boswell Company Balance Sheets at 31 December 2018 (£ millions)

	NinMount	Boswell
Cash	50	20
Receivables—net	70	45
Inventory	130	75
Total current assets	250	140
Property, plant, & equipment—net	1,570	930
Investment in Boswell	320	—
Total assets	2,140	1,070
Current liabilities	110	90
Long-term debt	600	400
Total liabilities	710	490
Common stock	850	535
Retained earnings	580	45
Total equity	1,430	580
Total liabilities and equity	2,140	1,070

Note: Balance sheets reflect the purchase price paid by NinMount, but do not yet consider the impact of the accounting method choice.

23 NinMount's current ratio on 31 December 2018 *most likely* will be highest if the results of the acquisition are reported using:

A the equity method.

B consolidation with full goodwill.

C consolidation with partial goodwill.

24 NinMount's long-term debt to equity ratio on 31 December 2018 *most likely* will be lowest if the results of the acquisition are reported using:

A the equity method.

B consolidation with full goodwill.

C consolidation with partial goodwill.

25 Based on Byron's forecast, if NinMount deems it has acquired control of Boswell, NinMount's consolidated 2019 depreciation and amortization expense (in £ millions) will be *closest* to:

A 102.

B 148.

C 204.

26 Based on Byron's forecast, NinMount's net profit margin for 2019 *most likely* will be highest if the results of the acquisition are reported using:

A the equity method.

B consolidation with full goodwill.

C consolidation with partial goodwill.

27 Based on Byron's forecast, NinMount's 2019 return on beginning equity *most likely* will be the same under:

A either of the consolidations, but different under the equity method.

B the equity method, consolidation with full goodwill, and consolidation with partial goodwill.

C none of the equity method, consolidation with full goodwill, or consolidation with partial goodwill.

28 Based on Byron's forecast, NinMount's 2019 total asset turnover ratio on beginning assets under the equity method is *most likely*:

A lower than if the results are reported using consolidation.

B the same as if the results are reported using consolidation.

C higher than if the results are reported using consolidation.

The following information relates to questions 29–35

John Thronen is an analyst in the research department of an international securities firm. Thronen is preparing a research report on Topmaker, Inc., a publicly-traded company that complies with IFRS. Thronen reviews two of Topmaker's recent transactions relating to investments in Blanco Co. and Rainer Co.

Investment in Blanca Co.

On 1 January 2016, Topmaker invested $11 million in Blanca Co. debt securities (with a 5.0% stated coupon rate on par value, payable each 31 December). The par value of the securities is $10 million, and the market interest rate in effect when the bonds were purchased was 4.0%. Topmaker designates the investment as held-to-maturity. On 31 December 2016, the fair value of the securities was $12 million.

Blanca Co. plans to raise $40 million in capital by borrowing against its financial receivables. Blanca plans to create a special purpose entity (SPE), invest $10 million in the SPE, have the SPE borrow $40 million, and then use the total funds to purchase $50 million of receivables from Blanca. Blanca meets the definition of control and plans to consolidate the SPE. Blanca's current balance sheet is presented in Exhibit 1.

Exhibit 1 Blanca Co. Balance Sheet at 31 December 2016 ($ millions)

Cash	20	Current liabilities	25
Accounts receivable	50	Noncurrent liabilities	30
Other assets	30	Shareholders' equity	45
Total assets	**100**	**Total liabilities and equity**	**100**

Investment in Rainer Co.

On 1 January 2016, Topmaker acquired a 15% equity interest with voting power in Rainer Co. for $300 million. Exhibit 2 presents selected financial information for Rainer on the acquisition date. Thronen notes that the plant and equipment are depreciated on a straight-line basis and have 10 years of remaining life. Topmaker has representation on Rainer's board of directors and participates in the associate's policy-making process.

Exhibit 2 Selected Financial Data for Rainer Co., 1 January 2018 (Acquisition Date) ($ millions)

	Book Value	Fair Value
Current assets	270	270
Plant and equipment	2,900	3,160
Total assets	3,170	3,430
Liabilities	1,830	1,830
Net assets	1,340	1,600

Thronen notes that, for fiscal year 2018, Rainer reported total revenue of $1,740 million and net income of $360 million, and paid dividends of $220 million.

Thronen is concerned about possible goodwill impairment for Topmaker due to expected changes in the industry effective at the end of 2017. He calculates the impairment loss based on selected data from the projected consolidated balance sheet data presented in Exhibit 3, assuming that the cash-generating unit and reporting unit of Topmaker are the same.

Exhibit 3 Selected Financial Data for Topmaker, Inc., Estimated Year Ending 31 December 2017 ($ millions)

Carrying value of cash-generating unit/reporting unit	15,200
Recoverable amount of cash-generating unit/reporting unit	14,900
Fair value of reporting unit	14,800
Identifiable net assets	14,400
Goodwill	520

Finally, Topmaker announces its plan to increase its ownership interest in Rainer to 80% effective 1 January 2018 and will account for the investment in Rainer using the partial goodwill method. Thronen estimates that the fair market value of the Rainer's shares on the expected date of exchange is $2 billion with the identifiable assets valued at $1.5 billion.

29 The carrying value of Topmaker's investment in Blanca's debt securities reported on the balance sheet at 31 December 2016 is:

A $10.94 million.

B $11.00 million.

C $12.00 million.

30 Based on Exhibit 1 and Blanca's plans to borrow against its financial receivables, the new consolidated balance sheet will show total assets of:

A $50 million.

B $140 million.

C $150 million.

31 Based on Exhibit 2, Topmaker's investment in Rainer resulted in goodwill of:

A $21 million.

B $60 million.

C $99 million.

32 Topmaker's influence on Rainer's business activities can be *best* described as:

A significant.

B controlling.

C shared control.

33 Using only the information from Exhibit 2, the carrying value of Topmaker's investment in Rainer at the end of 2016 is *closest* to:

A \$282 million.

B \$317 million.

C \$321 million.

34 Based on Exhibit 3, Topmaker's impairment loss under IFRS is:

A \$120 million.

B \$300 million.

C \$400 million.

35 Based on Thronen's value estimates on the acquisition date of 1 January 2018, the estimated value of the minority interest related to Rainer will be:

A \$300 million.

B \$400 million.

C \$500 million.

SOLUTIONS

1 B is correct. If Cinnamon is deemed to have control over Cambridge, it would use the acquisition method to account for Cambridge and prepare consolidated financial statements. Proportionate consolidation is used for joint ventures; the equity method is used for some joint ventures and when there is significant influence but not control.

2 A is correct. If Cinnamon is deemed to have control over Cambridge, consolidated financial statements would be prepared and Cinnamon's total shareholders' equity would increase and include the amount of the noncontrolling interest. If Cinnamon is deemed to have significant influence, the equity method would be used and there would be no change in the total shareholders' equity of Cinnamon.

3 C is correct. If Cinnamon is deemed to have significant influence, it would report half of Cambridge's net income as a line item on its income statement, but no additional revenue is shown. Its profit margin is thus higher than if it consolidated Cambridge's results, which would impact revenue and income, or if it only reported 19 percent of Cambridge's dividends (no change in ownership).

4 C is correct. The full and partial goodwill method will have the same amount of debt; however, shareholders' equity will be higher under full goodwill (and the debt to equity ratio will be lower). Therefore, the debt to equity will be higher under partial goodwill. If control is assumed, Cinnamon cannot use the equity method.

5 A is correct. Cambridge has a lower operating margin (88/1,100 = 8.0%) than Cinnamon (142/1,575 = 9.0%). If Cambridge's results are consolidated with Cinnamon's, the consolidated operating margin will reflect that of the combined company, or 230/2,675 = 8.6%.

6 A is correct. When a company is deemed to have control of another entity, it records all of the other entity's assets on its own consolidated balance sheet.

7 B is correct. If Zimt is deemed to have significant influence, it would use the equity method to record its ownership. Under the equity method, Zimt's share of Oxbow's net income would be recorded as a single line item. Net income of Zimt = 75 + 0.5(68) = 109.

8 B is correct. Under the proportionate consolidation method, Zimt's balance sheet would show its own total liabilities of €1,421 – 735 = €686 plus half of Oxbow's liabilities of €1,283 – 706 = €577. €686 + (0.5 × 577) = €974.5.

9 C is correct. Under the assumption of control, Zimt would record its own sales plus 100 percent of Oxbow's. €1,700 + 1,350 = €3,050.

10 C is correct. Net income is not affected by the accounting method used to account for active investments in other companies. "One-line consolidation" and consolidation result in the same impact on net income; it is the disclosure that differs.

11 B is correct. Under IFRS 9, FVPL and FVOCI securities are carried at market value, whereas amortized cost securities are carried at historical cost. €28,000 + 37,000 + 50,000 = €115,000.

12 C is correct. If Dumas had been classified as a FVPL security, its carrying value would have been the €55,000 fair value rather than the €50,000 historical cost.

13 B is correct. The coupon payment is recorded as interest income whether securities are amortized cost or FVPL. No adjustment is required for amortization since the bonds were bought at par.

14 C is correct. Unrealized gains and losses are included in income when securities are classified as FVPL. During 2018 there was an unrealized gain of €1,000.

15 B is correct. The difference between historical cost and par value must be amortized under the effective interest method. If the par value is less than the initial cost (stated interest rate is greater than the effective rate), the interest income would be lower than the interest received because of amortization of the premium.

16 B is correct. Under IFRS, SPEs must be consolidated if they are conducted for the benefit of the sponsoring entity. Further, under IFRS, SPEs cannot be classified as qualifying. Under US GAAP, qualifying SPEs (a classification which has been eliminated) do not have to be consolidated.

17 B is correct. Statewide Medical was accounted for under the pooling of interest method, which causes all of Statewide's assets and liabilities to be reported at historical book value. The excess of assets over liabilities generally is lower using the historical book value method than using the fair value method (this latter method must be used under currently required acquisition accounting). It would have no effect on revenue.

18 A is correct. Under the equity method, BetterCare would record its interest in the joint venture's net profit as a single line item, but would show no line-by-line contribution to revenues or expenses.

19 C is correct. Net income will be the same under the equity method and proportional consolidation. However, sales, cost of sales, and expenses are different because under the equity method the net effect of sales, cost of sales, and expenses is reflected in a single line.

20 B is correct. Under the proportionate consolidation method, Supreme Healthcare's consolidated financial statements will include its 50 percent share of the joint venture's total assets.

21 C is correct. The choice of equity method or proportionate consolidation does not affect reported shareholders' equity.

22 A is correct. Revenue will not be higher for 2018 because Supreme Healthcare controls the SPE and thus eliminates intra-entity transactions and balances in consolidation. Consolidated revenue will thus present the results as if this transaction did not occur.

23 A is correct. The current ratio using the equity method of accounting is Current assets/Current liabilities = £250/£110 = 2.27. Using consolidation (either full or partial goodwill), the current ratio = £390/£200 = 1.95. Therefore, the current ratio is highest using the equity method.

24 A is correct. Using the equity method, long-term debt to equity = £600/£1,430 = 0.42. Using the consolidation method, long-term debt to equity = long-term debt/equity = £1,000/£1,750 = 0.57. Equity includes the £320 non-controlling interest under either consolidation. It does not matter if the full or partial goodwill method is used since there is no goodwill.

25 C is correct. The projected depreciation and amortization expense will include NinMount's reported depreciation and amortization (£102), Boswell's reported depreciation and amortization (£92), and amortization of Boswell's licenses (£10 million). The licenses have a fair value of £60 million. £320 purchase price indicates a fair value of £640 for the net assets of Boswell. The net book (fair) value of the recorded assets is £580. The previously unrecorded licenses have

a fair value of £60 million. The licenses have a remaining life of six years; the amortization adjustment for 2018 will be £10 million. Therefore, Projected depreciation and amortization = £102 + £92 + £10 = £204 million.

26 A is correct. Net income is the same using any of the methods but under the equity method, net sales are only £950; Boswell's sales are not included in the net sales figure. Therefore, net profit margin is highest using the equity method.

27 A is correct. Net income is the same using any of the choices. Beginning equity under the equity method is £1,430. Under either of the consolidations, beginning equity is £1,750 since it includes the £320 noncontrolling interest. Return on beginning equity is highest under the equity method.

28 A is correct. Using the equity method, Total asset turnover = Net sales/ Beginning total assets = £950/£2,140 = 0.444. Total asset turnover on beginning assets using consolidation = £1,460/£2,950 = 0.495. Under consolidation, Assets = £2,140 − 320 + 1,070 + 60 = £2,950. Therefore, total asset turnover is lowest using the equity method.

29 A is correct. Since the investment is designated as held-to-maturity, it is reported at amortized cost at 31 December 2016 using the effective interest method where the amortization is calculated as the difference between the amount received and the interest income.

The interest payment each period is $500,000, which is calculated as the product of the par value of $10 million and the stated 5% coupon rate. The interest income of $440,000 is the product of the 4.0% market rate in effect when the bonds were purchased and the initial fair value of $11 million. The difference between the interest payment of $500,000 and the interest income of $440,000, equal to $60,000, is the amortization amount for 2016.

So, the initial fair value of $11 million is reduced by the amortization amount of $60,000, resulting in an amortized cost of $10.94 million at 31 December 2016.

30 B is correct. The SPE balance sheet will show accounts receivable of $50 million, long-term debt of $40 million and equity of $10 million. When the balance sheets of Blanca and the SPE are consolidated, Blanca's cash will increase by $40 million due to the sale of the receivables to the SPE (net of its $10 million cash investment in the SPE). Long-term debt (non-current liabilities) will also increase by $40 million. So, the consolidated balance sheet will show total assets of $140 million and will look the same as if Blanca borrowed directly against the receivables.

Blanca Co. Current Balance Sheet (before consolidation)			
Cash	20	Current liabilities	25
Accounts receivable	50	Noncurrent liabilities	30
Other assets	30	Shareholders' equity	45
Total assets	**100**	**Total liabilities and equity**	**100**

SPE Balance Sheet ($ Millions)			
		Long-term debt	$40
Accounts receivable	$50	Equity	$10
Total assets	**$50**	**Total liabilities and equity**	**$50**

Blanca Co. Consolidated Balance Sheet ($ Millions)			
Cash	$60	Current liabilities	$25
Accounts receivable	$50	Noncurrent liabilities	$70
Other assets	$30	Shareholder's equity	$45
Total assets	**$140**	**Total liabilities and equity**	**$140**

31 B is correct. The goodwill in Topmaker's $300 million purchase of Rainer's common shares using the equity method is $60 million, calculated as:

	$ Millions
Purchase price	$300
Less: 15% of book value of Rainer: (15% x $1,340)	201
Excess purchase price	99
Attributable to net assets	39
Plant and equipment (15% x ($3,160 – $2,900))	
Goodwill (residual)	60
	99

32 A is correct. Topmaker's representation on the Rainer board of directors and participation in Rainer's policymaking process indicate significant influence. Significant influence is generally assumed when the percentage of ownership interest is between 20% and 50%. Topmaker's representation on the board of directors and participation in the policymaking process, however, demonstrate significant influence despite its 15% equity interest.

33 B is correct. The carrying value of Topmaker's investment in Rainer using the equity method is $317 million and is calculated as:

	$ Millions
Purchase price	$300
Plus: Topmaker's share of Rainer's net income (15% x $360)	54
Less: Dividends received (15% x $220)	33
Less: Amortization of excess purchase price attributable to plant and equipment (15% x ($3,160 – $2,900))) / 10 years	3.9
Investment in associate (Rainer) at the end of 2016	$317.1

34 B is correct. The goodwill impairment loss under IFRS is $300 million, calculated as the difference between the recoverable amount of a cash-generating unit and the carrying value of the cash-generating unit. Topmaker's recoverable

amount of the cash-generating unit is $14,900 million, which is less than the carrying value of the cash-generating unit of $15,200 million. This results in an impairment loss of $300 million ($14,900 – $15,200).

35 A is correct. According to IFRS, under the partial goodwill method, the value of the minority interest is equal to the non-controlling interest's proportionate share of the subsidiary's identifiable net assets. Rainer's proportionate share is 20% and the value of its identifiable assets on the acquisition date is $1.5 billion. The value of the minority interest is $300 million (20% x $1.5 billion).

READING

10

Employee Compensation: Post-Employment and Share-Based

by Elaine Henry, PhD, CFA, and Elizabeth A. Gordon, PhD, MBA, CPA

Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA). Elizabeth A. Gordon, PhD, MBA, CPA, is at Temple University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe the types of post-employment benefit plans and implications for financial reports;
☐	**b.** explain and calculate measures of a defined benefit pension obligation (i.e., present value of the defined benefit obligation and projected benefit obligation) and net pension liability (or asset);
☐	**c.** describe the components of a company's defined benefit pension costs;
☐	**d.** explain and calculate the effect of a defined benefit plan's assumptions on the defined benefit obligation and periodic pension cost;
☐	**e.** explain and calculate how adjusting for items of pension and other post-employment benefits that are reported in the notes to the financial statements affects financial statements and ratios;
☐	**f.** interpret pension plan note disclosures including cash flow related information;
☐	**g.** explain issues associated with accounting for share-based compensation;
☐	**h.** explain how accounting for stock grants and stock options affects financial statements, and the importance of companies' assumptions in valuing these grants and options.

© 2013 CFA Institute. All rights reserved.

1 INTRODUCTION

This reading covers two complex aspects of employee compensation: post-employment (retirement) benefits and share-based compensation. Retirement benefits include pensions and other post-employment benefits, such as health insurance. Examples of share-based compensation are stock options and stock grants.

A common issue underlying both of these aspects of employee compensation is the difficulty in measuring the value of the compensation. One factor contributing to the difficulty is that employees earn the benefits in the periods that they provide service but typically receive the benefits in future periods, so measurement requires a significant number of assumptions.

This reading provides an overview of the methods companies use to estimate and measure the benefits they provide to their employees and how this information is reported in financial statements. There has been some convergence between International Financial Reporting Standards (IFRS) and US generally accepted accounting principles (US GAAP) in the measurement and accounting treatment for pensions, other post-employment benefits, and share-based compensation, but some differences remain. Although this reading focuses on IFRS as the basis for discussion, instances where US GAAP significantly differ are discussed.

The reading is organized as follows: Sections 2–11 address pensions and other post-employment benefits, and Sections 12–15 cover share-based compensation with a primary focus on the accounting for and analysis of stock options. A summary and practice problems conclude the reading.

2 PENSIONS AND OTHER POST-EMPLOYMENT BENEFITS

a describe the types of post-employment benefit plans and implications for financial reports;

This section discusses the accounting and reporting of pensions and other post-employment benefits by the companies that provide these benefits (accounting and reporting by pension and other retirement funds are not covered in this reading). Under IFRS, IAS 19, *Employee Benefits*, provides the principal source of guidance in accounting for pensions and other post-employment benefits.[1] Under US GAAP, the guidance is spread across several sections of the FASB Codification.[2]

The discussion begins with an overview of the types of benefits and measurement issues involved, including the accounting treatment for defined contribution plans. It then continues with financial statement reporting of pension plans and other post-employment benefits, including an overview of critical assumptions used to value these benefits. The section concludes with a discussion of evaluating defined benefit pension plan and other post-employment benefit disclosures.

1 This reading describes IFRS requirements contained in IAS 19 as updated in June 2011 and effective beginning January 2013.

2 Guidance on pension and other post-employment benefits is included in FASB ASC Topic 712 [Compensation-Nonretirement Postemployment Benefits], FASB ASC Topic 715 [Compensation-Retirement Benefits], FASB ASC Topic 960 [Plan Accounting-Defined Benefit Pension Plans], and FASB ASC Topic 965 [Plan Accounting-Health and Welfare Benefit Plans].

2.1 Types of Post-Employment Benefit Plans

Companies may offer various types of benefits to their employees following retirement, including pension plans, health care plans, medical insurance, and life insurance. Some of these benefits involve payments in the current period, but many are promises of future benefits. The objectives of accounting for employee benefits is to measure the cost associated with providing these benefits and to recognise these costs in the sponsoring company's financial statements during the employees' periods of service. Complexity arises because the sponsoring company must make assumptions to estimate the value of future benefits. The assumptions required to estimate and recognise these future benefits can have a significant impact on the company's reported performance and financial position. In addition, differences in assumptions can reduce comparability across companies.

Pension plans, as well as other post-employment benefits, may be either defined contribution plans or defined benefit plans. Under **defined contribution pension plans**, specific (or agreed-upon) contributions are made to an employee's pension plan. The agreed upon amount is the pension expense. Typically, in a defined contribution (DC) pension plan, an individual account is established for each participating employee. The accounts are generally invested through a financial intermediary, such as an investment management company or an insurance company. The employees and the employer may each contribute to the plan. After the employer makes its agreed-upon contribution to the plan on behalf of an employee—generally in the same period in which the employee provides the service—the employer has no obligation to make payments beyond this amount. The future value of the plan's assets depends on the performance of the investments within the plan. Any gains or losses related to those investments accrue to the employee. Therefore, in DC pension plans, the employee bears the risk that plan assets will not be sufficient to meet future needs. The impact on the company's financial statements of DC pension plans is easily assessed because the company has no obligations beyond the required contributions.

In contrast to DC pension plans, **defined benefit pension plans** are essentially promises by the employer to pay a defined amount of pension in the future. As part of total compensation, the employee works in the current period in exchange for a pension to be paid after retirement. In a defined benefit (DB) pension plan, the amount of pension benefit to be provided is defined, usually by reference to age, years of service, compensation, etc. For example, a DB pension plan may provide for the retiree to be paid, annually until death, an amount equal to 1 percent of the final year's salary times the number of years of service. The future pension payments represent a liability or obligation of the employer (i.e., the sponsoring company). To measure this obligation, the employer must make various actuarial assumptions (employee turnover, average retirement age, life expectancy after retirement) and computations. It is important for an analyst to evaluate such assumptions for their reasonableness and to analyse the impact of these assumptions on the financial reports of the company.

Under IFRS and US GAAP, all plans for pensions and other post-employment benefits other than those explicitly structured as DC plans are classified as DB plans.[3] DB plans include both formal plans and those informal arrangements that create a constructive obligation by the employer to its employees.[4] The employer must estimate the total cost of the benefits promised and then allocate these costs to the periods in

3 Multi-employer plans are an exception under IFRS. These are plans to which many different employers contribute on behalf of their employees, such as an industry association pension plan. For multi-employer plans, the employer accounts for its proportionate share of the plan. If, however, the employer does not have sufficient information from the plan administrator to meet the reporting requirement for a defined benefit plan, IFRS allow the employer to account for the plan as if it were a defined contribution plan.

4 For example, a company has a constructive obligation if the benefits it promises are not linked solely to the amount of its contributions or if it indirectly or directly guarantees a specified return on pension assets.

which the employees provide service. This estimation and allocation further increases the complexity of pension reporting because the timing of cash flows (contributions into the plan and payments from the plan) can differ significantly from the timing of accrual-basis reporting. Accrual-basis reporting is based on when the services are rendered and the benefits are earned.

Most DB pension plans are funded through a separate legal entity, typically a pension trust, and the assets of the trust are used to make the payments to retirees. The sponsoring company is responsible for making contributions to the plan. The company also must ensure that there are sufficient assets in the plan to pay the ultimate benefits promised to plan participants. Regulatory requirements usually specify minimum funding levels for DB pension plans, but those requirements vary by country. The funded status of a pension plan—overfunded or underfunded—refers to whether the amount of assets in the pension trust is greater than or less than the estimated liability. If the amount of assets in the DB pension trust exceeds the present value of the estimated liability, the DB pension plan is said to be overfunded; conversely, if the amount of assets in the pension trust is less than the estimated liability, the plan is said to be underfunded. Because the company has promised a defined amount of benefit to the employees, it is obligated to make those pension payments when they are due regardless of whether the pension plan assets generate sufficient returns to provide the benefits. In other words, the company bears the investment risk. Many companies are reducing the use of DB pension plans because of this risk.

Similar to DB pension plans, **other post-employment benefits** (OPB) are promises by the company to pay benefits in the future, such as life insurance premiums and all or part of health care insurance for its retirees. OPB are typically classified as DB plans, with accounting treatment similar to DB pension plans. However, the complexity in reporting for OPB may be even greater than for DB pension plans because of the need to estimate future increases in costs, such as health care, over a long time horizon. Unlike DB pension plans, however, companies may not be required by regulation to fund an OPB in advance to the same degree as DB pension plans. This is partly because governments, through some means, often insure DB pension plans but not OPB, partly because OPB may represent a much smaller financial liability, and partly because OPB are often easier to eliminate should the costs become burdensome. It is important that an analyst determine what OPB are offered by a company and the obligation they represent.

Types of post-employment benefits offered by employers differ across countries. For instance, in countries where government-sponsored universal health care plans exist (such as Germany, France, Canada, Brazil, Mexico, New Zealand, South Africa, India, Israel, Bhutan, and Singapore), companies are less likely to provide post-retirement health care benefits to employees. The extent to which companies offer DC or DB pension plans also varies by country.

Exhibit 1 summarizes these three types of post-employment benefits.

Exhibit 1 Types of Post-Employment Benefits

Type of Benefit	Amount of Post-Employment Benefit to Employee	Obligation of Sponsoring Company	Sponsoring Company's Pre-funding of its Future Obligation
Defined contribution pension plan	Amount of future benefit is not defined. Actual future benefit will depend on investment performance of plan assets. Investment risk is borne by employee.	Amount of the company's obligation (contribution) is defined in each period. The contribution, if any, is typically made on a periodic basis with no additional future obligation.	Not applicable.
Defined benefit pension plan	Amount of future benefit is defined, based on the plan's formula (often a function of length of service and final year's compensation). Investment risk is borne by company.	Amount of the future obligation, based on the plan's formula, must be estimated in the current period.	Companies typically pre-fund the DB plans by contributing funds to a pension trust. Regulatory requirements to pre-fund vary by country.
Other post-employment benefits (e.g., retirees' health care)	Amount of future benefit depends on plan specifications and type of benefit.	Eventual benefits are specified. The amount of the future obligation must be estimated in the current period.	Companies typically do not pre-fund other post-employment benefit obligations.

The following sections provide additional detail on how DB pension plan liabilities and periodic costs are measured, the financial statement impact of reporting pension and other post-employment benefits, and how disclosures in the notes to the financial statements can be used to gain insights about the underlying economics of a company's defined benefit plans. Section 3 describes how a DB pension plan's obligation is estimated and the key inputs into and assumptions behind the estimate. Sections 4–8 describe financial statement reporting of pension and OPB plans and demonstrate the calculation of defined benefit obligations and current costs and the effects of assumptions. Sections 9–11 describe disclosures in financial reports about pension and OPB plans. These include disclosures about assumptions that can be useful in analysing and comparing pension and OPB plans within and among companies.

3 MEASURING A DEFINED BENEFIT PENSION PLAN'S OBLIGATIONS

b explain and calculate measures of a defined benefit pension obligation (i.e., present value of the defined benefit obligation and projected benefit obligation) and net pension liability (or asset);

Both IFRS and US GAAP measure the **pension obligation** as the present value of future benefits earned by employees for service provided to date. The obligation is called the present value of the defined benefit obligation (PVDBO) under IFRS and the projected benefit obligation (PBO) under US GAAP.[5] This measure is defined as "the present value, without deducting any plan assets, of expected future payments required to settle the obligation arising from employee service in the current and prior periods" under IFRS and "the actuarial present value as of a date of all benefits attributed by the pension benefit formula to employee service rendered prior to that date" under US GAAP. In the remainder of this reading, the term "pension obligation" will be used to generically refer to PVDBO and PBO.

In determining the pension obligation, a company estimates the future benefits it will pay. To estimate the future benefits, the company must make a number of assumptions[6] such as future compensation increases and levels, discount rates, and expected vesting. For instance, an estimate of future compensation is made if the pension benefit formula is based on future compensation levels (examples include pay-related, final-pay, final-average-pay, or career-average-pay plans). The expected annual increase in compensation over the employee service period can have a significant impact on the defined benefit obligation. The determination of the benefit obligation implicitly assumes that the company will continue to operate in the future (the "going concern assumption") and recognises that benefits will increase with future compensation increases.

Another key assumption is the discount rate—the interest rate used to calculate the present value of the future benefits. This rate is based on current rates of return on high-quality corporate bonds (or government bonds in the absence of a deep market in corporate bonds) with currency and durations consistent with the currency and durations of the benefits.

Under both DB and DC pension plans, the benefits that employees earn may be conditional on remaining with the company for a specified period of time. "Vesting" refers to a provision in pensions plans whereby an employee gains rights to future benefits only after meeting certain criteria, such as a pre-specified number of years of service. If the employee leaves the company before meeting the criteria, he or she may be entitled to none or a portion of the benefits earned up until that point. However, once the employee has met the vesting requirements, he or she is entitled to receive the benefits earned in prior periods (i.e., once the employee has become vested, benefits are not forfeited if the employee leaves the company). In measuring the defined benefit obligation, the company considers the probability that some employees may not satisfy the vesting requirements (i.e., may leave before the vesting period) and uses this probability to calculate the current service cost and the present value of the obligation. Current service cost is the increase in the present value of a defined benefit obligation as a result of employee service in the current period. Current service cost is not the only cause of change in the present value of a defined benefit obligation.

The estimates and assumptions about future salary increases, the discount rate, and the expected vesting can change. Of course, any changes in these estimates and assumptions will change the estimated pension obligation. If the changes increase the

5 In addition to the projected benefit obligation, US GAAP identify two other measures of the pension liability. The **vested benefit obligation** (VBO) is the "actuarial present value of vested benefits" (FASB ASC Glossary). The **accumulated benefit obligation** (ABO) is "the actuarial present value of benefits (whether vested or non-vested) attributed, generally by the pension benefit formula, to employee service rendered before a specified date and based on employee service and compensation (if applicable) before that date. The accumulated benefit obligation differs from the projected benefit obligation in that it includes no assumption about future compensation levels" (FASB ASC Glossary). Both the vested benefit obligation and the accumulated benefit obligation are based on the amounts promised as a result of an employee's service up to a specific date. Thus, both of these measures will be less than the projected benefit obligation (VBO < ABO < PBO).

6 These assumptions are referred to as "actuarial assumptions." Thus, losses or gains due to changes in these assumptions, or due to differences between these assumptions and what actually occurs, are referred to as "actuarial gains or losses."

obligation, the increase is referred to as an actuarial loss. If the changes decrease the obligation, the change is referred to as an actuarial gain. Sections 7–8 further discuss estimates and assumptions and the effect on the pension obligation and expense.

FINANCIAL STATEMENT REPORTING OF PENSION PLANS AND OTHER POST-EMPLOYMENT BENEFITS: DEFINED CONTRIBUTION PENSION PLANS

4

a describe the types of post-employment benefit plans and implications for financial reports;

Sections 4–8 describe how pension plans and other post-employment benefits are reported in the financial statements of the sponsoring company and how assumptions affect the amounts reported. Disclosures related to pensions plans and OPB are described in Sections 9–11.

4.1 Defined Contribution Pension Plans

The accounting treatment for defined contribution pension plans is relatively simple. From a financial statement perspective, the employer's obligation for contributions into the plan, if any, is recorded as an expense on the income statement. Because the employer's obligation is limited to a defined amount that typically equals its contribution, no significant pension-related liability accrues on the balance sheet. An accrual (current liability) is recognised at the end of the reporting period only for any unpaid contributions.

FINANCIAL STATEMENT REPORTING OF PENSION PLANS: BALANCE SHEET REPORTING FOR DEFINED BENEFIT PENSION PLANS

5

b explain and calculate measures of a defined benefit pension obligation (i.e., present value of the defined benefit obligation and projected benefit obligation) and net pension liability (or asset);

The accounting treatment for defined benefit pension plans is more complex, primarily because of the complexities of measuring the pension obligation and expense.

5.1 Balance Sheet Presentation

Both IFRS and US GAAP require a pension plan's funded status to be reported on the balance sheet. The funded status is determined by netting the pension obligation against the fair value of the pension plan assets. If the pension obligation exceeds the pension plan assets, the plan has a deficit. If the plan assets exceed the pension obligation, the plan has a surplus. Summarizing this information in equation form gives

Funded status = Fair value of the plan assets − PV of the Defined benefit obligation

If the plan has a deficit, an amount equal to the net underfunded pension obligation is reported on the balance sheet as a net pension liability. If the plan has a surplus, an asset equal to the overfunded pension obligation is reported on the balance sheet as a net pension asset (except that the amount of reported assets is subject to a ceiling defined as the present value of future economic benefits, such as refunds from the plan or reductions of future contributions). Disclosures in the notes provide additional information about the net pension liability or asset reported on the balance sheet.

EXAMPLE 1

Determination of Amounts to be Reported on the Balance Sheet

The following information pertains to two hypothetical companies' defined benefit pension plans as of 31 December 2010:

- For company ABC, the present value of the company's defined benefit obligation is €6,723 and the fair value of the pension plan's assets is €4,880.
- For company DEF, the present value of the company's defined benefit obligation is €5,485 and the fair value of the pension plan assets is €5,998. In addition, the present value of available future refunds and reductions in future contributions is €326.

Calculate the amount each company would report as a pension asset or liability on its 2010 balance sheet.

Solution:

Company ABC would report the full underfunded status of its pension plan (i.e., the amount by which the present value of the defined benefit obligation exceeds the fair value of plan assets) as a liability. Specifically, the company would report a pension liability of €1,843.

Present value of defined benefit obligation	€6,723
Fair value of plan assets	(4,880)
Net pension liability	€1,843

Company DEF's pension plan is overfunded by €513, which is the amount by which the fair value of the plan's assets exceed the defined benefit obligation (€5,998 – €5,485). However, when a company has a surplus in a defined benefit plan, the amount of asset that can be reported is the lower of the surplus and the asset ceiling (the present value of future economic benefits, such as refunds from the plan or reductions of future contributions). In this case, the asset ceiling is given as €326, so the amount of company DEF's reported net pension asset would be limited to €326.

6 FINANCIAL STATEMENT REPORTING OF PENSION PLANS: PERIODIC PENSION COSTS FOR DEFINED BENEFIT PENSION PLANS

c describe the components of a company's defined benefit pension costs;

The periodic cost of a company's DB pension plan is the change in the net pension liability or asset adjusted for the employer's contributions. Each period, the periodic pension cost is recognised in profit or loss (P&L) and/or in other comprehensive income (OCI). (In some cases, amounts of pension costs may qualify for inclusion as part of the costs of such assets as inventories and thus be included in P&L as part of cost of goods sold when those inventories are later sold. The focus here is on the amounts not capitalised.) IFRS and US GAAP differ in the way that the periodic pension cost is divided between P&L and OCI.

Under IFRS, the periodic pension cost is viewed as having three components, two of which are recognised in P&L and one of which is recognised in OCI.

1 *Service costs.* The first component of periodic pension cost is service cost. Current service cost is the amount by which a company's pension obligation increases as a result of employees' service in the current period. Past service cost is the amount by which a company's pension obligation relating to employees' service in prior periods changes as a result of plan amendments or a plan curtailment.[7] Under IFRS, service costs (including both current service costs and past service costs) are recognised as an expense in P&L.

2 *Net interest expense/income.* The second component of periodic pension cost is net interest expense or income, which we will refer to as "net interest expense/income." Net interest expense/income is calculated by multiplying the net pension liability or net pension asset by the discount rate used in determining the present value of the pension liability. A net interest expense represents the financing cost of deferring payments related to the plan, and a net interest income represents the financing income from prepaying amounts related to the plan. Under IFRS, the net interest expense/income is recognised in P&L.

3 *Remeasurement.* The third component of periodic pension cost is remeasurement of the net pension liability or asset. Remeasurement includes (a) actuarial gains and losses and (b) any differences between the actual return on plan assets and the amount included in the net interest expense/income calculation. Under IFRS, remeasurement amounts are recognised in OCI. Remeasurement amounts are not subsequently amortised to P&L.

Similar to IFRS, under US GAAP current service cost is recognised in P&L. However, under US GAAP, past service costs are reported in OCI in the period in which the change giving rise to the cost occurs. In subsequent periods, these past service costs are amortised to P&L over the average service lives of the affected employees.

Also similar to IFRS, under US GAAP the periodic pension cost for P&L includes interest expense on pension obligations (which increases the amount of the periodic cost) and returns on the pension plan assets (which reduce the amount of the periodic cost). Unlike IFRS, however, under US GAAP, the two components are not presented net. Also, under US GAAP, returns on plan assets included in the P&L recognition of pension costs (pension expense) use an expected return rather than the actual return. (Under IFRS, returns on plan assets included in the P&L recognition of pension costs (pension expense) use the discount rate as the expected return.) Thus, under US GAAP, differences between the expected return and the actual return on plan assets represent another source of actuarial gains or losses. As noted, actuarial gains and losses can also result from changes in the actuarial assumptions used in determining the benefit obligation. Under US GAAP, all actuarial gains and losses are included in the net pension liability or net pension asset and can be reported either in P&L or

7 A curtailment occurs when there is a significant reduction by the entity either in the number of employees covered by a plan or in benefits.

in OCI. Typically, companies report actuarial gains and losses in OCI and recognise gains and losses in P&L only when certain conditions are met under a so-called corridor approach.

Under the corridor approach, the net cumulative unrecognised actuarial gains and losses at the beginning of the reporting period are compared with the defined benefit obligation and the fair value of plan assets at the beginning of the period. If the cumulative amount of unrecognised actuarial gains and losses becomes too large (i.e., exceeds 10 percent of the greater of the defined benefit obligation or the fair value of plan assets), then the excess is amortised over the expected average remaining working lives of the employees participating in the plan and is included as a component of periodic pension cost in P&L. The term "corridor" refers to the 10 percent range, and only amounts in excess of the corridor must be amortised.

To illustrate the corridor approach, assume that the beginning balance of the defined benefit obligation is $5,000,000, the beginning balance of fair value of plan assets is $4,850,000, and the beginning balance of unrecognised actuarial losses is $610,000. The expected average remaining working lives of the plan employees is 10 years. In this scenario, the corridor is $500,000, which is 10 percent of the defined benefit obligation (selected as the greater of the defined benefit obligation or the fair value of plan assets). Because the balance of unrecognised actuarial losses exceeds the $500,000 corridor, amortisation is required. The amount of the amortisation is $11,000, which is the excess of the unrecognised actuarial loss over the corridor divided by the expected average remaining working lives of the plan employees [($610,000 − $500,000) ÷ 10 years]. Actuarial gains or losses can also be amortised more quickly than under the corridor method; companies may use a faster recognition method, provided the company applies the method of amortisation to both gains and losses consistently in all periods presented.

To summarize, under IFRS, the periodic pension costs recognised in P&L include service costs (both current and past) and net interest expense/income. The periodic pension costs recognised in OCI include remeasurements that comprise net return on plan assets and actuarial gains and losses. Under US GAAP, the periodic pension costs recognised in P&L include current service costs, interest expense on plan liabilities, expected returns on plan assets (which is a reduction of the cost), the amortisation of past service costs, and actuarial gains and losses to the extent not reported in OCI. The components of a company's defined benefit periodic pension costs are summarized in Exhibit 2.

Exhibit 2 Components of a Company's Defined Benefit Pension Periodic Costs

IFRS Component	IFRS Recognition	US GAAP Component	US GAAP Recognition
Service costs	Recognised in P&L.	Current service costs	Recognised in P&L.
		Past service costs	Recognised in OCI and subsequently amortised to P&L over the service life of employees.

Exhibit 2 (Continued)

IFRS Component	IFRS Recognition	US GAAP Component	US GAAP Recognition
Net interest income/expense	Recognised in P&L as the following amount: Net pension liability or asset × interest rate[a]	Interest expense on pension obligation	Recognised in P&L.
		Expected return on plan assets	Recognised in P&L as the following amount: Plan assets × expected return.
Remeasurements: Net return on plan assets and actuarial gains and losses	Recognised in OCI and not subsequently amortised to P&L. ■ Net return on plan assets = Actual return − (Plan assets × Interest rate). ■ Actuarial gains and losses = Changes in a company's pension obligation arising from changes in actuarial assumptions.	Actuarial gains and losses including differences between the actual and expected returns on plan assets	Recognised immediately in P&L *or*, more commonly, recognised in OCI and subsequently amortised to P&L using the corridor or faster recognition method.[b] ■ Difference between expected and actual return on assets = Actual return − (Plan assets × Expected return). ■ Actuarial gains and losses = Changes in a company's pension obligation arising from changes in actuarial assumptions.

[a] The interest rate used is equal to the discount rate used to measure the pension liability (the yield on high-quality corporate bonds.)
[b] If the cumulative amount of unrecognised actuarial gains and losses exceeds 10 percent of the greater of the value of the plan assets or of the present value of the DB obligation (under US GAAP, the projected benefit obligation), the difference must be amortised over the service lives of the employees.

Reporting the Periodic Pension Cost. As noted above, some amounts of pension costs may qualify for capitalisation as part of the costs of self-constructed assets, such as inventories. Pension costs included in inventories would thus be recognised in P&L as part of cost of goods sold when those inventories are sold. For pension costs that are not capitalised, IFRS do not specify where companies present the various components of periodic pension cost beyond differentiating between components included in P&L and in OCI. In contrast, for pension costs that are not capitalised, US GAAP require all components of periodic pension cost that are recognised in P&L to be aggregated and presented as a net amount within the same line item on the income statement. Both IFRS and US GAAP require total periodic pension cost to be disclosed in the notes to the financial statements.

7 MORE ON THE EFFECT OF ASSUMPTIONS AND ACTUARIAL GAINS AND LOSSES ON PENSION AND OTHER POST-EMPLOYMENT BENEFIT COSTS

d explain and calculate the effect of a defined benefit plan's assumptions on the defined benefit obligation and periodic pension cost;

As noted, a company's pension obligation for a DB pension plan is based on many estimates and assumptions. The amount of future pension payments requires assumptions about employee turnover, length of service, and rate of increase in compensation levels. The length of time the pension payments will be made requires assumptions about employees' life expectancy post-employment. Finally, the present value of these future payments requires assumptions about the appropriate discount rate (which is used as the rate at which interest expense or income will subsequently accrue on the net pension liability or asset).

Changes in any of the assumptions will increase or decrease the pension obligation. An increase in pension obligation resulting from changes in actuarial assumptions is considered an actuarial loss, and a decrease is considered an actuarial gain. The estimate of a company's pension liability also affects several components of periodic pension costs, apart from actuarial gains and losses. First, the service cost component of annual pension cost is essentially the amount by which the pension liability increases as a result of the employees' service during the year. Second, the interest expense component of annual pension cost is based on the amount of the liability. Third, the past service cost component of annual pension cost is the amount by which the pension liability increases because of changes to the plan.

Estimates related to plan assets can also affect annual pension cost reported in P&L (pension expense), primarily under US GAAP. Because a company's periodic pension cost reported in P&L under US GAAP includes the *expected* return on pension assets rather than the actual return, the assumptions about the expected return on plan assets can have a significant impact. Also, the expected return on plan assets requires estimating in which future period the benefits will be paid. As noted above, a divergence of actual returns on pension assets from expected returns results in an actuarial gain or loss.

Understanding the effect of assumptions on the estimated pension obligation and on periodic pension costs is important both for interpreting a company's financial statements and for evaluating whether a company's assumptions appear relatively conservative or aggressive.

The projected unit credit method is the IFRS approach to measuring the DB obligation. Under the projected unit credit method, each period of service (e.g., year of employment) gives rise to an additional unit of benefit to which the employee is entitled at retirement. In other words, for each period in which an employee provides service, they earn a portion of the post-employment benefits that the company has promised to pay. An equivalent way of thinking about this is that the amount of eventual benefit increases with each additional year of service. The employer measures each unit of service as it is earned to determine the amount of benefits it is obligated to pay in future reporting periods.

The objective of the projected unit credit method is to allocate the entire expected retirement costs (benefits) for an employee over the employee's service periods. The defined benefit obligation represents the actuarial present value of all units of benefit (credit) to which the employee is entitled (i.e., those that the employee has earned) as a result of prior and current periods of service. This obligation is based on actuarial assumptions about demographic variables, such as employee turnover and life expectancy, and on estimates of financial variables, such as future inflation and the discount rate. If the pension benefit formula is based on employees' future compensation levels, then the unit of benefit earned each period will reflect this estimate.

Under both IFRS and US GAAP, the assumed rate of increase in compensation—the expected annual increase in compensation over the employee service period—can have a significant impact on the defined benefit obligation. Another key assumption is the discount rate used to calculate the present value of the future benefits. It represents the rate at which the defined benefit obligation could be effectively settled. This rate is based on current rates of return on high quality corporate bonds with durations consistent with the durations of the benefit.

CALCULATION OF DEFINED BENEFIT PENSION OBLIGATION AND CURRENT SERVICE COSTS

8

d explain and calculate the effect of a defined benefit plan's assumptions on the defined benefit obligation and periodic pension cost;

The following example illustrates the calculation of the defined benefit pension obligation and current service costs, using the projected unit credit method, for an individual employee under four different scenarios. Interest on the opening obligation also increases the obligation and is part of current costs. The fourth scenario is used to demonstrate the impact on a company's pension obligation of changes in certain key estimates. Examples 2 and 3 focus on the pension obligation. The change in pension obligation over the period is included in the calculation of pension expense (pension cost reported in P&L).

EXAMPLE 2

Calculation of Defined Benefit Pension Obligation for an Individual Employee

The following information applies to each of the four scenarios. Assume that a (hypothetical) company establishes a DB pension plan. The employee has a salary in the coming year of €50,000 and is expected to work five more years before retiring. The assumed discount rate is 6 percent, and the assumed annual compensation increase is 4.75 percent. For simplicity, assume that there are no changes in actuarial assumptions, all compensation increases are awarded on the first day of the service year, and no additional adjustments are made to reflect the possibility that the employee may leave the company at an earlier date.

Current salary	€50,000.00
Years until retirement	5
Annual compensation increases	4.75%
Discount rate	6.00%
Final year's estimated salary[a]	€60,198.56

[a] Final year's estimated salary = Current year's salary × [(1 + Annual compensation increase)$^{\text{Years until retirement} - 1}$].

At the end of Year 1, the final year's estimated salary = €50,000 × [(1 + 0.0475)^4] = €60,198.56, assuming that the employee's salary increases by 4.75 percent each year. With no change in assumption about the rate of increase in compensation or the date of retirement, the estimate of the final year's salary will remain unchanged.

At the end of Year 2, assuming the employee's salary actually increased by 4.75 percent, the final year's estimated salary = €52,375 × [(1 + 0.0475)^3] = €60,198.56.

Scenario 1: Benefit is paid as a lump sum amount upon retirement.

The plan will pay a lump sum pension benefit equal to 1.5 percent of the employee's final salary for each year of service beyond the date of establishment. The lump sum payment to be paid upon retirement = (Final salary × Benefit formula) × Years of service = (€60,198.56 × 0.015) × 5 = €4,514.89.

Annual unit credit (benefit) per service year = Value at retirement/Years of service = €4,514.89/5 = €902.98.

If the discount rate (the interest rate at which the defined benefit obligation could be effectively settled) is assumed to be 0 percent, the amount of annual unit credit per service year is the amount of the company's annual obligation and the closing obligation each year is simply the annual unit credit multiplied by the number of past and current years of service. However, because the assumed discount rate must be based on the yield on high-quality corporate bonds and will thus not equal 0 percent, the future obligation resulting from current and prior service is discounted to determine the value of the obligation at any point in time.

The following table shows how the obligation builds up for this employee.

Year	1	2	3	4	5
Estimated annual salary	€50,000.00	€52,375.00	€54,862.81	€57,468.80	€60,198.56
Benefits attributed to:					
Prior years[a]	€0.00	€902.98	€1,805.96	€2,708.94	€3,611.92
Current year[b]	902.98	902.98	902.98	902.98	902.97*
Total benefits earned	€902.98	€1,805.96	€2,708.94	€3,611.92	€4,514.89
Opening obligation[c]	€0.00	€715.24	€1,516.31	€2,410.94	€3,407.47
Interest cost at 6 percent[d]	0.00	42.91	90.98	144.66	204.45
Current service costs[e]	715.24	758.16	803.65	851.87	902.97
Closing obligation[f]	€715.24	€1,516.31	€2,410.94	€3,407.47	€4,514.89

*Final amounts may differ slightly to compensate for rounding in earlier years.

[a] The benefit attributed to prior years = Annual unit credit × Years of prior service.
For Year 2, €902.98 × 1 = €902.98.
For Year 3, €902.98 × 2 = €1,805.96.

[b] The benefit attributed to current year = Annual unit credit based on benefit formula = Final year's estimated salary × Benefit formula = Value at retirement date/Years of service = (€60,198.56 × 1.5%) = €4,514.89/5 = €902.98.

[c] The opening obligation is the closing obligation at the end of the previous year, but can also be viewed as the present value of benefits earned in prior years:
Benefits earned in prior years/[(1 + Discount rate)$^{\text{Years until retirement}}$].
Opening obligation Year 1 = €0.
Opening obligation Year 2 = €902.98/[(1 + 0.06)^4] = €715.24.
Opening obligation Year 3 = €1,805.96/[(1 + 0.06)^3] = €1,516.32.

[d] The interest cost is the increase in the present value of the defined benefit obligation due to the passage of time:
Interest cost = Opening obligation × Discount rate.
For Year 2 = €715.24 × 0.06 = €42.91.
For Year 3 = €1,516.32 × 0.06 = €90.98.

[e] Current service costs are the present value of annual unit credits earned in the current period:
Annual unit credit/[(1 + Discount rate)$^{\text{Years until retirement}}$].
For Year 1 = €902.98/[(1 + 0.06)^4] = €715.24.
For Year 2 = €902.98/[(1 + 0.06)^3] = €758.16.
Note: Given no change in actuarial assumptions and estimates of financial growth, the current service costs in any year (except the first) are the previous year's current service costs increased by the discount rate; the current service costs increase with the passage of time.

[f] The closing obligation is the opening obligation plus the interest cost and the current service costs but can also be viewed as the present value of benefits earned in prior and current years. There is a slight difference due to rounding.
Total benefits earned/[(1 + Discount rate)$^{\text{Years until retirement}}$].
Closing obligation Year 1 = €902.98/[(1 + 0.06)^4] = €715.24.
Closing obligation Year 2 = €1,805.96/[(1 + 0.06)^3] = €1,516.32.

Closing obligation Year 3 = €2,708.94/[$(1 + 0.06)^2$] = €2,410.95.

Note: Assuming no past service costs or actuarial gains/losses, the closing obligation less the fair value of the plan assets represents both the funded status of the plan and the net pension liability/asset. The change in obligation is the amount of expense for pensions on the income statement.

Scenario 2: Prior years of service, and benefit paid as a lump sum upon retirement.

The plan will pay a lump sum pension benefit equal to 1.5 percent of the employee's final salary for each year of service beyond the date of establishment. In addition, at the time the pension plan is established, the employee is given credit for 10 years of prior service with immediate vesting. The lump sum payment to be paid upon retirement = (Final salary × Benefit formula) × Years of service = (€60,198.56 × 0.015) × 15 = €13,544.68.

Annual unit credit = Value at retirement date/Years of service = €13,544.68/15 = €902.98.

The following table shows how the obligation builds up for this employee.

Year	1	2	3	4	5
Benefits attributed to:					
Prior years[a]	€9,029.78	€9,932.76	€10,835.74	€11,738.72	€12,641.70
Current years	902.98	902.98	902.98	902.98	902.98
Total benefits earned	€9,932.76	€10,835.74	€11,738.72	€12,641.70	€13,544.68
Opening obligation[b]	€6,747.58	€7,867.67	€9,097.89	€10,447.41	€11,926.13
Interest at 6 percent	404.85	472.06	545.87	626.85	715.57
Current service costs	715.24	758.16	803.65	851.87	902.98
Closing obligation	€7,867.67	€9,097.89	€10,447.41	€11,926.13	€13,544.68

[a] Benefits attributed to prior years of service = Annual unit credit × Years of prior service. At beginning of Year 1 = (€60,198.56 × 0.015) × 10 = €9,029.78.

[b] Opening obligation is the present value of the benefits attributed to prior years = Benefits attributed to prior years/$(1 + \text{Discount rate})^{\text{Number of years to retirement}}$.

At beginning of Year 1 = €9,029.78/$(1.06)^5$ = €6,747.58. This is treated as past service costs in Year 1 because there was no previous recognition and there is immediate vesting.

Scenario 3: Employee to receive benefit payments for 20 years (no prior years of service).

Years of receiving pension = 20.

Estimated annual payment (end of year) for each of the 20 years = (Estimated final salary × Benefit formula) × Years of service = (€60,198.56 × 0.015) × 5 = €4,514.89.

Value at the end of Year 5 (retirement date) of the estimated future payments = PV of €4,514.89 for 20 years at 6 percent = €51,785.46.[8]

Annual unit credit = Value at retirement date/Years of service = €51,785.46/5 = €10,357.09.

8 This is a simplification of the valuation process for illustrative purposes. For example, the actuarial valuation would use mortality rates, not just assumed life expectancy. Additionally, annualizing the present value of an ordinary annuity probably understates the liability because the actual benefit payments are usually made monthly or bi-weekly rather than annually.

Year	1	2	3	4	5
Benefit attributed to:					
Prior years	€0.00	€10,357.09	€20,714.18	€31,071.27	€41,428.36
Current year	10,357.09	10,357.09	10,357.09	10,357.09	10,357.10
Total benefits earned	€10,357.09	€20,714.18	€31,071.27	€41,428.36	€51,785.46
Opening obligation	€0.00	€8,203.79	€17,392.03	€27,653.32	€39,083.36
Interest at 6 percent	0.00	492.23	1,043.52	1,659.20	2,345.00
Current service costs	8,203.79	8,696.01	9,217.77	9,770.84	10,357.10
Closing obligation	€8,203.79	€17,392.03	€27,653.32	€39,083.36	€51,785.46

In this scenario, the pension obligation at the end of Year 3 is €27,653.32 and the portion of pension expense (pension costs reported in P&L) attributable to interest and current service costs for Year 3 is €10,261.29 (= €1,043.52 + €9,217.77). The total pension expense would include other items such as a reduction for return on plan assets.

Scenario 4: Employee to receive benefit payments for 20 years and is given credit for 10 years of prior service with immediate vesting.

Estimated annual payment (end of year) for each of the 20 years = (Estimated final salary × Benefit formula) × Years of service = (€60,198.56 × 0.015) × (10 + 5) = €13,544.68.

Value at the end of Year 5 (retirement date) of the estimated future payments = PV of €13,544.68 for 20 years at 6 percent = €155,356.41.

Annual unit credit = Value at retirement date/Years of service = €155,356.41/15 = €10,357.09.

Year	1	2	3	4	5
Benefit attributed to:					
Prior years	€103,570.94	€113,928.03	€124,285.12	€134,642.21	€144,999.30
Current year	10,357.09	10,357.09	10,357.09	10,357.09	10,357.11
Total benefits earned	€113,928.03	€124,285.12	€134,642.21	€144,999.30	€155,356.41
Opening obligation[a]	€77,394.23	€90,241.67	€104,352.18	€119,831.08	€136,791.79
Interest at 6 percent	4,643.65	5,414.50	6,261.13	7,189.87	8,207.51
Current service costs	8,203.79	8,696.01	9,217.77	9,770.84	10,357.11
Closing obligation	€90,241.67	€104,352.18	€119,831.08	€136,791.79	€155,356.41

[a] This is treated as past service costs in Year 1 because there was no previous recognition and there is immediate vesting.

EXAMPLE 3

The Effect of a Change in Assumptions

Based on Scenario 4 of Example 2 (10 years of prior service and the employee receives benefits for 20 years after retirement):

1 What is the effect on the Year 1 closing pension obligation of a 100 basis point increase in the assumed discount rate—that is, from 6 percent to 7 percent? What is the effect on pension cost in Year 1?

2 What is the effect on the Year 1 closing pension obligation of a 100 basis point increase in the assumed annual compensation increase—that is, from 4.75 percent to 5.75 percent? Assume this is independent of the change in Question 1.

Solution to 1:

The estimated final salary and the estimated annual payments after retirement are unchanged at €60,198.56 and €13,544.68, respectively. However, the value at the retirement date is changed. Value at the end of Year 5 (retirement date) of the estimated future payments = PV of €13,544.68 for 20 years at 7 percent = €143,492.53. Annual unit credit = Value at retirement date/Years of service = €143,492.53/15 = €9,566.17.

Year	1
Benefit attributed to:	
Prior years	€95,661.69
Current year	9,566.17
Total benefits earned	€105,227.86
Opening obligation[a]	€68,205.46
Interest at 7 percent	4,774.38
Current service costs	7,297.99
Closing obligation	€80,277.83

[a] Opening obligation = Benefit attributed to prior years discounted for the remaining time to retirement at the assumed discount rate = $95{,}661.69/(1 + 0.07)^5$.

A 100 basis point increase in the assumed discount rate (from 6 percent to 7 percent) will *decrease* the Year 1 closing pension obligation by €90,241.67 – €80,277.83 = €9,963.84. The Year 1 pension cost declined from €12,847.44 (= 4,643.65 + 8,203.79) to €12,072.37 (= 4,774.38 + 7,297.99). The change in the interest component is a function of the decline in the opening obligation (which will decrease the interest component) and the increased discount rate (which will increase the interest component). In this case, the increase in the discount rate dominated and the interest component increased. The current service costs and the opening obligation both declined because of the increase in the discount rate.

Solution to 2:

The estimated final salary is [€50,000 × $[(1 + 0.0575)^4]$ = €62,530.44. Estimated annual payment for each of the 20 years = (Estimated final salary × Benefit formula) × Years of service = (€62,530.44 × 0.015) × (10 + 5) = €14,069.35. Value at the end of Year 5 (retirement date) of the estimated future payments = PV of €14,069.35 for 20 years at 6 percent = €161,374.33. Annual unit credit = Value at retirement date/Years of service = €161,374.33/15 = €10,758.29.

Year	**1**
Benefit attributed to:	
Prior years	€107,582.89
Current year	10,758.29
Total benefits earned	€118,341.18
Opening obligation	€80,392.19
Interest at 6 percent	4,823.53
Current service costs	8,521.57
Closing obligation	€93,737.29

A 100 basis point increase in the assumed annual compensation increase (from 4.75 percent to 5.75 percent) will *increase* the pension obligation by €93,737.29 – €90,241.67 = €3,495.62.

Example 3 illustrates that an increase in the assumed discount rate will *decrease* a company's pension obligation. In the Solution to 1, there is a slight increase in the interest component of the pension obligation and periodic pension cost (from €4,643.65 in Scenario 4 of Example 2 to €4,774.38 in Example 3). Depending on the pattern and duration of the annual benefits being projected, however, it is possible that the amount of the interest component could decrease because the decrease in the opening obligation may more than offset the effect of the increase in the discount rate.

Example 3 also illustrates that an increase in the assumed rate of annual compensation increase will *increase* a company's pension obligation when the pension formula is based on the final year's salary. In addition, a higher assumed rate of annual compensation increase will increase the service components and the interest component of a company's periodic pension cost because of an increased annual unit credit and the resulting increased obligation. An increase in life expectancy also will increase the pension obligation unless the promised pension payments are independent of life expectancy—for example, paid as a lump sum or over a fixed period.

Finally, under US GAAP, because the expected return on plan assets reduces periodic pension costs reported in P&L, a higher expected return will decrease pension cost reported in P&L (pension expense). Exhibit 3 summarizes the impact of some key estimates on the balance sheet and the periodic pension cost.

Exhibit 3 Impact of Key DB Pension Assumptions on Balance Sheet and Periodic Costs

Assumption	**Impact of Assumption on Balance Sheet**	**Impact of Assumption on Periodic Cost**
Higher discount rate.	Lower obligation.	Periodic pension costs will typically be lower because of lower opening obligation and lower service costs.
Higher rate of compensation increase.	Higher obligation.	Higher service costs.

Exhibit 3 (Continued)

Assumption	Impact of Assumption on Balance Sheet	Impact of Assumption on Periodic Cost
Higher expected return on plan assets.	No effect, because fair value of plan assets is used on balance sheet.	Not applicable for IFRS. Lower periodic pension expense under US GAAP.

Accounting for other post-employment benefits also requires assumptions and estimates. For example, assumed trends in health care costs are an important component of estimating costs of post-employment health care plans. A higher assumed medical expense inflation rate will result in a higher post-employment medical obligation. Companies also estimate various patterns of health care cost trend rates—for example, higher in the near term but becoming lower after some point in time. For post-employment health plans, an increase in the assumed inflationary trends in health care costs or an increase in life expectancy will increase the obligation and associated periodic expense of these plans.

The sections above have explained how the amounts to be reported on the balance sheet are calculated, how the various components of periodic pension cost are reflected in income, and how changes in assumptions can affect pension-related amounts. The next section evaluates disclosures of pension and other post-employment benefits, including disclosures about key assumptions.

9 DISCLOSURES OF PENSION AND OTHER POST-EMPLOYMENT BENEFITS: ASSUMPTIONS

e explain and calculate how adjusting for items of pension and other post-employment benefits that are reported in the notes to the financial statements affects financial statements and ratios;

f interpret pension plan note disclosures including cash flow related information;

Several aspects of the accounting for pensions and other post-employment benefits described above can affect comparative financial analysis using ratios based on financial statements.

- Differences in key assumptions can affect comparisons across companies.
- Amounts disclosed in the balance sheet are net amounts (plan liabilities minus plan assets). Adjustments to incorporate gross amounts would change certain financial ratios.
- Periodic pension costs recognized in P&L (pension expense) may not be comparable. IFRS and US GAAP differ in their provisions about costs recognised in P&L versus in OCI.
- Reporting of periodic pension costs in P&L may not be comparable. Under US GAAP, all of the components of pension costs in P&L are reported in operating expense on the income statement even though some of the components are

of a financial nature (specifically, interest expense and the expected return on assets). However, under IFRS, the components of periodic pension costs in P&L can be included in various line items.

- Cash flow information may not be comparable. Under IFRS, some portion of the amount of contributions might be treated as a financing activity rather than an operating activity; under US GAAP, the contribution is treated as an operating activity.

Information related to pensions can be obtained from various portions of the financial statement note disclosures, and appropriate analytical adjustments can be made. In the following sections, we examine pension plan note disclosures and highlight analytical issues related to each of the points listed above.

9.1 Assumptions

Companies disclose their assumptions about discount rates, expected compensation increases, medical expense inflation, and—for US GAAP companies—expected return on plan assets. Comparing these assumptions over time and across companies provides a basis to assess any conservative or aggressive biases. Some companies also disclose the effects of a change in their assumptions.

Exhibit 4 presents the assumed discount rates (Panel A) and assumed annual compensation increases (Panel B) to estimate pension obligations for four companies operating in the automotive and equipment manufacturing sector. Fiat S.p.A. (an Italy-based company) and the Volvo Group[9] (a Sweden-based company) use IFRS. General Motors and Ford Motor Company are US-based companies that use US GAAP. All of these companies have both US and non-US defined benefit pension plans, which facilitates comparison.

Exhibit 4

Panel A. Assumed discount rates used to estimate pension obligations (percent)

	2009	2008	2007	2006	2005
Fiat S.p.A. (Italy)	5.02	5.10	4.70	3.98	3.53
The Volvo Group (Sweden)	4.00	4.50	4.50	4.00	4.00
General Motors (non-US plans)	5.31	6.22	5.72	4.76	4.72
Ford Motor Company (non-US plans)	5.93	5.58	5.58	4.91	4.58
Fiat S.p.A. (US plans)	5.50	5.10	5.80	5.80	5.50
The Volvo Group (US plans)	4.00–5.75	5.75–6.25	5.75–6.25	5.50	5.75
General Motors (US plans)	5.52	6.27	6.35	5.90	5.70
Ford Motor Company (US plans)	6.50	6.25	6.25	5.86	5.61

Panel B. Assumed annual compensation increases used to estimate pension obligations (percent)

	2009	2008	2007	2006	2005
Fiat S.p.A. (Italy)	4.02	4.65	4.60	3.65	2.58
The Volvo Group (Sweden)	3.00	3.50	3.20	3.20	3.20

9 The Volvo Group primarily manufactures trucks, buses, construction equipment, and engines and engine components for boats, industry, and aircraft. The Volvo car division was sold to Ford Motor Company in 1999, and Ford sold Volvo Car Corporation to the Zhejiang Geely Holding Group in 2010.

Exhibit 4 (Continued)

Panel B. Assumed annual compensation increases used to estimate pension obligations (percent)

	2009	2008	2007	2006	2005
General Motors (non-US plans)	3.23	3.59	3.60	3.00	3.10
Ford Motor Company (non-US plans)	3.13	3.21	3.21	3.30	3.44
Fiat S.p.A. (US plans)*	na	na	na	na	na
The Volvo Group (US plans)	3.00	3.50	3.50	3.50	3.50
General Motors (US plans)	3.94	5.00	5.25	5.00	4.90
Ford Motor Company (US plans)	3.80	3.80	3.80	3.80	4.00

* In the United States, Fiat has obligations to former employees under DB pension plans but no longer offers DB plans. As a result, annual compensation increases are not applicable (na).

The assumed discount rates used to estimate pension obligations are generally based on the market interest rates of high-quality corporate fixed-income investments with a maturity profile similar to the timing of a company's future pension payments. The trend in discount rates across the companies (in both their non-US plans and US plans) is generally similar. In the non-US plans, discount rates increased from 2005 to 2008 and then decreased in 2009 except for Ford, which increased discount rates in 2009. In the US plans, discount rates increased from 2005 to 2007 and held steady or decreased in 2008. In 2009, Fiat and Ford's discount rates increased while Volvo and GM's discount rates decreased. Ford had the highest assumed discount rates for both its non-US and US plans in 2009. Recall that a higher discount rate assumption results in a lower estimated pension obligation. Therefore, the use of a higher discount rate compared with its peers may indicate a less conservative bias.

Explanations for differences in the level of the assumed discount rates, apart from bias, are differences in the regions/countries involved and differences in the timing of obligations (for example, differences in the percentage of employees covered by the DB pension plan that are at or near retirement). In this example, the difference in regions/countries might explain the difference in rates used for the non-US plans but would not explain the difference in the rates shown for the companies' US plans. The timing of obligations under the companies' DB pension plans likely varies, so the relevant market interest rates selected as the discount rate will vary accordingly. Because the timing of the pension obligations is not disclosed, differences in timing cannot be ruled out as an explanation for differences in discount rates.

An important consideration is whether the assumptions are internally consistent. For example, do the company's assumed discount rates and assumed compensation increases reflect a consistent view of inflation? For Volvo, both the assumed discount rates and the assumed annual compensation increases (for both its non-US and US plans) are lower than those of the other companies, so the assumptions appear internally consistent. The assumptions are consistent with plans located in lower-inflation regions. Recall that a lower rate of compensation increase results in a lower estimated pension obligation.

In Ford's US and non-US pension plans, the assumed discount rate is increasing and the assumed rate of compensation increase is decreasing or holding steady in 2009. Each of these will reduce the pension obligation. Therefore, holding all else equal, Ford's pension liability is decreasing because of the higher assumed discount rate and the reduced assumed rate of compensation increase.

Another relevant assumption—for US GAAP companies but not for IFRS companies—is the expected return on pension plan assets. Under US GAAP, a higher expected return on plan assets lowers the periodic pension cost. (Of course, a higher expected return on plan assets presumably reflects riskier investments, so it would not be advisable for a company to simply invest in riskier investments to reduce periodic pension expense.) Because companies are also required to disclose the target asset allocation for their pension plan assets, analysts can assess reasonableness of those assumptions by comparing companies' assumed expected return on plan assets in the context of the plans' asset allocation. For example, a higher expected return is consistent with a greater proportion of plan assets being allocated to riskier asset classes.

Companies with other post-employment benefits also disclose information about these benefits, including assumptions made to estimate the obligation and expense. For example, companies with post-employment health care plans disclose assumptions about increases in health care costs. The assumptions are typically that the inflation rate in health care costs will taper off to some lower, constant rate at some year in the future. That future inflation rate is known as the ultimate health care trend rate. Holding all else equal, each of the following assumptions would result in a higher benefit obligation and a higher periodic cost:

- A higher assumed near-term increase in health care costs,
- A higher assumed ultimate health care cost trend rate, and
- A later year in which the ultimate health care cost trend rate is assumed to be reached.

Conversely, holding all else equal, each of the following assumptions would result in a lower benefit obligation and a lower periodic cost:

- A lower assumed near-term increase in health care costs,
- A lower assumed ultimate health care cost trend rate, and
- An earlier year in which the ultimate health care cost trend rate is assumed to be reached.

Example 4 examines two companies' assumptions about trends in US health care costs.

EXAMPLE 4

Comparison of Assumptions about Trends in US Health Care Costs

In addition to disclosing assumptions about health care costs, companies also disclose information on the sensitivity of the measurements of both the obligation and the periodic cost to changes in those assumptions. Exhibit 5 presents information obtained from the notes to the financial statements for CNH Global N.V. (a Dutch manufacturer of construction and mining equipment) and Caterpillar Inc. (a US manufacturer of construction and mining equipment, engines, and turbines). Each company has US employees for whom they provide post-employment health care benefits.

Panel A shows the companies' assumptions about health care costs and the amounts each reported for post-employment health care benefit plans. For example, CNH assumes that the initial year's (2010) increase in health care costs will be 9 percent, and this rate of increase will decline to 5 percent over the next seven years to 2017. Caterpillar assumes a lower initial-year increase of 7 percent and a decline to the ultimate health care cost trend rate of 5 percent in 2016.

Panel B shows the effect of a 100 basis point increase or decrease in the assumed health care cost trend rates. A 1 percentage point increase in the assumed health care cost trend rates would increase Caterpillar's 2009 service and interest cost component of the other post-employment benefit costs by $23 million and the related obligation by $220 million. A 1 percentage point increase in the assumed health care cost trend rates would increase CNH Global's 2009 service and interest cost component of the other post-employment benefit costs by $8 million and the related obligation by $106 million.

Exhibit 5 Post-Employment Health Care Plan Disclosures

Panel A. Assumptions and Reported Amounts for US Post-Employment Health Care Benefit Plans

	Assumptions about Health Care Costs			Amounts Reported for Other Post-Employment Benefits ($ Millions)	
	Initial Health Care Cost Trend Rate 2010	Ultimate Health Care Cost Trend Rate	Year Ultimate Trend Rate Attained	Accumulated Benefit Obligation Year-End 2009	Periodic Expense for Benefits for 2009
CNH Global N.V.	9.0%	5%	2017	$1,152	$65
Caterpillar Inc.	7.0%	5%	2016	$4,537	$287

Panel B. Effect of 1 Percentage Point Increase (Decrease) in Assumed Health Care Cost Trend Rates on 2009 Total Accumulated Post-Employment Benefit Obligations and Periodic Expense

	1 Percentage Point Increase	1 Percentage Point Decrease
CNH Global N.V.	+$106 million (Obligation) +$8 million (Expense)	–$90 million (Obligation) –$6 million (Expense)
Caterpillar Inc.	+$220 million (Obligation) +$23 million (Expense)	–$186 million (Obligation) –$20 million (Expense)

Sources: Caterpillar information is from the company's Form 10-K filed 19 February 2010, Note 14 (pages A-36 and A-42). CNH Global information is from the company's 2009 Form 20-F, Note 12 (pages F-41, F-43, and F-45).

Based on the information in Exhibit 5, answer the following questions:

1 Which company's assumptions about health care costs appear less conservative?

2 What would be the effect of adjusting the post-employment benefit obligation and the periodic post-employment benefit expense of the less conservative company for a 1 percentage point increase in health care cost trend rates? Does this make the two companies more comparable?

3 What would be the change in each company's 2009 ratio of debt to equity assuming a 1 percentage point increase in the health care cost trend rate? Assume the change would have no impact on taxes. Total liabilities and total equity at 31 December 2009 are given below.

At 31 December 2009 (US$ millions)	CNH Global N.V.	Caterpillar Inc.
Total liabilities	$16,398	$50,738
Total equity	$6,810	$8,823

Solution to 1:

Caterpillar's assumptions about health care costs appear less conservative (the assumptions will result in lower health care costs) than CNH's. Caterpillar's initial assumed health care cost increase of 7 percent is significantly lower than CNH's assumed 9 percent. Further, Caterpillar assumes that the ultimate health care cost trend rate of 5 percent will be reached a year earlier than assumed by CNH.

Solution to 2:

The sensitivity disclosures indicate that a 1 percentage point increase in the assumed health care cost trend rate would increase Caterpillar's post-employment benefit obligation by $220 million and its periodic cost by $23 million. However, Caterpillar's initial health care cost trend rate is 2 percentage points lower than CNH's. Therefore, the impact of a 1 percentage point change for Caterpillar multiplied by 2 provides an approximation of the adjustment required for comparability to CNH. Note, however, that the sensitivity of the pension obligation and expense to a change of more than 1 percentage point in the assumed health care cost trend rate cannot be assumed to be exactly linear, so this adjustment is only an approximation. Further, there may be justifiable differences in the assumptions based on the location of their US operations.

Solution to 3:

A 1 percentage point increase in the health care cost trend rate increases CNH's ratio of debt to equity by about 2 percent, from 2.41 to 2.46. A 1 percentage point increase in the health care cost trend rate increases Caterpillar's ratio of debt to equity by about 3 percent, from 5.75 to 5.92.

CNH Global N.V. ($ millions)	Reported	Adjustment for 1 percentage point increase in health care cost trend rate	Adjusted
Total liabilities	$16,398	+ $106	$16,504
Total equity	$6,810	– $106	$6,704
Ratio of debt to equity	2.41		2.46

Caterpillar Inc. ($ millions)	Reported	Adjustment for 1 percentage point increase in health care cost trend rate	Adjusted
Total liabilities	$50,738	+ $220	$50,958
Total equity	$8,823	– $220	$8,603
Ratio of debt to equity	5.75		5.92

This section has explored the use of pension and other post-employment benefit disclosures to assess a company's assumptions and explore how the assumptions can affect comparisons across companies. The following sections describe the use of disclosures to further analyse a company's pension and other post-employment benefits.

DISCLOSURES OF PENSION AND OTHER POST-EMPLOYMENT BENEFITS: NET PENSION LIABILITY (OR ASSET) AND PERIODIC PENSION COSTS

10

e explain and calculate how adjusting for items of pension and other post-employment benefits that are reported in the notes to the financial statements affects financial statements and ratios;

f interpret pension plan note disclosures including cash flow related information;

Under both IFRS and US GAAP standards, the amount disclosed in the balance sheet is a net amount. Analysts can use information from the notes to adjust a company's assets and liabilities for the gross amount of the benefit plan assets and the gross amount of the benefit plan liabilities. An argument for making such adjustments is that they reflect the underlying economic liabilities and assets of a company; however, it should be recognised that actual consolidation is precluded by laws protecting a pension or other benefit plan as a separate legal entity.

At a minimum, an analyst will compare the gross benefit obligation (i.e., the benefit obligation without deducting related plan assets) with the sponsoring company's total assets, including the gross amount of the benefit plan assets, shareholders' equity, and earnings. Although presumably infrequent in practice, if the gross benefit obligation is large relative to these items, a small change in the pension liability can have a significant financial impact on the sponsoring company.

10.1 Total Periodic Pension Costs

The total periodic cost of a company's DB pension plan is the change in the net pension liability or asset—excluding the effect of the employer's periodic contribution into the plan. To illustrate this point, assume a company has a completely new DB pension plan. At inception, the net pension liability equals \$0 (\$0 plan assets minus \$0 obligations). In the first period, the plan obligation increases by \$500 because of service costs. If the employer makes no contribution to the plan, then the net pension liability would increase to \$500 (\$0 plan assets minus \$500 obligations) and the periodic service costs would be exactly equal to that change. If, however, the employer contributes \$500 to the plan in that period, then the net pension liability would remain at \$0 (\$500 plan assets minus \$500 obligations). In this situation, although the change in net pension liability is \$0, the periodic pension cost is \$500.

Thus, the total periodic pension cost in a given period is calculated by summing the periodic components of cost or, alternatively, by adjusting the change in the net pension liability or asset for the amount of employer contributions. The relationship between the periodic pension cost and the plan's funded status can be expressed as Periodic pension cost = Ending funded status – Employer contributions – Beginning funded status.[10]

Note that, unlike employer contributions into the plan's assets, the payment of cash out of a DB plan to a retiree does not affect the net pension liability or asset. Payment of cash out of a DB plan to a retiree reduces plan assets and plan obligations in an equal amount.

10 Note that a net pension liability is treated as a negative funded status in this relationship.

10.2 Periodic Pension Costs Recognised in P&L vs. OCI

Each period, the components of periodic pension cost—other than any amounts that qualify for capitalisation as part of the costs of such assets as inventories—are recognised either in P&L (an expense) or in OCI. To understand the total pension cost of the period, an analyst should thus consider the amounts shown both in P&L and in OCI.

IFRS and US GAAP differ in their provisions about which periodic pension costs are recognised in P&L versus in OCI. These differences can be relevant to an analyst in comparing the reported profitability of companies that use different sets of standards. Under IFRS, P&L for the period includes both current and past service costs; in contrast, under US GAAP, P&L for the period includes only current service costs (and any amortisation of past service costs.) Under IFRS, P&L incorporates a return on plan assets set equal to the discount rate used in estimating the pension obligation; in contrast, under US GAAP, P&L incorporates an expected return on plan assets. Under US GAAP, P&L may show the impact of amortising actuarial gains or losses that were recognised in previous periods' OCI. Under IFRS, P&L would not show any similar impact because amortising amounts from OCI into P&L is not permitted.

An analyst comparing an IFRS-reporting company with a US GAAP–reporting company could adjust the reported amounts of P&L to achieve comparability. For example, the analyst could adjust the US GAAP company's P&L to make it similar to an IFRS company by including past service costs arising during the period, excluding amortisation of past service costs arising in previous periods, and including an amount of return on plan assets at the discount rate rather than the expected rate. Alternatively, the analyst could use comprehensive income (net income from P&L plus OCI) as the basis for comparison.

10.3 Classification of Periodic Pension Costs Recognised in P&L

Amounts of periodic pension costs recognised in P&L (pension expense) are generally treated as operating expenses. An issue with the reported periodic pension expense is that conceptually the components of this expense could be classified as operating and/or non-operating expenses. It can be argued that only the current service cost component is an operating expense, whereas the interest component and asset returns component are both non-operating. The interest expense component of pension expense is conceptually similar to the interest expense on any of the company's other liabilities. The pension liability is essentially equivalent to borrowing from employees, and the interest expense of that borrowing can be considered a financing cost. Similarly, the return on pension plan assets is conceptually similar to returns on any of the company's other financial assets. These classification issues apply equally to OPB costs.

To better reflect a company's operating performance, an adjustment can be made to operating income by adding back the full amount of pensions costs reported in the P& L (pension expense) and then subtracting only the service costs (or the total of service costs and settlements and curtailments). Note that this adjustment excludes from operating income the amortisation of past service costs and the amortisation of net actuarial gains and losses. This adjustment also eliminates the interest expense component and the return on plan assets component from the company's operating income. The interest expense component would be added to the company's interest expense, and the return on plan assets would be treated as non-operating income.

In addition to adjusting for the classification of different components of pension costs, an adjustment can be made to incorporate the *actual return* on plan assets. Recall that under IFRS, the net interest expense/income calculation effectively includes a return on plan assets calculated using the discount rate used to determine the present value of the pension liability and any difference from the actual return is shown as a component of OCI. Under US GAAP, the *expected* return on plan assets is included as a component of periodic pension cost in P&L and any difference between the actual and expected return is shown as a component of OCI. Under either set of standards, an adjustment can incorporate the actual return. This adjustment changes net income and potentially introduces earnings volatility. The reclassification of interest expense would not change net income. Example 5 illustrates adjustments to operating and non-operating incomes.

EXAMPLE 5

Adjusting Periodic Costs Expensed to P&L and Reclassifying Components between Operating and Non-Operating Income

SABMiller plc is a UK-based company that brews and distributes beer and other beverages. The following information was taken from the company's 2010 Annual Report. Note that in 2010, IFRS required the use of expected return on plan assets, similar to US GAAP. All amounts are in millions of US dollars.

Summary information from the Consolidated Income Statement
For the year ended 31 March 2010

Revenue	$18,020
Net operating expenses	(15,401)
Operating profit	2,619
Interest payable and similar charges*	(879)
Interest receivable and similar income*	316
Share of post-tax results of associates	873
Profit before taxation	$2,929

* *Note*: This is the terminology used in the income statement. The solution to question 2 below uses *interest expense* and *interest and investment income*.

Excerpt from Note 31: Pensions and post-retirement benefits

	Pension	OPB	Total
Current service costs	$(8)	$(3)	$(11)
Interest costs	(29)	(10)	(39)
Expected return on plan assets	14		14
Total	$(23)	$(13)	$(36)
Actual return (loss) on plan assets	$47		

(Components of the amount recognised in net operating expenses for pension and other post-retirement benefits.)

Based on the information above,

1 Adjust pre-tax income for the actual rather than expected return on plan assets.

2 Adjust the individual line items on the company's income statement to reclassify the components of the pension and other post-retirement benefits expense as operating expense, interest expense, or interest income.

Solution to 1:

The total amount of periodic pension cost reported in P&L as an expense is $23. If the actual return on plan assets of $47 is used instead of the expected return on plan assets, the total P&L expense (income) will be $(10) [(= 8 + 29 – 47) or (= 23 + 14 – 47)]. Use of the actual rather than expected return on plan assets provides an estimate of the economic expense (income) for the pension. Profit before taxation adjusted for actual rather than expected return on plan assets will be higher by $33 ($47 – $14) and will total $2,962.

Solution to 2:

All adjustments are summarized below.

	Reported	Adjustments	Adjusted
Revenue	$18,020		$18,020
Net operating expenses	–15,401	+ 36 – 11[a]	–15,376
Operating profit	2,619		2,644
Interest expense	–879	– 39[b]	–918
Interest and investment income	316	+ 47[c]	363
Share of post-tax results of associates	873		873
Profit before taxation	$2,929	$33	$2,962

[a] Operating income is adjusted to include only the current service costs. The $36 total of pension and OPB expenses are excluded from operating expenses, and only the $11 current service cost component is included in operating expenses.
[b] The $39 interest cost component is reclassified as interest expense.
[c] The *actual* return on plan assets is added as investment income.

11 DISCLOSURES OF PENSION AND OTHER POST-EMPLOYMENT BENEFITS: CASH FLOW INFORMATION

e explain and calculate how adjusting for items of pension and other post-employment benefits that are reported in the notes to the financial statements affects financial statements and ratios;

f interpret pension plan note disclosures including cash flow related information;

For a sponsoring company, the cash flow impact of pension and other post-employment benefits is the amount of contributions that the company makes to fund the plan—or for plans without funding requirements, the amount of benefits paid. The amount of contributions a company makes to fund a pension or other post-employment benefit plan is partially determined by the regulations of the countries in which the company operates. In the United States, for example, the amount of contributions to DB pension plans is governed by ERISA (the Employee Retirement and Income Security Act) and depends on the funded status of the plan. Companies may choose to make contributions in excess of those required by regulation.

If a sponsoring company's periodic contributions to a plan exceed the total pension costs of the period, the excess can be viewed from an economic perspective as a reduction of the pension obligation. The contribution covers not only the pension obligation arising in the current period but also the pension obligations of another period. Such a contribution would be similar in concept to making a principal payment on a loan in excess of the scheduled principal payment. Conversely, a periodic contribution that is less than the total pension cost of the period can be viewed as a source of financing. Where the amounts of benefit obligations are material, an analyst may choose to adjust the cash flows that a company presents in its statement of cash flows. Example 6 describes such an adjustment.

EXAMPLE 6

Adjusting Cash Flow

Vassiliki Doukas is analysing the cash flow statement of a hypothetical company, GeoRace plc, as part of a valuation. Doukas suggests to her colleague, Dimitri Krontiras, that the difference between the company's contributions to the pension plan and the total pension costs incurred during a period is similar to a form of borrowing or a repayment of borrowing, depending on the direction of the difference; this affects the company's reported cash from operating activities and cash from financing activities. Based on information from the company's 2009 annual report (currency in £ millions), she determines that the company's total pension cost was £437; however, the company also disclosed that it made a contribution of £504. GeoRace reported cash inflow from operating activities of £6,161 and cash outflow from financing activities of £1,741. The company's effective tax rate was 28.7 percent.

Use the information provided to answer the following questions:

1. How did the company's 2009 contribution to the pension plan compare with the total pension cost for the year?
2. How would cash from operating activities and financing activities be adjusted to illustrate Doukas' interpretation of the difference between the company's contribution and the total pension cost?

Solution to 1:

The company's contribution to the pension plan in 2009 was £504, which was £67 more than the total pension cost of £437. The £67 difference is approximately £48 on an after-tax basis, using the effective tax rate of 28.7 percent.

Total pension costs	£437	
Company's contribution	£504	
Amount by which the sponsoring company's contribution exceeds total pension cost (pre-tax)	£67	
Tax rate	28.7%	
After-tax amount by which the sponsoring company's contribution exceeds total pension cost	£48	[= £67 × (1 − 0.2870)]

Solution to 2:

The company's contribution to the pension plan in 2009 was £67 (£48 after tax) greater than the 2009 total pension cost. Interpreting the excess contribution as similar to a repayment of borrowing (financing use of funds) rather than as an operating cash flow would increase the company's cash outflow from financing activities by £48, from £1,741 to £1,789, and increase the cash inflow from operations by £48, from £6,161 to £6,209.

12 SHARE-BASED COMPENSATION

g explain issues associated with accounting for share-based compensation;

In this section, we provide an overview of executive compensation other than pension plans and other post-retirement benefits, focusing on share-based compensation. First, we briefly discuss common components of executive compensation packages, their objectives, and advantages and disadvantages of share-based compensation. The discussion of share-based compensation then moves to accounting for and reporting of stock grants and stock options. The explanation includes a discussion of fair value accounting, the choice of valuation models, the assumptions used, common disclosures, and important dates in measuring and reporting compensation expense.

Employee compensation packages are structured to achieve varied objectives, including satisfying employees' needs for liquidity, retaining employees, and motivating employees. Common components of employee compensation packages are salary, bonuses, non-monetary benefits, and share-based compensation.[11] The salary component provides for the liquidity needs of an employee. Bonuses, generally in the form of cash, motivate and reward employees for short- or long-term performance or goal achievement by linking pay to performance. Non-monetary benefits, such as medical care, housing, and cars, may be provided to facilitate employees performing their jobs. Salary, bonuses, and non-monetary benefits are short-term employee benefits.

Share-based compensation is intended to align employees' interests with those of the shareholders and is typically a form of deferred compensation. Both IFRS and US GAAP[12] require a company to disclose in their annual report key elements of management compensation. Regulators may require additional disclosure. The disclosures enable analysts to understand the nature and extent of compensation, including the share-based payment arrangements that existed during the reporting period. Below are examples of descriptions of the components and objectives of executive compensation programs for companies that report under IFRS and under US GAAP. Exhibit 6 shows excerpts of the disclosure for the executive compensation program of SABMiller plc; SABMiller plc reports under IFRS and includes a nine-page remuneration report as part of its annual report.

11 An extensive overview of different employee compensation mechanisms can be found in Lynch and Perry (2003).

12 IAS 24 *Related Party Disclosures*, paragraph 17; FASB ASC Section 718-10-50 [Compensation-Stock Compensation-Overall-Disclosure].

Exhibit 6

Excerpts from Remuneration Report of SABMiller plc

… On balance, the committee concluded that its policy of agreeing a total remuneration package for each executive director comprising an annual base salary, a short-term incentive in the form of an annual cash bonus, long-term incentives through participation in share incentive plans, pension contributions, other usual security and health benefits, and benefits in kind, continued to be appropriate.…

The committee's policy continues to be to ensure that executive directors and members of the executive committee are rewarded for their contribution to the group's operating and financial performance at levels which take account of industry, market and country benchmarks, and that their remuneration is appropriate to their scale of responsibility and performance, and will attract, motivate and retain individuals of the necessary calibre. The committee takes account of the need to be competitive in the different parts of the world in which the company operates.…

The committee considers that alignment with shareholders' interests and linkage to SABMiller's long-term strategic goals is best achieved through a twin focus on earnings per share and, from 2010 onwards, additional value created for shareholders, and a blend of absolute and relative performance.

Source: SABMiller plc, Annual Report 2010.

In the United States, similar disclosures are required in a company's proxy statement that is filed with the SEC. Exhibit 7 shows the disclosure of American Eagle Outfitters, Inc.'s executive compensation program, including a description of the key elements and objectives.

Exhibit 7 Excerpts from Executive Compensation Disclosures of American Eagle Outfitters, Inc.

Compensation Program Elements

Our executive compensation program is designed to place a sizeable amount of pay at risk for all executives and this philosophy is intended to cultivate a pay-for-performance environment. Our executive compensation plan design has six key elements:

- Base Salary
- Annual Incentive Bonus
- Long-term Incentive Cash Plan—in place for the Chief Executive Officer and Vice Chairman, Executive Creative Director only
- Restricted Stock ("RS")—issued as Units ("RSUs") and Awards ("RSAs")
- Performance Shares ("PS")
- Non-Qualified Stock Options ("NSOs")

(continued)

Exhibit 7 (Continued)

Two of the elements (Annual Incentive Bonus and LTICP) were entirely "at risk" based on the Company's performance in Fiscal 2009 and were subject to forfeiture if the Company did not achieve threshold performance goals. Performance Shares are entirely "at risk" and subject to forfeiture if the Company does not achieve threshold performance goals by the close of Fiscal 2011, as described below. At threshold performance, the CEO's total annual compensation declines by 46% relative to target performance. The NEO's total annual compensation declines by an average of 33% relative to target performance. Company performance below threshold levels results in forfeiture of all elements of direct compensation other than base salary, RSUs and NSOs. NSOs provide compensation only to the extent that vesting requirements are satisfied and our share price appreciates.

We strategically allocate compensation between short-term and long-term components and between cash and equity in order to maximize executive performance and retention. Long-term compensation and equity awards comprise an increasingly larger proportion of total compensation as position level increases. The portion of total pay attributable to long-term incentive cash and equity compensation increases at successively higher levels of management. This philosophy ensures that executive compensation closely aligns with changes in stockholder value and achievement of performance objectives while also ensuring that executives are held accountable for results relative to position level.

Source: American Eagle Outfitters, Inc. Proxy Statement (Form Def 14A) filed 26 April 2010.

Share-based compensation, in addition to theoretically aligning the interests of employees (management) with shareholders, has the advantage of potentially requiring no cash outlay.[13] Share-based compensation arrangements can take a variety of forms, including those that are equity-settled and those that are cash-settled. However, share-based compensation is treated as an expense and thus as a reduction of earnings even when no cash changes hands. In addition to decreasing earnings through compensation expense, stock options have the potential to dilute earnings per share.

Although share-based compensation is generally viewed as motivating employees and aligning managers' interests with those of the shareholders, there are several disadvantages of share-based compensation. One disadvantage is that the recipient of the share-based compensation may have limited influence over the company's market value (consider the scenario of overall market decline), so share-based compensation does not necessarily provide the desired incentives. Another disadvantage is that the increased ownership may lead managers to be risk averse. In other words, fearing a large market value decline (and loss in individual wealth), managers may seek less risky (and less profitable) projects. An opposite effect, excessive risk taking, can also occur with the awarding of options. Because options have skewed payouts that reward excessive risk taking, managers may seek more risky projects. Finally, when share-based compensation is granted to employees, existing shareholders' ownership is diluted.

For financial reporting, a company reports compensation expense during the period in which employees earn that compensation. Accounting for cash salary payments and cash bonuses is relatively straightforward. When the employee has earned the salary or bonus, an expense is recorded. Typically, compensation expense for managers is reported in sales, general, and administrative expenses on the income statement.

13 Although issuing employee stock options requires no initial cash outlay, the company implicitly forgoes issuing new shares of stock at the then-current market price (and receiving cash) when the options are exercised.

Share-based compensation is more varied and includes such items as stock, stock options, stock appreciation rights, and phantom shares. By granting shares or share options in addition to other compensation, companies are paying additional compensation for services rendered by employees. Under both IFRS and US GAAP, companies use the fair value of the share-based compensation granted to measure the value of the employees' services for purposes of reporting compensation expense. However, the specifics of the accounting depend on the type of share-based compensation given to the employee. Under both IFRS and US GAAP, the usual disclosures required for share-based compensation include (1) the nature and extent of share-based compensation arrangements during the period, (2) how the fair value of a share-based compensation arrangement was determined, and (3) the effect of share-based compensation on the company's income for the period and on its financial position.

Two common forms of equity-settled share-based compensation, stock grants and stock options, are discussed below.

STOCK GRANTS

13

h explain how accounting for stock grants and stock options affects financial statements, and the importance of companies' assumptions in valuing these grants and options.

A company can grant stock to employees outright, with restrictions, or contingent on performance. For an outright stock grant, compensation expense is reported on the basis of the fair value of the stock on the grant date—generally the market value at grant date. Compensation expense is allocated over the period benefited by the employee's service, referred to as the service period. The employee service period is presumed to be the current period unless there are some specific requirements, such as three years service in the future, before the employee is vested (has the right to receive the compensation).

Another type of stock award is a restricted stock, which requires the employee to return ownership of those shares to the company if certain conditions are not met. Common restrictions include the requirements that employees remain with the company for a specified period or that certain performance goals are met. Compensation expense for restricted stock grants is measured as the fair value (usually market value) of the shares issued at the grant date. This compensation expense is allocated over the employee service period.

Shares granted contingent on meeting performance goals are called performance shares. The amount of the grant is usually determined by performance measures other than the change in stock price, such as accounting earnings or return on assets. Basing the grant on accounting performance addresses employees' potential concerns that the stock price is beyond their control and thus should not form the basis for compensation. However, performance shares can potentially have the unintended impact of providing incentives to manipulate accounting numbers. Compensation expense is equal to the fair value (usually market value) of the shares issued at the grant date. This compensation expense is allocated over the employee service period.

14 STOCK OPTIONS

h explain how accounting for stock grants and stock options affects financial statements, and the importance of companies' assumptions in valuing these grants and options.

Like stock grants, compensation expense related to option grants is reported at fair value under both IFRS and US GAAP. Both require that fair value be estimated using an appropriate valuation model.

Whereas the fair value of stock grants is usually based on the market value at the date of the grant, the fair value of option grants must be estimated. Companies cannot rely on market prices of options to measure the fair value of employee stock options because features of employee stock options typically differ from traded options. To measure the fair value of employee stock options, therefore, companies must use a valuation model. The choice of valuation or option pricing model is one of the critical elements in estimating fair value. Several models are commonly used, such as the Black–Scholes option pricing model or a binomial model. Accounting standards do not prescribe a particular model. Generally, though, the valuation method should (1) be consistent with fair value measurement, (2) be based on established principles of financial economic theory, and (3) reflect all substantive characteristics of the award.

Once a valuation model is selected, a company must determine the inputs to the model, typically including exercise price, stock price volatility, estimated life of each award, estimated number of options that will be forfeited, dividend yield, and the risk-free rate of interest.[14] Some inputs, such as the exercise price, are known at the time of the grant. Other critical inputs are highly subjective—such as stock price volatility or the estimated life of stock options—and can greatly change the estimated fair value and thus compensation expense. Higher volatility, a longer estimated life, and a higher risk-free interest rate increase the estimated fair value, whereas a higher assumed dividend yield decreases the estimated fair value.

Combining different assumptions with alternative valuation models can significantly affect the fair value of employee stock options. Below is an excerpt from GlaxoSmithKline, plc explaining the assumptions and model used in valuing its stock options. (Although not discussed in the disclosure, from 2007 to 2009 the trends of decreasing interest rates, lower share price, and increasing dividend yield would decrease estimated fair values and thus lower option expense. In contrast, the trend of increasing volatility would increase the estimated fair values.)

Exhibit 8 Assumptions Used in Stock Option Pricing Models: Excerpts from Financial Statements of GlaxoSmithKline, plc

Note 42—Employee share schemes *[excerpt]*

Option pricing

For the purposes of valuing options and awards to arrive at the share based payment charge, the Black–Scholes option pricing model has been used. The assumptions used in the model for 2007, 2008 and 2009 are as follows:

14 The estimated life of an option award incorporates such assumptions as employee turnover and is usually shorter than the expiration period.

Exhibit 8 (Continued)

	2009	2008	2007
Risk-free interest rate	1.4% – 2.9%	1.3% – 4.8%	4.7% – 5.3%
Dividend yield	5.20%	4.80%	4.00%
Volatility	23% – 29%	19% – 24%	17% – 25%
Expected lives of options granted under:			
Share option schemes	5 years	5 years	5 years
Savings-related share option and share award schemes	3–4 years	3 years	3 years
Weighted average share price for grants in the year:			
Ordinary Shares	£11.72	£11.59	£14.41
ADS*	$33.73	$45.02	$57.59

* American Depositary Shares

Volatility is determined based on the three and five year share price history where appropriate. The fair value of performance share plan grants take into account market conditions. Expected lives of options were determined based on weighted average historic exercises of options.

Source: GlaxoSmithKline Annual Report 2009.

In accounting for stock options, there are several important dates, including the grant date, the vesting date, the exercise date, and the expiration date. The **grant date** is the day that options are granted to employees. The **service period** is usually the period between the grant date and the vesting date.

The **vesting date** is the date that employees can first exercise the stock options. The vesting can be immediate or over a future period. If the share-based payments vest immediately (i.e., no further period of service is required), then expense is recognised on the grant date. If the share-based awards do not vest until a specified service period is completed, compensation expense is recognised and allocated over the service period. If the share-based awards are conditional upon the achievement of a performance condition or a market condition (i.e., a target share price), then compensation expense is recognised over the estimated service period. The **exercise date** is the date when employees actually exercise the options and convert them to stock. If the options go unexercised, they may expire at some pre-determined future date, commonly 5 or 10 years from the grant date.

The grant date is also usually the date that compensation expense is measured if both the number of shares and the option price are known. If facts affecting the value of options granted depend on events after the grant date, then compensation expense is measured at the exercise date. In the example below, Coca Cola, Inc. reported, in the 2009 Form 10-K, $241 million of compensation expense from option grants.

EXAMPLE 7

Disclosure of Stock Options' Current Compensation Expense, Vesting, and Future Compensation Expense

Using information from Coca Cola, Inc.'s Note 9 to financial statements, given below, determine the following:

1 Total compensation expense relating to options already granted that will be recognised in future years as options vest.

2 Approximate compensation expense in 2010 and 2011 relating to options already granted.

Excerpts from Note 9: Stock Compensation Plans in the Notes to Financial Statements of Coca Cola, Inc.

NOTE 9: STOCK COMPENSATION PLANS

Our Company grants stock options and restricted stock awards to certain employees of the Company. Total stock-based compensation expense was approximately $241 million in 2009, $266 million in 2008 and $313 million in 2007 and was included as a component of selling, general and administrative expenses in our consolidated statements of income. The total income tax benefit recognized in our consolidated statements of income for share-based compensation arrangements was approximately $68 million, $72 million and $91 million for 2009, 2008 and 2007, respectively.

As of December 31, 2009, we had approximately $335 million of total unrecognised compensation cost related to nonvested share-based compensation arrangements granted under our plans. This cost is expected to be recognized over a weighted-average period of 1.7 years as stock-based compensation expense. This expected cost does not include the impact of any future stock-based compensation awards.

Source: Coca Cola, Inc. Form 10-K filed 26 February 2010.

Solution to 1:

Coca Cola, Inc. discloses that unrecognised compensation expense relating to stock options already granted but not yet vested totals $335 million.

Solution to 2:

The options already granted will vest over the next 1.7 years. Compensation expense related to stock options already granted will be $197 million ($335/1.7 years) in 2010 and $138 million in 2011 ($335 total less $197 expensed in 2010). New options granted in the future will likely raise the total reported compensation expense.

As the option expense is recognised over the relevant vesting period, the impact on the financial statements is to ultimately reduce retained earnings (as with any other expense). The offsetting entry is an increase in paid-in capital. Thus, the recognition of option expense has no net impact on total equity.

OTHER TYPES OF SHARE -BASED COMPENSATION

15

g explain issues associated with accounting for share-based compensation;

Both stock grants and stock options allow the employee to obtain direct ownership in the company. Other types of share-based compensation, such as stock appreciation rights (SARs) or phantom stock, compensate an employee on the basis of changes in the value of shares without requiring the employee to hold the shares. These are referred to as cash-settled share-based compensation. With SARs, an employee's compensation is based on increases in a company's share price. Like other forms of share-based compensation, SARs serve to motivate employees and align their interests with shareholders. The following are two additional advantages of SARs:

- The potential for risk aversion is limited because employees have limited downside risk and unlimited upside potential similar to employee stock options, and
- Shareholder ownership is not diluted.

Similar to other share-based compensation, SARs are valued at fair value and compensation expense is allocated over the service period of the employee. While phantom share plans are similar to other types of share-based compensation, they differ somewhat because compensation is based on the performance of hypothetical stock rather than the company's actual stock. Unlike SARs, phantom shares can be used by private companies or business units within a company that are not publicly traded or by highly illiquid companies.

SUMMARY

This reading discussed two different forms of employee compensation: post-employment benefits and share-based compensation. Although different, the two are similar in that they are forms of compensation outside of the standard salary arrangements. They also involve complex valuation, accounting, and reporting issues. Although IFRS and US GAAP are converging on accounting and reporting, it is important to note that differences in a country's social system, laws, and regulations can result in differences in a company's pension and share-based compensation plans that may be reflected in the company's earnings and financial reports.

Key points include the following:

- Defined contribution pension plans specify (define) only the amount of contribution to the plan; the eventual amount of the pension benefit to the employee will depend on the value of an employee's plan assets at the time of retirement.
- Balance sheet reporting is less analytically relevant for defined contribution plans because companies make contributions to defined contribution plans as the expense arises and thus no liabilities accrue for that type of plan.
- Defined benefit pension plans specify (define) the amount of the pension benefit, often determined by a plan formula, under which the eventual amount of the benefit to the employee is a function of length of service and final salary.
- Defined benefit pension plan obligations are funded by the sponsoring company contributing assets to a pension trust, a separate legal entity. Differences exist in countries' regulatory requirements for companies to fund defined benefit pension plan obligations.

- Both IFRS and US GAAP require companies to report on their balance sheet a pension liability or asset equal to the projected benefit obligation minus the fair value of plan assets. The amount of a pension asset that can be reported is subject to a ceiling.
- Under IFRS, the components of periodic pension cost are recognised as follows: Service cost is recognised in P&L, net interest income/expense is recognised in P&L, and remeasurements are recognised in OCI and are not amortised to future P&L.
- Under US GAAP, the components of periodic pension cost recognised in P&L include current service costs, interest expense on the pension obligation, and expected returns on plan assets (which reduces the cost). Other components of periodic pension cost—including past service costs, actuarial gains and losses, and differences between expected and actual returns on plan assets—are recognised in OCI and amortised to future P&L.
- Estimates of the future obligation under defined benefit pension plans and other post-employment benefits are sensitive to numerous assumptions, including discount rates, assumed annual compensation increases, expected return on plan assets, and assumed health care cost inflation.
- Employee compensation packages are structured to fulfill varied objectives, including satisfying employees' needs for liquidity, retaining employees, and providing incentives to employees.
- Common components of employee compensation packages are salary, bonuses, and share-based compensation.
- Share-based compensation serves to align employees' interests with those of the shareholders. It includes stocks and stock options.
- Share-based compensation has the advantage of requiring no current-period cash outlays.
- Share-based compensation expense is reported at fair value under IFRS and US GAAP.
- The valuation technique, or option pricing model, that a company uses is an important choice in determining fair value and is disclosed.
- Key assumptions and input into option pricing models include such items as exercise price, stock price volatility, estimated life of each award, estimated number of options that will be forfeited, dividend yield, and the risk-free rate of interest. Certain assumptions are highly subjective, such as stock price volatility or the expected life of stock options, and can greatly change the estimated fair value and thus compensation expense.

REFERENCES

Lynch, L.J., and S.E. Perry. 2003. "An Overview of Management Compensation." *Journal of Accounting Education*, vol. 21, no. 1 (1st Quarter):43–60.

PRACTICE PROBLEMS

The following information relates to Questions 1–7

Kensington plc, a hypothetical company based in the United Kingdom, offers its employees a defined benefit pension plan. Kensington complies with IFRS. The assumed discount rate that the company used in estimating the present value of its pension obligations was 5.48 percent. Information on Kensington's retirement plans is presented in Exhibit 1.

Exhibit 1 Kensington plc Defined Benefit Pension Plan

(in millions)	**2010**
Components of periodic benefit cost	
Service cost	£228
Net interest (income) expense	273
Remeasurements	–18
Periodic pension cost	£483
Change in benefit obligation	
Benefit obligations at beginning of year	£28,416
Service cost	228
Interest cost	1,557
Benefits paid	–1,322
Actuarial gain or loss	0
Benefit obligations at end of year	£28,879
Change in plan assets	
Fair value of plan assets at beginning of year	£23,432
Actual return on plan assets	1,302
Employer contributions	693
Benefits paid	–1,322
Fair value of plan assets at end of year	£24,105
Funded status at beginning of year	–£4,984
Funded status at end of year	–£4,774

1 At year-end 2010, £28,879 million represents:

© 2019 CFA Institute. All rights reserved.

A the funded status of the plan.

B the defined benefit obligation.

C the fair value of the plan's assets.

2 For the year 2010, the net interest expense of £273 represents the interest cost on the:

A ending benefit obligation.

B beginning benefit obligation.

C beginning net pension obligation.

3 For the year 2010, the remeasurement component of Kensington's periodic pension cost represents:

A the change in the net pension obligation.

B actuarial gains and losses on the pension obligation.

C actual return on plan assets minus the amount of return on plan assets included in the net interest expense.

4 Which of the following is *closest* to the actual rate of return on beginning plan assets and the rate of return on beginning plan assets that is included in the interest income/expense calculation?

A The actual rate of return was 5.56 percent, and the rate included in interest income/expense was 5.48 percent.

B The actual rate of return was 1.17 percent, and the rate included in interest income/expense was 5.48 percent.

C Both the actual rate of return and the rate included in interest income/expense were 5.48 percent.

5 Which component of Kensington's periodic pension cost would be shown in OCI rather than P&L?

A Service cost

B Net interest (income) expense

C Remeasurements

6 The relationship between the periodic pension cost and the plan's funded status is *best* expressed in which of the following?

A Periodic pension cost of –£483 = Ending funded status of –£4,774 – Employer contributions of £693 – Beginning funded status of –£4,984.

B Periodic pension cost of £1,322 = Benefits paid of £1,322.

C Periodic pension cost of £210 = Ending funded status of –£4,774 – Beginning funded status of –£4,984.

7 An adjustment to Kensington's statement of cash flows to reclassify the company's excess contribution for 2010 would *most likely* entail reclassifying £210 million (excluding income tax effects) as an outflow related to:

A investing activities rather than operating activities.

B financing activities rather than operating activities.

C operating activities rather than financing activities.

The following information relates to Questions 8–12

XYZ SA, a hypothetical company, offers its employees a defined benefit pension plan. Information on XYZ's retirement plans is presented in Exhibit 1. It also grants stock options to executives. Exhibit 2 contains information on the volatility assumptions used to value stock options.

Exhibit 1 XYZ SA Retirement Plan Information 2009

Employer contributions	1,000
Current service costs	200
Past service costs	120
Discount rate used to estimate plan liabilities at beginning of year	7.00%
Benefit obligation at beginning of year	42,000
Benefit obligation at end of year	41,720
Actuarial loss due to increase in plan obligation	460
Plan assets at beginning of year	39,000
Plan assets at end of year	38,700
Actual return on plan assets	2,700
Expected rate of return on plan assets	8.00%

Exhibit 2 XYZ SA Volatility Assumptions Used to Value Stock Option Grants

Grant Year	Weighted Average Expected Volatility
2009 valuation assumptions	
2005–2009	21.50%
2008 valuation assumptions	
2004–2008	23.00%

8 The total periodic pension cost is *closest* to:

A 320.

B 1,020.

C 1,320.

9 The amount of periodic pension cost that would be reported in P&L under IFRS is *closest* to:

A 28.

B 538.

C 1,020.

10 Assuming the company chooses not to immediately recognise the actuarial loss and assuming there is no amortisation of past service costs or actuarial gains and losses, the amount of periodic pension cost that would be reported in P&L under US GAAP is *closest* to:

A 28.

B 59.

C 530.

11 Under IFRS, the amount of periodic pension cost that would be reported in OCI is *closest* to:

A 20.

B 490.

C 1,020.

12 Compared to 2009 net income as reported, if XYZ had used the same volatility assumption for its 2009 option grants that it had used in 2008, its 2009 net income would have been:

A lower.

B higher.

C the same.

The following information relates to Questions 13–18

Stereo Warehouse is a US retailer that offers employees a defined benefit pension plan and stock options as part of its compensation package. Stereo Warehouse prepares its financial statements in accordance with US GAAP.

Peter Friedland, CFA, is an equity analyst concerned with earnings quality. He is particularly interested in whether the discretionary assumptions the company is making regarding compensation plans are contributing to the recent earnings growth at Stereo Warehouse. He gathers information from the company's regulatory filings regarding the pension plan assumptions in Exhibit 1 and the assumptions related to option valuation in Exhibit 2.

Exhibit 1 Assumptions Used for Stereo Warehouse Defined Benefit Plan

	2009	2008	2007
Expected long-term rate of return on plan assets	6.06%	6.14%	6.79%
Discount rate used to estimate PBO at beginning of year	4.85	4.94	5.38
Estimated future salary increases	4.00	4.44	4.25
Inflation	3.00	2.72	2.45

Exhibit 2 Option Valuation Assumptions

	2009	2008	2007
Risk-free rate	4.6%	3.8%	2.4%
Expected life	5.0 yrs	4.5 yrs	5.0 yrs
Dividend yield	1.0%	0.0%	0.0%
Expected volatility	29%	31%	35%

13 Compared to the 2009 reported financial statements, if Stereo Warehouse had used the same expected long-term rate of return on plan assets assumption in 2009 as it used in 2007, its year-end 2009 pension obligation would *most likely* have been:

A lower.

B higher.

C the same.

14 Compared to the reported 2009 financial statements, if Stereo Warehouse had used the same discount rate as it used in 2007, it would have *most likely* reported lower:

A net income.

B total liabilities.

C cash flow from operating activities.

15 Compared to the assumptions Stereo Warehouse used to compute its periodic pension cost in 2008, earnings in 2009 were *most favorably* affected by the change in the:

A discount rate.

B estimated future salary increases.

C expected long-term rate of return on plan assets.

16 Compared to the pension assumptions Stereo Warehouse used in 2008, which of the following pairs of assumptions used in 2009 is *most likely* internally inconsistent?

A Estimated future salary increases, inflation

B Discount rate, estimated future salary increases

C Expected long-term rate of return on plan assets, discount rate

17 Compared to the reported 2009 financial statements, if Stereo Warehouse had used the 2007 volatility assumption to value its employee stock options, it would have *most likely* reported higher:

A net income.

B compensation expense.

C deferred compensation liability.

18 Compared to the assumptions Stereo Warehouse used to value stock options in 2008, earnings in 2009 were most favorably affected by the change in the:

A expected life.

B risk-free rate.

C dividend yield.

The following information relates to question 19–25

The board of directors at Sallie-Kwan Industrials (SKI), a publicly traded company, is meeting with various committees following the release of audited financial statements prepared in accordance with US GAAP. The finance committee (FC) is next on the agenda to review retirement benefits funding and make recommendations to the board.

SKI's three retirement benefit plans are described as follows:

Plan A

- Benefit: Annual payments for life equal to 1% of the employee's final salary for each year of service beyond the date of the plan's establishment
- The employer makes regular contributions to the plan in order to meet the future obligation
- Closed to new participants; benefits accrue for existing participants
- Fair value of assets: €5.98 billion
- Present value of obligation: €4.80 billion
- Present value of reductions in future contributions: €1.50 billion
- Ten-year vesting schedule; 70% of the participants are fully vested

Plan B

- Benefit: Discretionary retirement withdrawals; amounts depend on the plan's investment performance
- Employer makes its agreed-upon contribution to the plan on behalf of the employee in the same period during which the employee provides the service; SKI is current on this obligation
- The employee may also contribute to the plan during employment years
- Available to all employees after one year of service; 80% of the employees are fully vested

Plan C

- Benefit: Medical, prescription drug, and dental coverage for the retiree, spouse, and dependents under age 18
- 80% funded
- Available to all employees on day one of service

The FC chair reviews Plan A's funded status and the amount recorded on the balance sheet with the board, explaining that the current service cost change from last quarter has primarily resulted from a higher percentage of employees that are expected to leave before the full vesting period.

A board member inquires how Plan A's periodic pension costs affect SKI's operating performance. The FC chair reviews the adjustments needed to account for individual pension components that are considered operating costs and those considered non-operating costs, when calculating profit before taxation. Note 16 in the income statement lists the following: current service costs of €40 million, interest costs of €263 million, expected return on plan assets of €299 million, and actual return on plan assets of €205 million. Moreover, Note 16 indicates that SKI was required to use expected return on plan assets for the reporting year in question.

Next, the FC chairman presents the following case study data to illustrate SKI's current pension obligation for an average fully vested participant in Plan A with 10 years of prior service:

- Current annual salary: €100,000
- Years to retirement: 17
- Retirement life expectancy: 20 years
- Current plan assumptions:
  - Annual compensation increase: 6%
  - Discount rate: 4%
  - Compensation increases are awarded on the first day of the service year; no adjustments are made to reflect the possibility that the employee may leave the firm at an earlier date.

A discussion ensues regarding the effect on the pension obligation, for an average participant, of changing Plan A's annual compensation increase to 5%.

Lastly, the FC chair recommends that the board consider modifying some key assumptions affecting Plan A in response to recent market trends. The chair also reviews how these changes will alter SKI's plan obligation.

Recommendation 1: Change the assumed discount rate to 5%.

Recommendation 2: Increase the retirement life expectancy assumption by eight years.

Recommendation 3: Reduce investment risk by decreasing the expected return to 3%.

19 The participant bears the greatest amount of investment risk under which plan?

A Plan A

B Plan B

C Plan C

20 The plan for which the amount of SKI's financial obligation is defined in the current period with no obligation for future retirement benefits is:

A Plan A.

B Plan B.

C Plan C.

21 For Plan A, SKI should report a net pension:

A asset of €1.50 billion.

B asset of €1.18 billion.

C liability of €1.18 billion.

22 Based on the FC chair's explanation about the current service cost change, the present value of Plan A's obligation:

A decreased.

B stayed the same.

C increased.

23 Based on Note 16, after reclassifying pension components to reflect economic income or expense, the net adjustment to profit before taxation is:

A –€205 million.

B –€94 million.

C +€129 million.

24 Based on the case study illustration and the effect of changing the annual compensation rate, the annual unit credit for the average participant would decrease by an amount *closest* to:

A €4,349.

B €4,858.

C €5,446.

25 All else being equal, which of the following FC recommendations will increase the plan's obligation?

A Recommendation 1

B Recommendation 2

C Recommendation 3

The following information relates to question 26–31

Natalie Holmstead, a senior portfolio manager, works with Daniel Rickards, a junior analyst. Together they are evaluating the financial statements of Company XYZ (XYZ) with a focus on post-employment benefits. XYZ has a defined benefit pension plan and prepares financial statements according to IFRS requirements.

Rickards calculates the current service cost for a single employee's defined benefit pension obligation using the projected unit credit method. The employee is expected to work for 7 years before retiring and has 15 years of vested service. Rickards assumes a discount rate of 4.00% and a lump sum value of the employee's benefit at retirement of $393,949.

Next, Holmstead and Rickards discuss the present value of the defined benefit obligation (PVDBO). Rickards makes the following statements to clarify his understanding:

Statement 1 An actuarial loss is accompanied by an increase in the PVDBO.

Statement 2 The PVDBO measures the present value of future benefits earned by plan participants and includes plan assets.

Statement 3 The company should use the expected long-term rate of return on plan assets as the discount rate to calculate the PVDBO.

XYZ's pension plan offers benefits based on the employee's final year's salary. Rickards calculates the PVDBO as of the end of the current period, based on the information presented in Exhibit 1.

Exhibit 1 Select XYZ Defined Benefit Pension Plan Data

	Current Period	Prior Period
Assumed future compensation growth rate	2.5%	3.0%
Plan assets (in $ millions)	3,108	
Net pension liability (in $ millions)	525	
Present value of reductions of future contributions (in $ millions)	48	

Rickards adjusts the balance sheet and cash flow statement information presented in Exhibit 2 to better reflect the economic nature of certain items related to the pension plan.

Exhibit 2 Select XYZ Balance Sheet and Cash Flow Data (in $ millions)

Item	Current Period
Total assets	24,130
Total liabilities	17,560
Total equity	6,570
Total pension cost	96
Pension contribution	66
Financing cash flow	2,323
Operating cash flow	–1,087
Effective tax rate	30%

Finally, Rickards examines the data in Exhibit 3 and calculates the effect of a 100-basis-point increase in health care inflation on XYZ's debt-to-equity ratio.

Exhibit 3 Sensitivity of Accumulated Post-Employment Benefit Obligations to Changes in Assumed Health Care Inflation (in $ millions)

Item	100-bp Increase	100-bp Decrease
Benefit obligation change	$93	–$76
Benefit expense change	$12	–$10

26 The current service cost is *closest* to:

- **A** $14,152.
- **B** $15,758.
- **C** $17,907.

27 Which of Rickards's statements about the PVDBO is correct?

- **A** Statement 1
- **B** Statement 2
- **C** Statement 3

28 Based on Exhibit 1, the PVDBO is *closest* to:

- **A** $3,585 million.
- **B** $3,633 million.
- **C** $3,681 million.

29 Based on Exhibit 1 and the method XYZ uses to link pension benefits to salaries, the change in the compensation growth rate compared with the prior period will *most likely* result in:

- **A** lower periodic pension cost.

B no change in the periodic pension cost.

C higher periodic pension cost.

30 Based on Exhibit 2, Rickards should adjust the operating and financing cash flows by:

A \$21 million.

B \$30 million.

C \$96 million.

31 Based on Exhibits 2 and 3, as well as Holmstead's assumption about future health care inflation, the debt-to-equity ratio calculated by Rickards for XYZ should be *closest* to:

A 2.69.

B 2.71.

C 2.73.

SOLUTIONS

1 B is correct. The £28,879 million year-end benefit obligation represents the defined benefit obligation.

2 C is correct. The net interest expense of £273 million represents the interest cost on the beginning net pension obligation (beginning funded status) using the discount rate that the company uses in estimating the present value of its pension obligations. This is calculated as –£4,984 million times 5.48 percent = –£273 million; this represents an interest expense on the amount that the company essentially owes the pension plan.

3 C is correct. The remeasurement component of periodic pension cost includes both actuarial gains and losses on the pension obligation and net return on plan assets. Because Kensington does not have any actuarial gains and losses on the pension obligation, the remeasurement component includes only net return on plan assets. In practice, actuarial gains and losses are rarely equal to zero. The net return on plan assets is equal to actual returns minus beginning plan assets times the discount rate, or £1,302 million – (£23,432 million × 0.0548) = £18 million.

4 A is correct. The actual return on plan assets was 1,302/23,432 = 0.0556, or 5.56 percent. The rate of return included in the interest income/expense is the discount rate, which is given in this example as 5.48 percent.

   The rate of 1.17 percent, calculated as the net interest income divided by beginning plan assets, is not used in pension cost calculations.

5 C is correct. Under IFRS, the component of periodic pension cost that is shown in OCI rather than P&L is remeasurments.

6 A is correct. The relation between the periodic pension cost and the plan's funded status can be expressed as Periodic pension cost = Ending funded status – Employer contributions – Beginning funded status.

7 B is correct. Kensington's periodic pension cost was £483. The company's contributions to the plan were £693. The £210 difference between these two numbers can be viewed as a reduction of the overall pension obligation. To adjust the statement of cash flows to reflect this view, an analyst would reclassify the £210 million (excluding income tax effects) as an outflow related to financing activities rather than operating activities.

8 B is correct. The total periodic pension cost is the change in the net pension liability adjusted for the employer's contribution into the plan. The net pension liability increased from 3,000 to 3,020, and the employer's contribution was 1,000. The total periodic pension cost is 1,020. This will be allocated between P&L and OCI.

9 B is correct. Under IFRS, the components of periodic pension cost that would be reported in P&L are the service cost (composed of current service and past service costs) and the net interest expense or income, calculated by multiplying the net pension liability or net pension asset by the discount rate used to measure the pension liability. Here, the service costs are 320 (= 200 + 120) and the net interest expense is 210 [= (42,000 + 120 – 39,000) × 7%]. Thus, the total periodic pension cost is equal to 538.

10 A is correct. Under US GAAP—assuming the company chooses not to immediately recognise the actuarial loss and assuming there is no amortisation of past service costs or actuarial gains and losses—the components of periodic pension cost that would be reported in P&L include the current service cost of 200,

the interest expense on the pension obligation at the beginning of the period of 2,940 (= 7.0% × 42,000), and the expected return on plan assets, which is a reduction of the cost of 3,120 (= 8.0% × 39,000). Summing these three components gives 20.

11 B is correct. The component of periodic pension cost that would be reported in OCI is the remeasurements component. It consists of actuarial gains and losses on the pension obligation and net return on plan assets. Here, the actuarial loss was 460. In addition, the actual return on plan assets was 2,700, which was 30 lower than the return of 2,730 (= 39,000 × 0.07) incorporated in the net interest income/expense. Therefore, the total remeasurements are 490.

12 A is correct. In 2009, XYZ used a lower volatility assumption than it did in 2008. Lower volatility reduces the fair value of an option and thus the reported expense. Using the 2008 volatility estimate would have resulted in higher expense and thus lower net income.

13 C is correct. The assumed long-term rate of return on plan assets is not a component that is used in calculating the pension obligation, so there would be no change.

14 B is correct. A higher discount rate (5.38 percent instead of 4.85 percent) will reduce the present value of the pension obligation (liability). In most cases, a higher discount rate will decrease the interest cost component of the net periodic cost because the decrease in the obligation will more than offset the increase in the discount rate (except if the pension obligation is of short duration). Therefore, periodic pension cost would have been lower and reported net income higher. Cash flow from operating activities should not be affected by the change.

15 B is correct. In 2009, the three relevant assumptions were lower than in 2008. Lower expected salary increases reduce the service cost component of the periodic pension cost. A lower discount rate will increase the defined benefit obligation and increase the interest cost component of the periodic pension cost (the increase in the obligation will, in most cases, more than offset the decrease in the discount rate). Reducing the expected return on plan assets typically increases the periodic pension cost.

16 A is correct. The company's inflation estimate rose from 2008 to 2009. However, it lowered its estimate of future salary increases. Normally, salary increases will be positively related to inflation.

17 B is correct. A higher volatility assumption increases the value of the stock option and thus the compensation expense, which, in turn, reduces net income. There is no associated liability for stock options.

18 C is correct. A higher dividend yield reduces the value of the option and thus option expense. The lower expense results in higher earnings. Higher risk-free rates and expected lives result in higher call option values.

19 B is correct. Plan B is a defined contribution (DC) pension plan because the amount of future benefit is not defined and SKI has an obligation to make only agreed-upon contributions. The actual future benefits depend on the investment performance of the individual's plan assets, and the employee bears the investment risk.

A is incorrect because Plan A is a defined benefit (DB) pension plan. In a DB plan, the amount of future benefit is defined based on the plan's formula (i.e., 1% of the employee's final salary for each year of service). With a DB pension plan, SKI bears the investment risk.

C is incorrect because Plan C is a health care plan and is classified as a DB plan. Under IFRS and US GAAP, all plans for pensions and other post-employment benefits (OPB) other than those explicitly structured as DC plans are classified as DB plans. The amount of future benefit depends on plan specifications and type of benefit, and it represents a promise by the firm to pay benefits in the future. SKI, not the employee, is responsible for estimating future increases in costs, such as health care, over a long time horizon.

20 B is correct. Plan B is a DC pension plan. SKI's financial obligation is defined in each period, and the employer makes its agreed-upon contribution to the plan on behalf of the employee in the same period during which the employee provides the service. SKI is current on this obligation and has no additional financial obligation for the current period.

21 B is correct. SKI's DB pension plan is overfunded by €1.18 billion, the amount by which the fair value of the pension plan assets exceeds the defined benefit obligation (€5.98 billion – €4.80 billion). When a company has a surplus in a DB pension plan, the amount of assets that can be reported is the lower of the surplus or the asset ceiling (the present value of future economic benefits, such as refunds from the plan or reductions in future contributions). In this case, the asset ceiling is given as €1.50 billion, so the amount of SKI's reported net pension asset is the amount of the surplus, because this amount is lower than the asset ceiling.

22 A is correct. A higher percentage of employees is expected to leave before the full 10-year vesting period, which would decrease the present value of the DB obligation. If the employee leaves the company before meeting the 10-year vesting requirement, she may be entitled to none or a portion of the benefits earned up until that point. In measuring the DB obligation, the company considers the probability that some employees may not satisfy the vesting requirements (i.e., may leave before the vesting period) and use this probability to calculate the current service cost and the present value of the obligation.

23 B is correct. Operating income is adjusted to include only the current service costs, the interest cost component is reclassified as interest expense, and the actual return on plan assets is added as investment income. Profit before taxation adjusted for actual rather than expected return on plan assets will decrease by €94 million (205 – 299).

	Total (€ millions)
Current service costs	– €40
Interest costs	– €263
Expected return on plan assets	+ €299
Total of pension and OPB expenses	– €4 million
Actual return (loss) on plan assets	€205 million

Because the actual return on plan assets is less than the expected return on plan assets, operating income will be adjusted downward by 299 – 205 = 94. Alternatively, the adjustments to the individual pension cost components are as follows:

Line Items to Adjust	Adjustments (€ millions)
Revenue	—
Net operating expenses	+4 − 40 = −36
Operating profit	—
Interest expense	−263
Interest and investment income	+205
Share of post-tax results of associates	—
Adjustment to profit before taxation	−€94 million

24 B is correct. The final year's estimated earnings at the end of Year 1 for the average participant would decrease by approximately €35,747.71.

	Current Assumptions	Case Study Assumptions
Current salary	€100,000	€100,000
Years until retirement	17	17
Years of service (includes prior 10)	27	27
Retirement life expectancy	20	20
Annual compensation increases	6%	5%
Discount rate	4%	4%
Final year's estimated earnings	€254,035.17	€218,287.46
Estimated annual payment for each of the 20 years	€68,589.50	€58,937.61
Value at the end of year 17 (retirement date) of the estimated future payments	€932,153.69	€800,981.35
Annual unit credit	€34,524.21	€29,665.98

Because there are now 17 years until retirement, there are 16 years until retirement from the end of Year 1. The final year's estimated earnings, estimated at the end of Year 1, are as follows:

Current year's salary × [(1 + Annual compensation increase)$^{\text{Years until retirement}}$]

Annual compensation increase of 6%: €100,000 × [$(1.06)^{16}$] = €254,035.17

Annual compensation increase of 5%: €100,000 × [$(1.05)^{16}$] = €218,287.46

The estimated annual payment for each of the 20 years (retirement life expectancy) is

(Estimated final salary × Benefit formula) × Years of service

Annual compensation increase of 6%: (€254,035.17 × 0.01) × (10 + 17) = €68,589.50

Annual compensation increase of 5%: (€218,287.46 × 0.01) × (10 + 17) = €58,937.61

The value at the end of Year 17 (retirement date) of the estimated future payments is the PV of the estimated annual payment for each of the 20 years at the discount rate of 4%:

Annual compensation increase of 6%: PV of €68,589.50 for 20 years at 4% = €932,153.69

Annual compensation increase of 5%: PV of €58,937.61 for 20 years at 4% = €800,981.35

The annual unit credit = Value at retirement/Years of service:

Annual compensation increase of 6%: €932,153.69/27 = €34,524.21

Annual compensation increase of 5%: €800,981.35/27 = €29,665.98

The annual unit credit for the average participant would decrease by €34,524.21 − €29,665.98 = €4,858.23.

25 B is correct. An increase in the retirement life expectancy (from 20 to 28 years) will increase the DB pension obligation, because Plan A pays annual payments for life.

26 A is correct. Current service cost is the present value of annual unit credit earned in the current period.

Annual unit credit (benefit) per service year = Value at retirement / Years of service.

Years of service = 15 (vested years of past service) + 7 (expected years until retirement) = 22.

Annual unit credit = \$393,949 / 22 = \$17,906.77.

$$\begin{aligned}&\text{Current service cost (for 1 year)}\\&\quad= \text{Annual unit credit}\Big/\big[(1+\text{Discount rate})^{(\text{Years until retirement at the end of Year 1})}\\&\quad= \$17{,}906.77\Big/(1+0.04)^{6} = \$14{,}151.98.\end{aligned}$$

27 A is correct. To estimate the PVDBO, the company must make a number of assumptions, such as future compensation increases, discount rates, and expected vesting. If changes in assumptions increase the obligation, the increase is referred to as an actuarial loss.

B is incorrect because the PVDBO does not include the value of plan assets in the calculation.

C is incorrect because the expected long-term rate of return on plan assets is not used to calculate the PVDBO. The interest rate used to calculate the PVDBO is based on current rates of return on high-quality corporate bonds (or government bonds, in the absence of a deep market in corporate bonds) with currency and durations consistent with the currency and durations of the benefits.

28 B is correct. The funded status of a pension plan is calculated as follows:

Funded status = Fair value of the plan assets − PVDBO

Based on the information provided in Exhibit 1, the PVDBO is calculated as follows:

PVDBO = Funded status (Net pension liability) + Plan assets

\$525 + \$3,108 = \$3,633 million

29 A is correct. A decrease in the assumed future compensation growth rate will decrease a company's pension obligation when the pension formula is based on the final year's salary. Lowering the assumed future compensation growth rate decreases the service and interest components of periodic pension costs because of a decreased annual unit credit.

30 A is correct. Rickards' task is to adjust the balance sheet and cash flow statement information to better reflect the economic nature of certain items related to the pension plan. When a company's periodic contribution to a plan is lower than the total pension cost of the period, it can be viewed as a source of financing. To reflect this event, the deficit amount is adjusted by the effective tax rate and should be reclassified from an operating cash flow to a financing cash flow. The company's contribution to the pension plan was \$66 million, which is \$30 million less than the pension cost of \$96 million. The \$30 million difference is \$21 million on an after-tax basis, using the effective tax rate of 30%. Therefore, \$21 million should be classified as an operating cash outflow (negative value) and a financing cash inflow (positive value).

31 C is correct. To calculate the debt-to-equity ratio, both liabilities and total equity need to be adjusted for the estimated impact of a 100-bp increase in health care costs. The proposed increase in health care costs will increase total liabilities and decrease equity by the same amount. Consequently, the debt-to-equity ratio changes as follows:

Sensitivity of benefit obligation to 100-bp increase = \$93

Adjusted liabilities = \$17,560 + \$93 = \$17,653

Adjusted equity = \$6,570 − \$93 = \$6,477

Adjusted debt-to-equity ratio = \$17,653/\$6,477 = 2.7255 ≈ 2.73

Consequently, a 100-bp increase in health care costs increases the debt-to-equity ratio to approximately 2.73.

READING

11

Multinational Operations

by Timothy S. Doupnik, PhD, and Elaine Henry, PhD, CFA

Timothy S. Doupnik, PhD (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** compare and contrast presentation in (reporting) currency, functional currency, and local currency;
☐	**b.** describe foreign currency transaction exposure, including accounting for and disclosures about foreign currency transaction gains and losses;
☐	**c.** analyze how changes in exchange rates affect the translated sales of the subsidiary and parent company;
☐	**d.** compare the current rate method and the temporal method, evaluate how each affects the parent company's balance sheet and income statement, and determine which method is appropriate in various scenarios;
☐	**e.** calculate the translation effects and evaluate the translation of a subsidiary's balance sheet and income statement into the parent company's presentation currency;
☐	**f.** analyze how the current rate method and the temporal method affect financial statements and ratios;
☐	**g.** analyze how alternative translation methods for subsidiaries operating in hyperinflationary economies affect financial statements and ratios;
☐	**h.** describe how multinational operations affect a company's effective tax rate;
☐	**i.** explain how changes in the components of sales affect the sustainability of sales growth;
☐	**j.** analyze how currency fluctuations potentially affect financial results, given a company's countries of operation.

© 2013 CFA Institute. All rights reserved.

1 INTRODUCTION AND FOREIGN CURRENCY TRANSACTIONS: FOREIGN CURRENCY TRANSACTION EXPOSURE TO FOREIGN EXCHANGE RISK AND ANALYTICAL ISSUES

a compare and contrast presentation in (reporting) currency, functional currency, and local currency;

b describe foreign currency transaction exposure, including accounting for and disclosures about foreign currency transaction gains and losses;

According to the World Trade Organization, merchandise exports worldwide were nearly US$15 trillion in 2010.[1] The amount of worldwide merchandise exports in 2010 was more than twice the amount in 2003 (US$7.4 trillion) and more than four times the amount in 1993 (US$3.7 trillion). The top five exporting countries in 2010, in order, were China, the United States, Germany, Japan, and the Netherlands. In the United States alone, 293,131 companies were identified as exporters in 2010, but only 2.2% of those companies were large (more than 500 employees).[2] The vast majority of US companies with export activity were small or medium-sized entities.

The point illustrated by these statistics is that many companies engage in transactions that cross national borders. The parties to these transactions must agree on the currency in which to settle the transaction. Generally, this will be the currency of either the buyer or the seller. Exporters that receive payment in foreign currency and allow the purchaser time to pay must carry a foreign currency receivable on their books. Conversely, importers that agree to pay in foreign currency will have a foreign currency account payable. To be able to include them in the total amount of accounts receivable (payable) reported on the balance sheet, these foreign currency denominated accounts receivable (payable) must be translated into the currency in which the exporter (importer) keeps its books and presents financial statements.

The prices at which foreign currencies can be purchased or sold are called foreign exchange rates. Because foreign exchange rates fluctuate over time, the value of foreign currency payables and receivables also fluctuates. The major accounting issue related to foreign currency transactions is how to reflect the changes in value for foreign currency payables and receivables in the financial statements.

Many companies have operations located in foreign countries. For example, the Swiss food products company Nestlé SA reports that it has factories in 83 countries and a presence in almost every country in the world. US-based Procter & Gamble's annual filing discloses more than 400 subsidiaries located in more than 80 countries around the world. Foreign subsidiaries are generally required to keep accounting records in the currency of the country in which they are located. To prepare consolidated financial statements, the parent company must translate the foreign currency financial statements of its foreign subsidiaries into its own currency. Nestlé, for example, must translate the assets and liabilities its various foreign subsidiaries carry in foreign currency into Swiss francs to be able to consolidate those amounts with the Swiss franc assets and liabilities located in Switzerland.

A multinational company like Nestlé is likely to have two types of foreign currency activities that require special accounting treatment. Most multinationals (1) engage in transactions that are denominated in a foreign currency and (2) invest in foreign

1 World Trade Organization, *International Trade Statistics 2011*, Table I4, page 21.

2 US Census Bureau, Department of Commerce. *A Profile of US Importing and Exporting Companies, 2009–2010.* Released 12 April 2012.

subsidiaries that keep their books in a foreign currency. To prepare consolidated financial statements, a multinational company must translate the foreign currency amounts related to both types of international activities into the currency in which the company presents its financial statements.

This reading presents the accounting for foreign currency transactions and the translation of foreign currency financial statements. The conceptual issues related to these accounting topics are discussed, and the specific rules embodied in International Financial Reporting Standards (IFRS) and US GAAP are demonstrated through examples. Fortunately, differences between IFRS and US GAAP with respect to foreign currency translation issues are minimal.

Analysts need to understand the effects of foreign exchange rate fluctuations on the financial statements of a multinational company and how a company's financial statements reflect foreign currency gains and losses, whether realized or not.

1.1 Foreign Currency Transactions

When companies from different countries agree to conduct business with one another, they must decide which currency will be used. For example, if a Mexican electronic components manufacturer agrees to sell goods to a customer in Finland, the two parties must agree whether the Finnish company will pay for the goods in Mexican pesos, euro, or perhaps even a third currency such as the US dollar. If the transaction is denominated in Mexican pesos, the Finnish company has a foreign currency transaction but the Mexican company does not. To account for the inventory being purchased and the account payable in Mexican pesos, the Finnish company must translate the Mexican peso amounts into euro using appropriate exchange rates. Although the Mexican company also has entered into an international transaction (an export sale), it does not have a foreign currency transaction and no translation is necessary. It simply records the sales revenue and account receivable in Mexican pesos, which is the currency in which it keeps its books and prepares financial statements.

The currency in which financial statement amounts are presented is known as the **presentation currency**. In most cases, a company's presentation currency will be the currency of the country where the company is located. Finnish companies are required to keep accounting records and present financial results in euro, US companies in US dollars, Chinese companies in Chinese yuan, and so on.

Another important concept in accounting for foreign currency activities is the **functional currency**, which is the currency of the primary economic environment in which an entity operates. Normally, the functional currency is the currency in which an entity primarily generates and expends cash. In most cases, an organization's functional currency will be the same as its presentation currency. And, because most companies primarily generate and expend cash in the currency of the country where they are located, the functional and presentation currencies are most often the same as the **local currency** where the company operates.

Because the local currency generally is an entity's functional currency, a multinational corporation with subsidiaries in a variety of different countries is likely to have a variety of different functional currencies. The Thai subsidiary of a Japanese parent company, for example, is likely to have the Thai baht as its functional currency, whereas the Japanese parent's functional currency is the Japanese yen. But in some cases, the foreign subsidiary could have the parent's functional currency as its own. For example, prior to its 2011 acquisition of McAfee, Intel Corporation had determined that the US dollar was the functional currency for all of its significant foreign subsidiaries. However, subsequent to the acquisition of McAfee, as stated in Intel Corporation's 2011 Annual Report, Note 1: Basis of Presentation, "Certain of the operations acquired from McAfee have a functional currency other than the US dollar."

By definition, for accounting purposes, a foreign currency is any currency other than a company's functional currency, and **foreign currency transactions** are those denominated in a currency other than the company's functional currency. Foreign currency transactions occur when a company (1) makes an import purchase or an export sale that is denominated in a foreign currency or (2) borrows or lends funds where the amount to be repaid or received is denominated in a foreign currency. In each of these cases, the company has an asset or a liability denominated in a foreign currency.

1.1.1 *Foreign Currency Transaction Exposure to Foreign Exchange Risk*

Assume that FinnCo, a Finland-based company, imports goods from Mexico in January under 45-day credit terms, and the purchase is denominated in Mexican pesos. By deferring payment until April, FinnCo runs the risk that from the date the purchase is made until the date of payment, the value of the Mexican peso might increase relative to the euro. FinnCo would then need to spend more euro to settle its Mexican peso account payable. In this case, FinnCo is said to have an **exposure to foreign exchange risk**. Specifically, FinnCo has a foreign currency **transaction exposure**. Transaction exposure related to imports and exports can be summarized as follows:

- *Import purchase.* A transaction exposure arises when the importer is obligated to pay in foreign currency and is allowed to defer payment until sometime after the purchase date. The importer is exposed to the risk that from the purchase date until the payment date the foreign currency might increase in value, thereby increasing the amount of functional currency that must be spent to acquire enough foreign currency to settle the account payable.
- *Export sale.* A transaction exposure arises when the exporter agrees to be paid in foreign currency and allows payment to be made sometime after the purchase date. The exporter is exposed to the risk that from the purchase date until the payment date, the foreign currency might decrease in value, thereby decreasing the amount of functional currency into which the foreign currency can be converted when it is received.

The major issue in accounting for foreign currency transactions is how to account for the foreign currency risk—that is, how to reflect in the financial statements the change in value of the foreign currency asset or liability. Both IFRS and US GAAP require the change in the value of the foreign currency asset or liability resulting from a foreign currency transaction to be treated as a gain or loss reported on the income statement.[3]

1.1.1.1 Accounting for Foreign Currency Transactions with Settlement before Balance Sheet Date Example 1 demonstrates FinnCo's accounting, assuming that it purchased goods on account from a Mexican supplier that required payment in Mexican pesos, and that it made payment before the balance sheet date. The basic principle is that all transactions are recorded at the spot rate on the date of the transaction. The foreign currency risk on *transactions*, therefore, arises only when the transaction date and the payment date are different.

3 International standards are presented in International Accounting Standard (IAS) 21, "The Effects of Changes in Foreign Exchange Rates," and US GAAP standards are presented in FASB ASC Topic 830, "Foreign Currency Matters."

EXAMPLE 1

Accounting for Foreign Currency Transactions with Settlement before the Balance Sheet Date

FinnCo purchases goods from its Mexican supplier on 1 November 20X1; the purchase price is 100,000 Mexican pesos. Credit terms allow payment in 45 days, and FinnCo makes payment of 100,000 pesos on 15 December 20X1. FinnCo's functional and presentation currency is the euro. Spot exchange rates between the euro (EUR) and Mexican peso (MXN) are as follows:

1 November 20X1	MXN1 = EUR0.0684
15 December 20X1	MXN1 = EUR0.0703

FinnCo's fiscal year end is 31 December. How will FinnCo account for this foreign currency transaction, and what effect will it have on the 20X1 financial statements?

Solution:

The euro value of the Mexican peso account payable on 1 November 20X1 was EUR6,840 (MXN100,000 × EUR0.0684). FinnCo could have paid for its inventory on 1 November by converting 6,840 euro into 100,000 Mexican pesos. Instead, the company purchases 100,000 Mexican pesos on 15 December 20X1, when the value of the peso has increased to EUR0.0703. Thus, FinnCo pays 7,030 euro to purchase 100,000 Mexican pesos. The net result is a loss of 190 euro (EUR7,030 – EUR6,840).

Although the cash outflow to acquire the inventory is EUR7,030, the cost included in the inventory account is only EUR6,840. This cost represents the amount that FinnCo could have paid if it had not waited 45 days to settle its account. By deferring payment, and because the Mexican peso increased in value between the transaction date and settlement date, FinnCo has to pay an additional 190 euro. The company will report a foreign exchange loss of EUR190 in its net income in 20X1. This is a realized loss because FinnCo actually spent an additional 190 euro to purchase its inventory. The net effect on the financial statements, in EUR, can be seen as follows:

Balance Sheet					Income Statement		
Assets		= Liabilities +	Stockholders' Equity		Revenues and Gains	Expenses and Losses	
Cash	–7,030		Retained			Foreign	
Inventory	+6,840		earnings	–190		exchange loss	–190
	–190						

1.1.1.2 Accounting for Foreign Currency Transactions with Intervening Balance Sheet Dates

Another important issue related to the accounting for foreign currency transactions is what, if anything, should be done if a balance sheet date falls between the initial transaction date and the settlement date. For foreign currency transactions whose settlement dates fall in subsequent accounting periods, both IFRS and US GAAP require adjustments to reflect intervening changes in currency exchange rates. Foreign currency transaction gains and losses are reported on the income statement, creating one of the few situations in which accounting rules allow, indeed require, companies to include (recognize) a gain or loss in income before it has been realized.

Subsequent foreign currency transaction gains and losses are recognized from the balance sheet date through the date the transaction is settled. Adding together foreign currency transaction gains and losses for both accounting periods (transaction initiation to balance sheet date and balance sheet date to transaction settlement) produces an amount equal to the actual realized gain or loss on the foreign currency transaction.

EXAMPLE 2

Accounting for Foreign Currency Transaction with Intervening Balance Sheet Date

FinnCo sells goods to a customer in the United Kingdom for £10,000 on 15 November 20X1, with payment to be received in British pounds on 15 January 20X2. FinnCo's functional and presentation currency is the euro. Spot exchange rates between the euro (€) and British pound (£) are as follows:

15 November 20X1	£1 = €1.460
31 December 20X1	£1 = €1.480
15 January 20X2	£1 = €1.475

FinnCo's fiscal year end is 31 December. How will FinnCo account for this foreign currency transaction, and what effect will it have on the 20X1 and 20X2 financial statements?

Solution:

The euro value of the British pound account receivable at each of the three relevant dates is determined as follows:

		Account Receivable (£10,000)	
Date	**€/£ Exchange Rate**	**Euro Value**	**Change in Euro Value**
15 Nov 20X1	€1.460	14,600	N/A
31 Dec 20X1	€1.480	14,800	+ 200
15 Jan 20X2	€1.475	14,750	– 50

A change in the euro value of the British pound receivable from 15 November to 31 December would be recognized as a foreign currency transaction gain or loss on FinnCo's 20X1 income statement. In this case, the increase in the value of the British pound results in a transaction gain of €200 [£10,000 × (€1.48 – €1.46)]. Note that the gain recognized in 20X1 income is unrealized, and remember that this is one of few situations in which companies include an unrealized gain in income.

Any change in the exchange rate between the euro and British pound that occurs from the balance sheet date (31 December 20X1) to the transaction settlement date (15 January 20X2) will also result in a foreign currency transaction gain or loss. In our example, the British pound weakened slightly against the euro during this period, resulting in an exchange rate of €1.475/ £1 on 15 January 20X2. The £10,000 account receivable now has a value of €14,750, which is a decrease of €50 from 31 December 20X1. FinnCo will recognize a foreign currency transaction loss on 15 January 20X2 of €50 that will be included in the company's calculation of net income for the first quarter of 20X2.

From the transaction date to the settlement date, the British pound has increased in value by €0.015 (€1.475 – €1.460), which generates a realized foreign currency transaction gain of €150. A gain of €200 was recognized in 20X1 and a loss of €50 is recognized in 20X2. Over the two-month period, the net gain recognized in the financial statements is equal to the actual realized gain on the foreign currency transaction.

In Example 2, FinnCo's British pound account receivable resulted in a net foreign currency transaction gain because the British pound strengthened (increased) in value between the transaction date and the settlement date. In this case, FinnCo has an asset exposure to foreign exchange risk. This asset exposure benefited the company because the foreign currency strengthened. If FinnCo instead had a British pound account payable, a liability exposure would have existed. The euro value of the British pound account payable would have increased as the British pound strengthened, and FinnCo would have recognized a foreign currency transaction loss as a result.

Whether a change in exchange rate results in a foreign currency transaction gain or loss (measured in local currency) depends on (1) the nature of the exposure to foreign exchange risk (asset or liability) and (2) the direction of change in the value of the foreign currency (strengthens or weakens).

		Foreign Currency	
Transaction	**Type of Exposure**	**Strengthens**	**Weakens**
Export sale	Asset (account receivable)	Gain	Loss
Import purchase	Liability (account payable)	Loss	Gain

A foreign currency receivable arising from an export sale creates an asset exposure to foreign exchange risk. If the foreign currency strengthens, the receivable increases in value in terms of the company's functional currency and a foreign currency transaction gain arises. The company will be able to convert the foreign currency when received into more units of functional currency because the foreign currency has strengthened. Conversely, if the foreign currency weakens, the foreign currency receivable loses value in terms of the functional currency and a loss results.

A foreign currency payable resulting from an import purchase creates a liability exposure to foreign exchange risk. If the foreign currency strengthens, the payable increases in value in terms of the company's functional currency and a foreign currency transaction loss arises. The company must spend more units of functional currency to be able to settle the foreign currency liability because the foreign currency has strengthened. Conversely, if the foreign currency weakens, the foreign currency payable loses value in terms of the functional currency and a gain exists.

1.1.2 *Analytical Issues*

Both IFRS and US GAAP require foreign currency transaction gains and losses to be reported in net income (even if the gains and losses have not yet been realized), but neither standard indicates where on the income statement these gains and losses should be placed. The two most common treatments are either (1) as a component of other operating income/expense or (2) as a component of non-operating income/expense, in some cases as a part of net financing cost. The calculation of operating profit margin is affected by where foreign currency transaction gains or losses are placed on the income statement.

EXAMPLE 3

Placement of Foreign Currency Transaction Gains/Losses on the Income Statement—Effect on Operating Profit

Assume that FinnCo had the following income statement information in both 20X1 and 20X2, excluding a foreign currency transaction gain of €200 in 20X1 and a transaction loss of €50 in 20X2.

	20X1	20X2
Revenues	€20,000	€20,000
Cost of goods sold	12,000	12,000
Other operating expenses, net	5,000	5,000
Non-operating expenses, net	1,200	1,200

FinnCo is deciding between two alternatives for the treatment of foreign currency transaction gains and losses. Alternative 1 calls for the reporting of foreign currency transaction gains/losses as part of "Other operating expenses, net." Under Alternative 2, the company would report this information as part of "Non-operating expenses, net."

FinnCo's fiscal year end is 31 December. How will Alternatives 1 and 2 affect the company's gross profit margin, operating profit margin, and net profit margin for 20X1? For 20X2?

Solution:

Remember that a gain would serve to reduce expenses, whereas a loss would increase expenses.

20X1—Transaction Gain of €200				
	Alternative 1		Alternative 2	
Revenues	€20,000		€20,000	
Cost of goods sold	(12,000)		(12,000)	
Gross profit	8,000		8,000	
Other operating expenses, net	(4,800)	incl. gain	(5,000)	
Operating profit	3,200		3,000	
Non-operating expenses, net	(1,200)		(1,000)	incl. gain
Net profit	€2,000		€2,000	

Profit margins in 20X1 under the two alternatives can be calculated as follows:

	Alternative 1	Alternative 2
Gross profit margin	€8,000/€20,000 = 40.0%	€8,000/€20,000 = 40.0%
Operating profit margin	3,200/20,000 = 16.0%	3,000/20,000 = 15.0%
Net profit margin	2,000/20,000 = 10.0%	2,000/20,000 = 10.0%

20X2—Transaction Loss of €50		
	Alternative 1	Alternative 2
Revenues	€20,000	€20,000
Cost of goods sold	(12,000)	(12,000)

20X2—Transaction Loss of €50				
	Alternative 1		Alternative 2	
Gross profit	8,000		8,000	
Other operating expenses, net	(5,050)	incl. loss	(5,000)	
Operating profit	2,950		3,000	
Non-operating expenses, net	(1,200)		(1,250)	incl. loss
Net profit	€1,750		€1,750	

Profit margins in 20X2 under the two alternatives can be calculated as follows:

	Alternative 1	Alternative 2
Gross profit margin	€8,000/€20,000 = 40.0%	€8,000/€20,000 = 40.0%
Operating profit margin	2,950/20,000 = 14.75%	3,000/20,000 = 15.0%
Net profit margin	1,750/20,000 = 8.75%	1,750/20,000 = 8.75%

Gross profit and net profit are unaffected, but operating profit differs under the two alternatives. In 20X1, the operating profit margin is larger under Alternative 1, which includes the transaction gain as part of "Other operating expenses, net." In 20X2, Alternative 1 results in a smaller operating profit margin than Alternative 2. Alternative 2 has the same operating profit margin in both periods. Because exchange rates do not fluctuate by the same amount or in the same direction from one accounting period to the next, Alternative 1 will cause greater volatility in operating profit and operating profit margin over time.

Because accounting standards do not provide guidance on the placement of foreign currency transaction gains and losses on the income statement, companies are free to choose among the alternatives. Two companies in the same industry could choose different alternatives, which would distort the direct comparison of operating profit and operating profit margins between those companies.

A second issue that should be of interest to analysts relates to the fact that unrealized foreign currency transaction gains and losses are included in net income when the balance sheet date falls between the transaction and settlement dates. The implicit assumption underlying this accounting requirement is that the unrealized gain or loss as of the balance sheet date reflects the company's ultimate net gain or loss. In reality, though, the ultimate net gain or loss may vary dramatically because of the possibility for changes in trend and volatility of currency prices.

This effect was seen in the previous hypothetical Example 2 with FinnCo. Using given currency exchange rate data shows that the real-world effect can also be quite dramatic. Assume that a French company purchased goods from a Canadian supplier on 1 December 20X1, with payment of 100,000 Canadian dollars (C$) to be made on 15 May 20X2. Actual exchange rates between the Canadian dollar and euro (€) during the period 1 December 20X1 and 15 May 20X2, the euro value of the Canadian dollar account payable, and the foreign currency transaction gain or loss are shown below:

		Account Payable (C$100,000)	
	€/C$	€ Value	Change in € Value (Gain/Loss)
1 Dec X1	0.7285	72,850	N/A
31 Dec X1	0.7571	75,710	2,860 loss

(continued)

	€/C$	Account Payable (C$100,000)	
		€ Value	Change in € Value (Gain/Loss)
31 Mar X2	0.7517	75,170	540 gain
15 May X2	0.7753	77,530	2,360 loss

As the Canadian dollar strengthened against the euro in late 20X1, the French company would have recorded a foreign currency transaction loss of €2,860 in the fourth quarter of 20X1. The Canadian dollar reversed course by weakening over the first three months of 20X2, resulting in a transaction gain of €540 in the first quarter, and then strengthened against the euro in the second quarter of 20X2, resulting in a transaction loss of €2,360. At the time payment is made on 15 May 20X2, the French company realizes a net foreign currency transaction loss of €4,680 (€77,530 – €72,850).

2 DISCLOSURES RELATED TO FOREIGN CURRENCY TRANSACTION GAINS AND LOSSES

b describe foreign currency transaction exposure, including accounting for and disclosures about foreign currency transaction gains and losses;

Because accounting rules allow companies to choose where they present foreign currency transaction gains and losses on the income statement, it is useful for companies to disclose both the amount of transaction gain or loss that is included in income and the presentation alternative they have selected. IFRS require disclosure of "the amount of exchange differences recognized in profit or loss," and US GAAP require disclosure of "the aggregate transaction gain or loss included in determining net income for the period," but neither standard specifically requires disclosure of the line item in which these gains and losses are located.

Exhibit 1 provides disclosures from BASF AG's 2011 annual report that the German company made related to foreign currency transaction gains and losses. Exhibit 2 presents similar disclosures found in the Netherlands-based Heineken NV's 2011 Annual Report. Both companies use IFRS to prepare their consolidated financial statements.

BASF's income statement in Exhibit 1 does not include a separate line item for foreign currency gains and losses. From Note 6 in Exhibit 1, an analyst can determine that BASF has chosen to include "Income from foreign currency and hedging transactions" in "Other operating income." Of the total amount of €2,008 million reported as "Other operating income" in 2011, €170 million is attributable to foreign currency and hedging transaction income. It is not possible to determine from BASF's financial statements whether or not these gains were realized in 2011, and any unrealized gain reported in 2011 income might or might not be realized in 2012.

Note 7 in Exhibit 1 indicates that "Expenses from foreign currency and hedging transactions as well as market valuation" in 2011 were €399 million, making up 15% of Other operating expenses. Combining foreign currency transaction gains and losses results in a net loss of €229 million, which is equal to 2.55% of BASF's "Income before taxes and minority interests."

Exhibit 1 Excerpts from BASF AG's 2011 Annual Report Related to Foreign Currency Transactions

Consolidated Statements of Income Million €	Explanation in Notes	2011	2010
Sales	(4)	73,497	63,873
Cost of sales		(53,986)	(45,310)
Gross profit on sales		**19,511**	**18,563**
Selling expenses		(7,323)	(6,700)
General and administrative expenses		(1,315)	(1,138)
Research and development expenses		(1,605)	(1,492)
Other operating income	(6)	2,008	1,140
Other operating expenses	(7)	(2,690)	(2,612)
Income from operations	(4)	**8,586**	**7,761**
(detail omitted)			
Financial result	(8)	**384**	**(388)**
Income before taxes and minority interests		**8,970**	**7,373**
Income taxes	(9)	(2,367)	(2,299)
Income before minority interests		**6,603**	**5,074**
Minority interests	(10)	(415)	(517)
Net income		**6,188**	**4,557**

Notes:

1 Summary of Accounting Policies

Foreign currency transactions: The cost of assets acquired in foreign currencies and revenues from sales in foreign currencies are recorded at the exchange rate on the date of the transaction. Foreign currency receivables and liabilities are valued at the exchange rates on the balance sheet date.

6 Other Operating Income

Million €	2011	2010
Reversal and adjustment of provisions	170	244
Revenue from miscellaneous revenue-generating activities	207	142
Income from foreign currency and hedging transactions	170	136
Income from the translation of financial statements in foreign currencies	42	76
Gains on the disposal of property, plant and equipment and divestitures	666	101
Reversals of impairments of property, plant and equipment	—	40
Gains on the reversal of allowance for doubtful business-related receivables	77	36
Other	676	365
	2,008	1,140

Income from foreign currency and hedging transactions concerned foreign currency transactions, the measurement at fair value of receivables and payables in foreign currencies, as well as currency derivatives and other hedging transactions.

(continued)

Exhibit 1 (Continued)

7 Other Operating Expenses

Million €	**2011**	**2010**
Restructuring measures	233	276
Environmental protection and safety measures, costs of demolition and removal, and planning expenses related to capital expenditures that are not subject to mandatory capitalization	203	98
Valuation adjustments on tangible and intangible assets	366	247
Costs from miscellaneous revenue-generating activities	220	180
Expenses from foreign currency and hedging transactions as well as market valuation	399	601
Losses from the translation of the financial statements in foreign currencies	56	63
Losses from the disposal of property, plant and equipment and divestitures	40	24
Oil and gas exploration expenses	184	190
Expenses from additions to allowances for business-related receivables	124	107
Expenses from the use of inventories measured at market value and the derecognition of obsolete inventory	233	188
Other	632	638
	2,690	2,612

Expenses from foreign currency and hedging transactions as well as market valuation concern foreign currency translations of receivables and payables as well as changes in the fair value of currency derivatives and other hedging transactions.

In Exhibit 2, Heineken's Note 2, Basis of Preparation, part (c) explicitly states that the euro is the company's functional currency. Note 3(b)*(i)* indicates that monetary assets and liabilities denominated in foreign currencies at the balance sheet date are translated to the functional currency and that foreign currency differences arising on the translation (i.e., translation gains and losses) are recognized on the income statement. Note 3(r) discloses that foreign currency gains and losses are included on a net basis in the other net finance income and expenses. Note 12, "Net finance income and expense," shows that a net foreign exchange loss of €107 million existed in 2011 and a net gain of €61 million arose in 2010. The net foreign currency transaction gain in 2010 amounted to 3.1% of Heineken's profit before income tax that year, and the net translation loss in 2011 represented 5.3% of the company's profit before income tax in that year. Note 12 also shows gains and losses related to changes in the fair value of derivatives, some of which related to foreign currency derivatives.

Exhibit 2 Excerpts from Heineken NV's 2011 Annual Report Related to Foreign Currency Transactions

Consolidated Income Statement for the Year Ended 31 December in Millions of EUR	**Note**	**2011**	**2010**
Revenue	5	**17,123**	**16,133**
Other income	8	**64**	**239**
Raw materials, consumables, and services	9	(10,966)	(10,291)
Personnel expenses	10	(2,838)	(2,665)

Exhibit 2 (Continued)

Consolidated Income Statement for the Year Ended 31 December in Millions of EUR	Note	2011	2010
Amortisation, depreciation, and impairments	11	(1,168)	(1,118)
Total expenses		**(14,972)**	**(14,074)**
Results from operating activities		**2,215**	**2,298**
Interest income	12	70	100
Interest expenses	12	(494)	(590)
Other net finance income/(expenses)	12	(6)	(19)
Net finance expenses		**(430)**	**(509)**
Share of profit of associates and joint ventures and impairments thereof (net of income tax)	16	240	193
Profit before income tax		**2,025**	**1,982**
Income tax expenses	13	(465)	(403)
Profit		**1,560**	**1,579**
Attributable to:			
Equity holders of the Company (net profit)		1,430	1,447
Minority interest		130	132
Profit		**1,560**	**1,579**

Notes:

2 Basis of preparation

c Functional and presentation currency

These consolidated financial statements are presented in euro, which is the Company's functional currency. All financial information presented in euro has been rounded to the nearest million unless stated otherwise.

3 Significant accounting policies

b Foreign currency

i. Foreign currency transactions

Transactions in foreign currencies are translated to the respective functional currencies of Heineken entities at the exchange rates at the dates of the transactions. Monetary assets and liabilities denominated in foreign currencies at the reporting date are retranslated to the functional currency at the exchange rate at that date. . . . Foreign currency differences arising on retranslation are recognised in profit or loss, except for differences arising on the retranslation of available-for-sale (equity) investments and foreign currency differences arising on the retranslation of a financial liability designated as a hedge of a net investment, which are recognised in other comprehensive income.[4]

r Interest income, interest expenses and other net finance income and expenses

(continued)

4 Note that this excerpt uses "retranslation" in the same way that "translation" is used throughout the rest of this reading. The translation of currency for foreign subsidiaries will be covered in the next section.

Exhibit 2 (Continued)

…Foreign currency gains and losses are reported on a net basis in the other net finance income and expenses.

12 Net finance income and expense

Recognised in profit or loss

In millions of EUR	**2011**	**2010**
Interest income	**70**	**100**
Interest expenses	**(494)**	**(590)**
Dividend income on available-for-sale investments	2	1
Dividend income on investments held for trading	11	7
Net gain/(loss) on disposal of available-for-sale investments	1	—
Net change in fair value of derivatives	96	(75)
Net foreign exchange gain/(loss)	(107)	61
Impairment losses on available-for-sale investments	—	(3)
Unwinding discount on provisions	(7)	(7)
Other net financial income/(expenses)	(2)	(3)
Other net finance income/(expenses)	**(6)**	**(19)**
Net finance income/(expenses)	**(430)**	**(509)**

Disclosures related to foreign currency are commonly found both in the Management Discussion & Analysis (MD&A) and the Notes to Financial Statements sections of an annual report. In applying US GAAP to account for its foreign currency transactions, Yahoo! Inc. reported the following in the Quantitative and Qualitative Disclosures about Market Risk section of its 2011 annual report:

> Our exposure to foreign currency transaction gains and losses is the result of assets and liabilities, (including inter-company transactions) that are denominated in currencies other than the relevant entity's functional currency.... We may enter into derivative instruments, such as foreign currency forward contracts or other instruments to minimize the short-term foreign currency fluctuations on such assets and liabilities. The gains and losses on the forward contracts may not offset any or more than a portion of the transaction gains and losses on certain foreign currency receivables, investments and payables recognized in earnings. Transaction gains and losses on these foreign exchange contracts are recognized each period in other income, net included on the consolidated statements of income. During the years ended December 31, 2011, 2010, and 2009, we recorded net realized and unrealized foreign currency transaction gains of $9 million and $13 million, and a transaction loss of $1 million, respectively.

Yahoo!'s disclosure clearly explains that both realized and unrealized foreign currency transaction gains and losses are reflected in income, specifically as a part of non-operating activities. The net foreign currency transaction gain in 2011 of $9 million represented only 1.1% of the company's pretax income ($827.5 million) for the year.

Some companies may choose not to disclose either the location or the amount of their foreign currency transaction gains and losses, presumably because the amounts involved are immaterial. There are several reasons why the amount of transaction gains and losses can be immaterial for a company:

1 The company engages in a limited number of foreign currency transactions that involve relatively small amounts of foreign currency.

2 The exchange rates between the company's functional currency and the foreign currencies in which it has transactions tend to be relatively stable.

3 Gains on some foreign currency transactions are naturally offset by losses on other transactions, such that the net gain or loss is immaterial. For example, if a US company sells goods to a customer in Canada with payment in Canadian dollars to be received in 90 days and at the same time purchases goods from a supplier in Canada with payment to be made in Canadian dollars in 90 days, any loss that arises on the Canadian dollar receivable due to a weakening in the value of the Canadian dollar will be exactly offset by a gain of equal amount on the Canadian dollar payable.

4 The company engages in foreign currency hedging activities to offset the foreign exchange gains and losses that arise from foreign currency transactions. Hedging foreign exchange risk is a common practice for many companies engaged in foreign currency transactions.

The two most common types of hedging instruments used to minimize foreign exchange transaction risk are foreign currency forward contracts and foreign currency options. Nokia Corporation describes its foreign exchange risk management approach in its 2011 Form 20-F annual report in Note 34, Risk Management. An excerpt from that note follows:

> Nokia operates globally and is thus exposed to foreign exchange risk arising from various currencies. Foreign currency denominated assets and liabilities together with foreign currency denominated cash flows from highly probable or probable purchases and sales contribute to foreign exchange exposure. These transaction exposures are managed against various local currencies because of Nokia's substantial production and sales outside the Euro zone.
>
> According to the foreign exchange policy guidelines of the Group, which remains the same as in the previous year, material transaction foreign exchange exposures are hedged unless hedging would be uneconomical due to market liquidity and/or hedging cost. Exposures are defined using nominal values of the transactions. Exposures are mainly hedged with derivative financial instruments such as forward foreign exchange contracts and foreign exchange options. The majority of financial instruments hedging foreign exchange risk have duration of less than a year. The Group does not hedge forecasted foreign currency cash flows beyond two years.

Elsewhere in its annual report, Nokia provides additional disclosures about the currencies to which it has exposure and the accounting for different types of hedges. The company also summarizes the effect of material exchange rate movements. For example, the 4.2% appreciation of the US dollar in 2011 had a positive effect on net sales expressed in euro (40% of Nokia's net sales are in US dollars or currencies closely following the US dollar) and a negative effect on product cost (60% of Nokia's components are sourced in US dollars); this resulted in a slightly negative effect on operating profit.

3 TRANSLATION OF FOREIGN CURRENCY FINANCIAL STATEMENTS AND TRANSLATION CONCEPTUAL ISSUES

c analyze how changes in exchange rates affect the translated sales of the subsidiary and parent company;

Many companies have operations in foreign countries. Most operations located in foreign countries keep their accounting records and prepare financial statements in the local currency. For example, the US subsidiary of German automaker BMW AG keeps its books in US dollars. IFRS and US GAAP require parent companies to prepare consolidated financial statements in which the assets, liabilities, revenues, and expenses of both domestic and foreign subsidiaries are added to those of the parent company. To prepare worldwide consolidated statements, parent companies must translate the foreign currency financial statements of their foreign subsidiaries into the parent company's presentation currency. BMW AG, for example, must translate both the US dollar financial statements of its US subsidiary and the South African rand financial statements of its South African subsidiary into euro to consolidate these foreign operations. If, for example, the US dollar and South African rand appreciate against the euro over the course of a given year, the amount of sales translated into euro will be greater than if the subsidiary's currencies weaken against the euro.

IFRS and US GAAP have similar rules for the translation of foreign currency financial statements. To fully understand the results from applying these rules, however, several conceptual issues must first be examined.

3.1 Translation Conceptual Issues

In translating foreign currency financial statements into the parent company's presentation currency, two questions must be addressed:

1 What is the appropriate exchange rate to use in translating each financial statement item?

2 How should the translation adjustment that inherently arises from the translation process be reflected in the consolidated financial statements? In other words, how is the balance sheet brought back into balance?

These issues and the basic concepts underlying the translation of financial statements are demonstrated through the following example.

Spanco is a hypothetical Spain-based company that uses the euro as its presentation currency. Spanco establishes a wholly owned subsidiary, Amerco, in the United States on 31 December 20X1 by investing €10,000 when the exchange rate between the euro and the US dollar is €1 = US$1. The equity investment of €10,000 is physically converted into US$10,000 to begin operations. In addition, Amerco borrows US$5,000 from local banks on 31 December 20X1. Amerco purchases inventory that costs US$12,000 on 31 December 20X1 and retains US$3,000 in cash. Amerco's balance sheet at 31 December 20X1 thus appears as follows:

Amerco Balance Sheet, 31 December 20X1 (in US Dollars)

Cash	$3,000	Notes payable	$5,000
Inventory	12,000	Common stock	10,000
Total	$15,000	Total	$15,000

To prepare a consolidated balance sheet in euro as of 31 December 20X1, Spanco must translate all of the US dollar balances on Amerco's balance sheet at the €1 = US$1 exchange rate. The translation worksheet as of 31 December 20X1 is as follows:

Translation Worksheet for Amerco, 31 December 20X1

	USD	Exchange Rate (€)	EUR
Cash	$3,000	1.00	€3,000
Inventory	12,000	1.00	12,000
Total	$15,000		€15,000
Notes payable	5,000	1.00	5,000
Common stock	10,000	1.00	10,000
Total	$15,000		€15,000

By translating each US dollar balance at the same exchange rate (€1.00), Amerco's translated balance sheet in euro reflects an equal amount of total assets and total liabilities plus equity and remains in balance.

During the first quarter of 20X2, Amerco engages in no transactions. During that period, however, the US dollar weakens against the euro such that the exchange rate on 31 March 20X2 is €0.80 = US$1.

To prepare a consolidated balance sheet at the end of the first quarter of 20X2, Spanco now must choose between the current exchange rate of €0.80 and the historical exchange rate of €1.00 to translate Amerco's balance sheet amounts into euro. The original investment made by Spanco of €10,000 is a historical fact, so the company wants to translate Amerco's common stock in such a way that it continues to reflect this amount. This goal is achieved by translating common stock of US$10,000 into euro using the historical exchange rate of €1 = US$1.

Two approaches for translating the foreign subsidiary's assets and liabilities are as follows:

1 All assets and liabilities are translated at the **current exchange rate** (the spot exchange rate on the balance sheet date).

2 Only **monetary assets and liabilities** are translated at the current exchange rate; **non-monetary assets and liabilities** are translated at **historical exchange rates** (the exchange rates that existed when the assets and liabilities were acquired). Monetary items are cash and receivables (payables) that are to be received (paid) in a fixed number of currency units. Non-monetary assets include inventory, fixed assets, and intangibles, and non-monetary liabilities include deferred revenue.

These two different approaches are demonstrated and the results analyzed in turn.

3.1.1 *All Assets and Liabilities Are Translated at the Current Exchange Rate*

The translation worksheet on 31 March 20X2, in which all assets and liabilities are translated at the current exchange rate (€0.80), is as follows:

Translation Worksheet for Amerco, 31 March 20X2

	US Dollar	Exchange Rate (€)	Euro	Change in Euro Value since 31 Dec 20X1
Cash	$3,000	0.80 C	€2,400	−€600
Inventory	12,000	0.80 C	9,600	−2,400
Total	$15,000		€12,000	−€3,000

(continued)

Translation Worksheet for Amerco, 31 March 20X2

	US Dollar	Exchange Rate (€)	Euro	Change in Euro Value since 31 Dec 20X1
Notes payable	5,000	0.80 C	4,000	−1,000
Common stock	10,000	1.00 H	10,000	0
Subtotal	$15,000		14,000	−1,000
Translation adjustment			(2,000)	−2,000
Total			€12,000	−€3,000

Note: C = current exchange rate; H = historical exchange rate

By translating all assets at the lower current exchange rate, total assets are written down from 31 December 20X1 to 31 March 20X2 in terms of their euro value by €3,000. Liabilities are written down by €1,000. To keep the euro translated balance sheet in balance, a *negative* translation adjustment of €2,000 is created and included in stockholders' equity on the consolidated balance sheet.

Those foreign currency balance sheet accounts that are translated using the current exchange rate are revalued in terms of the parent's functional currency. This process is very similar to the revaluation of foreign currency receivables and payables related to foreign currency transactions. The net translation adjustment that results from translating individual assets and liabilities at the current exchange rate can be viewed as the *net* foreign currency translation gain or loss caused by a change in the exchange rate:

(€600)	loss on cash
(€2,400)	loss on inventory
€1,000	gain on notes payable
(€2,000)	net translation loss

The negative translation adjustment (net translation loss) does not result in a cash outflow of €2,000 for Spanco and thus is unrealized. The loss could be realized, however, if Spanco were to sell Amerco at its book value of US$10,000. The proceeds from the sale would be converted into euro at €0.80 per US$1, resulting in a cash inflow of €8,000. Because Spanco originally invested €10,000 in its US operation, a *realized* loss of €2,000 would result.

The second conceptual issue related to the translation of foreign currency financial statements is whether the unrealized net translation loss should be included in the determination of consolidated net income currently or deferred in the stockholders' equity section of the consolidated balance sheet until the loss is realized through sale of the foreign subsidiary. There is some debate as to which of these two treatments is most appropriate. This issue is discussed in more detail after considering the second approach for translating assets and liabilities.

3.1.2 *Only Monetary Assets and Monetary Liabilities Are Translated at the Current Exchange Rate*

Now assume only monetary assets and monetary liabilities are translated at the current exchange rate. The worksheet at 31 March 20X2, in which only monetary assets and liabilities are translated at the current exchange rate (€0.80), is as follows:

Translation Worksheet for Amerco, 31 March 20X2

	US Dollar	Exchange Rate (€)	Euro	Change in Euro Value since 31 Dec 20X1
Cash	$3,000	0.80 C	€2,400	−€600
Inventory	12,000	1.00 H	12,000	0
Total	$15,000		€14,400	−€600
Notes payable	5,000	0.80 C	4,000	−1,000
Common stock	10,000	1.00 H	10,000	0
Subtotal	$15,000		14,000	−1,000
Translation adjustment			400	400
Total			€14,400	−€600

Note: C = current exchange rate; H = historical exchange rate

Using this approach, cash is written down by €600 but inventory continues to be carried at its euro historical cost of €12,000. Notes payable is written down by €1,000. To keep the balance sheet in balance, a positive translation adjustment of €400 must be included in stockholders' equity. The translation adjustment reflects the *net* translation gain or loss related to monetary items only:

(€600)	loss on cash
€1,000	gain on notes payable
€400	net translation gain

The positive translation adjustment (net translation gain) also is *unrealized*. The gain could be *realized*, however, if:

1. The subsidiary uses its cash (US$3,000) to pay as much of its liabilities as possible, and
2. The parent sends enough euro to the subsidiary to pay its remaining liabilities (US$5,000 − US$3,000 = US$2,000). As of 31 December 20X1, at the €1.00 per US$1 exchange rate, Spanco will have sent €2,000 to Amerco to pay liabilities of US$2,000. On 31 March 20X2, given the €0.80 per US$1 exchange rate, the parent needs to send only €1,600 to pay US$2,000 of liabilities. As a result, Spanco would enjoy a foreign exchange gain of €400.

The second conceptual issue again arises under this approach. Should the unrealized foreign exchange gain be recognized in current period net income or deferred on the balance sheet as a separate component of stockholders' equity? The answer to this question, as provided by IFRS and US GAAP, is described in Section 4, Translation Methods.

3.1.3 *Balance Sheet Exposure*

Those assets and liabilities translated at the *current* exchange rate are revalued from balance sheet to balance sheet in terms of the parent company's presentation currency. These items are said to be *exposed* to translation adjustment. Balance sheet items translated at *historical* exchange rates do not change in parent currency value and therefore are not exposed to translation adjustment. Exposure to translation adjustment is referred to as balance sheet translation exposure, or accounting exposure.

A foreign operation will have a **net asset balance sheet exposure** when assets translated at the current exchange rate are greater than liabilities translated at the current exchange rate. A **net liability balance sheet exposure** exists when liabilities translated at the current exchange rate are greater than assets translated at the current exchange rate. Another way to think about the issue is to realize that there is a net asset balance sheet exposure when exposed assets are greater than exposed liabilities and a net liability balance sheet exposure when exposed liabilities are greater than exposed assets. The sign (positive or negative) of the current period's translation adjustment is a function of two factors: (1) the nature of the balance sheet exposure (asset or liability) and (2) the direction of change in the exchange rate (strengthens or weakens). The relationship between exchange rate fluctuations, balance sheet exposure, and the current period's translation adjustment can be summarized as follows:

Balance Sheet Exposure	Foreign Currency (FC)	
	Strengthens	**Weakens**
Net asset	Positive translation adjustment	Negative translation adjustment
Net liability	Negative translation adjustment	Positive translation adjustment

These relationships are the same as those summarized in Section 1 with respect to foreign currency transaction gains and losses. In reference to the example in Section 3, for instance, the amount of exposed assets (the US$3,000 cash) was less than the amount of exposed liabilities (US$5,000 of notes payable), implying a net liability exposure. Further, in the example the foreign currency (US$) weakened, resulting in a positive translation adjustment.

The combination of balance sheet exposure and direction of exchange rate change determines whether the current period's translation adjustment will be positive or negative. After the initial period of operations, a cumulative translation adjustment is required to keep the translated balance sheet in balance. The cumulative translation adjustment will be the sum of the translation adjustments that arise over successive accounting periods. For example, assume that Spanco translates all of Amerco's assets and liabilities using the current exchange rate (a net asset balance sheet exposure exists), which, because of a weakening US dollar in the first quarter of 20X2, resulted in a negative translation adjustment of €2,000 on 31 March 20X2 (as shown in Section 3). Assume further that in the second quarter of 20X2, the US dollar strengthens against the euro and there still is a net asset balance sheet exposure, which results in a *positive* translation adjustment of €500 for that quarter. Although the current period translation adjustment for the second quarter of 20X2 is positive, the cumulative translation adjustment as of 30 June 20X2 still will be negative, but the amount now will be only €1,500.

4 TRANSLATION METHODS

d compare the current rate method and the temporal method, evaluate how each affects the parent company's balance sheet and income statement, and determine which method is appropriate in various scenarios;

The two approaches to translating foreign currency financial statements described in the previous section are known as (1) the **current rate method** (all assets and liabilities are translated at the current exchange rate), and (2) the **monetary/non-monetary method** (only monetary assets and liabilities are translated at the current exchange rate). A variation of the monetary/non-monetary method requires not only monetary assets and liabilities but also non-monetary assets and liabilities that are measured at their current value on the balance sheet date to be translated at the current exchange rate. This variation of the monetary/non-monetary method sometimes is referred to as the **temporal method**.

The basic idea underlying the temporal method is that assets and liabilities should be translated in such a way that the measurement basis (either current value or historical cost) in the foreign currency is preserved after translating to the parent's presentation currency. To achieve this objective, assets and liabilities carried on the foreign currency balance sheet at a current value should be translated at the current exchange rate, and assets and liabilities carried on the foreign currency balance sheet at historical costs should be translated at historical exchange rates. Although neither the IASB nor the FASB specifically refer to translation methods by name, the procedures specified by IFRS and US GAAP for translating foreign currency financial statements essentially require the use of either the current rate or the temporal method.

Which method is appropriate for an individual foreign entity depends on that entity's functional currency. As noted earlier, the functional currency is the currency of the primary economic environment in which an entity operates. A foreign entity's functional currency can be either the parent's presentation currency or another currency, typically the currency of the country in which the foreign entity is located. Exhibit 3 lists the factors that IFRS indicate should be considered in determining a foreign entity's functional currency. Although not identical, US GAAP provide similar indicators for determining a foreign entity's functional currency.

When the functional currency indicators listed in Exhibit 3 are mixed and the functional currency is not obvious, IFRS indicate that management should use its best judgment in determining the functional currency. In this case, however, indicators 1 and 2 should be given priority over indicators 3 through 9.

Exhibit 3 Factors Considered in Determining the Functional Currency

In accordance with IFRS, the following factors should be considered in determining an entity's functional currency:

1 The currency that mainly influences sales prices for goods and services.
2 The currency of the country whose competitive forces and regulations mainly determine the sales price of its goods and services.
3 The currency that mainly influences labour, material, and other costs of providing goods and services.
4 The currency in which funds from financing activities are generated.
5 The currency in which receipts from operating activities are usually retained.

Additional factors to consider in determining whether the foreign entity's functional currency is the same as the parent's functional currency are

6 Whether the activities of the foreign operation are an extension of the parent's or are carried out with a significant amount of autonomy.
7 Whether transactions with the parent are a large or a small proportion of the foreign entity's activities.

(continued)

Exhibit 3 (Continued)

8 Whether cash flows generated by the foreign operation directly affect the cash flow of the parent and are available to be remitted to the parent.

9 Whether operating cash flows generated by the foreign operation are sufficient to service existing and normally expected debt or whether the foreign entity will need funds from the parent to service its debt.

The following three steps outline the functional currency approach required by accounting standards in translating foreign currency financial statements into the parent company's presentation currency:

1 Identify the functional currency of the foreign entity.

2 Translate foreign currency balances into the foreign entity's functional currency.

3 Use the current exchange rate to translate the foreign entity's functional currency balances into the parent's presentation currency, if they are different.

To illustrate how this approach is applied, consider a US parent company with a Mexican subsidiary that keeps its accounting records in Mexican pesos. Assume that the vast majority of the subsidiary's transactions are carried out in Mexican pesos, but it also has an account payable in Guatemalan quetzals. In applying the three steps, the US parent company first determines that the Mexican peso is the functional currency of the Mexican subsidiary. Second, the Mexican subsidiary translates its foreign currency balances (i.e., the Guatemalan quetzal account payable), into Mexican pesos using the current exchange rate. In step 3, the Mexican peso financial statements (including the translated account payable) are translated into US dollars using the current rate method.

Now assume, alternatively, that the primary operating currency of the Mexican subsidiary is the US dollar, which thus is identified as the Mexican subsidiary's functional currency. In that case, in addition to the Guatemalan quetzal account payable, all of the subsidiary's accounts that are denominated in Mexican pesos also are considered to be foreign currency balances (because they are not denominated in the subsidiary's functional currency, which is the US dollar). Along with the Guatemalan quetzal balance, each of the Mexican peso balances must be translated into US dollars as if the subsidiary kept its books in US dollars. Assets and liabilities carried at current value in Mexican pesos are translated into US dollars using the current exchange rate, and assets and liabilities carried at historical cost in Mexican pesos are translated into US dollars using historical exchange rates. After completing this step, the Mexican subsidiary's financial statements are stated in terms of US dollars, which is both the subsidiary's functional currency and the parent's presentation currency. As a result, there is no need to apply step 3.

The following two sections describe the procedures to be followed in applying the functional currency approach in more detail.

4.1 Foreign Currency Is the Functional Currency

In most cases, a foreign entity will operate primarily in the currency of the country where it is located, which will differ from the currency in which the parent company presents its financial statements. For example, the Japanese subsidiary of a French parent company is likely to have the Japanese yen as its functional currency, whereas the French parent company must prepare consolidated financial statements in euro.

When a foreign entity has a functional currency that differs from the parent's presentation currency, the foreign entity's foreign currency financial statements are translated into the parent's presentation currency using the following procedures:

1 All assets and liabilities are translated at the current exchange rate at the balance sheet date.
2 Stockholders' equity accounts are translated at historical exchange rates.
3 Revenues and expenses are translated at the exchange rate that existed when the transactions took place. For practical reasons, a rate that approximates the exchange rates at the dates of the transactions, such as an average exchange rate, may be used.

These procedures essentially describe the *current rate method*.

When the current rate method is used, the cumulative translation adjustment needed to keep the translated balance sheet in balance is reported as a separate component of stockholders' equity.

The basic concept underlying the current rate method is that the entire investment in a foreign entity is exposed to translation gain or loss. Therefore, all assets and all liabilities must be revalued at each successive balance sheet date. The net translation gain or loss that results from this procedure is unrealized, however, and will be realized only when the entity is sold. In the meantime, the unrealized translation gain or loss that accumulates over time is deferred on the balance sheet as a separate component of stockholders' equity. When a specific foreign entity is sold, the cumulative translation adjustment related to that entity is reported as a realized gain or loss in net income.

The current rate method results in a net asset balance sheet exposure (except in the rare case in which an entity has negative stockholders' equity):

Items Translated at Current Exchange Rate

Total assets > Total liabilities → Net asset balance sheet exposure

When the foreign currency increases in value (i.e., strengthens), application of the current rate method results in an increase in the positive cumulative translation adjustment (or a decrease in the negative cumulative translation adjustment) reflected in stockholders' equity. When the foreign currency decreases in value (i.e., weakens), the current rate method results in a decrease in the positive cumulative translation adjustment (or an increase in the negative cumulative translation adjustment) in stockholders' equity.

4.2 Parent's Presentation Currency Is the Functional Currency

In some cases, a foreign entity might have the parent's presentation currency as its functional currency. For example, a Germany-based manufacturer might have a 100%-owned distribution subsidiary in Switzerland that primarily uses the euro in its day-to-day operations and thus has the euro as its functional currency. As a Swiss company, however, the subsidiary is required to record its transactions and keep its books in Swiss francs. In that situation, the subsidiary's Swiss franc financial statements must be translated into euro as if the subsidiary's transactions had originally been recorded in euro. US GAAP refer to this process as *remeasurement*. IFRS do not refer to this process as remeasurement but instead describe this situation as "reporting foreign currency transactions in the functional currency." To achieve the objective of translating to the parent's presentation currency as if the subsidiary's transactions had been recorded in that currency, the following procedures are used:

1 a Monetary assets and liabilities are translated at the current exchange rate.
  b Non-monetary assets and liabilities measured at historical cost are translated at historical exchange rates.

c Non-monetary assets and liabilities measured at current value are translated at the exchange rate at the date when the current value was determined.

2 Stockholders' equity accounts are translated at historical exchange rates.

3 a Revenues and expenses, other than those expenses related to non-monetary assets (as explained in 3.b. below), are translated at the exchange rate that existed when the transactions took place (for practical reasons, average rates may be used).

b Expenses related to non-monetary assets, such as cost of goods sold (inventory), depreciation (fixed assets), and amortization (intangible assets), are translated at the exchange rates used to translate the related assets.

These procedures essentially describe the *temporal method*.

Under the temporal method, companies must keep record of the exchange rates that exist when non-monetary assets (inventory, prepaid expenses, fixed assets, and intangible assets) are acquired, because these assets (normally measured at historical cost) are translated at historical exchange rates. Keeping track of the historical exchange rates for these assets is not necessary under the current rate method. Translating these assets (and their related expenses) at historical exchange rates complicates application of the temporal method.

The historical exchange rates used to translate inventory (and cost of goods sold) under the temporal method will differ depending on the cost flow assumption—first in, first out (FIFO); last in, first out (LIFO); or average cost—used to account for inventory. Ending inventory reported on the balance sheet is translated at the exchange rate that existed when the inventory's acquisition is assumed to have occurred. If FIFO is used, ending inventory is assumed to be composed of the most recently acquired items and thus inventory will be translated at relatively recent exchange rates. If LIFO is used, ending inventory is assumed to consist of older items and thus inventory will be translated at older exchange rates. The weighted-average exchange rate for the year is used when inventory is carried at weighted-average cost. Similarly, cost of goods sold is translated using the exchange rates that existed when the inventory items assumed to have been sold during the year (using FIFO or LIFO) were acquired. If weighted-average cost is used to account for inventory, cost of goods sold will be translated at the weighted-average exchange rate for the year.

Under both international and US accounting standards, when the temporal method is used, the translation adjustment needed to keep the translated balance sheet in balance is reported as a gain or loss in net income. US GAAP refer to these as *remeasurement* gains and losses. The basic assumption underlying the recognition of a translation gain or loss in income relates to timing. Specifically, if the foreign entity primarily uses the parent company's currency in its day-to-day operations, then the foreign entity's monetary items that are denominated in a foreign currency generate translation gains and losses that will be realized in the near future and thus should be reflected in current net income.

The temporal method generates either a net asset or a net liability balance sheet exposure, depending on whether assets translated at the current exchange rate—that is, monetary assets and non-monetary assets measured on the balance sheet date at current value (exposed assets)—are greater than or less than liabilities translated at the current exchange rate—that is, monetary liabilities and non-monetary liabilities measured on the balance sheet date at current value (exposed liabilities):

Items Translated at Current Exchange Rate

Exposed assets > Exposed liabilities → Net asset balance sheet exposure

Exposed assets < Exposed liabilities → Net liability balance sheet exposure

Most liabilities are monetary liabilities. Only cash and receivables are monetary assets, and non-monetary assets generally are measured at their historical cost. As a result, liabilities translated at the current exchange rate (exposed liabilities) often exceed assets translated at the current exchange rate (exposed assets), which results in a net liability balance sheet exposure when the temporal method is applied.

4.3 Translation of Retained Earnings

Stockholders' equity accounts are translated at historical exchange rates under both the current rate and the temporal methods. This approach creates somewhat of a problem in translating retained earnings (R/E), which are the accumulation of previous years' income less dividends over the life of the company. At the end of the first year of operations, foreign currency (FC) retained earnings are translated into the parent's currency (PC) as follows:

Net income in FC		[Translated according to the method used to translate the income statement]	=	Net income in PC
– Dividends in FC	×	Exchange rate when dividends declared	=	– Dividends in PC
R/E in FC				R/E in PC

Retained earnings in parent currency at the end of the first year become the beginning retained earnings in parent currency for the second year, and the translated retained earnings in the second year (and subsequent years) are then calculated in the following manner:

Beginning R/E in FC		[From last year's translation]	→	Beginning R/E in PC
+ Net income in FC		[Translated according to method used to translate the income statement]	=	+ Net income in PC
– Dividends in FC	×	Exchange rate when dividends declared	=	– Dividends in PC
Ending R/E in FC				Ending R/E in PC

Exhibit 4 summarizes the translation rules as discussed in Section 4.

Exhibit 4 Rules for the Translation of a Foreign Subsidiary's Foreign Currency Financial Statements into the Parent's Presentation Currency under IFRS and US GAAP

	Foreign Subsidiary's Functional Currency	
	Foreign Currency	**Parent's Presentation Currency**
Translation method:	**Current Rate Method**	**Temporal Method**
Exchange rate at which financial statement items are translated from the foreign subsidiary's bookkeeping currency to the parent's presentation currency:		
Assets		

(continued)

Exhibit 4 (Continued)

	Foreign Subsidiary's Functional Currency	
	Foreign Currency	**Parent's Presentation Currency**
Translation method:	**Current Rate Method**	**Temporal Method**
Monetary, such as cash and receivables	Current rate	Current rate
Non-monetary		
■ measured at current value (e.g., marketable securities and inventory measured at market value under the lower of cost or market rule)	Current rate	Current rate
■ measured at historical costs, (e.g., inventory measured at cost under the lower of cost or market rule; property, plant & equipment; and intangible assets)	Current rate	Historical rates
Liabilities		
Monetary, such as accounts payable, accrued expenses, long-term debt, and deferred income taxes	Current rate	Current rate
Non-monetary		
■ measured at current value	Current rate	Current rate
■ not measured at current value, such as deferred revenue	Current rate	Historical rates
Equity		
Other than retained earnings	Historical rates	Historical rates
Retained earnings	Beginning balance plus translated net income less dividends translated at historical rate	Beginning balance plus translated net income less dividends translated at historical rate
Revenues	Average rate	Average rate
Expenses		
Most expenses	Average rate	Average rate
Expenses related to assets translated at historical exchange rate, such as cost of goods sold, depreciation, and amortization	Average rate	Historical rates
Treatment of the translation adjustment in the parent's consolidated financial statements	Accumulated as a separate component of equity	Included as gain or loss in net income

4.4 Highly Inflationary Economies

When a foreign entity is located in a highly inflationary economy, the entity's functional currency is irrelevant in determining how to translate its foreign currency financial statements into the parent's presentation currency. IFRS require that the foreign entity's financial statements first be restated for local inflation using the procedures outlined in IAS 29, "Financial Reporting in Hyperinflationary Economies." Then, the inflation-restated foreign currency financial statements are translated into the parent's presentation currency using the current exchange rate.

US GAAP require a very different approach for translating the foreign currency financial statements of foreign entities located in highly inflationary economies. US GAAP do not allow restatement for inflation but instead require the foreign entity's financial statements to be remeasured as if the functional currency were the reporting currency (i.e., the temporal method).

US GAAP define a highly inflationary economy as one in which the cumulative three-year inflation rate exceeds 100% (but note that the definition should be applied with judgment, particularly because the trend of inflation can be as important as the absolute rate). A cumulative three-year inflation rate of 100% equates to an average of approximately 26% per year. IAS 21 does not provide a specific definition of high inflation, but IAS 29 indicates that a cumulative inflation rate approaching or exceeding 100% over three years would be an indicator of hyperinflation. If a country in which a foreign entity is located ceases to be classified as highly inflationary, the functional currency of that entity must be identified to determine the appropriate method for translating the entity's financial statements.

The FASB initially proposed that companies restate for inflation and then translate the financial statements, but this approach met with stiff resistance from US multinational corporations. Requiring the temporal method ensures that companies avoid a "disappearing plant problem" that exists when the current rate method is used in a country with high inflation. In a highly inflationary economy, as the local currency loses purchasing power within the country, it also tends to weaken in value in relation to other currencies. Translating the historical cost of assets such as land and buildings at progressively weaker exchange rates causes these assets to slowly disappear from the parent company's consolidated financial statements. Example 4 demonstrates the effect of three different translation approaches when books are kept in the currency of a highly inflationary economy. Example 4 pertains to Turkey in the period 2000 to 2002, when it was recognized as one of the few highly inflationary countries. Turkey is no longer viewed as having a highly inflationary economy. (In 2010, the International Practices Task Force of the Center for Audit Quality SEC Regulations Committee indicated that Venezuela had met the thresholds for being considered highly inflationary.)

EXAMPLE 4

Foreign Currency Translation in a Highly Inflationary Economy

Turkey was one of the few remaining highly inflationary countries at the beginning of the 21st century. Annual inflation rates and selected exchange rates between the Turkish lira (TL) and US dollar during the 2000–2002 period were as follows:

Date	Exchange Rates	Year	Inflation Rate (%)
01 Jan 2000	TL542,700 = US$1		
31 Dec 2000	TL670,800 = US$1	2000	38
31 Dec 2001	TL1,474,525 = US$1	2001	69
31 Dec 2002	TL1,669,000 = US$1	2002	45

Assume that a US-based company established a subsidiary in Turkey on 1 January 2000. The US parent sent the subsidiary US$1,000 on 1 January 2000 to purchase a piece of land at a cost of TL542,700,000 (TL542,700/US$ × US$1,000 = TL542,700,000). Assuming no other assets or liabilities, what are the annual and cumulative translation gains or losses that would be reported under each of three possible translation approaches?

Solution:

Approach 1: Translate Using the Current Rate Method

The historical cost of the land is translated at the current exchange rate, which results in a new translated amount at each balance sheet date.

Date	Carrying Value	Current Exchange Rate	Translated Amount in US$	Annual Translation Gain (Loss)	Cumulative Translation Gain (Loss)
01 Jan 2000	TL542,700,000	542,700	$1,000	N/A	N/A
31 Dec 2000	542,700,000	670,800	809	($191)	($191)
31 Dec 2001	542,700,000	1,474,525	368	(441)	(632)
31 Dec 2002	542,700,000	1,669,000	325	(43)	(675)

At the end of three years, land that was originally purchased with US$1,000 would be reflected on the parent's consolidated balance sheet at US$325 (and remember that land is not a depreciable asset). A cumulative translation loss of US$675 would be reported as a separate component of stockholders' equity on 31 December 2002. Because this method accounts for adjustments in exchange rates but does not account for likely changes in the local currency values of assets, it does a poor job of accurately reflecting the economic reality of situations such as the one in our example. That is the major reason this approach is not acceptable under either IFRS or US GAAP.

Approach 2: Translate Using the Temporal Method (US GAAP ASC 830)

The historical cost of land is translated using the historical exchange rate, which results in the same translated amount at each balance sheet date.

Date	Carrying Value	Historical Exchange Rate	Translated Amount in US$	Annual Translation Gain (Loss)	Cumulative Translation Gain (Loss)
01 Jan 2000	TL542,700,000	542,700	$1,000	N/A	N/A
31 Dec 2000	542,700,000	542,700	1,000	N/A	N/A
31 Dec 2001	542,700,000	542,700	1,000	N/A	N/A
31 Dec 2002	542,700,000	542,700	1,000	N/A	N/A

Under this approach, land continues to be reported on the parent's consolidated balance sheet at its original cost of US$1,000 each year. There is no translation gain or loss related to balance sheet items translated at historical exchange rates. This approach is required by US GAAP and ensures that non-monetary assets do not disappear from the translated balance sheet.

Approach 3: Restate for Inflation/Translate Using Current Exchange Rate (IAS 21)

The historical cost of the land is restated for inflation, and then the inflation-adjusted historical cost is translated using the current exchange rate.

Date	Inflation Rate (%)	Restated Carrying Value	Current Exchange Rate	Translated Amount in US$	Annual Translation Gain (Loss)	Cumulative Translation Gain (Loss)
01 Jan 00		TL542,700,000	542,700	$1,000	N/A	N/A
31 Dec 00	38	748,926,000	670,800	1,116	$116	$116
31 Dec 01	69	1,265,684,940	1,474,525	858	(258)	(142)
31 Dec 02	45	1,835,243,163	1,669,000	1,100	242	100

Under this approach, land is reported on the parent's 31 December 2002 consolidated balance sheet at US$1,100 with a cumulative, unrealized gain of US$100. Although the cumulative translation gain on 31 December 2002 is unrealized, it could have been realized if (1) the land had appreciated in TL value by the rate of local inflation, (2) the Turkish subsidiary sold the land for TL1,835,243,163, and (3) the sale proceeds were converted into US$1,100 at the current exchange rate on 31 December 2002.

This approach is required by IAS 21. It is the approach that, apart from doing an appraisal, perhaps best represents economic reality, in the sense that it reflects both the likely change in the local currency value of the land as well as the actual change in the exchange rate.

ILLUSTRATION OF TRANSLATION METHODS (EXCLUDING HYPERINFLATIONARY ECONOMIES) 5

d compare the current rate method and the temporal method, evaluate how each affects the parent company's balance sheet and income statement, and determine which method is appropriate in various scenarios;

e calculate the translation effects and evaluate the translation of a subsidiary's balance sheet and income statement into the parent company's presentation currency;

To demonstrate the procedures required in translating foreign currency financial statements (excluding hyperinflationary economies), assume that Interco is a Europe-based company that has the euro as its presentation currency. On 1 January 20X1, Interco establishes a wholly owned subsidiary in Canada, Canadaco. In addition to Interco making an equity investment in Canadaco, a long-term note payable to a Canadian bank was negotiated to purchase property and equipment. The subsidiary begins operations with the following balance sheet in Canadian dollars (C$):

Canadaco Balance Sheet, 1 January 20X1	
Assets	
Cash	C$1,500,000
Property and equipment	3,000,000
	C$4,500,000

(continued)

(Continued)	
Liabilities and Equity	
Long-term note payable	C$3,000,000
Capital stock	1,500,000
	C$4,500,000

Canadaco purchases and sells inventory in 20X1, generating net income of C$1,180,000, out of which C$350,000 in dividends are paid. The company's income statement and statement of retained earnings for 20X1 and balance sheet at 31 December 20X1 follow:

Canadaco Income Statement and Statement of Retained Earnings, 20X1	
Sales	C$12,000,000
Cost of sales	(9,000,000)
Selling expenses	(750,000)
Depreciation expense	(300,000)
Interest expense	(270,000)
Income tax	(500,000)
Net income	C$1,180,000
Less: Dividends, 1 Dec 20X1	(350,000)
Retained earnings, 31 Dec 20X1	C$830,000

Canadaco Balance Sheet, 31 December 20X1

Assets		Liabilities and Equity	
Cash	C$980,000	Accounts payable	C$450,000
Accounts receivable	900,000	Total current liabilities	450,000
Inventory	1,200,000	Long-term notes payable	3,000,000
Total current assets	C$3,080,000	Total liabilities	C$3,450,000
Property and equipment	3,000,000	Capital stock	1,500,000
Less: accumulated depreciation	(300,000)	Retained earnings	830,000
Total	C$5,780,000	Total	C$5,780,000

Inventory is measured at historical cost on a FIFO basis.

To translate Canadaco's Canadian dollar financial statements into euro for consolidation purposes, the following exchange rate information was gathered:

Date	€ per C$
1 January 20X1	0.70
Average, 20X1	0.75
Weighted-average rate when inventory was acquired	0.74
1 December 20X1 when dividends were declared	0.78
31 December 20X1	0.80

During 20X1, the Canadian dollar strengthened steadily against the euro from an exchange rate of €0.70 at the beginning of the year to €0.80 at year-end.

The translation worksheet that follows shows Canadaco's translated financial statements under each of the two translation methods. Assume first that Canadaco's functional currency is the Canadian dollar, and thus the current rate method must be used. The Canadian dollar income statement and statement of retained earnings are translated first. Income statement items for 20X1 are translated at the average exchange rate for 20X1 (€0.75), and dividends are translated at the exchange rate that existed when they were declared (€0.78). The ending balance in retained earnings as of 31 December 20X1 of €612,000 is transferred to the Canadian dollar balance sheet. The remaining balance sheet accounts are then translated. Assets and liabilities are translated at the current exchange rate on the balance sheet date of 31 December 20X1 (€0.80), and the capital stock account is translated at the historical exchange rate (€0.70) that existed on the date that Interco made the capital contribution. A positive translation adjustment of €202,000 is needed as a balancing amount, which is reported in the stockholders' equity section of the balance sheet.

If instead Interco determines that Canadaco's functional currency is the euro (the parent's presentation currency), the temporal method must be applied as shown in the far right columns of the table. The differences in procedure from the current rate method are that inventory, property, and equipment (and accumulated depreciation), as well as their related expenses (cost of goods sold and depreciation), are translated at the historical exchange rates that existed when the assets were acquired: €0.70 in the case of property and equipment, and €0.74 for inventory. The balance sheet is translated first, with €472,000 determined as the amount of retained earnings needed to keep the balance sheet in balance. This amount is transferred to the income statement and statement of retained earnings as the ending balance in retained earnings as of 31 December 20X1. Income statement items then are translated, with cost of goods sold and depreciation expense being translated at historical exchange rates. A negative translation adjustment of €245,000 is determined as the amount needed to arrive at the ending balance in retained earnings of €472,000, and this adjustment is reported as a translation loss on the income statement.

The positive translation adjustment under the current rate method can be explained by the facts that Canadaco has a net asset balance sheet exposure (total assets exceed total liabilities) during 20X1 and the Canadian dollar strengthened against the euro. The negative translation adjustment (translation loss) under the temporal method is explained by the fact that Canadaco has a net liability balance sheet exposure under this method (because the amount of exposed liabilities [accounts payable plus notes payable] exceeds the amount of exposed assets [cash plus receivables]) during 20X1 when the Canadian dollar strengthened against the euro.

Canadaco Income Statement and Statement of Retained Earnings, 20X1

Canadaco's Functional Currency Is:		*Local Currency (C$)*		*Parent's Currency (€)*	
		Current Rate		**Temporal**	
	C$	**Exch. Rate**	**€**	**Exch. Rate**	**€**
Sales	12,000,000	0.75 A	9,000,000	0.75 A	9,000,000
Cost of goods sold	(9,000,000)	0.75 A	(6,750,000)	0.74 H	(6,660,000)
Selling expenses	(750,000)	0.75 A	(562,500)	0.75 A	(562,500)
Depreciation expense	(300,000)	0.75 A	(225,000)	0.70 H	(210,000)
Interest expense	(270,000)	0.75 A	(202,500)	0.75 A	(202,500)
Income tax	(500,000)	0.75 A	(375,000)	0.75 A	(375,000)
Income before trans. gain (loss)	1,180,000		885,000		990,000
Translation gain (loss)	N/A		N/A	to balance	(245,000)
Net income	1,180,000		885,000		745,000
Less: Dividends, 12/1/20X1	(350,000)	0.78 H	(273,000)	0.78 H	(273,000)
Retained earnings, 12/31/20X1	830,000		612,000	from B/S	472,000

Note: C = current exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate

Canadaco Balance Sheet, 31 December 20X1

Canadaco's Functional Currency Is:		*Local Currency (C$)*		*Parent's Currency (€)*	
		Current Rate		**Temporal**	
	C$	**Exch. Rate**	**€**	**Exch. Rate**	**€**
Assets					
Cash	980,000	0.80 C	784,000	0.80 C	784,000
Accounts receivable	900,000	0.80 C	720,000	0.80 C	720,000
Inventory	1,200,000	0.80 C	960,000	0.74 H	888,000
Total current assets	3,080,000		2,464,000		2,392,000
Property and equipment	3,000,000	0.80 C	2,400,000	0.70 H	2,100,000
Less: accumulated depreciation	(300,000)	0.80 C	(240,000)	0.70 H	(210,000)
Total assets	5,780,000		4,624,000		4,282,000
Liabilities and Equity					
Accounts payable	450,000	0.80 C	360,000	0.80 C	360,000
Total current liabilities	450,000		360,000		360,000

(Continued)

Canadaco's Functional Currency Is:		*Local Currency (C$)*		*Parent's Currency (€)*	
		Current Rate		**Temporal**	
	C$	**Exch. Rate**	**€**	**Exch. Rate**	**€**
Long-term notes payable	3,000,000	0.80 C	2,400,000	0.80 C	2,400,000
Total liabilities	3,450,000		2,760,000		2,760,000
Capital stock	1,500,000	0.70 H	1,050,000	0.70 H	1,050,000
Retained earnings	830,000	from I/S	612,000	to balance	472,000
Translation adjustment	N/A	to balance	202,000		N/A
Total	5,780,000		4,624,000		4,282,000

Note: C = current exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate

TRANSLATION ANALYTICAL ISSUES

6

f analyze how the current rate method and the temporal method affect financial statements and ratios;

The two different translation methods used to translate Canadaco's Canadian dollar financial statements into euro result in very different amounts to be included in Interco's consolidated financial statements. The chart below summarizes some of these differences:

Canadaco's Functional Currency Is:	*Local Currency (C$)*	*Parent's Currency (€)*	
	Translation Method		
Item	**Current Rate (€)**	**Temporal (€)**	**Difference (%)**
Sales	9,000,000	9,000,000	0.0
Net income	885,000	745,000	+18.8
Income before translation gain (loss)	885,000	990,000	−10.6
Total assets	4,624,000	4,282,000	+8.0
Total equity	1,864,000	1,522,000	+22.5

In this particular case, the current rate method results in a significantly larger net income than the temporal method. This result occurs because under the current rate method, the translation adjustment is not included in the calculation of income. If the translation loss were excluded from net income, the temporal method would result in a significantly larger amount of net income. The combination of smaller net income under the temporal method and a positive translation adjustment reported on the balance sheet under the current rate method results in a much larger amount of total equity under the current rate method. Total assets also are larger under the current rate method because all assets are translated at the current exchange rate, which is higher than the historical exchange rates at which inventory and fixed assets are translated under the temporal method.

To examine the effects of translation on the underlying relationships that exist in Canadaco's Canadian dollar financial statements, several significant ratios are calculated from the original Canadian dollar financial statements and the translated (euro) financial statements and presented in the table below.

Canadaco's Functional Currency Is:				***Local Currency (C$)***		***Parent's Currency (€)***
		C$		**Current Rate (€)**		**Temporal (€)**
Current ratio		6.84		6.84		6.64
Current assets		3,080,000		2,464,000		2,392,000
Current liabilities	=	450,000	=	360,000	=	360,000
Debt-to-assets ratio		0.52		0.52		0.56
Total debt		3,000,000		2,400,000		2,400,000
Total assets	=	5,780,000	=	4,624,000	=	4,282,000
Debt-to-equity ratio		1.29		1.29		1.58
Total debt		3,000,000		2,400,000		2,400,000
Total equity	=	2,330,000	=	1,864,000	=	1,522,000
Interest coverage		7.22		7.22		7.74
EBIT		1,950,000		1,462,500		1,567,500
Interest payments	=	270,000	=	202,500	=	202,500
Gross profit margin		0.25		0.25		0.26
Gross profit		3,000,000		2,250,000		2,340,000
Sales	=	12,000,000	=	9,000,000	=	9,000,000
Operating profit margin		0.16		0.16		0.17
Operating profit		1,950,000		1,462,500		1,567,500
Sales	=	12,000,000	=	9,000,000	=	9,000,000
Net profit margin		0.10		0.10		0.08
Net income		1,180,000		885,000		745,000
Sales	=	12,000,000	=	9,000,000	=	9,000,000
Receivables turnover		13.33		12.50		12.50
Sales		12,000,000		9,000,000		9,000,000
Accounts receivable	=	900,000	=	720,000	=	720,000
Inventory turnover		7.50		7.03		7.50
Cost of goods sold		9,000,000		6,750,000		6,660,000
Inventory	=	1,200,000	=	960,000	=	888,000
Fixed asset turnover		4.44		4.17		4.76
Sales		12,000,000		9,000,000		9,000,000
Property & equipment (net)	=	2,700,000	=	2,160,000	=	1,890,000
Return on assets		0.20		0.19		0.17
Net income		1,180,000		885,000		745,000
Total assets	=	5,780,000	=	4,624,000	=	4,282,000
Return on equity		0.51		0.47		0.49

Canadaco's Functional Currency Is:			*Local Currency (C$)*		*Parent's Currency (€)*
		C$	**Current Rate (€)**		**Temporal (€)**
Net income		1,180,000	885,000		745,000
Total equity	=	2,330,000	= 1,864,000	=	1,522,000

Comparing the current rate method (€) and temporal method (€) columns in the above table shows that financial ratios calculated from Canadaco's translated financial statements (in €) differ significantly depending on which method of translation is used. Of the ratios presented, only receivables turnover is the same under both translation methods. This is the only ratio presented in which there is no difference in the type of exchange rate used to translate the items that comprise the numerator and the denominator. Sales are translated at the average exchange rate and receivables are translated at the current exchange rate under both methods. For each of the other ratios, at least one of the items included in either the numerator or the denominator is translated at a different type of rate (current, average, or historical) under the temporal method than under the current rate method. For example, the current ratio has a different value under the two translation methods because inventory is translated at the current exchange rate under the current rate method and at the historical exchange rate under the temporal method. In this case, because the euro/Canadian dollar exchange rate on 31 December 20X1 (€0.80) is higher than the historical exchange rate when the inventory was acquired (€0.74), the current ratio is larger under the current rate method of translation.

Comparing the ratios in the Canadian dollar and current rate method (euro) columns of the above table shows that many of the underlying relationships that exist in Canadaco's Canadian dollar financial statements are preserved when the current rate method of translation is used (i.e., the ratio calculated from the Canadian dollar and euro translated amounts is the same). The current ratio, the leverage ratios (debt-to-assets and debt-to-equity ratios), the interest coverage ratio, and the profit margins (gross profit margin, operating profit margin, and net profit margin) are the same in the Canadian dollar and current rate method (euro) columns of the above table. This result occurs because each of the ratios is calculated using information from either the balance sheet or the income statement, but not both. Those ratios that compare amounts from the balance sheet with amounts from the income statement (e.g., turnover and return ratios) are different. In this particular case, each of the turnover and return ratios is larger when calculated from the Canadian dollar amounts than when calculated using the current rate (euro) amounts. The underlying Canadian dollar relationships are distorted when translated using the current rate method because the balance sheet amounts are translated using the current exchange rate while revenues and expenses are translated using the average exchange rate. (These distortions would not occur if revenues and expenses also were translated at the current exchange rate.)

Comparing the ratios in the Canadian dollar and temporal method (euro) columns of the table shows that translation using the temporal method distorts all of the underlying relationships that exist in the Canadian dollar financial statements, except inventory turnover. Moreover, it is not possible to generalize the direction of the distortion across ratios. In Canadaco's case, using the temporal method results in a larger gross profit margin and operating profit margin but a smaller net profit margin as compared with the values of these ratios calculated from the original Canadian dollar amounts. Similarly, receivables turnover is smaller, inventory turnover is the same, and fixed asset turnover is larger when calculated from the translated amounts.

In translating Canadaco's Canadian dollar financial statements into euro, the temporal method results in a smaller amount of net income than the current rate method only because IFRS and US GAAP require the resulting translation loss to be included in net income when the temporal method is used. The translation loss arises because the Canadian dollar strengthened against the euro and Canadaco has a larger amount of liabilities translated at the current exchange rate (monetary liabilities) than it has assets translated at the current exchange rate (monetary assets). If Canadaco had a net monetary asset exposure (i.e., if monetary assets exceeded monetary liabilities), a translation gain would arise and net income under the temporal method (including the translation gain) would be greater than under the current rate method. Example 5 demonstrates how different types of balance sheet exposure under the temporal method can affect translated net income.

EXAMPLE 5

Effects of Different Balance Sheet Exposures under the Temporal Method *(Canadaco's functional currency is the parent's functional currency)*

Canadaco begins operations on 1 January 20X1, with cash of C$1,500,000 and property and equipment of C$3,000,000. In Case A, Canadaco finances the acquisition of property and equipment with a long-term note payable and begins operations with net monetary liabilities of C$1,500,000 (C$3,000,000 long-term note payable less C$1,500,000 cash). In Case B, Canadaco finances the acquisition of property and equipment with capital stock and begins operations with net monetary assets of C$1,500,000. To isolate the effect that balance sheet exposure has on net income under the temporal method, assume that Canadaco continues to have C$270,000 in interest expense in Case B, even though there is no debt financing. This assumption is inconsistent with reality, but it allows us to more clearly see the effect of balance sheet exposure on net income. The only difference between Case A and Case B is the net monetary asset/liability position of the company, as shown in the following table:

Canadaco Balance Sheet, 1 January 20X1

	Case A	Case B
Assets		
Cash	C$1,500,000	C$1,500,000
Property and equipment	3,000,000	3,000,000
	C$4,500,000	C$4,500,000
Liabilities and Equity		
Long-term note payable	C$3,000,000	C$0
Capital stock	1,500,000	4,500,000
	C$4,500,000	C$4,500,000

Canadaco purchases and sells inventory in 20X1, generating net income of C$1,180,000, out of which dividends of C$350,000 are paid. The company has total assets of C$5,780,000 as of 31 December 20X1. Canadaco's functional currency is determined to be the euro (the parent's presentation currency), and the company's Canadian dollar financial statements are translated into euro using the temporal method. Relevant exchange rates are as follows:

Date	€ per C$
1 January 20X1	0.70
Average, 20X1	0.75
Weighted-average rate when inventory was acquired	0.74
1 December 20X1 when dividends were declared	0.78
31 December 20X1	0.80

What effect does the nature of Canadaco's net monetary asset or liability position have on the euro translated amounts?

Solution:

Translation of Canadaco's 31 December 20X1 balance sheet under the temporal method in Case A and Case B is shown in the following table:

Canadaco Balance Sheet on 31 December 20X1 under the Temporal Method

	Case A: Net Monetary Liabilities			Case B: Net Monetary Assets		
	C$	Exch. Rate	€	C$	Exch. Rate	€
Assets						
Cash	980,000	0.80 C	784,000	980,000	0.80 C	784,000
Accounts receivable	900,000	0.80 C	720,000	900,000	0.80 C	720,000
Inventory	1,200,000	0.74 H	888,000	1,200,000	0.74 H	888,000
Total current assets	3,080,000		2,392,000	3,080,000		2,392,000
Property and equipment	3,000,000	0.70 H	2,100,000	3,000,000	0.70 H	2,100,000
Less: accum. deprec.	(300,000)	0.70 H	(210,000)	(300,000)	0.70 H	(210,000)
Total assets	5,780,000		4,282,000	5,780,000		4,282,000
Liabilities and Equity						
Accounts payable	450,000	0.80 C	360,000	450,000	0.80 C	360,000
Total current liabilities	450,000		360,000	450,000		360,000
Long-term notes payable	3,000,000	0.80 C	2,400,000	0		0
Total liabilities	3,450,000		2,760,000	450,000		360,000
Capital stock	1,500,000	0.70 H	1,050,000	4,500,000	0.70 H	3,150,000
Retained earnings	830,000		472,000	830,000		772,000
Total	5,780,000		4,282,000	5,780,000		4,282,000

Note: C = current exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate.

To keep the balance sheet in balance, retained earnings must be €472,000 in Case A (net monetary liability exposure) and €772,000 in Case B (net monetary asset exposure). The difference in retained earnings of €300,000 is equal to the translation loss that results from holding a Canadian dollar–denominated note payable during a period in which the Canadian dollar strengthens against the euro. This difference is determined by multiplying the amount of long-term note payable in Case A by the change in exchange rate during the year [C$3,000,000 × (€0.80 − €0.70) = €300,000]. Notes payable are exposed to foreign exchange risk under the temporal method, whereas capital stock is not. Canadaco could avoid the €300,000 translation loss related to long-term debt by financing the acquisition of property and equipment with equity rather than debt.

Translation of Canadaco's 20X1 income statement and statement of retained earnings under the temporal method for Case A and Case B is shown in the following table:

Canadaco Income Statement and Statement of Retained Earnings for 20X1 under the Temporal Method

	Case A: Net Monetary Liabilities			Case B: Net Monetary Assets		
	C$	Exch. Rate	€	C$	Exch. Rate	€
Sales	12,000,000	0.75 A	9,000,000	12,000,000	0.75 A	9,000,000
Cost of goods sold	(9,000,000)	0.74 H	(6,660,000)	(9,000,000)	0.74 H	(6,660,000)
Selling expenses	(750,000)	0.75 A	(562,500)	(750,000)	0.75 A	(562,500)
Depreciation expense	(300,000)	0.70 H	(210,000)	(300,000)	0.70 H	(210,000)
Interest expense	(270,000)	0.75 A	(202,500)	(270,000)	0.75 A	(202,500)
Income tax	(500,000)	0.75 A	(375,000)	(500,000)	0.75 A	(375,000)
Income before translation gain (loss)	1,180,000		990,000	1,180,000		990,000
Translation gain (loss)	N/A		(245,000)	N/A		55,000
Net income	1,180,000		745,000	1,180,000		1,045,000
Less: Dividends on 1 December 20X1	(350,000)	0.78 H	(273,000)	(350,000)	0.78 H	(273,000)
Retained earnings on 31 December 20X1	830,000		472,000	830,000		772,000

Note: C = current exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate.

Income before translation gain (loss) is the same in both cases. To obtain the amount of retained earnings needed to keep the balance sheet in balance, a translation loss of €245,000 must be subtracted from net income in Case A (net monetary liabilities), whereas a translation gain of €55,000 must be added to net income in Case B (net monetary assets). The difference in net income between the two cases is €300,000, which equals the translation loss related to the long-term note payable.

When using the temporal method, companies can manage their exposure to translation gain (loss) more easily than when using the current rate method. If a company can manage the balance sheet of a foreign subsidiary such that monetary assets equal monetary liabilities, no balance sheet exposure exists. Elimination of balance sheet exposure under the current rate method occurs only when total assets equal total liabilities. This equality is difficult to achieve because it requires the foreign subsidiary to have no stockholders' equity.

For Canadaco, in 20X1, applying the current rate method results in larger euro amounts of total assets and total equity being reported in the consolidated financial statements than would result from applying the temporal method. The direction of these differences between the two translation methods is determined by the direction of change in the exchange rate between the Canadian dollar and the euro. For example, total exposed assets are greater under the current rate method because all assets are translated at the current exchange rate. The current exchange rate at 31 December 20X1 is greater than the exchange rates that existed when the non-monetary assets were acquired, which is the translation rate for these assets under the temporal method. Therefore, the current rate method results in a larger amount of total assets because the Canadian dollar strengthened against the euro. The current rate method would result in a smaller amount of total assets than the temporal method if the Canadian dollar had weakened against the euro.

Applying the current rate method also results in a much larger amount of stockholders' equity than the temporal method. A positive translation adjustment arises under the current rate method, which is included in equity, whereas a translation loss reduces total equity (through retained earnings) under the temporal method.

Example 6 shows the effect that the direction of change in the exchange rate has on the translated amounts. Canadaco's Canadian dollar financial statements are translated into euro, first assuming no change in the exchange rate during 20X1, and then assuming the Canadian dollar strengthens and weakens against the euro. Using the current rate method to translate the foreign currency financial statements into the parent's presentation currency, the foreign currency strengthening increases the revenues, income, assets, liabilities, and total equity reported on the parent company's consolidated financial statements. Likewise, smaller amounts of revenues, income, assets, liabilities, and total equity will be reported if the foreign currency weakens against the parent's presentation currency.

When the temporal method is used to translate foreign currency financial statements, foreign currency strengthening still increases revenues, assets, and liabilities reported in the parent's consolidated financial statements. Net income and stockholders' equity, however, translate into smaller amounts (assuming that the foreign subsidiary has a net monetary liability position) because of the translation loss. The opposite results are obtained when the foreign currency weakens against the parent's presentation currency.

EXAMPLE 6

Effect of Direction of Change in the Exchange Rate on Translated Amounts

Canadaco's Canadian dollar (C$) financial statements are translated into euro (€) under three scenarios: (1) the Canadian dollar remains stable against the euro, (2) the Canadian dollar strengthens against the euro, and (3) the Canadian dollar weakens against the euro. Relevant exchange rates are as follows:

	€ per C$		
Date	**Stable**	**Strengthens**	**Weakens**
1 January 20X1	0.70	0.70	0.70
Average, 20X1	0.70	0.75	0.65
Weighted-average rate when inventory was acquired	0.70	0.74	0.66
Rate when dividends were declared	0.70	0.78	0.62
31 December 20X1	0.70	0.80	0.60

What amounts will be reported on the parent's consolidated financial statements under the three different exchange rate assumptions if Canadaco's Canadian dollar financial statements are translated using the:

1 current rate method?

2 temporal method?

Solution to 1:

Current Rate Method: Using the current rate method, Canadaco's Canadian dollar financial statements would be translated into euro as follows under the three different exchange rate assumptions:

Canadaco Income Statement and Statement of Retained Earnings for 20X1 under the Current Rate Method

		C$ Stable		C$ Strengthens		C$ Weakens	
	C$	Exch. Rate	€	Exch. Rate	€	Exch. Rate	€
Sales	12,000,000	0.70	8,400,000	0.75 A	9,000,000	0.65 A	7,800,000
Cost of goods sold	(9,000,000)	0.70	(6,300,000)	0.75 A	(6,750,000)	0.65 A	(5,850,000)
Selling expenses	(750,000)	0.70	(525,000)	0.75 A	(562,500)	0.65 A	(487,500)
Deprec. expense	(300,000)	0.70	(210,000)	0.75 A	(225,000)	0.65 A	(195,000)
Interest expense	(270,000)	0.70	(189,000)	0.75 A	(202,500)	0.65 A	(175,500)
Income tax	(500,000)	0.70	(350,000)	0.75 A	(375,000)	0.65 A	(325,000)
Net income	1,180,000		826,000		885,000		767,000
Less: Dividends	(350,000)	0.70	(245,000)	0.78 H	(273,000)	0.62 H	(217,000)
Retained earnings	830,000		581,000		612,000		550,000

Note: C = current (period-end) exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate.

Compared with the translated amount of sales and net income under a stable Canadian dollar, a stronger Canadian dollar results in a larger amount of sales and net income being reported in the consolidated income statement. A weaker Canadian dollar results in a smaller amount of sales and net income being reported in consolidated net income.

Canadaco Balance Sheet on 31 December 20X1 under the Current Rate Method

		C$ Stable		C$ Strengthens		C$ Weakens	
	C$	Exch. Rate	€	Exch. Rate	€	Exch. Rate	€
Assets							
Cash	980,000	0.70	686,000	0.80 C	784,000	0.60 C	588,000

(Continued)

	C$	C$ Stable		C$ Strengthens		C$ Weakens	
	C$	Exch. Rate	€	Exch. Rate	€	Exch. Rate	€
Accounts receivable	900,000	0.70	630,000	0.80 C	720,000	0.60 C	540,000
Inventory	1,200,000	0.70	840,000	0.80 C	960,000	0.60 C	720,000
Total current assets	3,080,000		2,156,000		2,464,000		1,848,000
Property and equipment	3,000,000	0.70	2,100,000	0.80 C	2,400,000	0.60 C	1,800,000
Less: accum. deprec.	(300,000)	0.70	(210,000)	0.80 C	(240,000)	0.60 C	(180,000)
Total assets	5,780,000		4,046,000		4,624,000		3,468,000
Liabilities and Equity							
Accounts payable	450,000	0.70	315,000	0.80 C	360,000	0.60 C	270,000
Total current liabilities	450,000		315,000		360,000		270,000
Long-term notes pay	3,000,000	0.70	2,100,000	0.80 C	2,400,000	0.60 C	1,800,000
Total liabilities	3,450,000		2,415,000		2,760,000		2,070,000
Capital stock	1,500,000	0.70	1,050,000	0.70 H	1,050,000	0.70 H	1,050,000
Retained earnings	830,000		581,000		612,000		550,000
Translation adjustment	N/A		0		202,000		(202,000)
Total equity	2,330,000		1,631,000		1,864,000		1,398,000
Total	5,780,000		4,046,000		4,624,000		3,468,000

Note: C = current (period-end) exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate.

The translation adjustment is zero when the Canadian dollar remains stable for the year; it is positive when the Canadian dollar strengthens and negative when the Canadian dollar weakens. Compared with the amounts that would appear in the euro consolidated balance sheet under a stable Canadian dollar assumption, a stronger Canadian dollar results in a larger amount of assets, liabilities, and equity being reported on the consolidated balance sheet, and a weaker Canadian dollar results in a smaller amount of assets, liabilities, and equity being reported on the consolidated balance sheet.

Solution to 2:

Temporal Method: Using the temporal method, Canadaco's financial statements would be translated into euro as follows under the three different exchange rate scenarios:

Canadaco Balance Sheet on 31 December 20X1

		Temporal Method					
		C$ Stable		C$ Strengthens		C$ Weakens	
	C$	Exch. Rate	€	Exch. Rate	€	Exch. Rate	€
Assets							
Cash	980,000	0.70	686,000	0.80 C	784,000	0.60 C	588,000
Accounts receivable	900,000	0.70	630,000	0.80 C	720,000	0.60 C	540,000
Inventory	1,200,000	0.70	840,000	0.74 H	888,000	0.66 H	792,000
Total current assets	3,080,000		2,156,000		2,392,000		1,920,000
Property and equipment	3,000,000	0.70	2,100,000	0.70 H	2,100,000	0.70 H	2,100,000
Less: accum. deprec.	(300,000)	0.70	(210,000)	0.70 H	(210,000)	0.70 H	(210,000)
Total assets	5,780,000		4,046,000		4,282,000		3,810,000
Liabilities and Equity							
Accounts payable	450,000	0.70	315,000	0.80 C	360,000	0.60 C	270,000
Total current liabilities	450,000		315,000		360,000		270,000
Long-term notes pay	3,000,000	0.70	2,100,000	0.80 C	2,400,000	0.60 C	1,800,000
Total liabilities	3,450,000		2,415,000		2,760,000		2,070,000
Capital stock	1,500,000	0.70	1,050,000	0.70 H	1,050,000	0.70 H	1,050,000
Retained earnings	830,000		581,000		472,000		690,000
Total equity	2,330,000		1,631,000		1,522,000		1,740,000
Total	5,780,000		4,046,000		4,282,000		3,810,000

Note: C = current (period-end) exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate.

Compared with the stable Canadian dollar scenario, a stronger Canadian dollar results in a larger amount of assets and liabilities but a smaller amount of equity reported on the consolidated balance sheet. A weaker Canadian dollar results in a smaller amount of assets and liabilities but a larger amount of equity reported on the consolidated balance sheet.

Canadaco Income Statement and Statement of Retained Earnings for 2008 under the Temporal Method

		C$ Stable		C$ Strengthens		C$ Weakens	
	C$	Exch. Rate	€	Exch. Rate	€	Exch. Rate	€
Sales	12,000,000	0.70	8,400,000	0.75 A	9,000,000	0.65 A	7,800,000
Cost of sales	(9,000,000)	0.70	(6,300,000)	0.74 H	(6,660,000)	0.66 H	(5,940,000)
Selling expenses	(750,000)	0.70	(525,000)	0.75 A	(562,500)	0.65 A	(487,500)
Depreciation expense	(300,000)	0.70	(210,000)	0.70 H	(210,000)	0.70 H	(210,000)
Interest expense	(270,000)	0.70	(189,000)	0.75 A	(202,500)	0.65 A	(175,500)
Income tax	(500,000)	0.70	(350,000)	0.75 A	(375,000)	0.65 A	(325,000)
Income before translation gain (loss)	1,180,000		826,000		990,000		662,000
Translation gain (loss)	N/A		0		(245,000)		245,000
Net income	1,180,000		826,000		745,000		907,000
Less: Dividends	(350,000)	0.70	(245,000)	0.78 H	(273,000)	0.62 H	(217,000)
Retained earnings	830,000		581,000		472,000		690,000

Note: C = current (period-end) exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate.

No translation gain or loss exists when the Canadian dollar remains stable during the year. Because the subsidiary has a net monetary liability exposure to changes in the exchange rate, a stronger Canadian dollar results in a translation loss and a weaker Canadian dollar results in a translation gain. Compared with a stable Canadian dollar, a stronger Canadian dollar results in a larger amount of sales and a smaller amount of net income reported on the consolidated income statement. This difference in direction results from the translation loss that is included in net income. (As demonstrated in Example 5, a translation gain would have resulted if the subsidiary had a net monetary asset exposure.) A weaker Canadian dollar results in a smaller amount of sales but a larger amount of net income than if the Canadian dollar had remained stable.

Exhibit 5 summarizes the relationships illustrated in Examples 5 and 6, focusing on the typical effect that a strengthening or weakening of the foreign currency has on financial statement amounts compared with what the amounts would be if the foreign currency were to remain stable.

Exhibit 5 Effect of Currency Exchange Rate Movement on Financial Statements

	Temporal Method, Net Monetary Liability Exposure	Temporal Method, Net Monetary Asset Exposure	Current Rate Method
Foreign currency strengthens relative to parent's presentation currency	↑ Revenues ↑ Assets ↑ Liabilities ↓ Net income ↓ Shareholders' equity Translation loss	↑ Revenues ↑ Assets ↑ Liabilities ↑ Net income ↑ Shareholders' equity Translation gain	↑ Revenues ↑ Assets ↑ Liabilities ↑ Net income ↑ Shareholders' equity Positive translation adjustment
Foreign currency weakens relative to parent's presentation currency	↓ Revenues ↓ Assets ↓ Liabilities ↑ Net income ↑ Shareholders' equity Translation gain	↓ Revenues ↓ Assets ↓ Liabilities ↓ Net income ↓ Shareholders' equity Translation loss	↓ Revenues ↓ Assets ↓ Liabilities ↓ Net income ↓ Shareholders' equity Negative translation adjustment

7 TRANSLATION WHEN A FOREIGN SUBSIDIARY OPERATES IN AN HYPERINFLATIONARY ECONOMY

g analyze how alternative translation methods for subsidiaries operating in hyperinflationary economies affect financial statements and ratios;

As noted earlier, IFRS and US GAAP differ substantially in their approach to translating the foreign currency financial statements of foreign entities operating in the currency of a hyperinflationary economy. US GAAP simply require the foreign currency financial statements of such an entity to be translated as if the parent's currency is the functional currency (i.e., the temporal method must be used with the resulting translation gain or loss reported in net income). IFRS require the foreign currency financial statements first to be restated for inflation using the procedures of IAS 29, and then the inflation-adjusted financial statements are translated using the current exchange rate.

IAS 29 requires the following procedures in adjusting financial statements for inflation:

Balance Sheet

- Monetary assets and monetary liabilities are not restated because they are already expressed in terms of the monetary unit current at the balance sheet date. Monetary items consist of cash, receivables, and payables.
- Non-monetary assets and non-monetary liabilities are restated for changes in the general purchasing power of the monetary unit. Most non-monetary items are carried at historical cost. In these cases, the restated cost is determined by applying to the historical cost the change in the general price index from the

date of acquisition to the balance sheet date. Some non-monetary items are carried at revalued amounts; for example, property, plant, and equipment revalued according to the allowed alternative treatment in IAS 16, "Property, Plant and Equipment." These items are restated from the date of revaluation.

- All components of stockholders' equity are restated by applying the change in the general price level from the beginning of the period or, if later, from the date of contribution to the balance sheet date.

Income Statement

- All income statement items are restated by applying the change in the general price index from the dates when the items were originally recorded to the balance sheet date.
- The net gain or loss in purchasing power that arises from holding monetary assets and monetary liabilities during a period of inflation is included in net income.

The procedures for adjusting financial statements for inflation are similar in concept to the procedures followed when using the temporal method for translation. By restating non-monetary assets and liabilities along with stockholders' equity in terms of the general price level at the balance sheet date, these items are carried at their historical amount of purchasing power. Only the monetary items, which are not restated for inflation, are exposed to inflation risk. The effect of that exposure is reflected through the purchasing power gain or loss on the net monetary asset or liability position.

Holding cash and receivables during a period of inflation results in a **purchasing power loss**, whereas holding payables during inflation results in a **purchasing power gain**. This relationship can be demonstrated through the following examples.

Assume that the general price index (GPI) on 1 January 20X1 is 100; that is, a representative basket of goods and services can be purchased on that date for $100. At the end of 20X1, the same basket of goods and services costs $120; thus, the country has experienced an inflation rate of 20% [($120 – $100) ÷ $100]. Cash of $100 can be used to acquire one basket of goods on 1 January 20X1. One year later, however, when the GPI stands at 120, the same $100 in cash can now purchase only 83.3% of a basket of goods and services. At the end of 20X1, it now takes $120 to purchase the same amount as $100 could purchase at the beginning of the year. The difference between the amount of cash needed to purchase one market basket at year end ($120) and the amount actually held ($100) results in a purchasing power loss of $20 from holding cash of $100 during the year.

Borrowing money during a period of inflation increases purchasing power. Assume that a company expects to receive $120 in cash at the end of 20X1. If it waits until the cash is received, the company will be able to purchase exactly 1.0 basket of goods and services when the GPI stands at 120. If instead, the company borrows $120 on 1 January 20X1 when the GPI is 100, it can acquire 1.2 baskets of goods and services. This transaction results in a purchasing power gain of $20. Of course, there is an interest cost associated with the borrowing that offsets a portion of this gain.

A net purchasing power gain will arise when a company holds a greater amount of monetary liabilities than monetary assets, and a net purchasing power loss will result when the opposite situation exists. As such, purchasing power gains and losses are analogous to the translation gains and losses that arise when the currency is weakening in value and the temporal method of translation is applied.

Although the procedures required by IFRS and US GAAP for translating the foreign currency financial statements in high-inflation countries are fundamentally different, the results, in a rare occurrence, can be very similar. Indeed, if the exchange rate between two currencies changes by exactly the same percentage as the change in the general price index in the highly inflationary country, then the two methodologies produce the same results. Example 7 demonstrates this scenario.

EXAMPLE 7

Translation of Foreign Currency Financial Statements of a Foreign Entity Operating in a High Inflation Country

ABC Company formed a subsidiary in a foreign country on 1 January 20X1, through a combination of debt and equity financing. The foreign subsidiary acquired land on 1 January 20X1, which it rents to a local farmer. The foreign subsidiary's financial statements for its first year of operations, in foreign currency units (FC), are as follows:

Foreign Subsidiary Income Statement

(in FC)	20X1
Rent revenue	1,000
Interest expense	(250)
Net income	750

Foreign Subsidiary Balance Sheets

(in FC)	1 Jan 20X1	31 Dec 20X1
Cash	1,000	1,750
Land	9,000	9,000
Total	10,000	10,750
Note payable (5%)	5,000	5,000
Capital stock	5,000	5,000
Retained earnings	0	750
Total	10,000	10,750

The foreign country experienced significant inflation in 20X1, especially in the second half of the year. The general price index during the year was as follows:

1 January 20X1	100
Average, 20X1	125
31 December 20X1	200

The inflation rate in 20X1 was 100%, and the foreign country clearly meets the definition of a highly inflationary economy.

As a result of the high inflation rate in the foreign country, the FC weakened substantially during the year relative to other currencies. Relevant exchange rates between ABC's presentation currency (US dollars) and the FC during 20X1 were as follows:

	US$ per FC
1 January 20X1	1.00
Average, 20X1	0.80
31 December 20X1	0.50

What amounts will ABC Company include in its consolidated financial statements for the year ended 31 December 20X1 related to this foreign subsidiary?

Solution:

Assuming that ABC Company wishes to prepare its consolidated financial statements in accordance with IFRS, the foreign subsidiary's 20X1 financial statements will be restated for local inflation and then translated into ABC's presentation currency using the current exchange rate as follows:

	FC	Restatement Factor	Inflation-Adjusted FC	Exch. Rate	US$
Cash	1,750	200/200	1,750	0.50	875
Land	9,000	200/100	18,000	0.50	9,000
Total	10,750		19,750		9,875
Note payable	5,000	200/200	5,000	0.50	2,500
Capital stock	5,000	200/100	10,000	0.50	5,000
Retained earnings	750		4,750	0.50	2,375
Total	10,750		19,750		9,875
Revenues	1,000	200/125	1,600	0.50	800
Interest expense	(250)	200/125	(400)	0.50	(200)
Subtotal	750		1,200		600
Purchasing power gain/loss			3,550	0.50	1,775
Net income			4,750		2,375

All financial statement items are restated to the GPI at 31 December 20X1. The net purchasing power gain of FC3,550 can be explained as follows:

Gain from holding note payable	FC5,000 × (200 − 100)/100 =	FC5,000
Loss from holding beginning balance in cash	−1,000 × (200 − 100)/100 =	(1,000)
Loss from increase in cash during the year	−750 × (200 − 125)/125 =	(450)
Net purchasing power gain (loss)		FC3,550

Note that all inflation-adjusted FC amounts are translated at the current exchange rate, and thus no translation adjustment is needed.

Now assume alternatively that ABC Company wishes to comply with US GAAP in preparing its consolidated financial statements. In that case, the foreign subsidiary's FC financial statements are translated into US dollars using the temporal method, with the resulting translation gain/loss reported in net income, as follows:

	FC	Exch. Rate	US$
Cash	1,750	0.50 C	875
Land	9,000	1.00 H	9,000
Total	10,750		9,875
Note payable	5,000	0.50 C	2,500
Capital stock	5,000	1.00 H	5,000
Retained earnings	750		2,375
Total	10,750		9,875
Revenues	1,000	0.80 A	800
Interest expense	(250)	0.80 A	(200)
Subtotal	750		600
Translation gain*			1,775
Net income			2,375

* The dividend is US$0 and the increase in retained earnings is US$2,375 (from the balance sheet); so, net income is US$2,375, and thus the translation gain is US$1,775.
Note: C = current (period-end) exchange rate; A = average-for-the-year exchange rate; H = historical exchange rate

Application of the temporal method as required by US GAAP in this situation results in exactly the same US dollar amounts as were obtained under the restate/translate approach required by IFRS. The equivalence of results under the two approaches exists because of the exact one-to-one inverse relationship between the change in the foreign country's GPI and the change in the dollar value of the FC, as predicted by the theory of purchasing power parity. The GPI doubled and the FC lost half its purchasing power, which caused the FC to lose half its value in dollar terms. To the extent that this relationship does not hold, and it rarely ever does, the two different methodologies will generate different translated amounts. For example, if the 31 December 20X1 exchange rate had adjusted to only US$0.60 per FC1 (rather than US$0.50 per FC1), then translated net income would have been US$2,050 under US GAAP and US$2,850 under IFRS.

8 COMPANIES USE BOTH TRANSLATION METHODS AT THE SAME TIME AND DISCLOSURES RELATED TO TRANSLATION METHODS

f analyze how the current rate method and the temporal method affect financial statements and ratios;

Under both IFRS and US GAAP, a multinational corporation may need to use both the current rate and the temporal methods of translation at a single point in time. This situation will apply when some foreign subsidiaries have a foreign currency as their functional currency (and therefore are translated using the current rate method) and other foreign subsidiaries have the parent's currency as their functional currency (and therefore are translated using the temporal method). As a result, a multinational corporation's consolidated financial statements can reflect simultaneously both a net translation gain or loss that is included in the determination of net income (from foreign subsidiaries translated using the temporal method) and a separate cumulative translation adjustment reported on the balance sheet in stockholders' equity (from foreign subsidiaries translated using the current rate method).

Exxon Mobil Corporation is an example of a company that has a mixture of foreign currency and parent currency functional currency subsidiaries, as evidenced by the following excerpt from its 2011 annual report, Note 1 Summary of Accounting Policies:

> **Foreign Currency Translation.** The Corporation selects the functional reporting currency for its international subsidiaries based on the currency of the primary economic environment in which each subsidiary operates. Downstream and Chemical operations primarily use the local currency. However, the US dollar is used in countries with a history of high inflation (primarily in Latin America) and Singapore, which predominantly sells into the US dollar export market. Upstream operations which are relatively self-contained and integrated within a particular country, such as Canada, the United Kingdom, Norway and continental Europe, use the local currency. Some upstream operations, primarily in Asia and Africa, use the US dollar because they predominantly sell crude and natural gas production into US dollar–denominated markets. For all operations, gains or losses from remeasuring foreign currency transactions into the functional currency are included in income.

Because of the judgment involved in determining the functional currency of foreign operations, two companies operating in the same industry might apply this judgment differently. For example, although Exxon Mobil has identified the local currency as the functional currency for many of its international subsidiaries, Chevron Corporation has designated the US dollar as the functional currency for substantially all of its overseas operations, as indicated in its 2011 annual report, Note 1 Summary of Significant Accounting Policies:

> **Currency Translation**. The US dollar is the functional currency for substantially all of the company's consolidated operations and those of its equity affiliates. For those operations, all gains and losses from currency remeasurement are included in current period income. The cumulative translation effects for those few entities, both consolidated and affiliated, using functional currencies other than the US dollar are included in "Currency translation adjustment" on the Consolidated Statement of Equity.

Evaluating net income reported by Exxon Mobil against net income reported by Chevron presents a comparability problem. This problem can be partially resolved by adding the translation adjustments reported in stockholders' equity to net income for both companies. The feasibility of this solution depends on the level of detail disclosed by multinational corporations with respect to the translation of foreign currency financial statements.

8.1 Disclosures Related to Translation Methods

Both IFRS and US GAAP require two types of disclosures related to foreign currency translation:

1. the amount of exchange differences recognized in net income, and
2. the amount of cumulative translation adjustment classified in a separate component of equity, along with a reconciliation of the amount of cumulative translation adjustment at the beginning and end of the period.

US GAAP also specifically require disclosure of the amount of translation adjustment transferred from stockholders' equity and included in current net income as a result of the disposal of a foreign entity.

The amount of exchange differences recognized in net income consists of

- foreign currency *transaction* gains and losses, and
- *translation* gains and losses resulting from application of the temporal method.

Neither IFRS nor US GAAP require disclosure of the two separate amounts that constitute the total exchange difference recognized in net income, and most companies do not provide disclosure at that level of detail. However, BASF AG (shown earlier in Exhibit 1) is an exception. Note 6 in BASF's annual report separately discloses gains from foreign currency and hedging transactions and gains from translation of financial statements, both of which are included in the line item "Other Operating Income" on the income statement, as shown below:

6. Other Operating Income

Million €	2011	2010
Reversal and adjustment of provisions	170	244
Revenue from miscellaneous revenue-generating activities	207	142
Income from foreign currency and hedging transactions	170	136
Income from the translation of financial statements in foreign currencies	42	76
Gains on the disposal of property, plant and equipment and divestitures	666	101
Reversals of impairments of property, plant and equipment	—	40
Gains on the reversal of allowance for doubtful business-related receivables	77	36
Other	676	365
	2,008	1,140

The company provides a similar level of detail in Note 7 related to "Other Operating Expenses."

Disclosures related to foreign currency translation are commonly found in both the MD&A and the Notes to Financial Statements sections of an annual report. Example 8 uses the foreign currency translation–related disclosures made in 2011 by Yahoo! Inc.

EXAMPLE 8

Disclosures Related to Foreign Currency Translation: Yahoo! Inc. 2011 Annual Report

Yahoo! Inc. is a US-based digital media company that reports in US dollars and prepares financial statements in accordance with US GAAP.

The stockholders' equity section of Yahoo!'s consolidated balance sheets includes the following line items:

	31 December	
(in thousands)	**2010**	**2011**
Common stock	$1,306	$1,242
Additional paid-in capital	10,109,913	9,825,899
Treasury stock	—	(416,237)
Retained earnings	1,942,656	2,432,294
Accumulated other comprehensive income (loss)	504,254	697,869
Total Yahoo! Inc. stockholders' equity	12,558,129	12,541,067

The consolidated statement of stockholders' equity provides detail on the components comprising "Accumulated other comprehensive income." The relevant portion of that statement appears below:

	Years Ended 31 December		
	2009	**2010**	**2011**
Accumulated other comprehensive income			
Balance, beginning of year	120,276	369,236	504,254
Net change in unrealized gains/losses on available-for-sale securities, net of tax	(1,936)	3,813	(16,272)
Foreign currency translation adjustments, net of tax	250,896	131,205	209,887
Balance, end of year	369,236	504,254	697,869

Yahoo! reported the following net income in 2010 and 2011, as shown on the consolidated statement of income:

	2010	**2011**	**% Change**
Net income	$1,244,628	$1,062,699	−14.6%

Yahoo!'s disclosures for its three geographic segments are disclosed in a note to the financial statements. Revenue (excluding total acquisition costs) and direct segment operating costs are shown below:

	2009	**2010**	**2011**
Revenue ex-TAC by segment:			
Americas	3,656,752	3,467,850	3,142,879
EMEA	390,456	368,884	407,467
Asia Pacific	635,281	751,495	830,482
Total revenue ex-TAC	4,682,489	4,588,229	4,380,828

(continued)

	2009	2010	2011
Direct costs by segment:			
Americas	620,690	568,017	560,016
EMEA	115,778	118,954	135,266
Asia Pacific	138,739	146,657	194,394

In the MD&A section of the 2011 annual report, Yahoo! describes the source of its translation exposure:

> **Translation Exposure**
>
> We are also exposed to foreign exchange rate fluctuations as we convert the financial statements of our foreign subsidiaries and our investments in equity interests into US dollars in consolidation. If there is a change in foreign currency exchange rates, the conversion of the foreign subsidiaries' financial statements into US dollars results in a gain or loss which is recorded as a component of accumulated other comprehensive income which is part of stockholders' equity.
>
> Revenue ex-TAC (total acquisition costs) and related expenses generated from our international subsidiaries are generally denominated in the currencies of the local countries. The statements of income of our international operations are translated into US dollars at exchange rates indicative of market rates during each applicable period. To the extent the US dollar strengthens against foreign currencies, the translation of these foreign currency-denominated transactions results in reduced consolidated revenue and operating expenses. Conversely, our consolidated revenue and operating expenses will increase if the US dollar weakens against foreign currencies. Using the foreign currency exchange rates from the year ended December 31, 2010, revenue ex-TAC for the Americas segment for the year ended December 31, 2011 would have been lower than we reported by $6 million, revenue ex-TAC for the EMEA segment would have been lower than we reported by $16 million, and revenue ex-TAC for the Asia Pacific segment would have been lower than we reported by $59 million. Using the foreign currency exchange rates from the year ended December 31, 2010, direct costs for the Americas segment for the year ended December 31, 2011 would have been lower than we reported by $2 million, direct costs for the EMEA segment would have been lower than we reported by $5 million, and direct costs for the Asia Pacific segment would have been lower than we reported by $15 million.

Using the information above, address the following questions:

1. By how much did accumulated other comprehensive income change during the year ended 31 December 2011? Where can this information be found?
2. How much foreign currency translation adjustment was included in other comprehensive income for the year ended 31 December 2011? How does such an adjustment arise?
3. If foreign currency translation adjustment had been included in net income (rather than in other comprehensive income), how would the 2010/2011 change in income have been affected?

4 From what perspective does Yahoo! describe its foreign currency risk?

5 What percentage of total revenue ex-TAC was generated by the Asia-Pacific segment for the year ended 31 December 2011? What would this percentage have been if there had been no change in foreign currency exchange rates during the year?

Solutions:

1 Accumulated other comprehensive income increased by $193,615 thousand (from $504,254 thousand beginning balance to $697,869 thousand at the end of the year). This information can be found in two places: the stockholders' equity section of the balance sheet and the consolidated statement of stockholders' equity.

2 The amount of foreign currency translation adjustment included in other comprehensive income for 2011 was $209,887 thousand. The foreign currency translation adjustment arises from applying the current rate method to translate the foreign currency functional currency financial statements of foreign subsidiaries. Assuming that Yahoo!'s foreign subsidiaries have positive net assets, the positive translation adjustment in 2011 results from a strengthening in foreign currencies (weakening in the US dollar).

3 If foreign currency translation adjustment had been included in net income (rather than other comprehensive income), the percentage decrease in reported net income from 2010 to 2011 of 14.6% would have been smaller (7.5%).

	2010	2011	% Change
Net income	$1,244,628	$1,062,699	−14.6%
Foreign currency translation adjustment	131,205	209,887	
	$1,375,833	$1,272,586	−7.5%

4 Yahoo! describes its foreign currency risk from the perspective of how the US dollar fluctuates against foreign currencies because the dollar is the reporting currency. If the US dollar strengthens, then foreign currencies must weaken, which will result in reduced revenues, expenses, and income from foreign operations.

5 The Asia-Pacific segment represented 19.0% of total revenue ex-TAC. Information from the MD&A disclosure can be used to determine that if there had been no change in foreign currency exchange rates during the year, the segment would have represented a slightly lower percentage of total revenue (17.9%).

	2011, as reported			2011, if no change in exchange rates	
Revenue ex-TAC by segment:					
Americas	3,142,879	71.7%	6,000	3,136,879	73.0%
EMEA	407,467	9.3%	16,000	391,467	9.1%
Asia Pacific	830,482	19.0%	59,000	771,482	17.9%
Total revenue ex-TAC	4,380,828	100.0%		4,299,828	100.0%

As noted in the previous section, because of the judgment involved in determining the functional currency of foreign operations, two companies operating in the same industry might use different predominant translation methods. As a result, income reported by these companies may not be directly comparable. Exxon Mobil Corporation and Chevron Corporation, both operating in the petroleum industry, are an example of two companies for which this is the case. Whereas Chevron has identified the US dollar as the functional currency for substantially all of its foreign subsidiaries, Exxon Mobil indicates that its downstream and chemical operations, as well as some of its upstream operations, primarily use the local currency as the functional currency. As a result, Chevron primarily uses the temporal method with translation gains and losses included in income, while Exxon Mobil uses the current rate method to a much greater extent, with the resulting translation adjustments excluded from income. To make the income of these two companies more comparable, an analyst can use the disclosures related to translation adjustments to include these as gains and losses in determining an adjusted amount of income. Example 9 demonstrates this process for Exxon Mobil and Chevron.

EXAMPLE 9

Comparing Net Income for Exxon Mobil Corporation and Chevron Corporation

Exxon Mobil Corporation uses the current rate method to translate the foreign currency financial statements of a substantial number of its foreign subsidiaries and includes the resulting translation adjustments in the "Accumulated other non-owner changes in equity" line item in the stockholders' equity section of the consolidated balance sheet. Detail on the items composing "Accumulated other non-owner changes in equity," including "Foreign exchange translation adjustment," is provided in the consolidated statement of shareholders' equity.

Chevron Corporation uses the temporal method to translate the foreign currency financial statements of substantially all of its foreign subsidiaries. For those few entities using functional currencies other than the US dollar, however, the current rate method is used and the resulting translation adjustments are included in the "Accumulated other comprehensive loss" component of stockholders' equity. The consolidated statement of stockholders' equity provides detail on the changes in the component of stockholders' equity, including a "Currency translation adjustment."

Combining net income from the income statement and the change in the cumulative translation adjustment account from the statement of stockholders' equity, an adjusted net income in which translation adjustments are treated as gains and losses can be calculated for each company, as shown in the following table (amounts in millions of US dollars):

Exxon Mobil	2011	2010	2009
Reported net income	42,206	31,398	19,658
Translation adjustment	(867)	1,034	3,629
Adjusted net income	41,339	32,432	23,287

Chevron	2011	2010	2009
Reported net income	27,008	19,136	10,563
Translation adjustment	17	6	60
Adjusted net income	27,025	19,142	10,623

The direction, positive or negative, of the translation adjustment is the same for both companies in 2009 and 2010 but not in 2011. Overall, Exxon Mobil has significantly larger translation adjustments than Chevron because Exxon Mobil designates the local currency as functional currency for a substantially larger portion of its foreign operations.

A comparison of the relative amounts of net income generated by the two companies is different depending on whether reported net income or adjusted net income is used. Exxon Mobil's reported net income in 2009 is 1.90 times larger than Chevron's, whereas its adjusted net income is 2.2 times larger, as shown in the following table.

	2011	2010	2009
Exxon Mobil reported net income/ Chevron reported net income	1.6	1.6	1.9
Exxon Mobil adjusted net income/ Chevron adjusted net income	1.5	1.7	2.2

Including translation adjustments as gains and losses in the measurement of an adjusted net income provides a more comparable basis for evaluating the profitability of two companies that use different predominant translation methods. Bringing the translation adjustments into the calculation of adjusted net income still might not provide truly comparable measures, however, because of the varying effect that the different translation methods have on reported net income.

Some analysts believe that all non-owner changes in stockholders' equity, such as translation adjustments, should be included in the determination of net income. This approach is referred to as clean-surplus accounting, as opposed to dirty-surplus accounting, in which some income items are reported as part of stockholders' equity rather than as gains and losses on the income statement. One of the dirty-surplus items found in both IFRS and US GAAP financial statements is the translation adjustment that arises when a foreign currency is determined to be the functional currency of a foreign subsidiary. Disclosures made in accordance with IFRS and US GAAP provide analysts with the detail needed to calculate net income on a clean-surplus basis. In fact, both sets of standards now require companies to prepare a statement of comprehensive income in which unrealized gains and losses that have been deferred in stockholders' equity are included in a measure of comprehensive income.

MULTINATIONAL OPERATIONS AND A COMPANY'S EFFECTIVE TAX RATE 9

h describe how multinational operations affect a company's effective tax rate;

In general, multinational companies incur income taxes in the country in which the profit is earned. Transfer prices, the prices that related companies charge on intercompany transactions, affect the allocation of profit between the companies. An entity with operations in multiple countries with different tax rates could aim to set transfer prices such that a higher portion of its profit is allocated to lower tax rate jurisdictions. Countries have established various laws and practices to prevent aggressive transfer pricing practices. Transfer pricing has been defined as "the system of laws and practices used by countries to ensure that goods, services and intellectual property transferred between related companies are appropriately priced, based on market conditions, such that profits are correctly reflected in each jurisdiction."[5] Also, most countries are party to tax treaties that prevent double-taxation of corporate profits by granting a credit for taxes paid to another country.

Whether and when a company also pays income taxes in its home country depends on the specific tax regime. In the United States, for example, multinational companies are liable only for a residual tax on foreign income, after applying a credit for foreign taxes paid on that same income. The effect of the tax credit is that the multinational company owes taxes on the foreign income only to the extent that the US corporate tax rate exceeds the foreign rate of tax on that income. In addition, much of the foreign income earned by US multinationals is not taxed until it is repatriated.[6]

An analyst can obtain information about the effect of multinational operations from companies' disclosure on effective tax rates. Accounting standards require companies to provide an explanation of the relationship between tax expense and accounting profit. The explanation is presented as a reconciliation between the average effective tax rate (tax expense divided by pretax accounting profits) and the relevant statutory rate. The purpose of this disclosure is to enable users of financial statements to understand whether the relationship between tax expense and accounting profit in a particular fiscal period is unusual and to understand the significant factors—including the effect of foreign taxes—that could affect that relationship in the future.[7] Changes in the effective tax rate impact of foreign taxes could be caused by changes in the applicable tax rates and/or changes in the mix of profits earned in different jurisdictions.

EXAMPLE 10

Below are excerpts from the effective tax rate reconciliation disclosures by two companies: Heineken N.V., a Dutch brewer, and Colgate Palmolive, a US consumer products company. Use the disclosures to answer the following questions:

1 Which company's home country has a lower statutory tax rate?

2 What was the impact of multinational operations on each company's 2011 effective tax rate?

3 Changes in the tax rate impact of multinational operations can often be explained by changes of profit mix between countries with higher or lower marginal tax rates. What do Heineken's disclosures suggest about the geographic mix of its 2011 profit?

5 TP Analytics. http://www.tpanalytics.com.

6 United States Government Accountability Office (GAO) Report GAO-08-950. *US Multinational Corporations: Effective Tax Rates Are Correlated with Where Income Is Reported*. August 2008.

7 International Accounting Standard 12 *Income Taxes*, ¶84.

Heineken N.V. Annual Report 2011
Notes to the consolidated financial statements
13. Income tax expense (excerpt)

Reconciliation of the effective tax rate

In millions of EUR	2011	2010
Profit before income tax	2,025	1,982
Share of net profit of associates and joint ventures and impairments thereof	(240)	(193)
Profit before income tax excluding share of profit of associates and joint ventures (inclusive impairments thereof)	**1,785**	1,789

	%	2011	%	2010
Income tax using the Company's domestic tax rate	25.0	446	25.5	456
Effect of tax rates in foreign jurisdictions	3.5	62	1.9	34
Effect of non-deductible expenses	3.2	58	4	72
Effect of tax incentives and exempt income	(6.0)	−107	−8.2	−146
Recognition of previously unrecognised temporary differences	(0.5)	−9	−0.1	−2
Utilisation or recognition of previously unrecognised tax losses	(0.3)	−5	−1.2	−21
Unrecognised current year tax losses	1.0	18	0.8	15
Effect of changes in tax rate	0.1	1	0.2	3
Withholding taxes	1.5	26	1.4	25
Under/(over) provided in prior years	(1.5)	−27	−2.3	−42
Other reconciling items	0.1	2	0.5	9
	26.1	**465**	22.5	403

COLGATE-PALMOLIVE COMPANY Annual Report 2011
Notes to Consolidated Financial Statements
10. Income Taxes (excerpt)

The difference between the statutory US federal income tax rate and the Company's global effective tax rate as reflected in the Consolidated Statements of Income is as follows:

Percentage of Income before income taxes	2011	2010	2009
Tax at United States statutory rate	35.0%	35.0%	35.0%
State income taxes, net of federal benefit	0.4	1.1	0.5
Earnings taxed at other than United States statutory rate	(1.7)	(4.6)	(2.5)
Venezuela hyperinflationary transition charge	—	2.8	—

(continued)

(Continued)

Percentage of Income before income taxes	2011	2010	2009
Other, net	(1.1)	(1.7)	(0.8)
Effective tax rate	32.6%	32.6%	32.2%

Solution to 1:

Heineken's home country tax rate (25.0% in 2011) is lower than Colgate Palmolive's home country tax rate (35.0%).

Solution to 2:

The line item labeled "Effect of tax rates in foreign jurisdictions" indicates that multinational operations increased Heineken's effective tax rate by 3.5 percentage points. The line item labeled "Earnings taxed at other than United States statutory rate" indicates that multinational operations lowered Colgate Palmolive's effective tax rate by 1.7 percentage points in 2011.

Solution to 3:

Multinational operations increased Heineken's effective tax rate by 3.5 percentage points in 2011 but only 1.9 percentage points in 2010. This greater impact in 2011 could indicate that Heineken's profit mix in 2011 shifted to countries with higher marginal tax rates. (The change could also indicate that the marginal tax rates increased in the countries in which Heineken earns profits.)

10 ADDITIONAL DISCLOSURES ON THE EFFECTS OF FOREIGN CURRENCY

i. explain how changes in the components of sales affect the sustainability of sales growth;

j. analyze how currency fluctuations potentially affect financial results, given a company's countries of operation.

We turn now to the question of how an analyst can use multinational companies' disclosures to better understand the effects of foreign currency.

10.1 Disclosures Related to Sales Growth

Companies often make important disclosures about foreign currency effect on sales growth in the MD&A. Additional disclosures are also often made in financial presentations to the analyst community.

For a multinational company, sales growth is driven not only by changes in volume and price but also by changes in the exchange rates between the reporting currency and the currency in which sales are made. Arguably, growth in sales that comes from changes in volume or price is more sustainable than growth in sales that comes from changes in exchange rates. Further, management arguably has greater control over growth in sales resulting from greater volume or higher price than from changes in exchange rates. Thus, an analyst will consider the foreign currency effect on sales growth both for forecasting future performance and for evaluating a management team's historical performance.

Companies often include disclosures about the effect of exchange rates on sales growth in the MD&A. Such disclosures may also appear in other financial reports, such as company presentations to investors or earnings announcements. Exhibit 6 provides an example of disclosure from the MD&A, and Example 11 illustrates even more detailed disclosure from a company's report to analysts.

Exhibit 6

General Mills' 2011 annual report includes the following disclosures about the components of net sales growth in its international segment. The first excerpt is from the MD&A, and the second is from a supplementary schedule reconciling non-GAAP measures. Although the overall effect on international net sales growth was minimal "flat," the geographic detail provided in the supplementary schedule shows that the effects varied widely by region.

Excerpt from MD&A

Components of International Net Sales Growth

	Fiscal 2011 vs. 2010	Fiscal 2010 vs. 2009
Contributions from volume growth[a]	6 pts	Flat
Net price realization and mix	1 pt	3 pts
Foreign currency exchange	Flat	1 pt
Net sales growth	7 pts	4 pts

[a] Measured in tons based on the stated weight of our product shipments.

Excerpt from Supplementary Schedule on Non-GAAP Measures

International Segment and Region Sales Growth Rates Excluding Impact of Foreign Exchange

	Fiscal Year 2011		
	Percentage change in Net Sales as Reported	Impact of Foreign Currency Exchange	Percentage change in Net Sales on Constant Currency Basis
Europe	5%	−2%	7%
Canada	8	5	3
Asia/Pacific	14	5	9
Latin America	−5	−16	11
Total International segment	7%	Flat	7%

EXAMPLE 11

Use the information disclosed in Procter & Gamble Company's CAGNY [Consumer Analyst Group of New York] conference slides to answer the following questions:

1 Why does the company present "organic sales growth"?

2 On average, for the four quarters beginning October 2008 and ending September 2009, how did changes in foreign exchange rates affect P&G's reported sales growth?

The Procter & Gamble Company

2012 CAGNY CONFERENCE SLIDES

Reg G Reconciliation of Non-GAAP measures

In accordance with the SEC's Regulation G, the following provides definitions of the non-GAAP measures used in the earnings call and slides with the reconciliation to the most closely related GAAP measure.

1 *Organic Sales Growth:*

Organic sales growth is a non-GAAP measure of sales growth excluding the impacts of acquisitions, divestitures and foreign exchange from year-over-year comparisons. We believe this provides investors with a more complete understanding of underlying sales trends by providing sales growth on a consistent basis. "Organic sales" is also one of the measures used to evaluate senior management and is a factor in determining their at-risk compensation. The reconciliation of reported sales growth to organic sales is as follows:

Total P&G	Net Sales Growth	Foreign Exchange Impact	Acquisition/ Divestiture Impact	Organic Sales Growth
JAS 06	27%	−1%	−20%	6%
OND 06	8%	−3%	0%	5%
JFM07	8%	−2%	0%	6%
AMJ07	8%	−3%	0%	5%
JAS07	8%	−3%	0%	5%
OND07	9%	−5%	1%	5%
JFM08	9%	−5%	1%	5%
AMJ08	10%	−6%	1%	5%
JAS08	9%	−5%	1%	5%
Average–JAS 06–JAS 08	11%	−4%	−2%	5%
OND08	−3%	5%	0%	2%
JFM09	−8%	9%	0%	1%
AMJ09	−11%	9%	1%	−1%
JAS09	−6%	7%	1%	2%
Average–OND 08–JAS 09	−7%	8%	0%	1%
OND09	6%	−2%	1%	5%
JFM010	7%	−3%	0%	4%
AMJ010	5%	−1%	0%	4%

Total P&G	Net Sales Growth	Foreign Exchange Impact	Acquisition/ Divestiture Impact	Organic Sales Growth
JAS010	2%	3%	–1%	4%
OND010	2%	2%	–1%	3%
JFM011	5%	–1%	0%	4%
AMJ011	10%	–5%	0%	5%
JAS011	9%	–5%	0%	4%
OND011	4%	0%	0%	4%
Average–OND 09–OND 11	5%	–1%	0%	4%
JFM 12 (Estimate)	0% to 2%	3%	0%	3% to 5%
AMJ 12(Estimate)	–1% to 2%	5% to 4%	0%	4% to 6%

Solution to 1:

According to its disclosures, Procter & Gamble presents "organic sales growth" because the company believes it provides investors with a better understanding of underlying sales trends and because it is one of the measures used for management evaluation and compensation.

Solution to 2:

The average effect of foreign exchange changes during the period was negative: Although organic sales grew by 1%, the company reported net sales growth of –7% as a result of a negative 8% foreign exchange effect In other words, if no foreign exchange effect had occurred, reported sales growth and organic sales growth would have been equal, both at 1%.

10.2 Disclosures Related to Major Sources of Foreign Exchange Risk

Disclosures about the effects of currency fluctuations often include sensitivity analyses. For example, a company might describe the major sources of foreign exchange risk given its countries of operations and then disclose the profit impact of a given change in exchange rates.

Exhibit 7 includes two excerpts from the 2011 BMW AG annual report. The first excerpt, from the management report, describes the source of the company's currency risks and its approach to measuring and managing those risks. The second excerpt, from the additional disclosures section of the notes, presents the results of the company's sensitivity analysis.

Exhibit 7

Excerpts from 2011 BMW AG Annual Report

Excerpt from the management report describing the source of the company's currency risks and its approach to measuring and managing those risks:

"The sale of vehicles outside the euro zone gives rise to exchange risks. Three currencies (the Chinese renminbi, the US dollar and the British pound) accounted for approximately two-thirds of the BMW Group's foreign currency

(continued)

Exhibit 7 (Continued)

exposures in 2011. We employ cash-flow-at-risk models and scenario analyses to measure exchange rate risks. These tools provide information which serves as the basis for decision-making in the area of currency management.

"We manage currency risks both at a strategic (medium and long term) and at an operating level (short and medium term). In the medium and long term, foreign exchange risks are managed by "natural hedging", in other words by increasing the volume of purchases denominated in foreign currency or increasing the volume of local production. In this context, the expansion of the plant in Spartanburg, USA, and the new plant under construction in Tiexi* at the Shenyang site in China are helping to reduce foreign exchange risks in two major sales markets. For operating purposes (short and medium term), currency risks are hedged on the financial markets. Hedging transactions are entered into only with financial partners of good credit standing. Counterparty risk management procedures are carried out continuously to monitor the creditworthiness of those partners."

Excerpt, from the additional disclosures section of the notes, presenting the results of the company's sensitivity analysis risks:

"The BMW Group measures currency risk using a cash-flow-at-risk model. The starting point for analysing currency risk with this model is the identification of forecast foreign currency transactions or "exposures". At the end of the reporting period, the principal exposures for the coming year were as follows:

in € million	31.12.2011	31.12.2010
Euro/Chinese Renminbi	7,114	6,256
Euro/US Dollar	4,281	3,888
Euro/British Pound	3,266	3,056
Euro/Japanese Yen	1,334	1,086

"In the next stage, these exposures are compared to all hedges that are in place. The net cash flow surplus represents an uncovered risk position. The cash-flow-at-risk approach involves allocating the impact of potential exchange rate fluctuations to operating cash flows on the basis of probability distributions. Volatilities and correlations serve as input factors to assess the relevant probability distributions.

"The potential negative impact on earnings for the current period is computed on the basis of current market prices and exposures to a confidence level of 95% and a holding period of up to one year for each currency. Aggregation of these results creates a risk reduction effect due to correlations between the various portfolios.

"The following table shows the potential negative impact for the BMW Group—measured on the basis of the cash-flow-at-risk approach—attributable at the balance sheet date to unfavourable changes in exchange rates for the principal currencies."

in € million	31.12.2011	31.12.2010
Euro/Chinese Renminbi	180	265
Euro/US Dollar	121	103
Euro/British Pound	182	184
Euro/Japanese Yen	23	30

The level of detail varies in companies' disclosures about sensitivity of earnings to foreign currency fluctuations, with some companies providing information on the range of possible values of foreign exchange rates. An analyst can use sensitivity analysis disclosures in conjunction with his or her own forecast of exchange rates when developing forecasts of profit and cash flow. When detailed disclosures are provided, the analyst can explicitly incorporate foreign exchange impact. Alternatively, in the absence of detailed disclosures, the analyst can incorporate the sensitivity analysis when calibrating the downside risks to base-case profit and cash flow forecasts.

SUMMARY

The translation of foreign currency amounts is an important accounting issue for companies with multinational operations. Foreign exchange rate fluctuations cause the functional currency values of foreign currency assets and liabilities resulting from foreign currency transactions as well as from foreign subsidiaries to change over time. These changes in value give rise to foreign exchange differences that companies' financial statements must reflect. Determining how to measure these foreign exchange differences and whether to include them in the calculation of net income are the major issues in accounting for multinational operations.

- The local currency is the national currency of the country where an entity is located. The functional currency is the currency of the primary economic environment in which an entity operates. Normally, the local currency is an entity's functional currency. For accounting purposes, any currency other than an entity's functional currency is a foreign currency for that entity. The currency in which financial statement amounts are presented is known as the presentation currency. In most cases, the presentation currency will be the same as the local currency.
- When an export sale (import purchase) on an account is denominated in a foreign currency, the sales revenue (inventory) and foreign currency account receivable (account payable) are translated into the seller's (buyer's) functional currency using the exchange rate on the transaction date. Any change in the functional currency value of the foreign currency account receivable (account payable) that occurs between the transaction date and the settlement date is recognized as a foreign currency transaction gain or loss in net income.
- If a balance sheet date falls between the transaction date and the settlement date, the foreign currency account receivable (account payable) is translated at the exchange rate at the balance sheet date. The change in the functional currency value of the foreign currency account receivable (account payable) is recognized as a foreign currency transaction gain or loss in income. Analysts should understand that these gains and losses are unrealized at the time they are recognized and might or might not be realized when the transactions are settled.
- A foreign currency transaction gain arises when an entity has a foreign currency receivable and the foreign currency strengthens or it has a foreign currency payable and the foreign currency weakens. A foreign currency transaction loss arises when an entity has a foreign currency receivable and the foreign currency weakens or it has a foreign currency payable and the foreign currency strengthens.

- Companies must disclose the net foreign currency gain or loss included in income. They may choose to report foreign currency transaction gains and losses as a component of operating income or as a component of non-operating income. If two companies choose to report foreign currency transaction gains and losses differently, operating profit and operating profit margin might not be directly comparable between the two companies.
- To prepare consolidated financial statements, foreign currency financial statements of foreign operations must be translated into the parent company's presentation currency. The major conceptual issues related to this translation process are, What is the appropriate exchange rate for translating each financial statement item, and how should the resulting translation adjustment be reflected in the consolidated financial statements? Two different translation methods are used worldwide.
- Under the current rate method, assets and liabilities are translated at the current exchange rate, equity items are translated at historical exchange rates, and revenues and expenses are translated at the exchange rate that existed when the underlying transaction occurred. For practical reasons, an average exchange rate is often used to translate income items.
- Under the temporal method, monetary assets (and non-monetary assets measured at current value) and monetary liabilities (and non-monetary liabilities measured at current value) are translated at the current exchange rate. Non-monetary assets and liabilities not measured at current value and equity items are translated at historical exchange rates. Revenues and expenses, other than those expenses related to non-monetary assets, are translated at the exchange rate that existed when the underlying transaction occurred. Expenses related to non-monetary assets are translated at the exchange rates used for the related assets.
- Under both IFRS and US GAAP, the functional currency of a foreign operation determines the method to be used in translating its foreign currency financial statements into the parent's presentation currency and whether the resulting translation adjustment is recognized in income or as a separate component of equity.
- The foreign currency financial statements of a foreign operation that has a foreign currency as its functional currency are translated using the current rate method, and the translation adjustment is accumulated as a separate component of equity. The cumulative translation adjustment related to a specific foreign entity is transferred to net income when that entity is sold or otherwise disposed of. The balance sheet risk exposure associated with the current rate method is equal to the foreign subsidiary's net asset position.
- The foreign currency financial statements of a foreign operation that has the parent's presentation currency as its functional currency are translated using the temporal method, and the translation adjustment is included as a gain or loss in income. US GAAP refer to this process as remeasurement. The balance sheet exposure associated with the temporal method is equal to the foreign subsidiary's net monetary asset/liability position (adjusted for non-monetary items measured at current value).
- IFRS and US GAAP differ with respect to the translation of foreign currency financial statements of foreign operations located in a highly inflationary country. Under IFRS, the foreign currency statements are first restated for local inflation and then translated using the current exchange rate. Under US GAAP, the foreign currency financial statements are translated using the temporal method, with no restatement for inflation.

- Applying different translation methods for a given foreign operation can result in very different amounts reported in the parent's consolidated financial statements.
- Companies must disclose the total amount of translation gain or loss reported in income and the amount of translation adjustment included in a separate component of stockholders' equity. Companies are not required to separately disclose the component of translation gain or loss arising from foreign currency transactions and the component arising from application of the temporal method.
- Disclosures related to translation adjustments reported in equity can be used to include these as gains and losses in determining an adjusted amount of income following a clean-surplus approach to income measurement.
- Foreign currency translation rules are well established in both IFRS and US GAAP. Fortunately, except for the treatment of foreign operations located in highly inflationary countries, the two sets of standards have no major differences in this area. The ability to understand the impact of foreign currency translation on the financial results of a company using IFRS should apply equally well in the analysis of financial statements prepared in accordance with US GAAP.
- An analyst can obtain information about the tax impact of multinational operations from companies' disclosure on effective tax rates.
- For a multinational company, sales growth is driven not only by changes in volume and price but also by changes in the exchange rates between the reporting currency and the currency in which sales are made. Arguably, growth in sales that comes from changes in volume or price is more sustainable than growth in sales that comes from changes in exchange rates.

PRACTICE PROBLEMS

The following information relates to Questions 1–6

Pedro Ruiz is an analyst for a credit rating agency. One of the companies he follows, Eurexim SA, is based in France and complies with International Financial Reporting Standards (IFRS). Ruiz has learned that Eurexim used EUR220 million of its own cash and borrowed an equal amount to open a subsidiary in Ukraine. The funds were converted into hryvnia (UAH) on 31 December 20X1 at an exchange rate of EUR1.00 = UAH6.70 and used to purchase UAH1,500 million in fixed assets and UAH300 million of inventories.

Ruiz is concerned about the effect that the subsidiary's results might have on Eurexim's consolidated financial statements. He calls Eurexim's Chief Financial Officer, but learns little. Eurexim is not willing to share sales forecasts and has not even made a determination as to the subsidiary's functional currency.

Absent more useful information, Ruiz decides to explore various scenarios to determine the potential impact on Eurexim's consolidated financial statements. Ukraine is not currently in a hyperinflationary environment, but Ruiz is concerned that this situation could change. Ruiz also believes the euro will appreciate against the hryvnia for the foreseeable future.

1 If Ukraine's economy becomes highly inflationary, Eurexim will *most likely* translate inventory by:

A restating for inflation and using the temporal method.

B restating for inflation and using the current exchange rate.

C using the temporal method with no restatement for inflation.

2 Given Ruiz's belief about the direction of exchange rates, Eurexim's gross profit margin would be *highest* if it accounts for the Ukraine subsidiary's inventory using:

A FIFO and the temporal method.

B FIFO and the current rate method.

C weighted-average cost and the temporal method.

3 If the euro is chosen as the Ukraine subsidiary's functional currency, Eurexim will translate its fixed assets using the:

A average rate for the reporting period.

B rate in effect when the assets were purchased.

C rate in effect at the end of the reporting period.

4 If the euro is chosen as the Ukraine subsidiary's functional currency, Eurexim will translate its accounts receivable using the:

A rate in effect at the transaction date.

B average rate for the reporting period.

C rate in effect at the end of the reporting period.

5 If the hryvnia is chosen as the Ukraine subsidiary's functional currency, Eurexim will translate its inventory using the:

© 2018 CFA Institute. All rights reserved.

A average rate for the reporting period.

B rate in effect at the end of the reporting period.

C rate in effect at the time the inventory was purchased.

6 Based on the information available and Ruiz's expectations regarding exchange rates, if the hryvnia is chosen as the Ukraine subsidiary's functional currency, Eurexim will *most likely* report:

A an addition to the cumulative translation adjustment.

B a translation gain or loss as a component of net income.

C a subtraction from the cumulative translation adjustment.

The following information relates to Questions 7–12

Consolidated Motors is a US-based corporation that sells mechanical engines and components used by electric utilities. Its Canadian subsidiary, Consol-Can, operates solely in Canada. It was created on 31 December 20X1, and Consolidated Motors determined at that time that it should use the US dollar as its functional currency.

Chief Financial Officer Monica Templeton was asked to explain to the board of directors how exchange rates affect the financial statements of both Consol-Can and the consolidated financial statements of Consolidated Motors. For the presentation, Templeton collects Consol-Can's balance sheets for the years ended 20X1 and 20X2 (Exhibit 1), as well as relevant exchange rate information (Exhibit 2).

Exhibit 1 Consol-Can Condensed Balance Sheet for Fiscal Years Ending 31 December (C$ millions)

Account	20X2	20X1
Cash	135	167
Accounts receivable	98	—
Inventory	77	30
Fixed assets	100	100
Accumulated depreciation	(10)	—
Total assets	400	297
Accounts payable	77	22
Long-term debt	175	175
Common stock	100	100
Retained earnings	48	—
Total liabilities and shareholders' equity	400	297

Exhibit 2 Exchange Rate Information

	US$/C$
Rate on 31 December 20X1	0.86
Average rate in 20X2	0.92
Weighted-average rate for inventory purchases	0.92
Rate on 31 December 20X2	0.95

Templeton explains that Consol-Can uses the FIFO inventory accounting method and that purchases of C$300 million and the sell-through of that inventory occurred evenly throughout 20X2. Her presentation includes reporting the translated amounts in US dollars for each item, as well as associated translation-related gains and losses. The board responds with several questions.

- Would there be a reason to change the functional currency to the Canadian dollar?
- Would there be any translation effects for Consolidated Motors if the functional currency for Consol-Can were changed to the Canadian dollar?
- Would a change in the functional currency have any impact on financial statement ratios for the parent company?
- What would be the balance sheet exposure to translation effects if the functional currency were changed?

7 After translating Consol-Can's inventory and long-term debt into the parent company's currency (US$), the amounts reported on Consolidated Motor's financial statements on 31 December 20X2 would be *closest* to (in millions):

A $71 for inventory and $161 for long-term debt.

B $71 for inventory and $166 for long-term debt.

C $73 for inventory and $166 for long-term debt.

8 After translating Consol-Can's 31 December 20X2 balance sheet into the parent company's currency (US$), the translated value of retained earnings will be *closest* to:

A $41 million.

B $44 million.

C $46 million.

9 In response to the board's first question, Templeton would *most likely* reply that such a change would be justified if:

A the inflation rate in the United States became hyperinflationary.

B management wanted to flow more of the gains through net income.

C Consol-Can were making autonomous decisions about operations, investing, and financing.

10 In response to the board's second question, Templeton should reply that if the change is made, the consolidated financial statements for Consolidated Motors would begin to recognize:

A realized gains and losses on monetary assets and liabilities.

B realized gains and losses on non-monetary assets and liabilities.

C unrealized gains and losses on non-monetary assets and liabilities.

11 In response to the board's third question, Templeton should note that the change will *most likely* affect:

A the cash ratio.

B fixed asset turnover.

C receivables turnover.

12 In response to the board's fourth question, the balance sheet exposure (in C$ millions) would be *closest* to:

A –19.

B 148.

C 400.

The following information relates to Questions 13–18

Romulus Corp. is a US-based company that prepares its financial statements in accordance with US GAAP. Romulus Corp. has two European subsidiaries: Julius and Augustus. Anthony Marks, CFA, is an analyst trying to forecast Romulus's 20X2 results. Marks has prepared separate forecasts for both Julius and Augustus, as well as for Romulus's other operations (prior to consolidating the results.) He is now considering the impact of currency translation on the results of both the subsidiaries and the parent company's consolidated financials. His research has provided the following insights:

- The results for Julius will be translated into US dollars using the current rate method.
- The results for Augustus will be translated into US dollars using the temporal method.
- Both Julius and Augustus use the FIFO method to account for inventory.
- Julius had year-end 20X1 inventory of €340 million. Marks believes Julius will report €2,300 in sales and €1,400 in cost of sales in 20X2.

Marks also forecasts the 20X2 year-end balance sheet for Julius (Exhibit 1). Data and forecasts related to euro/dollar exchange rates are presented in Exhibit 2.

Exhibit 1 Forecasted Balance Sheet Data for Julius, 31 December 20X2 (€ millions)

Cash	50
Accounts receivable	100
Inventory	700
Fixed assets	1,450
Total assets	2,300
Liabilities	700
Common stock	1,500
Retained earnings	100
Total liabilities and shareholder equity	2,300

Exhibit 2 Exchange Rates ($/€)	
31 December 20X1	1.47
31 December 20X2	1.61
20X2 average	1.54
Rate when fixed assets were acquired	1.25
Rate when 20X1 inventory was acquired	1.39
Rate when 20X2 inventory was acquired	1.49

13 Based on the translation method being used for Julius, the subsidiary is *most likely:*

A a sales outlet for Romulus's products.

B a self-contained, independent operating entity.

C using the US dollar as its functional currency.

14 To account for its foreign operations, Romulus has *most likely* designated the euro as the functional currency for:

A Julius only.

B Augustus only.

C both Julius and Augustus.

15 When Romulus consolidates the results of Julius, any unrealized exchange rate holding gains on monetary assets should be:

A reported as part of operating income.

B reported as a non-operating item on the income statement.

C reported directly to equity as part of the cumulative translation adjustment.

16 When Marks translates his forecasted balance sheet for Julius into US dollars, total assets as of 31 December 20X2 (dollars in millions) will be *closest* to:

A $1,429.

B $2,392.

C $3,703.

17 When Marks converts his forecasted income statement data for Julius into US dollars, the 20X2 gross profit margin will be *closest* to:

A 39.1%.

B 40.9%.

C 44.6%.

18 Relative to the gross margins the subsidiaries report in local currency, Romulus's consolidated gross margin *most likely*:

A will not be distorted by currency translations.

B would be distorted if Augustus were using the same translation method as Julius.

C will be distorted because of the translation and inventory accounting methods Augustus is using.

The following information relates to Questions 19–24

Redline Products, Inc. is a US-based multinational with subsidiaries around the world. One such subsidiary, Acceletron, operates in Singapore, which has seen mild but not excessive rates of inflation. Acceletron was acquired in 2000 and has never paid a dividend. It records inventory using the FIFO method.

Chief Financial Officer Margot Villiers was asked by Redline's board of directors to explain how the functional currency selection and other accounting choices affect Redline's consolidated financial statements. Villiers gathers Acceletron's financial statements denominated in Singapore dollars (SGD) in Exhibit 1 and the US dollar/Singapore dollar exchange rates in Exhibit 2. She does not intend to identify the functional currency actually in use but rather to use Acceletron as an example of how the choice of functional currency affects the consolidated statements.

Exhibit 1 Selected Financial Data for Acceletron, 31 December 2007 (SGD millions)

Cash	SGD125
Accounts receivable	230
Inventory	500
Fixed assets	1,640
Accumulated depreciation	(205)
Total assets	SGD2,290
Accounts payable	185
Long-term debt	200
Common stock	620
Retained earnings	1,285
Total liabilities and equity	2,290
Total revenues	SGD4,800
Net income	SGD450

Exhibit 2 Exchange Rates Applicable to Acceletron

Exchange Rate in Effect at Specific Times	USD per SGD
Rate when first SGD1 billion of fixed assets were acquired	0.568
Rate when remaining SGD640 million of fixed assets were acquired	0.606
Rate when long-term debt was issued	0.588
31 December 2006	0.649
Weighted-average rate when inventory was acquired	0.654

(continued)

Exhibit 2 (Continued)

Exchange Rate in Effect at Specific Times	USD per SGD
Average rate in 2007	0.662
31 December 2007	0.671

19 Compared with using the Singapore dollar as Acceletron's functional currency for 2007, if the US dollar were the functional currency, it is *most likely* that Redline's consolidated:

A inventories will be higher.

B receivable turnover will be lower.

C fixed asset turnover will be higher.

20 If the US dollar were chosen as the functional currency for Acceletron in 2007, Redline could reduce its balance sheet exposure to exchange rates by:

A selling SGD30 million of fixed assets for cash.

B issuing SGD30 million of long-term debt to buy fixed assets.

C issuing SGD30 million in short-term debt to purchase marketable securities.

21 Redline's consolidated gross profit margin for 2007 would be *highest* if Acceletron accounted for inventory using:

A FIFO, and its functional currency were the US dollar.

B LIFO, and its functional currency were the US dollar.

C FIFO, and its functional currency were the Singapore dollar.

22 If the current rate method is used to translate Acceletron's financial statements into US dollars, Redline's consolidated financial statements will *most likely* include Acceletron's:

A USD3,178 million in revenues.

B USD118 million in long-term debt.

C negative translation adjustment to shareholder equity.

23 If Acceletron's financial statements are translated into US dollars using the temporal method, Redline's consolidated financial statements will *most likely* include Acceletron's:

A USD336 million in inventory.

B USD956 million in fixed assets.

C USD152 million in accounts receivable.

24 When translating Acceletron's financial statements into US dollars, Redline is *least likely* to use an exchange rate of USD per SGD:

A 0.671.

B 0.588.

C 0.654.

The following information relates to questions 25–33

Adrienne Yu is an analyst with an international bank. She analyzes Ambleu S.A. ("Ambleu"), a multinational corporation, for a client presentation. Ambleu complies with IFRS, and its presentation currency is the Norvoltian krone (NVK). Ambleu's two subsidiaries, Ngcorp and Cendaró, have different functional currencies: Ngcorp uses the Bindiar franc (₣B) and Cendaró uses the Crenland guinea (CRG).

Yu first analyzes the following three transactions to assess foreign currency transaction exposure:

Transaction 1:	Cendaró sells goods to a non-domestic customer that pays in dollars on the purchase date.
Transaction 2:	Ngcorp obtains a loan in Bindiar francs on 1 June 2016 from a European bank with the Norvoltian krone as its presentation currency.
Transaction 3:	Ambleu imports inventory from Bindiar under 45-day credit terms, and the payment is to be denominated in Bindiar francs.

Yu then reviews Transactions 2 and 3. She determines the method that Ambleu would use to translate Transaction 2 into its 31 December 2016 consolidated financial statements. While analyzing Transaction 3, Yu notes that Ambleu purchased inventory on 1 June 2016 for ₣B27,000/ton. Ambleu pays for the inventory on 15 July 2016. Exhibit 1 presents selected economic data for Bindiar and Crenland.

Exhibit 1 Selected Economic Data for Bindiar and Crenland

Date	Spot FB/NVK Exchange Rate	Bindiar Inflation Rate (%)	Spot CRG/NVK Exchange Rate	Crenland Inflation Rate (%)	Crenland GPI
31 Dec 2015	—	—	5.6780	—	100.0
1 Jun 2016	4.1779	—	—	—	—
15 Jul 2016	4.1790	—	—	—	—
31 Dec 2016	4.2374	3.1	8.6702	40.6	140.6
Average 2016	4.3450	—	—	—	—
31 Dec 2017	4.3729	2.1	14.4810	62.3	228.2
Average 2017	4.3618	—	11.5823	—	186.2

Prior to reviewing the 2016 and 2017 consolidated financial statements of Ambleu, Yu meets with her supervisor, who asks Yu the following two questions:

Question 1	Would a foreign currency translation loss reduce Ambleu's net sales growth?
Question 2	According to IFRS, what disclosures should be included relating to Ambleu's treatment of foreign currency translation for Ngcorp?

To complete her assignment, Yu analyzes selected information and notes from Ambleu's 2016 and 2017 consolidated financial statements, presented in Exhibit 2.

Exhibit 2 Selected Information and Notes from Consolidated Financial Statements of Ambleu S.A. (in NVK millions)

Income Statement	2017	2016	Balance Sheet	2017	2016
Revenue (1)	1,069	1,034	Cash (3)	467	425
Profit before tax	294	269	Intangibles (4)	575	570
Income tax expense (2)	–96	–94	—	—	—
Net profit	198	175	—	—	—

Note 1: Cendaro's revenue for 2017 is CRG125.23 million.
Note 2:

Reconciliation of Income Tax Expense	2017 (in NVK millions)	2016 (in NVK millions)
Income tax at Ambleu's domestic tax rate	102	92
Effect of tax rates on non-domestic jurisdictions	–14	–9
Unrecognized current year tax losses	8	11
Income tax expense	96	94

Note 3: The parent company transferred NVK15 million to Cendaró on 1 January 2016 to purchase a patent from a competitor for CRG85.17 million.
Note 4: The 2016 consolidated balance sheet includes Ngcorp's total intangible assets of NVK3 million, which were added to Ngcorp's balance sheet on 15 July 2016.

25 Which transaction would generate foreign currency transaction exposure for Ambleu?

A Transaction 1

B Transaction 2

C Transaction 3

26 Yu's determination regarding Transaction 2 should be based on the currency of the:

A loan.

B bank.

C borrower.

27 Based on Exhibit 1, what is the foreign exchange gain resulting from Transaction 3 on the 31 December 2016 financial statements?

A NVK1.70 per ton

B NVK90.75 per ton

C NVK248.54 per ton

28 What is the *best* response to Question 1?

A Yes

B No, because it would reduce organic sales growth

C No, because it would reduce net price realization and mix

29 Based on Exhibit 1, the *best* response to Question 2 is that Ambleu should disclose:

A a restatement for local inflation.

B that assets carried at historical cost are translated at historical rates.

C the amount of foreign exchange differences included in net income.

30 Based on Exhibit 1 and Note 1 in Exhibit 2, the amount that Ambleu should include in its 31 December 2017 revenue from Cendaró is *closest* to:

A NVK10.60 million.

B NVK13.25 million.

C NVK19.73 million.

31 Based on Exhibit 2 and Note 2, the change in Ambleu's consolidated income tax rate from 2016 to 2017 *most likely* resulted from a:

A decrease in Ambleu's domestic tax rate.

B more profitable business mix in its subsidiaries.

C stronger Norvoltian krone relative to the currencies of its subsidiaries.

32 Based on Exhibit 1 and Note 3 in Exhibit 2, the cumulative translation loss recognized by Ambleu related to the patent purchase on the 31 December 2017 financial statements is *closest* to:

A NVK0.39 million.

B NVK1.58 million

C NVK9.12 million.

33 Based on Exhibit 1 and Note 4 in Exhibit 2, the total intangible assets on Ngcorp's balance sheet as of 31 December 2016 are *closest* to:

A FB12.54 million.

B FB12.71 million.

C FB13.04 million.

The following information relates to questions 34–40

Triofind, Inc. (Triofind), based in the country of Norvolt, provides wireless services to various countries, including Norvolt, Borliand, Abuelio, and Certait. The company's presentation currency is the Norvolt euro (NER), and Triofind complies with IFRS. Triofind has two wholly owned subsidiaries, located in Borliand and Abuelio. The Borliand subsidiary (Triofind-B) was established on 30 June 2016, by Triofind both investing NER1,000,000, which was converted into Borliand dollars (BRD), and borrowing an additional BRD500,000.

Marie Janssen, a financial analyst in Triofind's Norvolt headquarters office, translates Triofind-B's financial statements using the temporal method. Non-monetary assets are measured at cost under the lower of cost or market rule. Spot BRD/NER exchange rates are presented in Exhibit 1, and the balance sheet for Triofind-B is presented in Exhibit 2.

Exhibit 1 Spot BRD/NER Exchange Rates

Date	BRD per NER
30 June 2016	1.15
Weighted-average rate when inventory was acquired (2016)	1.19
31 December 2016	1.20

(continued)

Exhibit 1 (Continued)

Date	BRD per NER
Weighted-average rate when inventory was acquired (2017)	1.18
30 June 2017	1.17

Exhibit 2 Triofind-B Balance Sheet for 2016 and 2017 (BRD)

Assets	31 December 2016	30 June 2017	Liabilities and Stockholders' Equity	31 December 2016	30 June 2017
Cash	900,000	1,350,000	Notes payable	500,000	500,000
Inventory	750,000	500,000	Common stock	1,150,000	1,150,000
			Retained earnings		200,000
Total	1,650,000	1,850,000	Total	1,650,000	1,850,000

Janssen next analyzes Triofind's Abuelio subsidiary (Triofind-A), which uses the current rate method to translate its results into Norvolt euros. Triofind-A, which prices its goods in Abuelio pesos (ABP), sells mobile phones to a customer in Certait on 31 May 2017 and receives payment of 1 million Certait rand (CRD) on 31 July 2017.

On 31 May 2017, Triofind-A also received NER50,000 from Triofind and used the funds to purchase a new warehouse in Abuelio. Janssen translates the financial statements of Triofind-A as of 31 July 2017 and must determine the appropriate value for the warehouse in Triofind's presentation currency. She observes that the cumulative Abuelio inflation rate exceeded 100% from 2015 to 2017. Spot exchange rates and inflation data are presented in Exhibit 3.

Exhibit 3 Spot Exchange Rates and Inflation Data for Triofind-A

Date	NER per CRD	NER per ABP	Abuelio Monthly Inflation Rate (%)
31 May 2017	0.2667	0.0496	—
30 June 2017	0.2703	0.0388	25
31 July 2017	0.2632	0.0312	22

Janssen gathers corporate tax rate data and company disclosure information to include in Triofind's annual report. She determines that the corporate tax rates for Abuelio, Norvolt, and Borliand are 35%, 34%, and 0%, respectively, and that Norvolt exempts the non-domestic income of multinationals from taxation. Triofind-B constitutes 25% of Triofind's net income, and Triofind-A constitutes 15%. Janssen also gathers data on components of net sales growth in different countries, presented in Exhibit 4.

Exhibit 4 Components of Net Sales Growth (%) Fiscal Year 2017

Country	Contribution from Volume Growth	Contribution from Price Growth	Foreign Currency Exchange	Net Sales Growth
Abuelio	7	6	–2	11
Borliand	4	5	4	13
Norvolt	7	3	—	10

34 Based on Exhibits 1 and 2 and Janssen's translation method, total assets for Triofind-B translated into Triofind's presentation currency as of 31 December 2016 are *closest* to:

A NER1,375,000.

B NER1,380,252.

C NER1,434,783.

35 Based on Exhibits 1 and 2, the translation adjustment for Triofind-B's liabilities into Triofind's presentation currency for the six months ended 31 December 2016 is:

A negative.

B zero.

C positive.

36 Based on Exhibits 1 and 2 and Janssen's translation method, retained earnings for Triofind-B translated into Triofind's presentation currency as of 30 June 2017 are *closest* to:

A NER150,225.

B NER170,940.

C NER172,414.

37 The functional currency for Triofind-A's sale of mobile phones to a customer in Certait is the:

A Certait real.

B Norvolt euro.

C Abuelio peso.

38 Based on Exhibit 3, the value of the new warehouse in Abuelio on Triofind's balance sheet as of 31 July 2017 is *closest* to:

A NER31,452.

B NER47,964.

C NER50,000.

39 Relative to its domestic tax rate, Triofind's effective tax rate is *most likely*:

A lower.

B the same.

C higher.

40 Based on Exhibit 4, the country with the highest sustainable sales growth is:

A Norvolt.

B Abuelio.

C Borliand.

SOLUTIONS

1 B is correct. IAS 21 requires that the financial statements of the foreign entity first be restated for local inflation using the procedures outlined in IAS 29, "Financial Reporting in Hyperinflationary Economies." Then, the inflation-restated foreign currency financial statements are translated into the parent's presentation currency using the current exchange rate. Under US GAAP, the temporal method would be used with no restatement.

2 B is correct. Ruiz expects the EUR to appreciate against the UAH and expects some inflation in the Ukraine. In an inflationary environment, FIFO will generate a higher gross profit than weighted-average cost. For either inventory choice, the current rate method will give higher gross profit to the parent company if the subsidiary's currency is depreciating. Thus, using FIFO and translating using the current rate method will generate a higher gross profit for the parent company, Eurexim SA, than any other combination of choices.

3 B is correct. If the parent's currency is chosen as the functional currency, the temporal method must be used. Under the temporal method, fixed assets are translated using the rate in effect at the time the assets were acquired.

4 C is correct. Monetary assets and liabilities such as accounts receivable are translated at current (end-of-period) rates regardless of whether the temporal or current rate method is used.

5 B is correct. When the foreign currency is chosen as the functional currency, the current rate method is used. All assets and liabilities are translated at the current (end-of-period) rate.

6 C is correct. When the foreign currency is chosen as the functional currency, the current rate method must be used and all gains or losses from translation are reported as a cumulative translation adjustment to shareholder equity. When the foreign currency decreases in value (weakens), the current rate method results in a negative translation adjustment in stockholders' equity.

7 B is correct. When the parent company's currency is used as the functional currency, the temporal method must be used to translate the subsidiary's accounts. Under the temporal method, monetary assets and liabilities (e.g., debt) are translated at the current (year-end) rate, non-monetary assets and liabilities measured at historical cost (e.g., inventory) are translated at historical exchange rates, and non-monetary assets and liabilities measured at current value are translated at the exchange rate at the date when the current value was determined. Because beginning inventory was sold first and sales and purchases were evenly acquired, the average rate is most appropriate for translating inventory and C$77 million × 0.92 = $71 million. Long-term debt is translated at the year-end rate of 0.95. C$175 million × 0.95 = $166 million.

8 B is correct. Translating the 20X2 balance sheet using the temporal method, as is required in this instance, results in assets of US$369 million. The translated liabilities and common stock are equal to US$325 million, meaning that the value for 20X2 retained earnings is US$369 million – US$325 million = US$44 million.

Temporal Method (20X2)

Account	C$	Rate	US$
Cash	135	0.95	128
Accounts receivable	98	0.95	93

Temporal Method (20X2)			
Account	**C$**	**Rate**	**US$**
Inventory	77	0.92	71
Fixed assets	100	0.86	86
Accumulated depreciation	(10)	0.86	(9)
Total assets	400		369
Accounts payable	77	0.95	73
Long-term debt	175	0.95	166
Common stock	100	0.86	86
Retained earnings	48	to balance	44
Total liabilities and shareholders' equity	400		369

9 C is correct. The Canadian dollar would be the appropriate reporting currency when substantially all operating, financing, and investing decisions are based on the local currency. The parent country's inflation rate is never relevant. Earnings manipulation is not justified, and at any rate changing the functional currency would take the gains off of the income statement.

10 C is correct. If the functional currency were changed from the parent currency (US dollar) to the local currency (Canadian dollar), the current rate method would replace the temporal method. The temporal method ignores unrealized gains and losses on non-monetary assets and liabilities, but the current rate method does not.

11 B is correct. If the Canadian dollar is chosen as the functional currency, the current rate method will be used and the current exchange rate will be the rate used to translate all assets and liabilities. Currently, only monetary assets and liabilities are translated at the current rate. Sales are translated at the average rate during the year under either method. Fixed assets are translated using the historical rate under the temporal method but would switch to current rates under the current rate method. Therefore, there will most likely be an effect on sales/fixed assets. Because the cash ratio involves only monetary assets and liabilities, it is unaffected by the translation method. Receivables turnover pairs a monetary asset with sales and is thus also unaffected.

12 B is correct. If the functional currency were changed, then Consol-Can would use the current rate method and the balance sheet exposure would be equal to net assets (total assets – total liabilities). In this case, 400 – 77 – 175 = 148.

13 B is correct. Julius is using the current rate method, which is most appropriate when it is operating with a high degree of autonomy.

14 A is correct. If the current rate method is being used (as it is for Julius), the local currency (euro) is the functional currency. When the temporal method is being used (as it is for Augustus), the parent company's currency (US dollar) is the functional currency.

15 C is correct. When the current rate method is being used, all currency gains and losses are recorded as a cumulative translation adjustment to shareholder equity.

16 C is correct. Under the current rate method, all assets are translated using the year-end 20X2 (current) rate of $1.61/€1.00. €2,300 × 1.61 = $3,703.

17 A is correct. Under the current rate method, both sales and cost of goods sold would be translated at the 20X2 average exchange rate. The ratio would be the same as reported under the euro. €2,300 – €1,400 = €900, €900/€2,300 = 39.1%. Or, $3,542 – $2,156 = $1,386, $1,386/$3,542 = 39.1%.

18 C is correct. Augustus is using the temporal method in conjunction with FIFO inventory accounting. If FIFO is used, ending inventory is assumed to be composed of the most recently acquired items, and thus inventory will be translated at relatively recent exchange rates. To the extent that the average weight used to translate sales differs from the historical rate used to translate inventories, the gross margin will be distorted when translated into US dollars.

19 C is correct. If the US dollar is the functional currency, the temporal method must be used. Revenues and receivables (monetary asset) would be the same under either accounting method. Inventory and fixed assets were purchased when the US dollar was stronger, so at historical rates (temporal method), translated they would be lower. Identical revenues/lower fixed assets would result in higher fixed-asset turnover.

20 A is correct. If the US dollar is the functional currency, the temporal method must be used, and the balance sheet exposure will be the net monetary assets of 125 + 230 − 185 − 200 = −30, or a net monetary liability of SGD30 million. This net monetary liability would be eliminated if fixed assets (non-monetary) were sold to increase cash. Issuing debt, either short-term or long-term, would increase the net monetary liability.

21 A is correct. Because the US dollar has been consistently weakening against the Singapore dollar, cost of sales will be lower and gross profit higher when an earlier exchange rate is used to translate inventory, compared with using current exchange rates. If the Singapore dollar is the functional currency, current rates would be used. Therefore, the combination of the US dollar (temporal method) and FIFO will result in the highest gross profit margin.

22 A is correct. Under the current rate method, revenue is translated at the average rate for the year, SGD4,800 × 0.662 = USD3,178 million. Debt should be translated at the current rate, SGD200 × 0.671 = USD134 million. Under the current rate method, Acceletron would have a net asset balance sheet exposure. Because the Singapore dollar has been strengthening against the US dollar, the translation adjustment would be positive rather than negative.

23 B is correct. Under the temporal method, inventory and fixed assets would be translated using historical rates. Accounts receivable is a monetary asset and would be translated at year-end (current) rates. Fixed assets are found as (1,000 × 0.568) + (640 × 0.606) = USD 956 million.

24 B is correct. USD0.671/SGD is the current exchange rate. That rate would be used regardless of whether Acceletron uses the current rate or temporal method. USD0.654 was the weighted-average rate when inventory was acquired. That rate would be used if the company translated its statements under the temporal method but not the current rate method. USD0.588/SGD was the exchange rate in effect when long-term debt was issued. As a monetary liability, long-term debt is always translated using current exchange rates. Consequently, that rate is not applicable regardless of how Acceletron translates its financial statements.

25 C is correct. In Transaction 3, the payment for the inventory is due in Bindiar francs, a different currency from the Norvoltian krone, which is Ambleu's presentation currency. Because the import purchase (account payable) is under 45-day credit terms, Ambleu has foreign currency transaction exposure. The payment is subject to fluctuations in the ₣B/NVK exchange rate during the 45-day period between the sale and payment dates. Thus, Ambleu is exposed to potential foreign currency gains if the Bindiar franc weakens against the Norvoltian krone or foreign currency losses if the Bindiar franc strengthens against the Norvoltian krone.

26 C is correct. The currency of Ngcorp as the borrowing foreign subsidiary, relative to that of Ambleu, determines Ambleu's choice of translation method for Transaction 2. Because Ngcorp's functional currency is the Bindiar franc and Ambleu's presentation currency is the Norvoltian krone, the current rate method rather than the temporal method should be used. Regardless of the currency in which the loan is denominated, the loan is first recorded in Ngcorp's financial statements. Then, Ngcorp's financial statements, which include the bank loan, are translated into Ambleu's consolidated financial statements.

27 A is correct. On Ambleu's balance sheet, the cost included in the inventory account is the translation of ₣B27,000/ton into Norvoltian krone on the purchase date. Ambleu could have paid this amount on the purchase date but chose to wait 45 days to settle the account. The inventory cost is determined using the ₣B/NVK exchange rate of 4.1779 on the purchase date of 1 June 2016. ₣B27,000/₣B4.1779/NVK = NVK6,462.58/ton

The cash outflow is the amount exchanged from the Norvoltian krone to the Bindiar franc to pay the ₣B27,000/ton owed for the inventory 45 days after the transaction date. This payment uses the ₣B/NVK exchange rate of 4.1790 on the settlement date of 15 July 2016.

₣B 27,000/₣B4.1790 per NVK = NVK6,460.88/ton

Foreign exchange gain = Inventory cost − Cash payment
= NVK6,462.58 − NVK6,460.88
= NVK1.70/ton

Thus, Ambleu's cash outflow is less than the cost included in the inventory account, and NVK1.70/ton is the realized foreign exchange gain relating to this transaction. By deferring payment for 45 days, and because the Bindiar franc decreased in value during this period, Ambleu pays NVK1.70/ton less than the inventory cost on the purchase date of 1 June 2016. Thus, Ambleu will report a foreign exchange gain in its 2016 net income.

28 A is correct. Net sales growth equals organic sales growth plus or minus the effects of acquisitions, divestitures, and foreign exchange. A foreign currency translation loss would reduce net sales growth. Thus the answer to Question 1 is yes.

29 C is correct. IFRS requires that Ambleu disclose "the amount of exchange differences recognized in profit or loss" when determining net income for the period. Because companies may present foreign currency transaction gains and losses in various places on the income statement, it is useful for companies to disclose both the amount of transaction gain or loss that is included in income as well as the presentation alternative used.

30 A is correct. Crenland experienced hyperinflation from 31 December 2015 to 31 December 2017, as shown by the General Price Index, with cumulative inflation of 128.2% during this period. According to IFRS, Cendaró's financial statements must be restated for local inflation, then translated into Norvoltian kroner using the current exchange rate. The 2017 revenue from Cendaró that should be included in Ambleu's income statement is calculated as follows:

Revenue in CRG × (GPI 31 December 2017/GPI average 2017) = Inflation-adjusted revenue in CRG

CRG125.23 million × (228.2/186.2) = CRG153.48 million

Inflation-adjusted revenue in CRG/31 December 2017 exchange rate (CRG/NVK) = Revenue in Norvoltian kroner

CRG153.48 million/14.4810 = NVK10.60 million

31 B is correct. The consolidated income tax rate is calculated as income tax expense divided by profit before tax. Note 2 shows that Ambleu's consolidated income tax rate decreases by 2.29%, from 34.94% (=94/269) in 2016 to 32.65% (=96/294) in 2017. The largest component of the decrease stems from the 1.42% change in the effect of tax rates in non-domestic jurisdictions, which lowers Ambleu's consolidated income tax rate in 2016 by 3.34% (=9/269) and in 2017 by 4.76% (=14/294). The decrease in 2017 could indicate that Ambleu's business mix shifted to countries with lower marginal tax rates, resulting in a lower consolidated income tax rate and more profit. (The change could also indicate that the marginal tax rates decreased in the countries in which Ambleu earns profits.)

32 B is correct. IAS 29 indicates that a cumulative inflation rate approaching or exceeding 100% over three years would be an indicator of hyperinflation. Because the cumulative inflation rate for 2016 and 2017 in Crenland was 128.2%, Cendaró's accounts must first be restated for local inflation. Then, the inflation-restated Crenland guinea financial statements can be translated into Ambleu's presentation currency, the Norvoltian krone, using the current exchange rate.

Using this approach, the cumulative translation loss on 31 December 2017 for the CRG85.17 million patent purchase is –NVK1.58 million, as shown in the following table.

Date	Inflation Rate (%)	Restated Carrying Value (CRG/MM)	Current Exchange Rate (CRG/NVK)	Translated Amount (NVK MM)	Annual Translation Gain/Loss (NVK MM)	Cumulative Translation Gain/Loss (NVK MM)
1 Jan 2016	—	85.17	5.6780	15.00	N/A	N/A
31 Dec 2016	40.6	119.75	8.6702	13.81	–1.19	–1.19
31 Dec 2017	62.3	194.35	14.4810	13.42	–0.39	–1.58

33 B is correct. Because Ngcorp has a functional currency that is different from Ambleu's presentation currency, the intangible assets are translated into Norvoltian kroner using the current rate method. The current ₣B/NVK exchange rate is 4.2374 as of 31 December 2016. Thus, the intangible assets on Ngcorp's 2016 balance sheet are NVK3 million × ₣B4.2374/NVK = ₣B12.71 million.

34 B is correct. Using the temporal method, monetary assets (i.e., cash) are translated using the current exchange rate (as of 31 December 2016) of BRD1.20/NER (or NER0.8333/BRD), and non-monetary assets are translated using the historical exchange rate when acquired. Inventory is translated at its 2016 weighted-average rate of BRD1.19/NER (or NER0.8403/BRD). Therefore, the total assets for Triofind-B translated into Norvolt euros (Triofind's presentation currency) as of 31 December 2016 are calculated as follows:

Assets	31 December 2016 (BRD)	Applicable Exchange Rate (NER/BRD)	Rate Used	NER
Cash	900,000	0.8333	Current	750,000
Inventory	750,000	0.8403	Average	630,252
Total	1,650,000			1,380,252

35 A is correct. The monetary balance sheet items for Triofind-B are translated at the current exchange rate, which reflects that the Borliand dollar weakened during the period relative to the Norvolt euro. The rate as of 30 June 2016 was BRD1.15/NER (or NER/BRD0.8696) and as of 31 December 2016 was BRD1.20/NER (or NER/BRD0.8333). Therefore, notes payable translates to NER416,667 (BRD500,000 × NER/BRD0.8333) as of 31 December 2016, compared with NER434,783 (BRD500,000 × NER/BRD0.8696) as of 30 June 2016. Thus, the translation adjustment for liabilities is negative.

36 A is correct. Triofind uses the temporal method to translate the financial statements of Triofind-B. The temporal method uses the current exchange rate for translating monetary assets and liabilities and the historical exchange rate (based on the date when the assets were acquired) for non-monetary assets and liabilities. Monetary assets and liabilities are translated using the current exchange rate (as of 30 June 2017) of NER1 = BRD1.17 (or NER0.8547/BRD), and non-monetary assets and liabilities are translated using the historical exchange rate (as of 30 June 2016) of NER1 = BRD1.15 (or NER0.8696/BRD). Inventory is translated at the 2017 weighted average rate of NER1 = BRD1.18 (or NER0.8475/BRD). The difference required to maintain equality between (a) total assets and (b) total liabilities and shareholder's equity is then recorded as retained earnings. The retained earnings for Triofind-B translated into Norvolt euros (Triofind's presentation currency) as of 30 June 2017 is calculated as follows:

Assets	30 June 2017 (BRD)	Exchange Rate (NER/BRD)	Rate Used	30 June 2017 (NER)	Liabilities and Stockholders' Equity	30 June 2017 (BRD)	Exchange Rate (NER/BRD)	Rate Used	30 June 2017 (NER)
Cash	1,350,000	0.8547	C	1,153,846	Notes Payable	500,000	0.8547	C	427,350
Inventory	500,000	0.8475	H	423,729	Common Stock	1,150,000	0.8696	H	1,000,000
					Retained Earnings	200,000			**150,225**
	1,850,000			1,577,575	Total	1,850,000			1,577,575

37 C is correct. The functional currency is the currency of the primary economic environment in which an entity operates. Abuelio is Triofind-A's primary economic environment, and its currency is the Abuelio peso (ABP). Another important factor used to determine the functional currency is the currency that mainly influences sales prices for goods and services. The fact that Triofind-A prices its goods in Abuelio pesos supports the case for the ABP to be the functional currency.

38 B is correct. Triofind complies with IFRS, and Abuelio can be considered a highly inflationary economy because its cumulative inflation rate exceeded 100% from 2015 to 2017. Thus, Triofind-A's financials must be restated to include local inflation rates and then translated using the current exchange rate into Norvolt euros, which is Triofind's presentation currency. This approach reflects both the likely change in the local currency value of the warehouse as well as the actual change in the exchange rate. The original purchase price is ABP1,008,065 (NER50,000/ABP0.0496). The value of the new warehouse in Abuelio as of 31 July 2017 is NER47,964, calculated as follows:

Date	Abuelio Monthly Inflation Rate (%)	Restated Warehouse Value (ABP)	NER/ABP	Warehouse Value (NER)
31 May 2017		1,008,065	0.0496	50,000
30 June 2017	25	1,260,081	0.0388	48,891
31 July 2017	22	1,537,298	0.0312	47,964

39 A is correct. Norvolt exempts the non-domestic income of multinationals from taxation. Because Norvolt has a corporate tax rate of 34%, the 0% tax rate in Borliand and the fact that 25% of Triofind's net income comes from Borliand should result in a lower effective tax rate on Triofind's consolidated financial statements compared with Triofind's domestic tax rate. Abuelio's tax rate of 35% is very close to that of Norvolt, and it constitutes only 15% of Triofind's net income, so its effect is unlikely to be significant.

40 B is correct. Although Borliand shows the highest growth in Norvolt euro terms, this result is partially because of currency fluctuations, which cannot be controlled. Abuelio had the highest change in sales resulting from price and volume at 13% (excluding foreign currency exchange). This growth is more sustainable than net sales growth, which includes currency fluctuations, because Triofind's management has more control over growth in sales resulting from greater volume or higher prices.

READING

12

Analysis of Financial Institutions

by Jack T. Ciesielski, CPA, CFA, and Elaine Henry, PhD, CFA

Jack T. Ciesielski, CPA, CFA, is at R.G. Associates, Inc., former publisher of The Analyst's Accounting Observer (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** describe how financial institutions differ from other companies;
☐	**b.** describe key aspects of financial regulations of financial institutions;
☐	**c.** explain the CAMELS (capital adequacy, asset quality, management, earnings, liquidity, and sensitivity) approach to analyzing a bank, including key ratios and its limitations;
☐	**d.** describe other factors to consider in analyzing a bank;
☐	**e.** analyze a bank based on financial statements and other factors;
☐	**f.** describe key ratios and other factors to consider in analyzing an insurance company.

INTRODUCTION TO FINANCIAL INSTITUTIONS

1

a describe how financial institutions differ from other companies;

b describe key aspects of financial regulations of financial institutions;

Financial institutions provide a wide range of financial products and services. They serve as intermediaries between providers and recipients of capital, facilitate asset and risk management, and execute transactions involving cash, securities, and other financial assets.

Given the diversity of financial services, it is unsurprising that numerous types of financial institutions exist. Types of financial institutions include deposit-taking, loan-making institutions (referred to as *banks* in this reading), investment banks, credit card companies, brokers, dealers, exchanges, clearing houses, depositories, investment managers, financial advisers, and insurance companies. In many situations, overlap of services exists across types of institutions. For example, banks not only

© 2018 CFA Institute. All rights reserved.

take deposits and make loans but also may undertake investment management and other securities-related activities and may offer such products as derivatives, which are effectively insurance against adverse effects of movements in the interest rate, equity, and foreign currency markets. As another example of overlap, life insurance companies not only provide mortality-related insurance products but also offer savings vehicles. This reading focuses primarily on two types of financial institutions: banks (broadly defined as deposit-taking, loan-making institutions) and insurance companies.

Section 1 explains what makes financial institutions different from other types of companies, such as manufacturers or merchandisers. Sections 2–4 discuss how to analyze a bank. Sections 5 and 6 focus on analyzing insurance companies. A summary of key points concludes the reading.

1.1 What Makes Financial Institutions Different?

A distinctive feature of financial institutions—in particular, banks—is their systemic importance, which means that their smooth functioning is essential to the overall health of an economy. The most fundamental role of banks is to serve as intermediaries, accepting deposits from capital providers and providing capital via loans to borrowers. Their role as intermediaries between and among providers and recipients of capital creates financial inter-linkages across all types of entities, including households, banks, corporations, and governments. The network of inter-linkages across entities means that the failure of one bank will negatively affect other financial and non-financial entities. The larger the bank and the more widespread its inter-linkages, the greater its potential impact on the entire financial system. If an extremely large bank were to fail, the negative impact of its failure could spread and potentially result in the failure of the entire financial system.

Systemic risk has been defined as "a risk of disruption to financial services that is (i) caused by an impairment of all or parts of the financial system and (ii) has the potential to have serious negative consequences for the economy as a whole. Fundamental to the definition is the notion of contagion across the economy from a disruption or failure in a financial institution, market or instrument. All types of financial intermediaries, markets and infrastructure can potentially be systemically important to some degree."[1] The problem of systemic risk (the risk of failure of the financial system as a result of the failure of a major financial institution) has emerged as an issue in many countries around the world in the aftermath of the 2008 global financial crisis. *Financial contagion* is a situation in which financial shocks spread from their place or sector of origin to other locales or sectors. Globally, a faltering economy may infect other, healthier economies.

Because of their systemic importance, financial institutions' activities are heavily regulated. Regulations attempt to constrain excessive risk taking that could cause an entity to fail. Regulations address various aspects of a financial institution's operations, including the amount of capital that must be maintained, the minimum liquidity, and the riskiness of assets.

The liabilities of most banks are made up primarily of deposits. For example, as of December 2016, deposits constituted over 80% of the total liabilities of domestically chartered commercial banks in the United States.[2] The failure of a bank to honor its deposits could have negative consequences across the economy. Even the expectation

1 "Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations," report to the G–20 finance ministers and central bank governors, prepared by the staff of the International Monetary Fund and the Bank for International Settlements and the secretariat of the Financial Stability Board (October 2009): https://www.imf.org/external/np/g20/pdf/100109.pdf.

2 "Assets and Liabilities of Commercial Banks in the United States - H.8," Federal Reserve statistical release (https://www.federalreserve.gov).

that a bank might not be able to honor its deposits could cause depositors to withdraw their money from the bank, and a large sudden withdrawal of deposits (a bank run) could cause an actual failure and financial contagion across the economy. Therefore, deposits are often insured (up to a stated limit) by the government of the country in which the bank operates.

Another distinctive feature of financial institutions is that their assets are predominantly financial assets, such as loans and securities. In contrast, the assets of most non-financial companies are predominantly tangible assets. Financial assets create direct exposure to a different variety of risks, including credit risks, liquidity risks, market risks, and interest rate risks. Unlike many tangible assets, financial assets are often measured at fair market value for financial reporting.

This reading focuses on the financial analysis of banks and insurers (property and casualty insurers and life and health insurers). There are many other types of financial institutions, including different types of depository institutions. Some of these other financial institutions are described briefly in Exhibit 1. Note that the list in Exhibit 1 includes types of entities that an analyst may evaluate for potential investment and, therefore, excludes supra-national organizations. Typically, supra-national entities are formed by member countries to focus on lending activities in support of specific missions. For example, the World Bank—whose mission is to reduce poverty and support development globally—comprises 189 member countries and provides loans and grants through the International Bank for Reconstruction and Development and the International Development Association.[3] Other prominent examples of supra-national entities are the Asian Development and Asian Infrastructure Investment Bank.

Exhibit 1 A Sampling of Financial Institutions

The list that follows is illustrative only and should not be viewed as comprehensive. The list is organized by primary activity, but many service overlaps exist. Additionally, the structure of financial service providers differs across countries, and state ownership of financial institutions is more common in some countries.

Institutions That Provide Basic Banking Services

- **Commercial banks.** This term generally refers to institutions whose business focuses on classic banking services, such as taking deposits, making loans, and facilitating payment transactions. Historically, regulation in some countries, such as the United States and France, created distinctions between commercial banking activities (e.g., deposit taking and loan making), insurance activities, and investment banking activities, such as securities underwriting, trading, and investing. In general, this distinction has been declining. For example, in France, regulations beginning in the mid-1980s eliminated many restrictions on banks' allowable types of activities, and in the United States, a 1999 law granted commercial banks the ability to undertake broad-based securities and insurance activities.[4] Germany's universal banks provide commercial banking, investment banking, insurance, and other financial and non-financial services, and Spain's leading commercial banks are "dominant in cross-selling mutual funds to their retail clients."[5] Japanese banks are permitted to engage in a

(continued)

3 www.worldbank.org.

4 Berger, Allen N., Phillip Molyneux, and John O.S. Wilson, *The Oxford Handbook of Banking* (Oxford, UK: Oxford University Press, 2009).

5 Berger et al., *The Oxford Handbook of Banking*.

Exhibit 1 (Continued)

range of activities including equity ownership in non-financial corporations (within limits) that strengthens their role in corporate governance beyond that typical of a creditor.[6]

- **Credit unions, cooperative and mutual banks.** These are depository institutions that function like banks and offer many of the same services as banks. They are owned by their members, rather than being publicly traded like many banks. Another difference from commercial banks is that these institutions are organized as non-profits and, therefore, do not pay income taxes.
- **Specialized financial service providers.**
  - **Building societies** and **savings and loan associations** are depository institutions that specialize in financing long-term residential mortgages.
  - **Mortgage banks** originate, sell, and service mortgages and are usually active participants in the securitization markets.
  - **Trust banks (Japan)** are commercial banks, and because their deposits are in the form of "money trusts" (typically with three- to five-year terms and one-year minimums), they can make long-term commercial loans and securities investments. Japan also has city banks (universal banks), regional banks, second regional banks, and Shinkin banks and credit cooperatives (which provide commercial banking services to their members—smaller enterprises and individuals).[7]
  - **Online payment companies**, such as Paypal (United States), Alipay (China), and other non-bank online payment companies, have expanded rapidly and continue to broaden service offerings.

Intermediaries within the Investment Industry

Within this category, services offered by different entities are particularly varied. A few of these are described briefly below.

- *Managers of pooled investment vehicles, such as open-end mutual funds, closed-end funds, and exchange-traded funds.* These financial institutions pool money from investors and buy and sell securities and other assets. The investors share ownership in the investment vehicle. Pooled investment vehicles, as required by regulation, disclose their investment policies, deposit and redemption procedures, fees and expenses, past performance statistics, and other information.
- *Hedge funds.* These funds also pool investors' money and invest it. They tend to follow more complex strategies; be less transparent, less liquid, and less regulated; and have higher fees and higher minimum investment amounts than open-end mutual funds, closed-end funds, and exchange-traded funds.
- *Brokers and dealers.* These firms facilitate trade in securities, earning a commission or spread on the trades.

6 Berger et al., *The Oxford Handbook of Banking.*
7 Berger et al., *The Oxford Handbook of Banking.*

Exhibit 1 (Continued)

Insurers

- **Property and casualty (P&C) insurance companies** provide protection against adverse events related to autos, homes, or commercial activities.
- **Life and health (L&H) insurers** provide mortality- and health-related insurance products. Life insurance companies also provide savings products.
- **Reinsurance companies** sell insurance to insurers. Rather than paying policyholder claims directly, they reimburse insurance companies for claims paid.[8]

1.2 Global Organizations

With respect to global systemic risk, important differences exist between the banking and insurance sectors.[9] Unlike banks, the overall insurance market has a smaller proportion of cross-border business, although the reinsurance business is largely international. The international aspect of the reinsurance business increases the importance of the insurance sector to the global financial system: Reinsurers may be an international link to financial institutions domiciled in different parts of the world, thereby increasing systemic vulnerability. Another important difference is that insurance companies' foreign branches are generally required to hold assets in a jurisdiction that are adequate to cover the related policy liabilities in that jurisdiction.

Aside from minimizing systemic risk, other reasons for the establishment of global and regional regulatory bodies include the harmonization and globalization of regulatory rules, standards, and oversight. Consistency of standards and regulations helps minimize regulatory arbitrage (whereby multinational companies capitalize on differences in jurisdictions' regulatory systems in order to avoid unfavorable regulation) around the world.

One of the most important global organizations focused on financial stability is the Basel Committee on Banking Supervision, which was established in 1974 and is a standing committee hosted and supported by the Bank for International Settlements. Members of the Basel Committee include central banks and entities responsible for supervising banks. The list of members of the Basel Committee in Exhibit 2 illustrates the range of entities involved with supervising banking activity in different countries and jurisdictions.

Exhibit 2 Members of the Basel Committee as of July 2017

Country/Jurisdiction	Institutional Representative
Argentina	Central Bank of Argentina
Australia	Reserve Bank of Australia Australian Prudential Regulation Authority
Belgium	National Bank of Belgium

(continued)

8 Insurance Information Institute (www.iii.org).

9 "Core Principles: Cross-Sectoral Comparison," report by the Joint Forum (Basel Committee on Banking Supervision, International Organization of Securities Commissions, and International Association of Insurance Supervisors; November 2001): https://www.iaisweb.org/page/supervisory-material/joint-forum//file/34300/core-principles-cross-sectoral-comparison.

Exhibit 2 (Continued)

Country/Jurisdiction	Institutional Representative
Brazil	Central Bank of Brazil
Canada	Bank of Canada Office of the Superintendent of Financial Institutions
Chinese mainland	People's Bank of China China Banking Regulatory Commission
European Union	European Central Bank European Central Bank Single Supervisory Mechanism
France	Bank of France Prudential Supervision and Resolution Authority
Germany	Deutsche Bundesbank (Central Bank of Germany) Federal Financial Supervisory Authority (BaFin)
Hong Kong SAR	Hong Kong Monetary Authority
India	Reserve Bank of India
Indonesia	Bank Indonesia Indonesia Financial Services Authority
Italy	Bank of Italy
Japan	Bank of Japan Financial Services Agency
Korea	Bank of Korea Financial Supervisory Service
Luxembourg	Surveillance Commission for the Financial Sector
Mexico	Bank of Mexico Comisión Nacional Bancaria y de Valores (National Banking and Securities Commission)
Netherlands	Netherlands Bank
Russia	Central Bank of the Russian Federation
Saudi Arabia	Saudi Arabian Monetary Agency
Singapore	Monetary Authority of Singapore
South Africa	South African Reserve Bank
Spain	Bank of Spain
Sweden	Sveriges Riksbank (Central Bank of Sweden) Finansinspektionen (Financial Supervisory Authority)
Switzerland	Swiss National Bank Swiss Financial Market Supervisory Authority FINMA
Turkey	Central Bank of the Republic of Turkey Banking Regulation and Supervision Agency
United Kingdom	Bank of England Prudential Regulation Authority
United States	Board of Governors of the Federal Reserve System Federal Reserve Bank of New York Office of the Comptroller of the Currency Federal Deposit Insurance Corporation

Exhibit 2 (Continued)

Observers

Country/Jurisdiction	Institutional representative
Chile	Central Bank of Chile Banking and Financial Institutions Supervisory Agency
Malaysia	Central Bank of Malaysia
United Arab Emirates	Central Bank of the United Arab Emirates

Source: www.bis.org.

The Basel Committee developed the international regulatory framework for banks known as Basel III, which is the enhanced framework succeeding Basel I and Basel II. The purposes of the measures contained in Basel III are the following: "to improve the banking sector's ability to absorb shocks arising from financial and economic stress, whatever the source, improve risk management and governance, and strengthen banks' transparency and disclosures."[10]

Three important highlights of Basel III are the minimum capital requirement, minimum liquidity, and stable funding. First, Basel III specifies the minimum percentage of its risk-weighted assets that a bank must fund with equity capital. This minimum capital requirement prevents a bank from assuming so much financial leverage that it is unable to withstand loan losses (asset write-downs). Second, Basel III specifies that a bank must hold enough high-quality liquid assets to cover its liquidity needs in a 30-day liquidity stress scenario. This minimum liquidity requirement ensures that a bank would have enough cash to cover a partial loss of funding sources (e.g., customers' deposits, other borrowings) or a cash outflow resulting from off-balance-sheet funding commitments. Third, Basel III requires a bank to have a minimum amount of stable funding relative to the bank's liquidity needs over a one-year horizon. Stability of funding is based on the tenor of deposits (e.g., longer-term deposits are more stable than shorter-term deposits) and the type of depositor (e.g., funds from consumers' deposits are considered more stable than funds raised in the interbank markets).

As a result of preventing banks from assuming excessive financial leverage, Basel III has prompted banks to focus on asset quality, hold capital against other types of risk (such as operational risk), and develop improved risk assessment processes. Basel III also presents fundamental changes regarding the quality and composition of the capital base of financial institutions. It has improved the ability of their capital base to sustain losses, so these are confined to the financial institutions' capital investors and are not transmitted to depositors, taxpayers, or other institutions in the financial system, thereby reducing risk of contagion.

Having developed the regulatory framework, the Basel Committee monitors the adoption and implementation of Basel III by member jurisdictions.

A number of other important organizations are involved in international cooperation in the area of financial stability. Some of these international organizations are described briefly below.

- The Financial Stability Board includes representatives from supervisory and regulatory authorities for the G–20 members plus Hong Kong SAR, Singapore, Spain, and Switzerland. Its overall goal is to strengthen financial stability. It aims to identify systemic risk in the financial sector and coordinate actions that jurisdictional authorities can take to address the risks.

10 www.bis.org.

- The International Association of Deposit Insurers' objective is to "enhance the effectiveness of deposit insurance systems."
- The International Association of Insurance Supervisors (IAIS) includes representatives from insurance regulators and supervisors from most countries around the world. Its overall goal is to promote effective supervision of the insurance industry globally.
- The International Organization of Securities Commissions (IOSCO) includes representatives from the regulators of the securities markets of various countries and jurisdictions. Its overall goals include maintaining fair and efficient securities markets.

The latter two organizations are part of a Joint Forum with the Basel Committee. The Joint Forum comprises representatives from the Basel Committee, IAIS, and IOSCO and works on issues common to the banking, insurance, and securities sectors.

1.3 Individual Jurisdictions' Regulatory Authorities

The global organizations described in the previous section aim to foster financial stability by working with individual jurisdictions' regulatory authorities. It is the individual jurisdictions' regulatory bodies that have authority over specific aspects of a financial institution's operations.

Globally, there are many regulators with overlapping and differing responsibilities over financial institutions; the global network of regulators and the resulting regulations are complex. Although there is some overlap between member institutions in the Basel Committee and other global organizations mentioned in the previous section, specific membership varies. For example, the 83 member organizations of the International Association of Deposit Insurers include some institutions that are Basel Committee members, such as the US Federal Deposit Insurance Corporation (FDIC), and some that are not Basel Committee members, such as the Singapore Deposit Insurance Corporation Ltd. and Germany's Bundesverband deutscher Banken (Deposit Protection Fund). In some countries, the same regulatory body oversees both banking and insurance—for example, Japan's Financial Services Agency. And in other countries, there is a separate regulatory body for insurance companies—for example, the US National Association of Insurance Commissioners (NAIC) and the China Insurance Regulatory Commission.

As a financial institution's operations expand globally, compliance requirements increase. One of the most global financial institutions, HSBC Holdings, discloses that their operations are "regulated and supervised by approximately 400 different central banks and other regulatory authorities in those jurisdictions in which we have offices, branches or subsidiaries. These authorities impose a variety of requirements and controls."[11]

2 ANALYZING A BANK: THE CAMELS APPROACH

c explain the CAMELS (capital adequacy, asset quality, management, earnings, liquidity, and sensitivity) approach to analyzing a bank, including key ratios and its limitations;

e analyze a bank based on financial statements and other factors;

11 HSBC Holdings Form 20-F (31 December 2016).

In this section, the term "bank" is used in its general sense and applies to entities whose primary business activities are taking deposits and making loans. This section first describes an approach widely used as a starting point to analyze a bank, known as CAMELS, and follows with a description of additional factors to consider when analyzing a bank. The section concludes with a case study analysis of a real bank.

2.1 The CAMELS Approach

"CAMELS" is an acronym for the six components of a widely used bank rating approach originally developed in the United States.[12] The six components are **C**apital adequacy, **A**sset quality, **M**anagement capabilities, **E**arnings sufficiency, **L**iquidity position, and **S**ensitivity to market risk.

A bank examiner using the CAMELS approach to evaluate a bank conducts an analysis and assigns a numerical rating of 1 through 5 to each component. A rating of 1 represents the best rating, showing the best practices in risk management and performance and generating the least concern for regulators. A rating of 5 is the worst rating, showing the poorest performance and risk management practices and generating the highest degree of regulatory concern.[13] After the components are rated, a composite rating for the entire bank is constructed from the component ratings. This is not a simple arithmetic mean of the six component ratings: Each component is weighted by the examiner performing the study. The examiner's judgment will affect the weighting accorded to each component's rating. Two examiners could evaluate the same bank on a CAMELS basis and even assign the same ratings to each component and yet arrive at different composite ratings for the entire bank.

Although the CAMELS system was developed as a tool for bank examiners, it provides a useful framework for other purposes, such as equity or debt investment analysis of banks. The following sections discuss each component of the rating system.

2.1.1 *Capital Adequacy*

It is important for a bank (as with any company) to have adequate capital so that potential losses can be absorbed without causing the bank to become financially weak or even insolvent. Losses reduce the amount of a bank's retained earnings, which is one component of capital. Large enough losses could even result in insolvency. A strong capital position lowers the probability of insolvency and bolsters public confidence in the bank.

Capital adequacy for banks is described in terms of the proportion of the bank's assets funded with capital. For purposes of determining capital adequacy, a bank's assets are adjusted based on their risk, with riskier assets requiring a higher weighting. The risk weightings are specified by individual countries' regulators, and these regulators typically take Basel III into consideration. The risk adjustment results in an amount for risk-weighted assets to use when determining the amount of capital required to fund those assets. For example, cash has a risk weighting of zero, so cash is not included in the risk-weighted assets. As a result, no capital is required to fund cash. Corporate loans have a risk weighting of 100%, and certain risky assets, such as loans on high-volatility commercial real estate and loans that are more than 90 days past due, have a weighting greater than 100%. As a simple example, consider a hypothetical bank with three assets: $10 in cash, $1,000 in performing loans, and $10

12 Information on the evolution of risk assessment can be found in "Supervisory Risk Assessment and Early Warning Systems," Ranjana Sahajwala and Paul Van den Bergh, Basel Committee on Banking Supervision Working Paper No. 4 (December 2000). Further information about the CAMELS rating system can be found in the FDIC's description of the Uniform Financial Institutions Rating System at www.fdic.gov.
13 Sahajwala and Van den Bergh, "Supervisory Risk Assessment and Early Warning Systems."

in non-performing loans. The bank's risk-weighted assets (RWAs) would equal ($10 × 0%) + ($1,000 × 100%) + ($10 × 150%) = $1,015. Also, off-balance-sheet exposures are assigned risk weights and included in the risk-weighted assets.

For purposes of determining a bank's capital and its capital adequacy, a bank's capital is classified into hierarchical tiers. The most important of these tiers is Common Equity Tier 1 Capital. According to the FDIC:

> Basel III capital standards emphasize common equity tier 1 capital as the predominant form of bank capital. Common equity tier 1 capital is widely recognized as the most loss-absorbing form of capital, as it is permanent and places shareholders' funds at risk of loss in the event of insolvency. Moreover, Basel III strengthens minimum capital ratio requirements and risk-weighting definitions, increases Prompt Corrective Action (PCA) thresholds, establishes a capital conservation buffer, and provides a mechanism to mandate counter-cyclical capital buffers.[14]

Common Equity Tier 1 Capital includes common stock, issuance surplus related to common stock, retained earnings, accumulated other comprehensive income, and certain adjustments including the deduction of intangible assets and deferred tax assets. Other Tier 1 Capital includes other types of instruments issued by the bank that meet certain criteria. The criteria require, for example, that the instruments be subordinate to such obligations as deposits and other debt obligations, not have a fixed maturity, and not have any type of payment of dividends or interest that is not totally at the discretion of the bank. Tier 2 Capital includes instruments that are subordinate to depositors and to general creditors of the bank, have an original minimum maturity of five years, and meet certain other requirements.

The minimum capital requirements set forth in Basel III are described here because they are global. However, it is the individual countries' regulators who have authority to establish the minimum capital requirements for institutions within their jurisdiction.

- Common Equity Tier 1 Capital must be at least 4.5% of risk-weighted assets.
- Total Tier 1 Capital must be at least 6.0% of risk-weighted assets.
- Total Capital (Tier 1 Capital plus Tier 2 Capital) must be at least 8.0% of risk-weighted assets.[15]

EXAMPLE 1

Capital Position

Exhibit 3 presents an excerpt from an annual report disclosure by HSBC Holdings plc about its capital position. The excerpt shows the group's capital ratios, amount of capital by tier, and risk-weighted assets by type.

---

14 FDIC, "Risk Management Manual of Examination Policies," Section 2.1 (www.fdic.gov). For a comprehensive description of capital tiers under Basel III, refer to "Basel III: A global regulatory framework for more resilient banks and banking systems" (pp. 13–27), available at www.bis.org.

15 www.bis.org.

Exhibit 3 Excerpt from Annual Report Disclosure of HSBC Holdings plc

Capital Ratios

	At 31 Dec.	
	2016 (%)	**2015 (%)**
Common equity tier 1 ratio	13.6	11.9
Tier 1 ratio	16.1	13.9
Total capital ratio	20.1	17.2

Total Regulatory Capital and Risk-Weighted Assets

	At 31 Dec.	
	2016 ($m)	**2015 ($m)**
Regulatory Capital		
Common equity tier 1 capital	116,552	130,863
Additional tier 1 capital	21,470	22,440
Tier 2 capital	34,336	36,530
Total regulatory capital	172,358	189,833
Risk-weighted assets	857,181	1,102,995

Risk-weighted assets (RWAs) by risk types

	RWAs ($bn)	**Capital required* ($bn)**
Credit risk	655.7	52.5
Counterparty credit risk	62.0	5.0
Market risk	41.5	3.3
Operational risk	98.0	7.8
At 31 Dec 2016	857.2	68.6

* "Capital required" represents the Pillar 1 capital charge at 8% of RWAs.
Source: HSBC Holdings plc Annual Report and Accounts 2016 (p. 127).

1. Based on Exhibit 3, did HSBC's capital ratios strengthen or weaken in 2016?
2. Based on Exhibit 3, what was the primary reason for the change in HSBC's capital ratios in 2016?

Solution to 1:

HSBC's capital ratios strengthened in 2016. Its Common Equity Tier 1 ratio increased from 11.9% of RWAs to 13.6% of RWAs. Its Tier 1 ratio also increased from 13.9% to 16.1%, and its Total Capital Ratio increased from 17.2% to 20.1%.

Solution to 2:

The primary reason for the change in HSBC's capital ratios in 2016 was a reduction in the amount of risk-weighted assets. Total risk-weighted assets declined from $1,102,995 million to $857,181 million.

2.1.2 *Asset Quality*

Asset quality pertains to the amount of existing and potential credit risk associated with a bank's assets, focusing primarily on financial assets. The concept of asset quality extends beyond the composition of a bank's assets and encompasses the strength of the overall risk management processes by which the assets are generated and managed.

Loans typically constitute the largest portion of a bank's assets. Asset quality for loans reported on the balance sheet depends on the creditworthiness of the borrowers and the corresponding adequacy of adjustments for expected loan losses. Loans are measured at amortized cost and are shown on the balance sheet net of allowances for loan losses.

Investments in securities issued by other entities, often another significant portion of a bank's assets, are measured differently, depending on how the security is categorized. Specifically, under International Financial Reporting Standards (IFRS),[16] financial assets are classified in one of three categories, depending on the company's business model for managing the asset and on the contractual cash flows of the asset. The financial asset's category specifies how it is subsequently measured (either amortized cost or fair value) and, for those measured based on fair value, how any changes in value are reported–either through other comprehensive income (OCI) or through profit and loss (PL). The three categories for financial assets are (1) measured at amortized cost, (2) measured at fair value through other comprehensive income (FVOCI), and (3) measured at fair value through profit and loss (FVTPL).

In contrast to IFRS, US GAAP require all equity investments "(except those accounted for under the equity method of accounting or those that result in consolidation of the investee) to be measured at fair value with changes in fair value recognized in net income."[17] Another exception to fair value measurement is that an equity investment without a readily determinable fair value can be measured at cost minus impairment. Thus, under US GAAP, the three categories used to classify and measure investments apply *only to debt securities*: held to maturity (measured at amortized cost), trading (measured at fair value through net income), and available for sale (measured at fair value through other comprehensive income).

The following example addresses asset quality from the perspective of overall asset composition. The example includes the asset portion of a bank's balance sheet. In practice, terminology used by different entities can vary, and an analyst should refer to the footnotes for further detail on a line item. Here, two comments can be helpful in interpreting the line items in the example. First, when determining the total amount of bank loans, two line items are clearly relevant: "Loans and advances to banks" and "Loans and advances to customers." In addition, note that "Reverse repurchase agreements" are a form of collateralized loan made by a bank to a client. In a repurchase agreement, a borrower (i.e., a bank client) sells a financial asset to a lender (i.e., a bank) and commits to repurchase the financial asset for a fixed price at a future date. The difference between the selling price and the higher repurchase price effectively constitutes interest on the borrowing. The borrower describes the transaction as a "repurchase agreement," and the lender describes the transaction as a "reverse repurchase agreement."[18] Second, the term "assets held for sale" is related to discontinued operations and specifically refers to long-term assets whose value is

16 IFRS 9 *Financial Instruments*, issued July 2014 and effective beginning January 2018.

17 Accounting Standards Update 2016-01 *Financial Instruments—Overall* (Subtopic 825-10) *Recognition and Measurement of Financial Assets and Financial Liabilities*. This Accounting Standards Update was issued in January 2016 and is effective for public business entities for fiscal years beginning after 15 December 2017.

18 The Office of Financial Research (part of the US Department of the Treasury) estimates that the size of the repurchase ("repo") market is $3.5 trillion.

driven mainly by their intended disposition rather than their continued use.[19] This term should not be confused with the securities-related term "available for sale" (described above).

EXAMPLE 2

Asset Quality: Composition of Assets

Exhibit 4 presents the asset portion of the balance sheet of HSBC Holdings, which is prepared according to IFRS.

Exhibit 4 Excerpt from Consolidated Balance Sheet

HSBC Holdings plc
Consolidated Balance Sheet [Excerpt]
at 31 December

Assets	2016 $m	2015 $m
Cash and balances at central banks	**128,009**	98,934
Items in the course of collection from other banks	**5,003**	5,768
Hong Kong Government certificates of indebtedness	**31,228**	28,410
Trading assets	**235,125**	224,837
Financial assets designated at fair value	**24,756**	23,852
Derivatives	**290,872**	288,476
Loans and advances to banks	**88,126**	90,401
Loans and advances to customers	**861,504**	924,454
Reverse repurchase agreements, non-trading	**160,974**	146,255
Financial investments	**436,797**	428,955
Assets held for sale	**4,389**	43,900
Prepayments, accrued income and other assets	**59,520**	54,398
Current tax assets	**1,145**	1,221
Interests in associates and joint ventures	**20,029**	19,139
Goodwill and intangible assets	**21,346**	24,605
Deferred tax assets	**6,163**	6,051
Total assets at 31 Dec	**2,374,986**	2,409,656

Source: HSBC Holdings plc Annual Report and Accounts 2016.

1 The following items are the most liquid: Cash and balances at central banks, Items in the course of collection from other banks, and Hong Kong Government certificates of indebtedness. What proportion of HSBC's total assets was invested in these liquid assets in 2015? In 2016? Did HSBC's balance sheet liquidity decrease or increase in 2016?

19 IFRS 5 *Non-Current Assets Held for Sale and Discontinued Operations.*

2 How did the percentage of investments to total assets change from 2015 to 2016? (Include trading assets, financial assets designated at fair value, and financial investments as investments.)

3 What proportion of HSBC's assets are loans? (As noted, the banks' loans include "Loans and advances to banks" and "Loans and advances to customers." In addition, "Reverse repurchase agreements" are a form of collateralized loan.)

Solution to 1:

HSBC's balance sheet liquidity increased in 2016.

In 2015, the proportion of HSBC's balance sheet invested in highly liquid assets was 5.5%

[($98,934 + $5,768 + $28,410)/$2,409,656 = 5.5%].

In 2016, the proportion of HSBC's balance sheet invested in highly liquid assets was 6.9%

[($128,009 + $5,003 + $31,228)/$2,374,986 = 6.9%].

Solution to 2:

The percentage of investments on HSBC's balance sheet increased in 2016.

In 2015, the percentage of investments to total assets was 28.1%

[($224,837 + $23,852 + $428,955)/$2,409,656 = 28.1%].

In 2016, the percentage of investments to total assets was 29.3%

[($235,125 + $24,756 + $436,797)/$2,374,986 = 29.3%].

Solution to 3:

In 2015, loans represented 48.2% [($90,401 + $924,454 + $146,255)/$2,409,656 = 48.2%] of HSBC's total assets, and in 2016, loans represented 46.8% [($88,126 + $861,504 + $160,974)/$2,374,986 = 46.8%] of HSBC's total assets.

The next example addresses asset quality from the perspective of credit quality. Assessment of credit risk is of course fundamental to banks' decisions about loans—the largest category of a banks' assets. As noted, investments in securities often constitute a significant portion of a bank's assets, and those activities also involve credit risk. Further, a bank's trading activities—including off-balance-sheet trading activities—create exposure to counterparty credit risk. Off-balance-sheet obligations such as guarantees, unused committed credit lines, and letters of credit represent potential assets (as well as potential liabilities) to the bank and thus involve credit risk. In addition to credit risk, other factors, such as liquidity, can also affect the value and marketability of a bank's assets. Diversification of credit risk exposure (and avoiding credit concentration) across the entire asset base—loans and investments—and among counterparties is an important aspect of asset quality.

EXAMPLE 3

Credit Quality of Assets

Exhibit 5 presents an excerpt from an annual report disclosure by HSBC Holdings plc about the credit quality of its financial instruments. The exhibit shows the distribution of financial instruments by credit quality.

Financial instruments included in the exhibit correspond to total amounts for some line items of assets listed on the balance sheet and to partial amounts for line items on the balance sheet where only a portion of the asset involves exposure to credit risk. Total amounts are included for the following balance sheet items: Cash and balances at central banks; Items in the course of collection from other banks; Hong Kong Government certificates of indebtedness; Derivatives; Loans and advances to banks; Loans and advances to customers; and Reverse repurchase agreements, non-trading. Partial amounts are included for the following balance sheet items: Trading assets; Financial assets designated at fair value; Financial investments; Assets held for sale; and Prepayments, accrued income and other assets.

Exhibit 5 Excerpt from Annual Report Disclosure of HSBC Holdings plc

		At 31 Dec. 2016 ($m)	At 31 Dec. 2015 ($m)
Neither past due nor impaired	Strong credit quality	$1,579,517	$1,553,830
	Good credit quality	$313,707	$331,141
	Satisfactory credit quality	$263,995	$293,178
	Sub-standard credit quality	$26,094	$26,199
	Past due but not impaired	$9,028	$13,030
	Impaired	$20,510	$28,058
	Total gross amount	$2,212,851	$2,245,436
	Impairment allowances	$(8,100)	$(11,027)
	Total	$2,204,751	$2,234,409

Source: HSBC Holdings plc Annual Report and Accounts 2016 (pp. 88–89).

1 Based on Exhibit 5, did the credit quality of HSBC's financial instruments improve or deteriorate in 2016? Specifically, how did the proportion of assets invested in strong credit quality instruments change from year to year?

2 Based on Exhibit 5, does the change in HSBC's impairment allowances in 2016 reflect the change in the credit quality of financial instruments (specifically the amount of impaired assets)?

Solutions Exhibit

		At 31 Dec. 2016 Percentage of total gross amount	At 31 Dec. 2015 Percentage of total gross amount	Percentage change in dollar amount
Neither past due nor impaired	Strong credit quality	71.4%	69.2%	1.7%
	Good credit quality	14.2%	14.7%	**–5.3%**
	Satisfactory credit quality	11.9%	13.1%	**–10.0%**
	Sub-standard credit quality	1.2%	1.2%	**–0.4%**
	Past due but not impaired	0.4%	0.6%	**–30.7%**
	Impaired	0.9%	1.2%	**–26.9%**
	Total gross amount	100.0%	100.0%	**–1.5%**
	Impairment allowances	**–0.4%**	**–0.5%**	**–26.5%**

Solution to 1:

Based on Exhibit 5, the credit quality of HSBC's financial instruments improved in 2016. As shown in the Solutions Exhibit, the percentage of total investment assets invested in strong credit quality instruments rose from 69.2% in 2015 to 71.4% in 2016 [$1,553,830/$2,245,436 = 69.2%; $1,579,517/$2,212,851 = 71.4%].

Solution to 2:

Yes. Based on Exhibit 5, the change in HSBC's impairment allowances in 2016 reflects the change in the credit quality of financial instruments. In general, it is expected that the amount of impairment allowances will be related to the amount of impaired assets. The 26.5% decrease in the amount of HSBC's impairment allowances in 2016 corresponds to the 26.9% decrease in impaired assets. As a corollary, the amount of impairment allowances as a percentage of impaired assets remained roughly constant in both years ($11,027/$28,058 = 39.3% for 2015 and $8,100/$20,510 = 39.5% for 2016).

2.1.3 *Management Capabilities*

Many of the attributes of effective management of financial institutions are the same as those for other types of entities. Effective management involves successfully identifying and exploiting appropriate profit opportunities while simultaneously managing risk. For all types of entities, compliance with laws and regulations is essential. A strong governance structure—with an independent board that avoids excessive compensation or self-dealing—is also critically important. Sound internal controls, transparent management communication, and financial reporting quality are indicators of management effectiveness. Across all entities, overall performance is ultimately the most reliable indicator of management effectiveness.

For financial institutions, a particularly important aspect of management capability is the ability to identify and control risk, including credit risk, market risk, operating risk, legal risk, and other risks. Directors of banks set overall guidance on risk exposure levels and appropriate implementation policies and provide oversight of bank management. Banks' senior managers must develop and implement effective procedures for measuring and monitoring risks consistent with that guidance.

2.1.4 *Earnings*

As with any entity, financial institutions should ideally generate an amount of earnings to provide an adequate return on capital to their capital providers and specifically to reward their stockholders through capital appreciation and/or distribution of the earnings. Further, all companies' earnings should ideally be high quality and trending upward. In general, high-quality earnings mean that accounting estimates are unbiased and the earnings are derived from sustainable rather than non-recurring items.

For banks, one important area involving significant estimates is loan impairment allowances. In estimating losses on the loan portfolio collectively, statistical analysis of historical loan losses can provide a basis for an estimation, but statistical analysis based on past data must be supplemented with management judgement about the potential for deviation in future. In estimating losses on individual loans, assessments are required concerning the likelihood of the borrower's default or bankruptcy and the value of any collateral. HSBC describes the complexity of estimating loan impairment allowances as follows: "The exercise of judgement requires the use of assumptions which are highly subjective and very sensitive to the risk factors, in particular to changes in economic and credit conditions across a large number of geographical areas. Many of the factors have a high degree of interdependency and there is no single factor to which our loan impairment allowances as a whole are sensitive."[20]

Banks also must use estimates in valuing some financial assets and liabilities that must be measured at fair value. When fair value of an investment is based on observable market prices, valuation requires little judgment. However, when fair values cannot be based on observable market prices, judgment is required.

Under both IFRS and US GAAP, fair value measurements of financial assets and liabilities are categorized on the basis of the type of inputs used to establish the fair value. Both sets of standards use the concept of a *fair value hierarchy*.[21] The three "levels" of the fair value hierarchy pertain to the observability of the inputs used to establish the fair value.

- Level 1 inputs are quoted prices for identical financial assets or liabilities in active markets.
- Level 2 inputs are observable but are not the quoted prices for identical financial instruments in active markets. Level 2 inputs include quoted prices for similar financial instruments in active markets, quoted prices for identical financial instruments in markets that are not active, and observable data such as interest rates, yield curves, credit spreads, and implied volatility. The inputs are used in a model to determine the fair value of the financial instrument.
- Level 3 inputs are unobservable. The fair value of a financial instrument is based on a model (or models) and unobservable inputs. Financial modeling, by its very nature, contains subjective estimates that are unobservable and will differ from one modeler to another. For example, a financial instrument's value might be based on an option-pricing model employing an unobservable and subjective estimate of the instrument's market volatility. Another example is that a financial instrument's value might be based on estimated future cash flows,

20 HSBC Holdings plc Annual Report and Accounts 2016, page 199: www.hsbc.com/investor-relations/group-results-and-reporting/annual-report

21 Refer to IFRS 13 *Fair Value Measurement* and Financial Accounting Standards Board ASC 820 *Fair Value Measurement*.

discounted to a present value. Neither the estimated future cash flows nor the discount rate can be observed objectively, because they depend on the determinations made by the modeler.

In practice, the "Level 1, 2, 3" fair value terminology can also refer to the valuation approach used. A Level 3 valuation technique is one that relies on one or more significant inputs that are unobservable. For example, as noted, a company might value a private equity investment using a model of estimated future cash flows.

Also, in practice, the "Level 1, 2, 3" terminology can refer to the assets or liabilities being valued using a given level of input. For example, investments can be referred to as "Level 1," "Level 2," or "Level 3" investments depending on whether their fair value is determined based on observable market prices for the exact instrument, observable market inputs for similar investments, or unobservable inputs, respectively.

Other areas involving significant estimates are common to non-financial and financial companies. Judging whether goodwill impairment exists requires estimating future cash flows of a business unit. Deciding to recognize a deferred tax asset relies on making assumptions about the probability of future taxes. Determining whether and how much of a liability to recognize in connection with contingencies (e.g., litigation) typically depends on professional expert advice but nonetheless requires some management judgment.

Regarding sustainability of a bank's earnings, it is important to examine the composition of earnings. Banks' earnings typically comprise (a) net interest income (the difference between interest earned on loans minus interest paid on the deposits supporting those loans), (b) service income, and (c) trading income. Of these three general sources, trading income is typically the most volatile. Thus, a greater proportion of net interest income and service income is typically more sustainable than trading income. In addition, lower volatility within net interest income is desirable: Highly volatile net interest income could indicate excessive interest rate risk exposure.

EXAMPLE 4

Composition of Earnings

An analyst has gathered the information in Exhibit 6 to evaluate how important each source of income is to HSBC.

Exhibit 6 Five-Year Summary of HSBC's Total Operating Income

	2016 ($m)	2015 ($m)	2014 ($m)	2013 ($m)	2012 ($m)
Net interest income	**$29,813**	$32,531	$34,705	$35,539	$37,672
Net fee income	**$12,777**	$14,705	$15,957	$16,434	$16,430
Net trading income	**$9,452**	$8,723	$6,760	$8,690	$7,091
Net income/(expense) from financial instruments designated at fair value	**($2,666)**	$1,532	$2,473	$768	($2,226)
Gains less losses from financial investments	**$1,385**	$2,068	$1,335	$2,012	$1,189
Dividend income	**$95**	$123	$311	$322	$221
Net insurance premium income	**$9,951**	$10,355	$11,921	$11,940	$13,044

Exhibit 6 (Continued)

	2016 ($m)	2015 ($m)	2014 ($m)	2013 ($m)	2012 ($m)
Gains on disposal of US branch network, US cards business, and Ping An Insurance (Group) Company of China, Ltd.	—	—	—	—	$7,024
Other operating income/(expense)	**($971)**	$1,055	$1,131	$2,632	$2,100
Total operating income	**$59,836**	$71,092	$74,593	$78,337	$82,545

Source: HSBC Holdings plc Annual Report and Accounts 2016 (p. 31).

1 Based on Exhibit 6, what is HSBC's primary source of operating income, and what proportion of total operating income was earned from this source in 2016?

2 Based on Exhibit 6, what proportion of total operating income did HSBC earn from trading income in 2016?

3 Based on Exhibit 6, describe the trend in HSBC's operating income.

Solution to 1:

HSBC's primary source of operating income is net interest income. In 2016, 49.8% ($29,813/$59,836 = 49.8%) of total operating income was earned from net interest income in 2016.

Solution to 2:

In 2016, HSBC earned 15.8% ($9,452/$59,836 = 15.8%) of total operating income from trading activities.

Solution to 3:

From 2012 to 2016, HSBC's operating income declined each year. The composition of operating income was fairly constant from 2012 to 2015, with around 46% from net interest income and 21% from fee income.

Exhibit 7 Five-Year Summary of HSBC's Total Operating Income: Common-Size Statement

	2016	2015	2014	2013	2012
	As a Percentage of Total Operating Income				
Net interest income	49.8%	45.8%	46.5%	45.4%	45.6%
Net fee income	21.4%	20.7%	21.4%	21.0%	19.9%
Net trading income	15.8%	12.3%	9.1%	11.1%	8.6%
Net income/(expense) from financial instruments designated at fair value	−4.5%	2.2%	3.3%	1.0%	−2.7%
Gains less losses from financial investments	2.3%	2.9%	1.8%	2.6%	1.4%
Dividend income	0.2%	0.2%	0.4%	0.4%	0.3%
Net insurance premium income	16.6%	14.6%	16.0%	15.2%	15.8%
Gains on disposal of US branch network, US cards business, and Ping An Insurance (Group) Company of China, Ltd.	—	—	—	—	8.5%

(continued)

Exhibit 7 (Continued)

	2016	2015	2014	2013	2012
Other operating income/(expense)	−1.6%	1.5%	1.5%	3.4%	2.5%
Total operating income	100.0%	100.0%	100.0%	100.0%	100.0%

2.1.5 *Liquidity Position*

Adequate liquidity is essential for any type of entity. Banks' systemic importance increases the importance of adequate liquidity. If a non-bank entity's insufficient liquidity prevents it from paying a current liability, the impact would primarily affect the entity's own supply chain. In contrast, because deposits constitute the primary component of a bank's current liabilities, the impact of a bank's failure to honor a current liability could affect an entire economy. Deposits in most banks are insured up to some specified amount by government insurers; thus, liquidity is a key focus of regulators.

The Basel III Regulatory Framework[22] cites the sudden illiquidity accompanying the financial crisis of 2008 as a main motivation for the introduction of a global liquidity standard. Because of the sudden pressures on liquidity at the inception of the financial crisis, some banks experienced difficulties, despite having an adequate capital base. Basel III thus introduced two minimum liquidity standards, both to be phased in over subsequent years.

- The Liquidity Coverage Ratio (LCR) is expressed as the minimum percentage of a bank's expected cash outflows that must be held in highly liquid assets. For this ratio, the expected cash outflows (the denominator) are the bank's anticipated one-month liquidity needs in a stress scenario, and the highly liquid assets (the numerator) include only those that are easily convertible into cash. The standards set a target minimum of 100%.
- The Net Stable Funding Ratio (NSFR) is expressed as the minimum percentage of a bank's *required* stable funding that must be sourced from *available* stable funding. For this ratio, required stable funding (the denominator) is a function of the composition and maturity of a bank's asset base, whereas available stable funding (the numerator) is a function of the composition and maturity of a bank's funding sources (i.e., capital and deposits and other liabilities). Under Basel III, the available stable funding is determined by assigning a bank's capital and liabilities to one of five categories presented in Exhibit 8, shown below. The amount assigned to each category is then multiplied by an available stable funding (ASF) factor, and the total available stable funding is the sum of the weighted amounts.[23]

22 Basel Committee on Banking Supervision, "Basel III: A Global Regulatory Framework For More Resilient Banks and Banking System": www.bis.org/publ/bcbs189.pdf.

23 Basel Committee on Banking Supervision, "Basel III: The Net Stable Funding Ratio" (October 2014, p. 3): www.bis.org/bcbs/publ/d295.pdf. Exhibit 8 is adapted from page 6 of this document.

Exhibit 8 Categories of Available Stable Funding

ASF Factor	Components of ASF Category
100%	■ Total regulatory capital (excluding Tier 2 instruments with residual maturity of less than one year) ■ Other capital instruments and liabilities with effective residual maturity of one year or more
95%	■ Stable non-maturity (demand) deposits and term deposits with residual maturity of less than one year provided by retail and small business customers
90%	■ Less stable non-maturity deposits and term deposits with residual maturity of less than one year provided by retail and small business customers
50%	■ Funding with residual maturity of less than one year provided by non-financial corporate customers ■ Operational deposits ■ Funding with residual maturity of less than one year from sovereigns, public sector entities, and multilateral and national development banks ■ Other funding with residual maturity between six months and less than one year not included in the above categories, including funding provided by central banks and financial institutions
0%	■ All other liabilities and equity not included in the above categories, including liabilities without a stated maturity (with a specific treatment for deferred tax liabilities and minority interests) ■ Net Stable Funding Ratio derivative liabilities net of Net Stable Funding Ratio derivative assets if Net Stable Funding Ratio derivative liabilities are greater than Net Stable Funding Ratio derivative assets ■ "Trade date" payables arising from purchases of financial instruments, foreign currencies, and commodities

The rationale for the Net Stable Funding Ratio is that it relates the liquidity needs of the financial institution's assets to the liquidity provided by the funding sources. With assets, for example, loans with long-dated maturities require stable funding whereas highly liquid assets do not. With funding sources, long-dated deposits and other liabilities are considered more stable than short-dated liabilities, and deposits from retail customers are considered more stable than deposits with the same maturity from other counterparties. The standards set a target minimum of greater than 100%.

Among the several liquidity-monitoring metrics described in Basel III,[24] two are discussed here: concentration of funding and contractual maturity mismatch. Concentration of funding refers to the proportion of funding that is obtained from a single source. Excessive concentration of funding exposes a bank to the risk that a single funding source could be withdrawn.

Contractual maturity mismatch refers to the maturity dates of a bank's assets compared to the maturity dates of a bank's funding sources. In a normal yield curve environment, where long-term interest rates are higher than short-term rates, a bank can maximize its net interest income—all else equal—by borrowing short term and lending long term. In doing so, the bank would minimize the interest paid to its

24 Basel Committee on Banking Supervision, "Basel III: A Global Regulatory Framework For More Resilient Banks and Banking System": www.bis.org/publ/bcbs189.pdf.

depositors and maximize interest earned on its loan assets. In excess, however, such maturity mismatches expose the bank to liquidity risk if the bank needs to return cash on its maturing deposits prior to the time that it receives cash repayment of loans from its borrowers. Monitoring maturity mismatch is thus an important tool in liquidity risk management.

EXAMPLE 5

The following excerpts from HSBC's annual report explain the bank's approach to management of its liquidity and funding risk. The disclosures state that the group's principal operating entities were within the risk tolerance levels established by the board for the Liquidity Coverage Ratio, the Net Stable Funding Ratio, depositor concentration, and term funding maturity concentration.

Exhibit 9 Liquidity Disclosure—Excerpts from HSBC's Annual Report

The management of liquidity and funding is primarily undertaken locally (by country) in our operating entities in compliance with the Group's LFRF [liquidity and funding risk management framework], and with practices and limits set by the GMB [Group Management Board] through the RMM [Risk Management Meeting of the Group Management Board] and approved by the Board. Our general policy is that each defined operating entity should be self-sufficient in funding its own activities. Where transactions exist between operating entities, they are reflected symmetrically in both entities.

As part of our asset, liability and capital management structure, we have established asset and liability committees ("ALCO") at Group level, in the regions and in operating entities. . . . The primary responsibility for managing liquidity and funding within the Group's framework and risk appetite resides with the local operating entities' ALCOs, Holdings ALCO and the RMM. . . .

The Liquidity Coverage Ratio ("LCR") aims to ensure that a bank has sufficient unencumbered high-quality liquid assets ("HQLA") to meet its liquidity needs in a 30-calendar-day liquidity stress scenario. HQLA consist of cash or assets that can be converted into cash at little or no loss of value in markets. We reported a Group European Commission ("EC") LCR at 31 December 2016 of 136% (31 December 2015: 116%) to the PRA [UK Prudential Regulation Authority]. . . . At 31 December 2016, all the Group's principal operating entities were within the LCR risk tolerance level established by the Board. . . . The liquidity position of the Group can also be represented by the stand-alone ratios of each of our principal operating entities. . . .

The Net Stable Funding Ratio ("NSFR") requires institutions to maintain sufficient stable funding relative to required stable funding, and reflects a bank's long-term funding profile (funding with a term of more than a year). It is designed to complement the LCR. At 31 December 2016, the Group's principal operating entities were within the NSFR risk tolerance level established by the Board and applicable under the LFRF.

The LCR and NSFR metrics assume a stressed outflow based on a portfolio of depositors within each deposit segment. The validity of these assumptions is challenged if the portfolio of depositors is not large enough to avoid depositor concentration. Operating entities are exposed to term re-financing concentration risk if the current maturity profile results in future maturities being overly concentrated in any defined period. At 31

Exhibit 9 (Continued)

December 2016, all principal operating entities were within the risk tolerance levels set for depositor concentration and term funding maturity concentration. These risk tolerances were established by the Board. . . .

[The table below displays the following liquidity metrics for HSBC's principal operating entities: individual LCR on an EC LCR basis and NSFR.]

Operating Entities' Liquidity Measures

	LCR		NSFR
	Dec-16 (%)	Dec-15 (%)	Dec-16 (%)
HSBC UK liquidity group	123	107	116
The Hongkong and Shanghai Banking Corporation, Hong Kong Branch	185	150	157
The Hongkong and Shanghai Banking Corporation, Singapore Branch	154	189	112
HSBC Bank USA	130	116	120
HSBC France	122	127	120
Hang Seng Bank	218	199	162
HSBC Canada	142	142	139
HSBC Bank China	253	183	49
HSBC Middle East, UAE Branch	241		141
HSBC Mexico	177		128
HSBC Private Bank	178		155

Source: HSBC Holdings plc Annual Report and Accounts 2016 (pp. 108, 143, and 144).

1 Based on the exhibit, in 2016, which of HSBC's operating entities had the highest level of liquid assets relative to its liquidity needs in a stress scenario?

2 Based on the exhibit, which of HSBC's operating entities had the most stable funding relative to its required need for stable funding?

3 Based on the exhibit, which of HSBC's operating entities is the furthest away from achieving the Basel III target for NSFR?

Solution to 1:

Based on the exhibit, HSBC Bank China had the highest level of liquid assets relative to its liquidity needs in a stress scenario. Its 2016 LCR of 253% is higher than that of any of the other HSBC entities.

Solution to 2:

Based on the exhibit, Hang Seng Bank had the most stable funding relative to its required need for stable funding. Its 2016 NSFR of 162% is higher than that of any of the other HSBC entities.

Solution to 3:

Based on the exhibit, HSBC Bank China is the furthest away from achieving the Basel III standard of NSFR greater than 100%. Its NSFR of 49% is lower than that of any of the other HSBC entities. (It is possible that these metrics result from RMB capital controls in China or jurisdictional issues; however, the example does not provide sufficient information to confirm the reason.)

2.1.6 *Sensitivity to Market Risk*

Almost every entity has some exposure to changes in interest rates, exchange rates, equity prices, or commodity prices. Every company in the United States, for example, is required to provide quantitative and qualitative disclosures in annual filings about exposure to market risk. The nature of banks' operations generally makes sensitivity of earnings to market risks a particularly important consideration for analysts. Mismatches in the maturity, repricing frequency, reference rates, or currency of banks' loans and deposits create exposure to market movements. Further, exposure to risk arises not only from loans and deposits on a bank's balance sheet but also from off-balance-sheet exposures, including, for example, guarantees or derivatives positions linked to interest rates, exchange rates, equities, or commodities. It is important to understand how an adverse change in any of these markets would affect a bank's earnings. It is also important to evaluate the strength of a bank's ability to manage market risks.

Banks disclose information about the sensitivity of earnings to different market conditions—namely, the earnings impact of a shift up or down in some market. Consider a bank's sensitivity to interest rate risk. Even in a purely hypothetical situation of a bank with assets and liabilities that are identical in terms of interest rates, maturity, and frequency of repricing, an increase in interest rates would cause the bank's net interest income to increase. This would occur simply because banks have more assets than liabilities. In reality, of course, the terms of a bank's assets and liabilities differ. Generally, the yield on a bank's loan assets is presumed to be higher than the rate it must pay its depositors, particularly consumer deposits. With respect to term structure, in a typical yield curve environment, longer-dated assets would have a higher yield *ceteris paribus* than shorter-dated funding sources, but another aspect of interest rate sensitivity is repricing frequency. For example, having assets with greater repricing frequency than liabilities would benefit earnings in a rising interest rate scenario. In sum, many structural factors affect interest rate sensitivity.

The following example includes an interest rate sensitivity disclosure showing the earnings impact of an upward and downward shift in interest rates. Disclosures such as these reflect the existing structure of a bank's assets and liabilities.

EXAMPLE 6

Market Risk

The following excerpts from HSBC's annual report explain the bank's approach to monitoring its market risk and illustrates one of the tools used by the bank: sensitivity analysis.

Exhibit 10 Excerpt from HSBC's Annual Report

Our objective is to manage and control market risk exposures while maintaining a market profile consistent with our risk appetite. We use a range of tools to monitor and limit market risk exposures including sensitivity analysis, value at risk and stress testing.

Exhibit 10 (Continued)

The following table sets out the assessed impact on our base case projected net interest income ("NII") for 2016 (excluding insurance) of a series of four quarterly parallel shocks of 25 basis points to the current market-implied path of interest rates worldwide at the beginning of each quarter from 1 January 2017. . . .

The sensitivities shown represent our assessment as to the change in expected base case net interest income under the two rate scenarios, assuming that all other non-interest rate risk variables remain constant, and there are no management actions. . . .

We expect NII to rise in the rising rate scenario and fall in the falling rate scenario. This is due to a structural mismatch between our assets and liabilities (on balance we would expect our assets to reprice more quickly, and to a greater extent, than our liabilities).

Net Interest Income Sensitivity (Audited)

	US dollar bloc ($m)	Rest of Americas bloc ($m)	Hong Kong dollar bloc ($m)	Rest of Asia bloc ($m)	Sterling bloc ($m)	Euro bloc ($m)	Total ($m)
Change in 2016 net interest income arising from a shift in yield curves of:							
+25 basis points at the beginning of each quarter	**605**	**47**	**504**	**280**	**61**	**212**	**1,709**
−25 basis points at the beginning of each quarter	**−1,024**	**−41**	**−797**	**−292**	**−261**	**9**	**−2,406**

Source: HSBC Holdings plc Annual Report and Accounts 2016 (pp. 78 and 117).

1 Based on the exhibit, by how much would HSBC's planned net interest income decrease if the yield curves shifted downward by 25 basis points at the beginning of each quarter for four quarters?

2 If a decrease in interest rates would hurt the earnings of banks such as HSBC, why would central banks lower interest rates so significantly following the financial crisis in order to prop up the financial sector?

Solution to 1.

HSBC's planned net interest income would decrease by $2,406 million if the yield curves shifted downward by 25 basis points at the beginning of each quarter.

Solution to 2.

An interest rate sensitivity table such as the one presented by HSBC is a static presentation and thus assumes that the relation between the structure of assets and liabilities in place at the time would remain stationary. Following the financial crisis, the central banks' actions reduced interest rates at which banks could borrow (effectively, to near zero), while the rates that banks were able to charge their loan customers were—while still low—far higher than their borrowing costs. Further, the central banks' actions were not intended solely to prop up banks' earnings but also to provide liquidity and stimulus to the overall economy.

As described in the example, another tool that HSBC uses to measure and monitor market risk is value at risk (VaR). Recall that VaR is a way to estimate the amount of potential loss based on simulations that incorporate historical pricing information. HSBC estimates its VaR using a 99% confidence level, a one-day holding period, and two prior years of pricing data on foreign exchange rates, interest rates, equity prices, commodity prices, and associated volatilities.

3 ANALYZING A BANK: NON-CAMELS FACTORS

d describe other factors to consider in analyzing a bank;

e analyze a bank based on financial statements and other factors;

While the CAMELS approach to assessing bank soundness is fairly comprehensive, there are important bank-specific attributes that it does not completely address. There are also important attributes not addressed by the CAMELS approach that apply to both banks and other types of companies.

3.1 Banking-Specific Analytical Considerations Not Addressed by CAMELS

The CAMELS acronym is useful as a composite of major factors, but it is neither comprehensive nor comprehensively integrated. Also, the ordering of the factors does not signify importance. For example, strong capital (the "C") and strong liquidity (the "L") are equally important in the Basel III standards.[25]

The following bank attributes are either unaddressed or not fully addressed by a CAMELS analysis:

- **Government support.** Governments do not normally strive to save a company or even an entire industry that may be facing failure. In capitalist societies, failure is the unfortunate occasional by-product of risk taking with capital, and bankruptcy laws and courts serve to administer the results of failed capital allocation. The banking industry is different from other industries, however, regarding government support. It is in a government's interest to have a healthy banking system because a nation's economy is affected by banks' lending activity, and a nation's central bank needs a healthy banking system for the effective transmission of monetary policy. A healthy banking system also facilitates commerce by providing adequate payment processing and instilling depositor confidence in the safekeeping of their deposits.

  Government agencies monitor the health of banks in the entire system and will close banks that might fail or will arrange mergers with healthy ones able to absorb them. This pruning activity addresses issues with banks that might otherwise weaken the banking system if left unattended. Alternatively, governments may directly assist banks to keep them afloat rather than closing them or

25 Basel Committee on Banking Supervision, "Basel III: A Global Regulatory Framework For More Resilient Banks and Banking System" (December 2010, p. 8, item B.34): www.bis.org/publ/bcbs189.pdf.

arranging for mergers with healthier banks. Visible examples of both assisting and pruning activities occurred during the financial crisis of 2008. For example, the US Treasury created the Troubled Asset Relief Program (TARP) to purchase loans held by banks and to provide equity injections to the banks. During the same period, the Treasury also arranged numerous mergers among banking giants, leading to even bigger banking giants.

CAMELS analysis will not provide an assessment of government support, but an investor can qualitatively assess whether a bank will enjoy the support of the government in times of economic distress. The following are factors to consider:

- *Size of the bank.* Is the bank large enough to bring damage to a significant part of the economy in the event of its failure? Is it "too big to fail"?
- *Status of the country's banking system.* Is the nation's banking system healthy enough to handle a particular bank's failure? Rather than force the banking system to cope with the failure of a particular bank, would it be a better solution for the government to intervene with taxpayer funds to support it? The global financial crisis of 2008–2009 led the US Federal Reserve to develop the concept of SIFIs: systemically important financial institutions, ones that would pose a significant risk to the economy in the event of a failure. Such institutions have been the target of an increased degree of regulation in the post-crisis era.

- **Government ownership.** Public ownership of banks may include a strong ownership representation by the government of their home country. Government ownership may exist for several reasons. A "development" view of government ownership incorporates a belief that government ownership aids financial development of the banks, leading to broad economic growth. A more pessimistic view is that a nation's banking system is not strong enough to stand on its own and attract large amounts of capital, because of low ethical standards within the industry or a lack of confidence in the banking system among the nation's public at large—an important source of funds for any bank.[26]

  Whatever the reason may be for a government's ownership stake in a bank, its presence adds another dimension of security for a bank investor. A government that owns a stake in a bank is likely to intervene on the bank's behalf in the event of economic distress. Conversely, a government that plans to reduce its ownership stake in a bank may directly reduce that dimension of security; however, that may not always be the case. During the global financial crisis of 2008–2009, some governments became reluctant owners of banks, which were ultimately supported by taxpayer funding. When government ownership of such banks was reduced after the crisis ended, markets viewed the reduction as a signal of renewed strength.

- **Mission of banking entity.** Not all banks share the same mission. For example, community banks primarily serve the needs of the immediate community in which they operate. That community's welfare could be driven by an economy based on farming, mining, or oil or could depend on a single large manufacturing entity. The fortunes of the banks and their borrowers and depositors would depend on economic factors that affect the primary industry or employer. Contrast that situation with a global banking entity absorbing deposits from all around the world while investing globally as well. The global bank is more diversified against a single risk than any community bank.

26 Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, "Government Ownership of Banks," NBER Working Paper No. 7620 (March 2000).

The mission of the bank and the economics of its constituents will affect the way the bank manages its assets and liabilities. That is a qualitative assessment that the bank investor needs to make, and it is not addressed by a CAMELS analysis.

- **Corporate culture.** A bank's culture may be very risk averse and cautious and make only loans perceived to be low risk, or alternatively, it may be risk seeking and willing to take risk in pursuit of high returns on investment. Or a bank's culture may be somewhere in the middle of those two extremes. An overly cautious culture may be too risk averse to provide adequate returns to shareholders for taking on the risk of ownership. A highly risk-hungry culture may lead to boom and bust results and volatility. Differences in the cultural environment are particularly important for banks operating in multiple countries, where there may be a disconnect between corporate culture and national culture.

  A bank investor can qualitatively assess a bank's cultural environment by considering factors such as these:

  - Has the bank generated recent losses resulting from a narrowly focused investment strategy, such as a large, outsized exposure to a particularly risky country or area of the economy?
  - Has the bank restated its financial statements owing to financial reporting internal control failures?
  - Does the bank award above-average equity-based compensation to its top managers, possibly incentivizing risk-taking behavior and short-termism?
  - What does the bank's experience with loss reserves say about its culture? Has it frequently been slow to provide for losses, only to record large asset write-downs later?

3.2 Analytical Considerations Not Addressed by CAMELS That Are Also Relevant for Any Company

There are other factors relevant to the analysis of a bank—and to any kind of company—that are not covered by the CAMELS approach. The following factors merit consideration by debt and equity investors in banks as well as investors in non-banking entities:

- **Competitive environment.** A bank's competitive position, relative to its peers, may affect how it allocates capital and assesses risks; it may also affect the aforementioned cultural mindset. A regional bank may have a near-monopolistic hold on a particular region and not take very many risks beyond maintaining its grip. A global bank may be affected by the actions of other global banks. Managers of a global bank may not be satisfied with following the lead of other banks and may pursue ambitious goals of growing market share at all costs and with little regard for risks, or they may be content with more profitable but slower growth. It depends on how the bank's managers perceive their competitive position and how they will react to the perception.
- **Off-balance-sheet items.** Off-balance-sheet assets and liabilities pose a risk to entities and their investors if they should unexpectedly drain resources. The global financial crisis of 2008–2009 was hastened by the Lehman Brothers bankruptcy, and the opacity of their involvement with such financial instruments as credit derivatives prevented concise pre-crisis analysis of the risks they shouldered. However difficult to examine, off-balance-sheet exposures need consideration whenever one analyzes a bank or financial institution.

Not all off-balance-sheet items involve exotic or highly engineered financial instruments. Operating leases are a low-risk example of off-balance-sheet liabilities: They are not a recognized liability of a company, yet they provide a creditor with a claim on a company's future cash flows. Fortunately, visibility into such future obligations is easily accessed by investors in the lease footnotes.

A financial institution analyst should be alert to the existence in the financial statements of an accounting construct known as variable interest entities, or VIEs. Variable interest entities are a form of "special-purpose entity" usually formed solely for one purpose: perhaps to hold only certain assets or assets that may be financed with specific debt instruments. Before the accounting for variable interest entities was developed, companies sometimes used outside parties to take a majority ownership stake in the special-purpose entity, ensuring that they would not have to consolidate the special-purpose entity's assets and liabilities. The accounting standard setters developed the VIE model to capture the consolidation of such special-purpose entities. By meeting generalized criteria for consolidation apart from clearly defined equity ownership tests, a company that is the primary beneficiary of a VIE's existence may be required to consolidate the VIE's financial statements with its own, even if it has no equity ownership in it. Yet a variable interest entity may also result in off-balance-sheet assets and liabilities for a bank if the bank has an interest in the VIE but is not required to consolidate it. If the VIE is not consolidated with the bank, its existence and certain financial information must be disclosed. Those non-consolidated VIEs should be of interest to investors: The reasons given for non-consolidation should be examined for reasonableness, and the implications to the bank of various scenarios affecting the VIE should be considered.

Benefit plans are another "off-balance-sheet" item for investors to examine. Although these are not completely off-balance-sheet items because the net benefit plan assets or obligations appear on the balance sheet, the economics that drive them are different from the bank's business. Shortfalls in assets due to market performance can cause rapid increases in required contributions to plans. Interest rate decreases, which drive plan obligations higher, can also cause rapid cash drains for required contributions to plans. Bank investors should examine benefits plan footnotes to determine the degree of risk posed by such plans.

One particular off-balance-sheet item that is found in financial companies only—sometimes in banks—is assets under management (AUM). Banks may have trust departments that generate management fees based on the assets under management. Those assets belong to the clients and are not consolidated with a bank's balance sheet accounts, yet they drive the returns of the bank. If such returns are material to a bank's results, the bank investor should be concerned with the size and growth or decline in assets under management.

- **Segment information.** Banks may be organized in different lines of business. They can be organized according to domestic and foreign markets; they can be organized along consumer or industrial lines of business; they may offer financial services, such as leasing or market making in securities; and they may have related businesses that are not strictly banking driven, such as trust operations. Regardless of the lines of business a bank (or any other company) may pursue, segment information should illustrate the information used by the chief operating decision maker in the entity. That information can help the investor decide whether capital is being allocated well within the bank's internally competing operations.

- **Currency exposure.** Although it may not be a problem for smaller, regional banks that operate in a single currency, floating currency exchange rates can create problems for global banks. Banks may finance and lend in a variety of currencies, resulting in foreign currency transaction exposure. Large banks may actively trade in foreign currencies and actively hedge using foreign exchange derivatives, leading to unforeseen gains or losses when world events affect currencies unexpectedly; not all banks may be successful currency traders. Global banks face the same balance sheet translation issues that affect other multinational corporations. When a bank's home currency strengthens against the functional currencies of its foreign subsidiaries, the translation of balance sheet accounts at the end of an accounting period may lead to currency translation adjustments that can reduce capital.
- **Risk factors.** Investors should review the risk factors presented in a company's annual filing. Sometimes derided as a mere list of worst-case scenarios created by a company's legal counsel, the risk factors section of a company's filing can also fill gaps in an investor's knowledge about legal and regulatory issues that might not otherwise be uncovered.
- **Basel III disclosures.** The Basel III requirements include extensive disclosures that complement the minimum risk-based capital requirements and other quantitative requirements with the goal of promoting market discipline by providing useful regulatory information to investors and other interested parties on a consistent, comparable basis.[27]

4 ANALYZING A BANK: EXAMPLE OF CAMELS APPROACH

c explain the CAMELS (capital adequacy, asset quality, management, earnings, liquidity, and sensitivity) approach to analyzing a bank, including key ratios and its limitations;

e analyze a bank based on financial statements and other factors;

This section illustrates the CAMELS approach using Citigroup's financial statements as an example. The CAMELS approach is based on the evidence gathered by the analyst in assessing each CAMELS component, and this evidence will vary from investor to investor. Some aspects of the CAMELS approach will matter more to certain investors than others: An equity investor may be far more concerned with earnings and earnings quality than with capital adequacy. A fixed-income investor might be far more concerned with capital adequacy and liquidity than earnings. The interests of each type of investor will determine what kind of analysis they perform to assess each CAMELS component. The following example of Citigroup is not intended to show all possible analyses.

It should also be understood that although the CAMELS approach entails quantitative aspects, it is not a wholly formulaic approach to analyzing a bank. An analyst's judgment and discretion also matter greatly in the application of the CAMELS approach. Judgment and discretion figure into the kind of testing done by an investor

27 Basel Committee on Banking Supervision, "Standards: Revised Pillar 3 Disclosure Requirements" (January 2015, p. 3): https://www.bis.org/bcbs/publ/d309.pdf.

to gather evidence for the various CAMELS components, and judgment and discretion also figure into the rating of the various CAMELS components once the evidence has been reviewed.

The following sections present examples of the relevant information for each component and conclude with a summary assessment. In each case, the summary assessment includes a rating, where a rating of 1 is the highest and a rating of 5 is the lowest.

4.1 Capital Adequacy

As noted above, capital adequacy relates to the proportion of a bank's assets funded by capital, with the assets accorded varying risk weightings. Not only are assets stratified into risk classes, but the bank capital funding those assets is also stratified into tiers: Common Equity Tier 1 Capital, Total Tier 1 Capital, and Tier 2 Capital.

Common Equity Tier 1 Capital includes common stock, issuance surplus related to common stock, retained earnings, accumulated other comprehensive income, and certain adjustments, including the deduction of intangible assets and deferred tax assets.

Exhibit 11 shows the calculation of Citigroup's Common Equity Tier 1 Capital, Risk-Weighted Assets, and Common Equity Tier 1 Capital Ratio at the end of 2016 and 2015. Citigroup's ratio is well within the required limits in both years. The ratio declined slightly in 2016, from 14.60% to 14.35%. The decline in the ratio is mostly attributable to the increase in deferred tax assets disallowed in the computation of Common Equity Tier 1 Capital.

Exhibit 11 Components of Citigroup Common Equity Tier 1 Capital under Current Regulatory Standards (Basel III Advanced Approaches with Transition Arrangements)

(In millions of dollars)	31 Dec. 2016	31 Dec. 2015
Citigroup common stockholders' equity	$206,051	$205,286
Add: Qualifying non-controlling interests	259	369
Regulatory capital adjustments and deductions:		
Less: Net unrealized gains (losses) on securities available for sale (AFS), net of tax	(320)	(544)
Less: Defined benefit plan liability adjustment, net of tax	(2,066)	(3,070)
Less: Accumulated net unrealized losses on cash flow hedges, net of tax (4)	(560)	(617)
Less: Cumulative unrealized net gain (loss) related to changes in fair value of financial liabilities attributable to own creditworthiness, net of tax	(37)	176
Less: Intangible assets:		
Goodwill, net of related deferred tax liabilities	20,858	21,980
Identifiable intangible assets other than mortgage servicing rights (MSRs), net of related deferred tax liabilities	2,926	1,434
Less: Defined benefit pension plan net assets	514	318
Less: Deferred tax assets (DTAs) arising from net operating loss, foreign tax credit, and general business credit carry-forwards	12,802	9,464
Less: Excess over 10%/15% limitations for other DTAs,certain common stock investments, and mortgage servicing rights	4,815	2,652
Total Common Equity Tier 1 Capital	$167,378	$173,862
Risk-Weighted Assets under Current Regulatory Standards:		
Credit risk	$773,483	$791,036
Market risk	64,006	74,817

(continued)

Exhibit 11 (Continued)

(In millions of dollars)	31 Dec. 2016	31 Dec. 2015
Operational risk	329,275	325,000
Total risk-weighted assets	$1,166,764	$1,190,853
Common Equity Tier 1 Capital Ratio (Tier 1 Capital/Total risk-weighted assets)	**14.35%**	**14.60%**
Stated minimum Common Equity Tier 1 Capital Ratio	4.50%	4.50%

Total Tier 1 Capital includes other instruments issued by the bank that meet certain criteria based on their subordination to deposit and other debt obligations, bear no fixed maturity, and carry no requirement to pay dividends or interest without full discretion of the bank. Preferred stocks can be constructed to meet these criteria.

Exhibit 12 shows the calculation of Citigroup's Total Tier 1 Capital and Total Tier 1 Capital Ratio at the end of 2016 and 2015. Again, Citigroup's ratio is well within the required limits in both years. The ratio improved in 2016, from 14.81% to 15.29%. The increase in this ratio is mostly attributable to additional perpetual preferred stock qualifying for inclusion in 2016 and the decrease in the amount of deferred tax assets disallowed in the computation of Total Tier 1 Capital.

Exhibit 12 Components of Citigroup Total Tier 1 Capital under Current Regulatory Standards (Basel III Advanced Approaches with Transition Arrangements)

(In millions of dollars)	31 Dec. 2016	31 Dec. 2015
Common Equity Tier 1 Capital (from Exhibit 11)	$167,378	$173,862
Additional Tier 1 Capital:		
Qualifying perpetual preferred stock	19,069	16,571
Qualifying trust preferred securities	1,371	1,707
Qualifying non-controlling interests	17	12
Regulatory capital adjustments and deductions:		
Less: Cumulative unrealized net gain (loss) related to changes in fair value of financial liabilities attributable to own creditworthiness, net of tax	(24)	265
Less: Defined benefit pension plan net assets	343	476
Less: DTAs arising from net operating loss, foreign tax credit and general business credit carry-forwards	8,535	14,195
Less: Permitted ownership interests in covered funds	533	567
Less: Minimum regulatory capital requirements of insurance underwriting subsidiaries	61	229
Total additional Tier 1 Capital	$11,009	$2,558
Total Tier 1 Capital (Common Equity Tier 1 Capital + Additional Tier 1 Capital)	$178,387	$176,420
Total risk-weighted assets (from Exhibit 11)	$1,166,764	$1,190,853
Tier 1 Capital Ratio	**15.29%**	**14.81%**
Minimum Tier 1 Capital Ratio	6.00%	6.00%

Tier 2 Capital includes, on a limited basis, portions of the allowance for loan and lease losses and other instruments that are subordinate to depositors and general creditors. Exhibit 13 shows the calculation of Citigroup's Tier 2 Capital and Total Capital Ratio at the end of 2016 and 2015. Consistent with the Common Equity Tier 1 Capital Ratio and the Total Tier 1 Capital Ratio, the 2016 Total Capital Ratio far exceeds the minimum requirement. The Total Capital Ratio improved from the 2015 level, from 16.69% to 17.33%. The improvement was mostly due to the increase in Total Tier 1 Capital and the amount of qualifying subordinated debt.

Exhibit 13 Components of Citigroup Tier 2 Capital under Current Regulatory Standards (Basel III Advanced Approaches with Transition Arrangements)

(In millions of dollars)	31 Dec. 2016	31 Dec. 2015
Total Tier 1 Capital (Common Equity Tier 1 Capital + Additional Tier 1 Capital)	$178,387	$176,420
Qualifying subordinated debt	22,818	21,370
Qualifying trust preferred securities	317	0
Qualifying non-controlling interests	22	17
Excess of eligible credit reserves over expected credit losses	660	1,163
Regulatory capital adjustments and deductions:		
Add: Unrealized gains on AFS equity exposures includable in Tier 2 Capital	3	5
Less: Minimum regulatory capital requirements of insurance underwriting subsidiaries	61	229
Total Tier 2 Capital	$23,759	$22,326
Total Capital (Tier 1 Capital + Tier 2 Capital)	$202,146	$198,746
Total risk-weighted assets	$1,166,764	$1,190,853
Total Capital Ratio	**17.33%**	**16.69%**
Minimum Capital Ratio	**8.00%**	**8.00%**

In summary, Citigroup's capital adequacy at the end of 2016 appears to be solidly positive. For each of the three chief capital ratios, the company has exceeded the minimum levels required for being considered to be a well-capitalized bank. A rating of 1 could be justified by their ratios, which far exceeded the minimum levels.

4.2 Asset Quality

Asset quality matters greatly to a bank. As financial intermediaries in an economy, banks owe their existence to the creation of loans. If a bank's credit policies are unsound, its capital base can be quickly eroded during economic downturns, creating strains on the bank's liquidity and its ability to generate earnings. Creating new loans becomes problematic.

A portion of bank assets are held in highly liquid financial instruments, such as cash, deposits held at other banks, and instruments that may convert into cash in a very short time frame, such as repurchase agreements and some receivables. These are not highly risky assets.

Increasing in riskiness are the investments made by the bank in financial instruments with cash deemed to be in excess of operating needs. Under US GAAP and IFRS, these investments may be classified as available-for-sale investments, which are reported at fair value, or held-to-maturity investments, which are reported at their amortized cost unless an impairment occurs. While these investments are riskier than the liquid securities and reflect an investment decision made by management, their value is quite transparent and their reported value reflects their realizability in cash—although it takes more analytical effort to make that assertion for held-to-maturity securities.

The riskiest, and often the largest, asset classes are the loans underwritten by the bank. Loans embody credit risk and the judgment of management in extending credit to customers. The underwriting risks and the management judgments in assessing them are reflected in the allowance for loan losses. It is here that the analyst faces some of the most difficult assessments in understanding the quality of banking assets and is at a disadvantage, because some information simply is unavailable to an analyst (or investor). Conversely, an examiner for a supervisory regulator has the ability to see the bank from the inside and assess the soundness of loan (and investment) policies and procedures. An examiner may also review the construction and workings of internal control procedures and may be able to examine how exceptions to credit policies are being handled.[28]

Although the analyst is interested in all of those inner workings, he/she can be concerned only with circumstantial evidence that the credit policies are sound and are being maintained. That circumstantial evidence can be found in the financial statements, but it is not completely obvious from merely looking at a balance sheet. There are ways to find evidence of asset quality, as will be shown with Citigroup. Exhibit 14 shows the asset side of Citigroup's balance sheet on a condensed basis at the end of 2016 and 2015.

Exhibit 14 Citigroup Asset Composition, 31 December 2016 and 2015

	31 December 2016		31 December 2015	
(In millions of dollars)	$	% Total Assets	$	% Total Assets
Liquid assets:				
Cash and due from banks	$23,043	*1.3%*	$20,900	*1.2%*
Deposits with banks	137,451	*7.7%*	112,197	*6.5%*
Federal funds sold and securities borrowed or purchased under resale agreements	236,813	*13.2%*	219,675	*12.7%*
Brokerage receivables	28,887	*1.6%*	27,683	*1.6%*
Trading account assets	243,925	*13.6%*	241,215	*13.9%*
Total liquid assets	670,119	*37.4%*	621,670	*35.9%*
Investments:				
Available-for-sale	299,424	*16.7%*	299,136	*17.3%*
Held-to-maturity	45,667	*2.5%*	36,215	*2.1%*
Non-marketable equity securities	8,213	*0.5%*	7,604	*0.4%*
Total investments	353,304	*19.7%*	342,955	*19.8%*
Loans:				
Consumer	325,366	*18.2%*	325,785	*18.8%*
Corporate	299,003	*16.7%*	291,832	*16.9%*
Loans, net of unearned income	624,369	*34.9%*	617,617	*35.7%*

28 See Section 3.1 of the FDIC's "RMS Manual of Examination Policies," available at https://www.fdic.gov/regulations/safety/manual/section3-1.pdf.

Exhibit 14 (Continued)

	31 December 2016		31 December 2015	
(In millions of dollars)	$	% Total Assets	$	% Total Assets
Allowance for loan losses	(12,060)	*−0.7%*	(12,626)	*−0.7%*
Total loans, net	612,309	*34.2%*	604,991	*35.0%*
Goodwill	21,659	*1.2%*	22,349	*1.3%*
Intangible assets (other than MSRs)	5,114	*0.3%*	3,721	*0.2%*
Mortgage servicing rights (MSRs)	1,564	*0.1%*	1,781	*0.1%*
Other assets	128,008	*7.1%*	133,743	*7.7%*
Total assets	$1,792,077	*100.0%*	$1,731,210	*100.0%*

Observations from the composition of the assets:

- Citigroup's liquid assets are the largest single group of all, at 37.4% in 2016, and slightly greater than the year before, indicating greater liquidity.
- The proportion of investments to total assets of 19.7% is practically unchanged from 2015; the majority of the investments are available-for-sale securities reported at fair value.
- Consumer and corporate loans are the highest-risk assets and in both years amount to more than one-third of all assets. They are the second largest class of assets after the liquid assets.

In assessing asset quality, an analyst would want to focus on the riskiest assets in the mix: the investments and the loans. He or she would want to determine that the investments, while transparent in value, represent sound investment decisions and that the loans result from similarly reasoned underwriting policies. The analyst would want assurance that the stated amount of loans is collectible and that the allowance for loan losses is reasonably stated.

First, take a look at the investments. Exhibit 15 shows Citigroup's available-for-sale securities by class at the end of 2016. Exhibit 15 was extracted from Note 13 of the 2016 10-K, which showed the amortized cost by investment instrument, gross unrealized gains, gross unrealized losses, and fair value as stated in the balance sheet. Added to the table were the gross unrealized gains and losses expressed as a percentage of amortized cost, which is the amount invested.

Exhibit 15 Citigroup Available-for-Sale (AFS) Securities at 31 December 2016

					% of Cost:	
		Gross Unrealized			Gross Unrealized	
(In millions of dollars)	Amortized Cost	Gains	Losses	Fair Value	Gains	Losses
Debt securities AFS						
Mortgage-backed securities:						
US government-sponsored agency guaranteed	$38,663	$248	$506	$38,405	*0.6%*	*1.3%*
Prime	2	—	—	2	—	—

(continued)

Exhibit 15 (Continued)

(In millions of dollars)	Amortized Cost	Gross Unrealized Gains	Gross Unrealized Losses	Fair Value	% of Cost: Gross Unrealized Gains	% of Cost: Gross Unrealized Losses
Alt-A	43	7	—	50	*16.3%*	—
Non-US residential	3,852	13	7	3,858	*0.3%*	*0.2%*
Commercial	357	2	1	358	*0.6%*	*0.3%*
Total mortgage-backed securities	$42,917	$270	$514	$42,673	*0.6%*	*1.2%*
US Treasury and federal agency securities						
US Treasury	$113,606	$629	$452	$113,783	*0.6%*	*0.4%*
Agency obligations	9,952	21	85	9,888	*0.2%*	*0.9%*
Total US Treasury and federal agency securities	$123,558	$650	$537	$123,671	*0.5%*	*0.4%*
State and municipal	$10,797	$80	$757	$10,120	*0.7%*	***7.0%***
Foreign government	98,112	590	554	98,148	*0.6%*	*0.6%*
Corporate	17,195	105	176	17,124	*0.6%*	*1.0%*
Asset-backed securities	6,810	6	22	6,794	*0.1%*	*0.3%*
Other debt securities	503	—	—	503	*0.0%*	*0.0%*
Total debt securities AFS	$299,892	$1,701	$2,560	$299,033	*0.6%*	*0.9%*
Marketable equity securities AFS	$377	$20	$6	$391	*5.3%*	*1.6%*
Total securities AFS	$300,269	$1,721	$2,566	$299,424	*0.6%*	*0.9%*

The fair value ($299,424 million) is less than the amortized cost ($300,269 million) in the aggregate, and the net difference is $845 million; the largest contributor to that loss is the state and municipal obligations, with a $757 million loss. At a 7% loss in value, those were the only securities to generate losses greater than 2%.

Observations from the AFS securities valuation table:

- Although Citigroup has not generated a net winning strategy with its available-for-sale investments, its losses do not suggest reckless abandon.
- In future years, new US GAAP standards will eliminate the AFS classification for marketable equity securities. They will still be measured at fair value, just as they were measured at year end 2016. Starting in 2018, however, the gains and losses resulting from remeasurement will be shown directly in the income statement instead of being recorded in other comprehensive income. As of 31 December 2016, Citigroup's unrealized gains on its AFS equity investments exceeded its unrealized losses. Based on market values at that point in time, the reclassification would benefit the group's income.

A closer look at the gross unrealized losses is possible, because Note 13 also contains a simple aging of the losses: It shows how much of the $2.566 billion of unrealized losses are less than 12 months old and how much of the losses are 12 months old or older, by category. The longer a loss position exists, the greater the possibility that a security is impaired on an "other-than-temporary" basis. It would be unusual for losses to exist for long periods of time and then suddenly reverse.

The aging for the losses in Citigroup's available-for-sale securities is shown in Exhibit 16. Observations from the aging of AFS unrealized losses table:

- A slight majority (54%) of the losses are less than 12 months old, making them of less concern than the rest.
- Of the $1.172 billion of gross unrealized losses 12 months old or older, 60% ($702 million) are related to state and municipal securities, which raises a concern that the largest class of losses may in fact become realized.

Exhibit 16 Citigroup Aging of Unrealized Losses on Available-for-Sale Securities at 31 December 2016

	Less than 12 months		12 months or longer		Total	
(In millions of dollars)	Fair value	Gross unrealized losses	Fair value	Gross unrealized losses	Fair value	Gross unrealized losses
Mortgage-backed securities						
US government-sponsored agency sponsored	$23,534	$436	$2,236	$70	$25,770	$506
Prime	1	—	—	—	1	—
Non-US residential	486	—	1,276	7	1,762	7
Commercial	75	1	58	—	133	1
Total mortgage-backed securities	$24,096	$437	$3,570	$77	$27,666	$514
US Treasury and federal agency securities						
US Treasury	$44,342	$445	$1,335	$7	$45,677	$452
Agency obligations	6,552	83	250	2	6,802	85
Total US Treasury and federal agency securities	$50,894	$528	$1,585	$9	$52,479	$537
State and municipal	$1,616	$55	$3,116	$702	$4,732	$757
Foreign government	38,226	243	8,973	311	47,199	554
Corporate	7,011	129	1,877	47	8,888	176
Asset-backed securities	411	—	3,213	22	3,624	22
Other debt securities	5	—	—	—	5	—
Marketable equity securities AFS	19	2	24	4	43	6
Total securities AFS	$122,278	$1,394	$22,358	$1,172	$144,636	$2,566

A similar analysis can be done for the held-to-maturity (HTM) securities. Even though they represent a much smaller proportion of total assets, they still provide evidence of the manager's investment acumen. The result of the HTM securities review of the losses and aging of the losses is consistent with the results of the available-for-sale securities review. Though not presented in exhibits because of space limitations, Citigroup's unrealized losses on its HTM securities totaled $457 million, which is 1.3% of the amount invested. Of that $457 million loss in value, 82% ($373 million) stemmed from held-to-maturity securities that were showing losses older than 12 months, of which $180 million related to state and municipal securities.

Observations on the HTM securities:

- The losses on the HTM securities are much less in dollar amount than the losses on the AFS securities, and although they are minor in percentage terms of a loss, they are troubling because of their age. Problem assets do not usually improve with age, and the fact that the bulk of the losses on the HTM securities are older than 12 months may indicate management reluctance to report economic reality.

- Because HTM securities are reported at amortized cost on the balance sheet, the classification obscures the fair value of the securities. The age of the securities generating the losses indicates that there may be more severe impairment than already recognized. The analysis of the HTM securities reinforces the observations noted in the analysis of the available-for-sale securities review.

Investment assets are not as significant in amount or as risky as the loans. The analyst wants to determine that the loans are the result of a sound credit policy and will be realized over their term. This cannot be determined without analyzing the allowance for loan losses. As was seen in Exhibit 14, **allowance for loan losses** is a balance sheet account; it is a contra asset account to loans. (It is analogous to an account common for non-financial institutions, allowance for bad debts, which is a contra asset account to accounts receivable.) **Provision for loan losses** is an income statement expense account that increases the amount of the allowance for loan losses. Actual loan losses (i.e., charge-offs—net of recoveries) reduce the amount of the allowance for loan losses.

The allowance for loan losses matters greatly to understanding loan quality, because total loans minus the allowance for loan losses represents the expected value of the loans. A bank's balance sheet will typically show the total amount of loans, the amount of allowance for loan losses, and the net amount. Importantly, the allowance for loan losses is discretionary by its very nature. Underestimating the allowance for loan losses would overstate the amounts reported for assets and net income. Almost every bank will disclose allowances for loan losses among its most critical accounting estimates.

To effectively assess the adequacy of the allowance for loan losses, an analyst can examine measures that involve less management discretion. Net charge-offs of loans are less discretionary indicators of loan quality than the allowance for loan losses but have the disadvantage of being a confirming event: The loan has already turned out to be a non-performing asset. Another disadvantage is that net charge-offs can be used in good times to pack away earnings to be brought into earnings later through recoveries of charge-offs. Non-performing loans are another measure that can help in assessing adequacy of the allowance for loan losses. Non-performing (i.e., non-accrual) loans are loans that are not currently paying their contractual amounts due, making them a more objective measure of the quality of loans in the portfolio.

Three ratios are helpful in assessing the quality of the allowance for loan losses:

- The ratio of the allowance for loan losses to non-performing loans
- The ratio of the allowance for loan losses to net loan charge-offs
- The ratio of the provision for loan losses to net loan charge-offs

In each ratio, a discretionary measure (such as the allowance for loan losses or provision for loan losses) is compared to a more objective measure.[29] In the case of Citigroup, the loans and the allowance for loan losses are stratified between consumer loans and corporate loans. Because the types of loan customers differ greatly, the analysis of each should be performed separately. Exhibit17 shows the variables required to compute the ratios for the last five years, selected from the management discussion and analysis of the relevant 10-Ks, and the resulting ratios.

29 For more discussion on the analysis of the allowance of loan loss reserves, see Stephen G. Ryan, *Financial Instruments and Institutions: Accounting and Disclosure Rules* (Hoboken, NJ: Wiley, 2002): 100–105.

Exhibit 17 Citigroup's Loan Loss Analysis Data at 31 December

(In millions of dollars)	2016	2015	2014	2013	2012
Data for Calculating Allowance for Loan Loss Ratios					
Allowance for loan losses:					
Consumer	$9,358	$9,835	$13,547	$16,974	$22,585
Corporate	$2,702	$2,791	$2,447	$2,674	$2,870
Provision for loan losses:					
Consumer	$6,323	$6,228	$6,695	$7,587	$10,312
Corporate	$426	$880	$133	$17	$146
Charge-offs:					
Consumer	$7,644	$8,692	$10,650	$12,400	$16,365
Corporate	$578	$349	$458	$369	$640
Recoveries:					
Consumer	$1,594	$1,634	$1,975	$2,138	$2,357
Corporate	$67	$105	$160	$168	$417
Net charge-offs:					
Consumer	$6,050	$7,058	$8,675	$10,262	$14,008
Corporate	$511	$244	$298	$201	$223
Non-accrual loans:					
Consumer	$3,158	$3,658	$5,905	$7,045	$9,136
Corporate	$2,421	$1,596	$1,202	$1,958	$2,394
Allowance for Loan Loss Ratios					
Allowance for loan losses to non-accrual loans:					
Consumer	2.96	2.69	2.29	2.41	2.47
Corporate	1.12	1.75	2.04	1.37	1.20
Allowance for loan losses to net loan charge-offs:					
Consumer	1.55	1.39	1.56	1.65	1.61
Corporate	5.29	11.44	8.21	13.30	12.87
Provision for loan losses to net loan charge-offs:					
Consumer	1.05	0.88	0.77	0.74	0.74
Corporate	0.83	3.61	0.45	0.08	0.65

Observations on the allowance for loan losses to non-accrual loans, which are loans that have experienced some non-payment from borrowers:

- For the consumer loans, the 2016 ratio of 2.96 is the highest level in the last five years, and this ratio has been increasing in the last two years. It indicates that the allowance (a discretionary amount) is increasing faster than the actual non-accrual loans, lending confidence to analysts that the allowance is being built in advance of loans turning out poorly.
- For the corporate loans, the 2016 ratio of 1.12 is less definitive. It might be expected that the ratio would be more volatile than for the consumer business because the corporate lending business is not homogeneous, and specific credits and their failures could cause spikes in the ratio. Still, the allowance has

declined in each of the last three years, and in 2016, it is at its lowest point in five years. This arouses concern that the allowance for loan losses may be a thin layer of protection against future losses.

Observations on the allowance for loan losses to net loan charge-offs:

- For the consumer loans, the 2016 ratio of 1.55 shows improvement from 2015 and indicates that there is a cushion between the allowance and the net loan charge-offs that has remained fairly constant over the last five years.
- For the corporate loans, the 2016 ratio of 5.29 shows an ample cushion between the allowance and the net loan charge-offs, although it declined greatly from 2015 and is much lower than at any time in the last five years.

Observations on the provision for loan losses to net loan charge-offs:

- The provision for loan losses is the amount added to the allowance each year, and one should expect that the provision correlates to the amount of net loan charge-offs.
- For the consumer loans, the 2016 ratio is the first ratio in five years where the provision exceeded the net loan charge-offs, and although it had been lower in the previous four years, the proportion of the provision to charge-offs had been increasing in the last three years. This indicates that the bank had become more conservative in its provisioning.
- For the corporate loans, the 2016 ratio significantly decreased from the previous year, and the ratio has been less than 1.0 in four of the last five years. This indicates that the provision for corporate loans has trailed the actual net charge-off experience. The large addition in 2015 gives the appearance of an urgent "catch-up" adjustment.

In summary, Citigroup's asset quality at the end of 2016 was mixed. The policies for investments appear to be fairly conservative, but the age of some of the investments with unrealized losses indicates a possible denial of impairment. With regard to loan quality, the ratio analysis of the allowance for loan losses suggests that the consumer loans appear to be well reserved, but the same ratio analysis for the corporate loans does not generate the same degree of comfort. A rating of 2.5—near the midpoint of the rating scale—could be assigned to the asset quality based on the mixed signals from the evidence.

4.3 Management Capabilities

External investors can observe only circumstantial evidence of management's quality. Some circumstantial evidence can be found through a review of the proxy statement.

Observations based on a review of Citigroup's 2016 proxy:

- Citigroup aims for two-thirds board representation of independent members, whereas the New York Stock Exchange requires only a majority of independent members.
- Citigroup has a separate CEO and chairman, often viewed as a good governance practice that avoids conflicts of interest. The positions have been separate since 2009.
- Citigroup's Risk Management Committee met frequently in 2016—14 times—providing evidence of attention to one of the most critical parts of a banking operation. Furthermore, the Risk Management Committee created a subcommittee in 2016 to provide oversight of data governance, data quality, and data integrity, and the subcommittee met seven times in 2016.

Although these are good practices, they do not constitute evidence of strong management capabilities. Rather, they provide evidence that an environment exists where strong management quality is permitted to flourish.

With a company as large as Citigroup, it is difficult to avoid related-party transactions. For example, BlackRock and Vanguard beneficially owned 5% or more of the outstanding shares of Citigroup's common stock as of 31 December 2016; during 2016, the company's subsidiaries provided ordinary course lending, trading, and other financial services to BlackRock and Vanguard. The proxy states that the transactions were on an arm's-length basis and contain customary terms that are substantially the terms of comparable transactions with unrelated third parties. Other related-party transactions exist and are discussed in the 10-K, but they are routine for a company of this size.

In terms of operational risk, evidence of the board's influence on management can be found in the unqualified opinion of Citigroup's auditor on the effectiveness of the system of internal controls. This is evidence of a minimally satisfactory environment in which a management should operate and not a clear signal of management competence. A qualified (or negative) opinion on the effectiveness of internal controls would be especially concerning for an investor.

In summary, although the board may be solidly constructed and appears to exert adequate control over the managers, the net performance of the company also speaks to the quality of management and directors. The asset quality, discussed above, was not overwhelmingly positive and detracts from the overall view of management quality. A rating of 2 could be assigned to management capabilities.

4.4 Earnings

Earnings ideally should be of high quality, and an indication of high-quality earnings is sustainability. Earnings are more sustainable if they are not dependent on the possibly opportunistic fine-tuning of discretionary estimates and not reliant on either non-recurring items or volatile sources of revenues.

As discussed above, allowance for loan losses and provisions for loan losses are estimated amounts that allow for management discretion. The provision for loan losses can have profound effects on the profitability of a bank in any single year and over long periods of time. Exhibit 18 shows the five-year change in Citigroup's pretax income through 2016 and the corresponding change in the consolidated total provisions for credit (i.e., loan loss reserves plus provisions for policyholder benefits and claims and unfunded lending commitments) drawn from the five-year selected financial data from the 2016 and 2015 10-Ks.

Exhibit 18 Historical Pretax Income and Total Provisions for Credit Losses

(In millions of dollars)	5- Year Net Change:	2016	2015	2014	2013	2012	2011
Pretax income		$21,477	$24,826	$14,701	$19,802	$8,165	$15,096
Change in pretax income	$6,381	($3,349)	$10,125	($5,101)	$11,637	($6,931)	—
Total provisions for credit losses		$6,982	$7,913	$7,467	$8,514	$11,329	$12,359

(continued)

Exhibit 18 (Continued)

(In millions of dollars)	5- Year Net Change:	2016	2015	2014	2013	2012	2011
Change in total provisions for credit losses	($5,377)	($931)	$446	($1,047)	($2,815)	($1,030)	—
Net difference	$1,004						

Observations on the provisions for credit losses:

- 2013 was the only year in which pretax income increased from the previous year while the total provisions for credit losses decreased. The $2.815 billion decrease in the credit loss provisions drove 24% of the increase.
- In 2016, 2014, and 2012, the pretax income declined from the previous year. The declines would have been more severe if they had not been buffered by decreases in the total provisions for credit losses in each year.
- Over the five-year span, the change in the total credit loss provisions contributed to improving the pretax earnings in four of the years. The only exception was 2015, when the total provisions increased only negligibly compared to the size of the decreases in the other years.
- On a longer-term basis, the five-year net change in the total provisions accounted for 84% of the net change in pretax income—an indication that not much profit growth happened elsewhere.

Another indicator of sustainability is the degree to which trading income is part of a bank's revenue stream. Trading income tends to be volatile and not necessarily sustainable. Higher-quality income would be net interest income and fee-based income: These provide sustainable, returning streams of income. An analyst should examine the composition of a bank's revenue stream to determine whether it is growing and to identify the drivers of growth or decline. The five-year summary of Citigroup's revenue stream, drawn from the five-year selected financial data from the 2016 and 2015 10-Ks, is shown in Exhibit 19.

Exhibit 19 Five-Year Summary of Composition of Citigroup's Revenue

(In millions of dollars)	2016	2015	2014	2013	2012
Net interest revenue	$45,104	$46,630	$47,993	$46,793	$46,686
Principal transactions (trading income)	7,585	6,008	6,698	7,302	4,980
All other non-interest revenue	17,186	23,716	22,528	22,629	17,864
Revenues, net of interest expense	$69,875	$76,354	$77,219	$76,724	$69,530
Percent attributable to trading income	10.9%	7.9%	8.7%	9.5%	7.2%
Percent of total:					
Net interest revenue	64.5%	61.1%	62.2%	61.0%	67.1%
All other non-interest revenue	24.6%	31.1%	29.2%	29.5%	25.7%

Observations on revenue composition:

- 2016 total revenues are almost unchanged from 2012 levels.
- At 10.9% of total revenues in 2016, trading income has been trending upward as a proportion of revenues in the last five years. Instead of increasing its sustainable, non-volatile revenues, Citigroup's principal transactions/trading income is moving in the opposite direction—increasing in absolute dollars and in relative importance.
- In 2016, all other non-interest revenue is at its lowest representative level since 2012.
- The net interest revenue proportion improved in 2016 but is still lower than it was in 2012.

A bank's net interest revenue results from the management of interest earned on loans and other interest-bearing assets and the management of interest paid on deposits and other interest-bearing liabilities. Thus, net interest revenue earned on average interest-bearing assets minus interest expense paid on average interest-bearing liabilities. Banks may create value through maturity transformation: They can borrow money on shorter terms than the terms for lending to customers. Although this can create value by lending for long terms at a higher rate than their short-term funding costs, it can also destroy value if the markets for short-term funding experience a dislocation or the yield curve unexpectedly inverts. Therefore, a bank's risk management practices, including its diversification practices, are integral to the maturity transformation process.

Analyzing the net interest revenue can provide an investor with a view of a bank management's activity and effectiveness in this area. To continue the example with Citigroup, the next two exhibits show the average balances (average volume column) for Citigroup's balance sheet accounts. Exhibit 20 shows Citigroup's average assets, as well as interest revenue and average interest rate earned on those assets. Exhibit 21 shows the company's average liabilities, the interest expense and average interest cost of those liabilities, and its equity accounts. It also includes the company's net interest revenue and net interest margin at the bottom.

Exhibit 20 Citigroup's Average Balances and Interest Rates—Assets

In millions of dollars, except rates	Average Volume			Interest Revenue			% Average Rate		
	2016	**2015**	**2014**	**2016**	**2015**	**2014**	**2016**	**2015**	**2014**
Assets									
Deposits with banks	**$131,925**	$133,853	$161,741	**$971**	$727	$959	**0.74%**	0.54%	0.59%
Federal funds sold and securities borrowed or purchased under agreements to resell									
In US offices	**$147,734**	$150,340	$153,703	**$1,483**	$1,215	$1,034	**1.00%**	0.81%	0.67%
In offices outside the United States	**85,142**	84,013	101,184	**1,060**	1,301	1,332	**1.24**	1.55	1.32
Total	**$232,876**	$234,353	$254,887	**$2,543**	$2,516	$2,366	**1.09%**	1.07%	0.93%
Trading account assets									
In US offices	**$103,610**	$113,475	$113,716	**$3,791**	$3,945	$3,471	**3.66%**	3.48%	3.05%
In offices outside the United States	**94,603**	96,333	113,563	**2,095**	2,140	2,540	**2.21**	2.22	2.24
Total	**$198,213**	$209,808	$227,279	**$5,886**	$6,085	$6,011	**2.97%**	2.90%	2.64%

(continued)

Exhibit 20 (Continued)

In millions of dollars, except rates	Average Volume			Interest Revenue			% Average Rate		
	2016	**2015**	**2014**	**2016**	**2015**	**2014**	**2016**	**2015**	**2014**
Investments									
In US offices									
Taxable	**$225,764**	$214,683	$188,909	**$3,980**	$3,812	$3,285	**1.76%**	1.78%	1.74%
Exempt from US income tax	**19,079**	20,034	20,383	**693**	443	626	**3.63**	2.21	3.07
In offices outside the United States	**106,159**	102,374	113,182	**3,157**	3,071	3,627	**2.97**	3.00	3.20
Total	**$351,002**	$337,091	$322,474	**$7,830**	$7,326	$7,538	**2.23%**	2.17%	2.34%
Loans (net of unearned income)									
In US offices	**$360,957**	$354,434	$361,773	**$24,240**	$25,082	$26,076	**6.72%**	7.08%	7.21%
In offices outside the United States	**262,715**	273,064	296,666	**15,578**	15,465	18,723	**5.93**	5.66	6.31
Total	**$623,672**	$627,498	$658,439	**$39,818**	$40,547	$44,799	**6.38%**	6.46%	6.80%
Other interest-earning assets	**$56,398**	$63,209	$48,954	**$1,029**	$1,839	$507	**1.82%**	2.91%	1.04%
Total interest-earning assets	**$1,594,086**	$1,605,812	$1,673,774	**$58,077**	$59,040	$62,180	**3.64%**	3.68%	3.71%
Non-interest-earning assets	**$214,642**	$218,025	$223,141						
Total assets	**$1,808,728**	$1,823,837	$1,896,915						

Observations from Citigroup's average assets table:

- The overall average interest rate earned declined slightly in 2016, from 3.68% to 3.64%. One reason is due to changes occurring within the loans, which are the single largest category of assets. Citigroup sold its OneMain Financial subsidiary at the end of 2015, which was engaged in US consumer installment lending and is a high-yielding loan business. That disposal pressured the interest income earned from US offices, decreasing the earned interest rate from 7.08% in 2015 to 6.72% in 2016.
- Average loans in US offices increased to $360,957 million in 2016 from $354,434 million in 2015, despite the OneMain disposal, because of the mid-2016 acquisition of Costco's credit card portfolio, which was insufficient to offset the OneMain interest income.
- Average loans in offices outside the United States decreased to $262,715 million in 2016 from $273,064 million in 2015, partly because Citigroup disposed of its retail banking and credit cards businesses in Japan in the fourth quarter of 2015.
- Although Citigroup realized better interest income from its trading account assets in 2016, earning 2.97% compared to 2.90% in 2015, it allocated less capital to trading and earned less absolute interest income from the trading account assets.
- Despite lower capital committed to deposits with banks ($131,925 million in 2016 compared to $133,853 million in 2015), the higher realized average interest rate increased Citigroup's overall interest income. The same is true for its tax-exempt investments in US offices.

Exhibit 21 Citigroup's Average Balances and Interest Rates—Liabilities, Equity, and Net Interest Revenue

In millions of dollars, except rates	Average Volume			Interest Expense			% Average Rate		
	2016	**2015**	**2014**	**2016**	**2015**	**2014**	**2016**	**2015**	**2014**
Liabilities									
Deposits in US offices	**$288,817**	$273,135	$292,062	**$1,630**	$1,291	$1,432	**0.56%**	0.47%	0.49%
In offices outside the United States	**429,608**	425,086	465,135	**3,670**	3,761	4,260	**0.85**	0.88	0.92
Total	**$718,425**	$698,221	$757,197	**$5,300**	$5,052	$5,692	**0.74%**	0.72%	0.75%
Federal funds purchased and securities loaned or sold under agreements to purchase									
In US offices	**$100,472**	$108,320	$102,672	**$1,024**	$614	$657	**1.02%**	0.57%	0.64%
In offices outside the United States	**57,588**	66,130	88,080	**888**	998	1,238	**1.54**	1.51	1.41
Total	**$158,060**	$174,450	$190,752	**$1,912**	$1,612	$1,895	**1.21%**	0.92%	0.99%
Trading account liabilities									
In US offices	**$29,481**	$24,711	$29,263	**$242**	$107	$74	**0.82%**	0.43%	0.25%
In offices outside the United States	**44,669**	45,252	47,904	**168**	110	94	**0.38**	0.24	0.20
Total	**$74,150**	$69,963	$77,167	**$410**	$217	$168	**0.55%**	0.31%	0.22%
Short-term borrowings									
In US offices	**$61,015**	$64,973	$77,967	**$202**	$224	$161	**0.33%**	0.34%	0.21%
In offices outside the United States	**19,184**	50,803	40,282	**275**	299	419	**1.43**	0.59	1.04
Total	**$80,199**	$115,776	$118,249	**$477**	$523	$580	**0.59%**	0.45%	0.49%
Long-term debt									
In US offices	**$175,342**	$182,347	$191,364	**$4,179**	$4,308	$5,093	**2.38%**	2.36%	2.66%
In offices outside the United States	**6,426**	7,642	7,346	**233**	209	262	**3.63**	2.73	3.57
Total	**$181,768**	$189,989	$198,710	**$4,412**	$4,517	$5,355	**2.43%**	2.38%	2.69%
Total interest-bearing liabilities	**$1,212,602**	$1,248,399	$1,342,075	**$12,511**	$11,921	$13,690	**1.03%**	0.95%	1.02%
Demand deposits									
In US offices	**$38,120**	$26,144	$26,227						
Other non-interest-bearing liabilities	**328,822**	330,104	316,061						
Total liabilities	**$1,579,544**	$1,604,647	$1,684,363						
Citigroup stockholders' equity	**$228,065**	$217,875	$210,863						
Non-controlling interest	**1,119**	1,315	1,689						
Total equity	**$229,184**	$219,190	$212,552						
Total liabilities and stockholders' equity	**$1,808,728**	$1,823,837	$1,896,915						

(continued)

Exhibit 21 (Continued)

In millions of dollars, except rates	Average Volume			Interest Expense			% Average Rate		
	2016	**2015**	**2014**	**2016**	**2015**	**2014**	**2016**	**2015**	**2014**
Net interest revenue as a percentage of average interest-earning assets	**Total Average Interest-Earning Assets**			**Net Interest Revenue**			**Net Interest Margin**		
In US offices	**$859,311**	$923,309	$953,394	**$27,929**	$28,495	$27,496	**3.25%**	3.09%	2.88%
In offices outside the United States	**734,775**	682,503	720,380	**17,637**	18,624	20,994	**2.40**	2.73	2.91
Total	**$1,594,086**	$1,605,812	$1,673,774	**$45,566**	$47,119	$48,490	**2.86%**	2.93%	2.90%

Observations from Exhibit 21:

- Citigroup's cost of funding its assets increased in every category of liability in 2016. That attribute was even more pronounced in offices outside the United States, with the exception of deposit liabilities.
- This difference between US and non-US asset and liability performance extends to the net interest margin, net interest revenue as a percentage of average interest-earning assets, shown at the bottom of the table. Although the net interest margin improved to 3.25% in 2016 from 3.09% in 2015 for assets in US offices, it declined significantly for assets in offices outside the United States, from 2.73% in 2015 to 2.40% in 2016. The net interest margin in offices outside the United States has been declining consistently since 2014, when the US dollar began strengthening. Citigroup has been experiencing negative foreign currency translation impacts since then.
- An investor might exercise increased caution when observing management's future actions in making foreign investments. These results do not provide assurance that all capital is well allocated overseas or that currency risk is adequately managed. The lower returns might also be due to macroeconomic factors, such as lower yield curves (and even negative rates) overseas, creating fewer profitable opportunities. An investor should factor those possibilities into his consideration.

Analyzing the net interest revenue resulting from average interest-bearing asset and liability balances can be useful for analyzing what happened within a bank for a given period but not necessarily useful for projecting future earnings. The interest earned or paid on an average balance for a given period may have no bearing on what a bank may actually earn or pay in the next period. End-of-period balances of balance sheet components and their associated interest rates may make a better starting point for projecting future earnings than the average balance information.

In summary, the quality of Citigroup's earnings is not exceedingly high. The fact that the decreases in the provision for loan losses has driven 84% of the pretax earnings increases over the last five years does nothing to relieve quality concerns, nor does the increase in trading income over the last five years instill more confidence in the earnings quality. The analysis of the net interest revenue shows declining net interest margin over the last three years, largely attributable to the non-US offices. A rating of 3 could be justified for earnings quality.

A Brief Overview of Accounting for Derivatives

Accounting rules for derivatives are extensive. The following points are a very brief summary of this complex topic and are generally applicable to both IFRS and US GAAP.

- At inception, many derivatives contracts do not give rise to an asset or liability on the balance sheet or to a gain or loss on the income statement. For example, an interest rate swap contract can involve the exchange of future cash flows with equivalent present value. Thus, at inception, the only accounting record required for every derivatives contract is a disclosure of the notional amount of the contract. This disclosure appears in the notes to the financial statements.
- Measurement of the mark-to-market value of a derivatives contract creates an asset or liability and subsequently increases or decreases the value of the asset or liability.
- Changes in the value of the asset or liability are recorded either as part of profit and loss on the income statement or as part of comprehensive income, depending on the classification.
- Derivative instruments can be classified as a hedge of a cash flow or a hedge of a net investment in a foreign subsidiary. Classification of a derivatives contract as a hedge requires substantiating its correlation with the risk being hedged. If a derivatives contract is classified as one of these two types of hedges, changes in its value are recorded as part of other comprehensive income and will be recognized in net income over the life of the hedged transaction.
- If such a derivatives contract fails classification as a hedge and is instead a free-standing derivative instrument or if the hedge is classified as a third type, a hedge of a fair value, then changes in its fair value are reported as income or expense in the income statement at each reporting period. The immediate recognition of a gain or loss in earnings, instead of reporting it in other comprehensive income, can lead to unexpected volatility of earnings and missed earnings targets. Depending on the nature of the derivative transaction, a secondary effect of a contract's failure to qualify as a hedge may also require additional posting of collateral or cash.

4.5 Liquidity Position

A bank's liquidity is an extremely important matter for its own well-being in times of financial stress. Given the interdependence of banks, through such transactions as interbank deposits and acting as counterparties in derivative transactions, a bank's liquidity also matters for the well-being of other banks—and possibly an entire economy.

Capital alone is not sufficient to assure liquidity; there must be enough capital available in cash or near-cash for the meeting of obligations. The Basel III Regulatory Framework introduced two liquidity standards to provide assurance that capital would be adequately liquid for meeting obligations under stressful conditions.

The first is the Liquidity Coverage Ratio, which is the minimum percentage of a bank's expected cash outflows to be held in highly liquid assets. Expected net cash outflows are the bank's anticipated 30-day liquidity needs in a stress scenario, and the highly liquid assets include only those of high quality and immediately convertible into cash. Expected net cash outflows are calculated by applying prescribed outflow factors to various liability categories, with any available offsets by inflows from assets maturing within the 30-day stress period. Additionally, banks must include an add-on amount to account for possible maturity mismatches between contractual cash outflows and inflows during the 30-day period to arrive at total net outflows. The minimum

LCR threshold is 100%; anything less would indicate an inability to meet the liquidity needs. Exhibit 22 shows the components of Citigroup's LCR at 31 December 2016, 30 September 2016, and 31 December 2015.

Exhibit 22 Citigroup's Liquidity Coverage Ratio

(In billions of dollars)	31 Dec. 2016	30 Sep. 2016	31 Dec. 2015
High-quality liquid assets	$403.7	$403.8	$389.2
Net outflows	332.5	335.3	344.4
HQLA in excess of net outflows	$71.2	$68.5	$44.8
Liquidity Coverage Ratio	121%	120%	113%

Observations from the Liquidity Coverage Ratio:

- Citigroup's Liquidity Coverage Ratio has improved in the last two years.
- Citigroup's 2016 LCR indicates it can withstand cash outflows that are 21% higher than its 30-day liquidity needs in a stress scenario or, equivalently, it can withstand a stress level volume of cash outflows for 36.3 days (121% times 30 days). Either way, the LCR indicates adequate liquidity even in the absence of any (likely) remedial management steps in an actual stress event.

The second Basel III liquidity standard is the Net Stable Funding Ratio: a minimum percentage of required stable funding that must be sourced from available stable funding. Required stable funding depends on the composition and maturity of a bank's asset base; available stable funding is a function of the composition and maturity of a bank's funding sources (capital and liabilities). The Net Stable Funding Ratio is a kind of inverted Liquidity Coverage Ratio. Where the Liquidity Coverage Ratio evaluates short-term liquidity, the Net Stable Funding Ratio is a measure of the available stable funding to cover funding of longer-term, less liquid assets, such as loans. Highly liquid assets do not enter the calculation of the Net Stable Funding Ratio. As with the Liquidity Coverage Ratio, a ratio of 100% is the minimum acceptable threshold.

The Net Stable Funding Ratio is not yet a required Basel III standard as of the end of 2016; final rules are expected in 2017. Still, a rough calculation may be made, without the various weightings for components of both available and required stable funding that will be part of the final rules. Exhibit 23 shows one possible calculation of a Net Stable Funding Ratio, based on Citigroup's consolidated balance sheet amounts at 31 December 2016, 30 September 2016, and 31 December 2015. The calculation divides the estimated, unweighted amount of available stable funding by the estimated required amount of stable funding.

Exhibit 23 Citigroup's Net Stable Funding Ratio

(In billions of dollars)	31 Dec. 2016	30 Sep. 2016	31 Dec. 2015
Available stable funding:			
Total deposits	$929.4	$940.3	$907.9
Long-term debt	206.2	209.1	201.3
Common equity	205.9	212.3	205.1
Total available stable funding	$1,341.5	$1,361.6	$1,314.3
Required stable funding:			

Exhibit 23 (Continued)

(In billions of dollars)	31 Dec. 2016	30 Sep. 2016	31 Dec. 2015
Total investments	$353.3	$354.9	$343.0
Total loans, net	612.3	626.0	605.0
Goodwill	21.7	22.5	22.3
Intangible assets (other than MSRs)	5.1	5.4	3.7
Mortgage servicing rights (MSRs)	1.6	1.3	1.8
Other assets	128.0	116.5	133.7
Total required stable funding	$1,122.0	$1,126.6	$1,109.5
Net Stable Funding Ratio	120%	121%	118%

Observations on the approximated Net Stable Funding Ratio:

- Citigroup's Net Stable Funding Ratio, as calculated, has stayed relatively stable since the end of 2015, and the available stable funding is well above the minimum required funding needed.

In summary, Citigroup's liquidity position is very good, based on its Liquidity Coverage and Net Stable Funding Ratios. A rating of 1 is justifiable based on the results of the two ratios.

4.6 Sensitivity to Market Risk

Bank assets and liabilities are constantly subject to market risk, which impacts their earnings performance and liquidity. Analysts need to understand how adverse changes in interest rates, exchange rates, and other market factors can affect a bank's earnings and balance sheet.

Required disclosures in banks' financial statements make it possible to assess various sensitivities. The value at risk disclosure is helpful for assessing a bank's exposure to market factors. VaR statistics can be effective indicators of trends in intra-company risk taking; because of differences in calculation assumptions across companies, VaR is not as useful for assessing risk-taking activities between different companies.

Using a 99% confidence level, Citigroup estimates the value at risk of a potential decline in the value of a position or a portfolio under normal market conditions for an assumed single-day holding period. Citigroup uses a Monte Carlo simulation VaR model to capture material risk sensitivities of various asset classes/risk types. Citigroup's VaR includes positions that are measured at fair value but excludes investment securities classified as AFS or HTM. Exhibit 24 is an excerpt from Citigroup's 2016 VaR disclosure.

Exhibit 24 Citigroup Year-End and Average Trading VaR and Trading and Credit Portfolio VaR

(In millions of dollars)	12/31/16	2016 Average	12/31/15	2015 Average
Interest rate	$37	$35	$37	$44
Credit spread	63	62	56	69
Covariance adjustment [(1)]	(17)	(28)	(25)	(26)

(continued)

Exhibit 24 (Continued)

(In millions of dollars)	12/31/16	2016 Average	12/31/15	2015 Average
Fully diversified interest rate and credit spread	$83	$69	$68	$87
Foreign exchange	32	24	27	34
Equity	13	14	17	17
Commodity	27	21	17	19
Covariance adjustment (1)	(70)	(58)	(53)	(65)
Total trading VaR—all market risk factors, including general and specific risk (excluding credit portfolios) (2)	$85	$70	$76	$92
Specific risk-only component (3)	$3	$7	$11	$6
Total trading VaR—general market risk factors only (excluding credit portfolios) (2)	$82	$63	$65	$86
Incremental impact of the credit portfolio (4)	$20	$22	$22	$25
Total trading and credit portfolio VaR	$105	$92	$98	$117
VaR Effects on Earnings & Capital:				
Total trading and credit portfolio VAR	$105	$92	$98	$117
Net income from continuing operations	$15,033		$17,386	
Common equity	$205,867		$205,139	
Total VaR as % of:				
Net income from continuing operations	0.7%	0.6%	0.6%	0.7%
Common equity	0.1%	0.0%	0.0%	0.1%

Notes:

1 Covariance adjustment reflects the fact that the risks within each and across risk types are not perfectly correlated and, consequently, the total VaR on a given day will be lower than the sum of the VARs relating to each individual risk type.

2 The total trading VaR includes mark-to-market and certain fair value option trading positions except for certain hedges. Available-for-sale and accrual exposures are not included.

3 The specific risk-only component represents the level of equity and fixed income issuer-specific risk embedded in VAR.

4 The credit portfolio is composed of mark-to-market positions associated with non-trading business units.

Observations from the VaR table:

- Citigroup's average trading VaR declined in 2016 to $70 million from $92 million in the previous year, mainly owing to changes in interest rate exposures from mark-to-market hedging activity.
- Average trading and credit portfolio VaR also declined in 2016 to $92 million from $117 million in the previous year.
- Although total trading and credit portfolio VaR increased at year end 2016 to $105 million, compared to $98 million at year end 2015, the magnitude of this worst-case single-day VaR is still less than 1% of net income from continuing operations in both years, on either an end-of-period basis (0.7%) or an average basis (0.6%). The magnitude is even more minor compared to equity, representing 0.1% on the end-of-period basis and less than 0.1% on average.
- Importantly, Citigroup's VaR is a single-day measure of market shocks that can affect a company. Market dislocations can linger for days, weeks, and even longer. Although VaR is useful for measuring the effects of very short-term shocks, it does not address the effects of longer-term market impacts.

Another useful disclosure in Citigroup's 10-K focuses on the estimated sensitivity of Citigroup's capital ratios to numerator changes of $100 million in Common Equity Tier 1 Capital, Tier 1 Capital, and Total Capital and changes of $1 billion in risk-weighted assets at the end of 2016. These sensitivities consider only a single change to either a component of capital or risk-weighted assets; an event affecting more than one factor at a time may have a far greater impact than Citigroup's estimate. Exhibit 25 shows an excerpt of the sensitivity table, along with the actual ratios calculated at the end of 2016.

Exhibit 25 Citigroup Capital Ratio Estimated Sensitivities at 31 December 2016

	Common Equity Tier 1 Capital Ratio		Tier 1 Capital Ratio		Total Capital Ratio	
(In basis points)	Impact of $100 Million Change in Common Equity Tier 1 Capital	Impact of $1 Billion Change in Risk-Weighted Assets	Impact of $100 Million Change in Tier 1 Capital	Impact of $1 Billion Change in Risk-Weighted Assets	Impact of $100 Million Change in Total Capital	Impact of $1 Billion Change in Risk-Weighted Assets
Citigroup						
Advanced Approach	0.90	1.20	0.90	1.30	0.90	1.50
Standardized Approach	0.90	1.30	0.90	1.40	0.90	1.70
Actual capital ratio	14.35%	14.35%	15.29%	15.29%	17.33%	17.33%
Minimum capital ratio	4.50%	4.50%	6.00%	6.00%	8.00%	8.00%

From Citigroup's description of its risk-based capital ratios, p. 33 of 2016 10-K: "Total risk-weighted assets under the Advanced Approaches, which are primarily models based, include credit, market, and operational risk-weighted assets. Conversely, the Standardized Approach excludes operational risk-weighted assets and generally applies prescribed supervisory risk weights to broad categories of credit risk exposures. As a result, credit risk-weighted assets calculated under the Advanced Approaches are more risk sensitive than those calculated under the Standardized Approach. Market risk-weighted assets are derived on a generally consistent basis under both approaches."

Observations from the capital ratio sensitivity table:

- Regardless of the calculation (advanced or standardized approach), the effect of a $100 million change in capital or a $1 billion change in risk-weighted assets is practically nil compared to the actual capital ratios calculated at year end.
- At the same time, these are static measures of sensitivity and adjust for only one impact at a time.

In summary, Citigroup's sensitivity to market impacts appears to be controlled and provides circumstantial evidence of effective risk management. Based on the evidence, Citigroup could be justifiably rated at 1 for its management of sensitivities.

4.7 Overall CAMELS Assessment

After each CAMELS component has been analyzed and rated, the overall CAMELS assessment can be completed. One approach to consolidating CAMELS components on an entity basis would be to simply add all the components' ratings. A bank earning the best CAMELS rating, a rating of 1, for each component would have a total score of 6, and a bank that received the worst ratings would have a composite CAMELS score of 30. To translate the score into the corresponding composite CAMELS rating, the score could be divided by 6. This approach arrives at an arithmetic mean rating as the composite rating for the bank. Note that if each component receives the same rating, the weighting of the components is irrelevant. The arithmetic mean approach, however, fails to take into account the fact that some components of the CAMELS approach are more important to some analysts than others, as discussed in Section 4. Depending on the focus of the analysis, the analyst-weighted composite CAMELS score and rating could be quite different from the unweighted score and arithmetic mean of the ratings.

Exhibit 26 presents the calculation of Citigroup's overall CAMELS score from the point of view of an equity analyst who places twice as much value on asset quality and earnings than on the other CAMELS components.

Exhibit 26 Citigroup Overall CAMELS Score

	Rating	Weighting	Weighted Rating
Capital adequacy	1.0	1	1.00
Asset quality	2.5	2	5.00
Management	2.0	1	2.00
Earnings	3.0	2	6.00
Liquidity	1.0	1	1.00
Sensitivity	1.0	1	1.00
Total score	**10.5**	**8**	**16.00**
Converted to CAMELS rating (score divided by 6)	**1.75**		**2.00**

Note that without the weighting, which helps the analyst quantify his or her priorities, Citigroup has an overall CAMELS rating of 1.75—not perfect, but indicating a bank that is generally showing strong performance and risk management. Once the ratings are weighted, however, the composite score is 2.00 (16/8 = 2.00). The weighted score indicates a slightly higher degree of flaws that management may need to address.

ANALYZING PROPERTY AND CASUALTY INSURANCE COMPANIES

5

f describe key ratios and other factors to consider in analyzing an insurance company.

Insurance companies provide protection against adverse events. Insurance companies earn revenues from **premiums** (amounts paid by the purchaser of insurance products) and from investment income earned on the **float** (amounts collected as premium and not yet paid out as benefits). Insurance companies are typically categorized as property and casualty (P&C) or life and health (L&H). The products of the two types of insurance companies differ in contract duration and variability of claims.[30] P&C insurers' policies are usually short term, and the final cost will usually be known within a year of occurrence of an insured event, whereas L&H insurers' policies are usually longer term. P&C insurers' claims are more variable and "lumpier" because they arise from accidents and other unpredictable events, whereas L&H insurers' claims are more predictable because they correlate closely with relatively stable actuarially based mortality rates when applied to large populations.

For both types of insurance companies, important areas for analysis include business profile, earnings characteristics, investment returns, liquidity, and capitalization. In addition, for P&C companies, analysis of reserves and the combined ratio, an indicator of overall underwriting profitability, are important.

Some countries, including, for example, the United States, require insurance companies to prepare financial reports according to statutory accounting rules, which differ from US GAAP and IFRS, and have a greater focus on solvency.[31] This section discusses analysis based on US GAAP and IFRS financial reports. A discussion of P&C insurers is followed by a discussion of L&H insurers.

5.1 Property and Casualty Insurance Companies

Property and casualty (P&C) insurers provide risk management services to their insured parties. For the price of an insurance premium, they protect the insured parties against losses many times greater than the premiums paid. Premiums are collected at the outset of the insurance contract, creating a float period between their receipt and the time of any payout to the insured party for losses. During the float period, the insurance company will invest the premiums, providing another income stream apart from the underwriting results. In addition to being risk managers, insurance companies also act as investment companies.

Exhibit 27 displays the revenue composition for Travelers Companies, Inc. The net investment income is the second-highest revenue source, after premiums earned, and is significant relative to total revenues.

30 Refer to the Insurance Information Institute's website: www.iii.org.

31 In the United States, the National Association of Insurance Commissioners (NAIC) has developed a system of analytical tools (i.e., ratios and guideline values) for solvency monitoring, known as the NAIC Insurance Regulatory Information System (IRIS). Ratios in IRIS are based on statutory accounting reports.

Exhibit 27 Travelers Companies, Inc., Revenues Composition

(For the year ended 31 December, in millions)	2016		2015		2014	
Premiums	**$24,534**	*88.8%*	$23,874	*89.0%*	$23,713	*87.3%*
Net investment income	**2,302**	*8.3%*	2,379	*8.9%*	2,787	*10.3%*
Fee income	**458**	*1.7%*	460	*1.7%*	450	*1.7%*
Net realized investment gains	**68**	*0.2%*	3	*0.0%*	79	*0.3%*
Other revenues	**263**	*1.0%*	99	*0.4%*	145	*0.5%*
Total revenues	**$27,625**	*100.0%*	$26,815	*100.0%*	$27,174	*100.0%*

Property and casualty insurers try to minimize their payouts to insured parties by exercising care in the underwriting process and charging an adequate price for the risk that they will bear. They may try to diversify the risks they accept by not concentrating excessively on one kind of policy, market, or customer type. They may also diversify their risk by transferring policies, in whole or in part, to reinsurers. Reinsurers deal only with risks insured by other insurers; they do not originate primary policies.

Property and casualty insurance companies differ from life insurance companies in that the length of their duty to perform is comparatively short. Policies are often offered on an annual basis, and the event being covered is often known with certainty during the policy period—fire or weather events, for example. Insured events can also take much longer to emerge: For instance, environmental harm occurring during the policy period may not be obvious until well after the expiration of the policy period.

5.1.1 *Operations: Products and Distribution*

Property insurance policies protect against loss or damage to property—buildings, automobiles, environmental damage, and other tangible objects of value. The events causing loss or damage vary and determine the kind of policy in force. Events may be attributed to accidents, fire, theft, or catastrophe. Casualty insurance, sometimes called *liability insurance*, protects against a legal liability related to an insured event. Casualty insurance covers the liability to a third party, such as passengers, employees, or bystanders. A single insured event may contain both property and casualty losses: For instance, an automobile accident may result in both the loss of the automobile and injury to passengers. Such policies may be referred to as *multiple peril policies*.

Property and casualty insurance may be considered as personal lines or commercial lines, depending on the customer; some products may be sold in both lines. Types of property and casualty insurance include automobile property and liability policies (an example of both personal and commercial lines selling the same product), homeowners' insurance, workers' compensation, marine insurance, and reinsurance.

There are two methods of distributing insurance: direct writing and agency writing. Direct writers of insurance have their own sales and marketing staff. Direct writers also may sell insurance policies via the internet; through direct response channels, such as mail; and through groups with a shared interest or bond, such as membership in a profession. Agency writers use independent agents, exclusive agents, and insurance brokers to sell policies.

5.1.2 *Earnings Characteristics*

In the macro view, the property and casualty insurance business is cyclical. It is a price-sensitive business, with many competitors unafraid to cut prices to obtain market share. According to A.M. Best, a US insurance rating agency, there are approximately

1,200 property and casualty groups in the United States, comprising approximately 2,650 property and casualty companies. Of those groups, the top 150 accounted for approximately 92% of the consolidated industry's total net written premiums in 2015. Once the price cutting drives out profitability, creating a "soft" pricing market for insurance premiums, the insurers reach an uncomfortably depleted level of capital. Competition lessens and underwriting standards tighten, creating a "hard" pricing market. Consequently, premiums rise and the insurers return to more reasonable levels of profitability. The increase in profitability once again attracts more entrants into the market, and the cycle repeats.

In the micro view, there are operating cost considerations that affect insurer profitability apart from the "softness" or "hardness" of the insurance market, depending on the method of distribution. Direct writers have higher fixed costs because of the in-house nature of their distribution method: The sales and marketing staff are salaried employees. Agency writers do not have this fixed cost; instead, the commissions paid to agents and brokers are a variable cost.

The underwriting cycle is driven largely by the expenses of the participants. When the industry's combined ratio—the total insurance expenses divided by the net premiums earned—is low, it indicates a hard insurance market, attracting new entrants who cut prices and push the cycle downward. The effect can be seen in the denominator of the combined ratio: The lower prices for premiums decreases the total net premiums earned, and the combined ratio increases, indicating a soft market. Competitors leave the market, either because they want to forgo unprofitable underwriting or because of their own failure.

For a single insurance company, a combined ratio higher than 100% indicates an underwriting loss. In the United States, Statutory Accounting Practices define the combined ratio as the sum of two ratios, using statutory financial statements: an underwriting loss ratio and an expense ratio. The underwriting loss ratio—losses [= claims paid plus (ending loss reserves minus beginning loss reserves)] divided by net premiums earned—is an indicator of the quality of a company's underwriting activities. Underwriting activities include decisions on whether to accept an application for insurance coverage and decisions on the premiums charged for any coverage extended. The expense ratio (underwriting expenses, including sales commissions and related employee expenses, divided by net premiums written) is an indicator of the efficiency of a company's operations in acquiring and managing underwriting business. For financial disclosures, companies sometimes report modified versions of the combined ratio. For example, the combined ratio reported by Travelers calculates the expense ratio with net earned premiums in the denominator, which is consistent with US GAAP.[32] Other companies may make different presentations.

P&C insurers' investment income is not as volatile as their operating income, because the investments are relatively low-return, low-risk holdings, as we will discuss in the next section.

One critical expense for property and casualty insurers results from the management of their loss reserves. Proper estimation of liabilities is essential to the pricing of policies. Underestimation of loss reserves may lead to undercharging for risks assumed. Development of the loss reserves is based on historical information, yet the process also incorporates estimates about future losses. It is a material account that is subject to management discretion, and its improper estimation can have consequences for the property and casualty insurer. If the loss reserves and the annual adjustments to them are too optimistic, the pricing of the insurance policies may be insufficient for the risk being borne by the insurer and insolvency may ensue. Another problematic attribute of the loss reserves is the fact that the longer the insurer's obligation runs, the

32 The Travelers Companies, Inc., Form 10-K for the year ended 31 December 2016 (p. 36).

more difficult it can be to estimate the loss reserve properly. For example, insurance policies covering asbestos liabilities written long before courts began awarding more generous payouts have been problematic for insurers. Their current experience is far different from what they expected when they issued the policies, and the rapid growth in the award sizes made it difficult to properly estimate the associated loss reserves.

Exhibit 28 shows the roll-forward schedule of activity in Travelers Companies' loss reserve balances, drawn from the insurance claims footnote in its 2016 financial statements. It provides a high-level view of the way the components affect the balance sheet and the income statement and offers insights into the way a property and casualty insurance company manages its assumed risks. The roll-forward activity is denominated in terms of the reserves, net of reinsurance recoverables expected to reduce Travelers' ultimate liability. The beginning and ending balances are shown at their gross amounts, reduced by the reinsurance recoverables to arrive at the net reserves.

Exhibit 28 Travelers Companies, Inc., Loss Reserve Balances and Activity

(At and for the year ended 31 December, in millions)	2016	2015	2014
Gross claims and claim adjustment expense reserves at beginning of year	$48,272	$49,824	$50,865
Less reinsurance recoverables on unpaid losses	(8,449)	(8,788)	(9,280)
Net reserves at beginning of year	39,823	41,036	41,585
Estimated claims and claim adjustment expenses for claims arising in the current year	15,675	14,471	14,688
Estimated decrease in claims and claim adjustment expenses for claims arising in prior years	(680)	(817)	(885)
Total increases	14,995	13,654	13,803
Claims and claim adjustment expense payments for claims arising in:			
Current year	(6,220)	(5,725)	(5,895)
Prior years	(8,576)	(8,749)	(8,171)
Total payments	(14,796)	(14,474)	(14,066)
Acquisition	—	2	—
Unrealized foreign exchange gain	(74)	(395)	(286)
Net reserves at end of year	39,948	39,823	41,036
Plus reinsurance recoverables on unpaid losses	7,981	8,449	8,788
Gross claims and claim adjustment expense reserves at end of year	$47,929	$48,272	$49,824
Reinsurance at end of year:			
Reinsurance recoverables on unpaid losses	$7,981	$8,449	$8,788
Gross claims and claim adjustment expense reserves at end of year	$47,929	$48,272	$49,824
Percentage of claims and claim adjustment expense reserves covered by reinsurance	16.7%	17.5%	17.6%
Revisions' effect on income before income taxes:			
Downward revisions of claims and claim adjustment expenses for claims arising in prior years	$680	$817	$885

Exhibit 28 (Continued)

(At and for the year ended 31 December, in millions)	2016	2015	2014
Income before income taxes	$4,053	$4,740	$5,089
Percentage of contributions of downward revisions to income before income taxes	16.8%	17.2%	17.4%

Observations from Exhibit 28:

- The 2016 claims paid of $6,220 million is 39.7% of the estimated claims and claim adjustments of $15,675 million, indicating that a major part of Travelers Companies' liability exposure is fairly short term. The two prior years show a similar exposure term.
- The company employs significant levels of reinsurance to control its risk exposure. In reinsurance, one insurance company transfers, or cedes, a portion of its risk to another insurer (the "reinsurer") for a premium. The ceding company expects to recover its losses from the reinsurer. As the table shows, Travelers has been ceding between 16.7% and 17.6% of its gross loss reserves to reinsurers.
- The total increases in loss reserves, net of decreases in claims and claim adjustment expenses for prior years' claims, affect the income statement more than any other expense. In 2016, the $14,995 million of total increases in loss reserves represented 63.6% of the $23,572 million of total claims and expenses in the income statement (which is not presented here because of space limitations).
- The company decreased its prior years' estimates of claims by $680 million in 2016, $817 million in 2015, and $885 million in 2014. Downward revisions indicate that a company is estimating its initial recognized reserves conservatively, but aggressive revisions may also be a tool for manipulating earnings. Travelers Companies' downward revisions may appear minor in comparison to the total increases, but they have a profound effect on income before taxes. This effect is shown in the bottom of the exhibit: Downward revisions of prior years' estimates contributed 16.8% to income before income taxes in 2016, 17.2% in 2015, and 17.4% in 2014.

Depending on the ratios used, the ratios of insurers' profitability may distinguish between net premiums written and net premiums earned. Net premiums written are an insurer's direct premiums written, net of any such premiums ceded to other insurers. Premiums are usually billed in advance—for example, twice per year—and they are earned over the period of coverage provided by the insurance policy. Only the net premiums written that are earned over a relevant accounting period—for example, quarterly—are considered to be the net premiums earned.

Useful ratios in analyzing property and casualty insurance companies' profitability include the following:

- *Loss and loss adjustment expense ratio = (Loss expense + Loss adjustment expense)/Net premiums earned.* This ratio indicates the degree of success an underwriter has achieved in estimating the risks insured. The lower the ratio, the greater the success.
- *Underwriting expense ratio = Underwriting expense/Net premiums written.* This ratio measures the efficiency of money spent in obtaining new premiums. A lower ratio indicates higher success.

- *Combined ratio = Loss and loss adjustment expense ratio + Underwriting expense ratio.* This ratio indicates the overall efficiency of an underwriting operation. A combined ratio of less than 100 is considered efficient.
- *Dividends to policyholders (shareholders) ratio = Dividends to policyholders (shareholders)/Net premiums earned.* This ratio is a measure of liquidity, in that it relates the cash outflow of dividends to the premiums earned in the same period.
- *Combined ratio after dividends = Combined ratio + Dividends to policyholders (shareholders) ratio.* This ratio is a stricter measure of efficiency than the ordinary combined ratio, in that it takes into account the cash satisfaction of policyholders or shareholders after consideration of the total underwriting efforts. Dividends are discretionary cash outlays, and factoring them into the combined ratio presents a fuller description of total cash requirements.[33]

Exhibit 29 displays the calculation of these ratios for a group of property and casualty insurers based on their 2016 financial reports. Notice the wide variation in the results. Markel Corp. performed the best (combined ratio of 89%), and Hartford Financial Services Group performed relatively poorly (combined ratio of 131%). The high loss and loss adjustment expense ratio (82.2%) and underwriting expense ratio (48.8%) suggest its underwriting business requires additional management attention. A review of the three ratios related to operations shows that Travelers ranks as the median with respect to loss and loss adjustment expense ratio and below median for underwriting expense and combined ratios. This finding indicates that Travelers' operations are in the better-performing half of this group. After taking into account the dividend distribution policy in the combined ratio after dividends to policyholders (shareholders), Travelers' overall performance remains in the better-performing half of the group.

Exhibit 29 2016 Ratios Calculated for Selected Property and Casualty Insurers

($ millions)	Travelers Companies	Hartford Financial Services Group	W. R. Berkley Corp.	CNA Financial Corp.	Markel Corp.
Loss and loss adjustment expense ratio:					
Loss expense and loss adjustment expense	$15,070	$11,351	$3,846	$5,270	$2,051
Net premiums earned	$24,534	$13,811	$6,293	$6,924	$3,866
Loss and loss adjustment expense ratio	**61.4%**	**82.2%**	**61.1%**	**76.1%**	**53.1%**
Underwriting expense ratio:					
Underwriting expense	$8,139	$5,156	$2,396	$2,787	$1,437
Net premiums written	$24,958	$10,568	$6,424	$6,988	$4,001
Underwriting expense ratio	**32.6%**	**48.8%**	**37.3%**	**39.9%**	**35.9%**
Combined ratio:					
Loss and loss adjustment expense ratio	61.4%	82.2%	61.1%	76.1%	53.0%
Underwriting expense ratio	32.6%	48.8%	37.3%	39.9%	35.9%
Combined ratio	**94.0%**	**131.0%**	**98.4%**	**116.0%**	**89.0%**
Dividends to policyholders (shareholders) ratio:					

33 "Annual Report on the Insurance Industry," Federal Insurance Office, US Department of the Treasury (September 2015), available at www.treasury.gov.

Exhibit 29 (Continued)

($ millions)	Travelers Companies	Hartford Financial Services Group	W. R. Berkley Corp.	CNA Financial Corp.	Markel Corp.
Dividends to policyholders (shareholders)	$757	$334	$184	$813	$0
Net premiums earned	$24,534	$13,811	$6,293	$6,924	$3,866
Dividends to policyholders (shareholders) ratio	**3.1%**	**2.4%**	**2.9%**	**11.7%**	**0.0%**
Combined ratio after dividends:					
Combined ratio	94.0%	131.0%	98.4%	116.0%	89.0%
Dividends to policyholders (shareholders) ratio	3.1%	2.4%	2.9%	11.7%	0.0%
Combined ratio after dividends	**97.1%**	**133.4%**	**101.3%**	**127.7%**	**89.0%**

5.1.3 *Investment Returns*

Property and casualty insurance companies face much uncertainty in the risks they insure, and their business is enormously competitive when insurance pricing moves into its "hard" stage. To counteract the environment of uncertainty, property and casualty insurers conservatively invest the collected premiums. They typically favor steady-return, low-risk assets, while shunning low-liquidity investments.

An illustration is found in Exhibit 30, which is the investment portion of the assets shown in the Travelers Companies' 2016 balance sheet. Investments represent 70% of total assets in 2016 and 2015. In both years, approximately 86% of the total investment portfolio is composed of fixed-maturity investments, and nearly another 7% of investments are short-term securities, which can be considered proxies for cash. Equity securities are only 1% of investments in both years, and real estate is also a very minor component of investments in both years.

Exhibit 30 The Travelers Companies, Inc., Portfolio Composition, 2016 and 2015

At 31 December *($ millions)*	2016		2015	
Fixed maturities, available for sale, at fair value (amortized cost $59,650 and $58,878)	$60,515	*85.9%*	$60,658	*86.1%*
Equity securities, available for sale, at fair value (cost $504 and $528)	732	*1.0%*	705	*1.0%*
Real estate investments	928	*1.3%*	989	*1.4%*
Short-term securities	4,865	*6.9%*	4,671	*6.6%*
Other investments	3,448	*4.9%*	3,447	*4.9%*
Total investments	$70,488	*100.0%*	$70,470	*100.0%*

As with any kind of company, the concentrations of assets merit attention. When considering the investments of a property and casualty insurer, the concentration of investments by type, maturity, credit quality, industry, or geographic location or within single issuers should be evaluated.

Investment performance can be estimated by dividing total investment income by invested assets (cash and investments). This metric can also be calculated on two different bases, by using investment income with and without unrealized capital gains, thus showing the relative importance of unrealized capital gains to the total investment income.

Given that property and casualty insurance companies stand ready to meet obligations for policy payouts, liquidity is a priority in the selection of assets. It will be addressed further in the following section.

5.1.4 *Liquidity*

The uncertainty of the payouts involved in the property and casualty business requires a high degree of liquidity so loss obligations can be met. Because the investments are typically low-risk, steady-return types of financial instruments, their nature is typically liquid. An analysis of the portfolio investments should take into account overall quality of the investments and the ease with which the investments can be converted into cash without affecting their value.

Evidence of the investment liquidity can be found by examining their status in the hierarchy of fair value reporting. Level 1 reported values are based on readily available prices for securities traded in liquid markets and thus indicate the most liquid of securities. Level 2 reported values are based on less liquid conditions: Prices for such securities are not available from a liquid market and may be inferred from similar securities trading in an active market. Thus, these securities are likely to be less liquid than those reported as Level 1 securities. Finally, Level 3 reported values are based on models and assumptions because there is no active market for the securities, implying illiquidity.

Exhibit 31 shows the fair value hierarchy for investment securities held by the Travelers Companies at 31 December 2016.

Exhibit 31 The Travelers Companies, Inc., Portfolio Composition by Fair Value Hierarchy

(at 31 December 2016, in millions)	Total	Level 1	Level 2	Level 3
Fixed maturities:				
US Treasury securities and obligations of US government and government agencies and authorities	$2,035	$2,035	$0	$0
Obligations of states, municipalities, and political subdivisions	31,910	—	31,898	12
Debt securities issued by foreign governments	1,662	—	1,662	—
Mortgage-backed securities, collateralized mortgage obligations, and pass-through securities obligations	1,708	—	1,704	4
All other corporate bonds	23,107	—	22,939	168
Redeemable preferred stock	93	3	90	—
Total fixed maturities	$60,515	$2,038	$58,293	$184
% of security class	*100.0%*	*3.4%*	*96.3%*	*0.3%*
Equity securities:				
Public common stock	$603	$603	$0	$0
Non-redeemable preferred stock	129	51	78	—
Total equity securities	$732	$654	$78	$0
% of security class	*100.0%*	*89.3%*	*10.7%*	*0.0%*

Travelers has very little of its portfolio invested in Level 1 assets—only 4.4% [($2,038 + $654)/($60,515 + $732) = 4.4%] on a combined fixed-income securities and equity securities basis. The majority is classified as Level 2 assets, implying less liquidity than Level 1, yet not implying illiquidity. The fair value footnote from the 10-K provides some assurance that the Level 2 assets are not illiquid (underline added by authors):

> The Company utilized a pricing service to estimate fair value measurements for approximately 98% of its fixed maturities at both December 31, 2016 and 2015. The pricing service utilizes market quotations for fixed maturity securities that have quoted prices in active markets. Since fixed maturities other than US Treasury securities generally do not trade on a daily basis, the pricing service prepares estimates of fair value measurements for these securities using its proprietary pricing applications, which include available relevant market information, benchmark curves, benchmarking of like securities, sector groupings and matrix pricing.
>
> Additionally, the pricing service uses an Option Adjusted Spread model to develop prepayment and interest rate scenarios. The pricing service evaluates each asset class based on relevant market information, relevant credit information, perceived market movements and sector news. The market inputs utilized in the pricing evaluation, listed in the approximate order of priority, include: benchmark yields, reported trades, broker/dealer quotes, issuer spreads, two-sided markets, benchmark securities, bids, offers, reference data, and industry and economic events. The extent of the use of each market input depends on the asset class and the market conditions. Depending on the security, the priority of the use of inputs may change or some market inputs may not be relevant. For some securities, additional inputs may be necessary.

The information does not provide an investor with absolute assurance of constant liquidity for the investments; instead, it provides persuasive evidence that the reported values are fair. The fact that the pricing service considers market information relating to liquidity (reported trades, broker/dealer quotes, issuer spreads, two-sided markets) in developing its price estimates increases an investor's confidence that the recognized values would reflect the prices Travelers might achieve if it liquidated the securities at year end 2016.

5.1.5 *Capitalization*

Unlike the banking sector, where international risk-based capital standards have existed since 1988, as of mid-2016, no such global standard exists for the insurance sector (although the IAIS is in the process of developing a risk-based global insurance capital standard).[34] The standard is expected to include a target minimum capital adequacy ratio. The ratio will be calculated as the amount of qualifying capital divided by the amount of risk-based capital required.

Although no risk-based global insurance capital standard exists, capital standards do exist in various jurisdictions. For example, in Europe, the EU adopted the "Solvency II regime" in 2014, which (among other provisions) establishes minimum capital requirements such that if an insurer falls below the requirements, the supervisory entity in the relevant country will be required to intervene.[35] In the United States, the NAIC risk-based capital requirements, begun in the 1990s, establish a minimum

34 See https://www.iaisweb.org/page/supervisory-material/insurance-capital-standard.

35 See http://europa.eu/rapid/press-release_MEMO-15-3120_en.htm.

amount of capital an insurer must have, based on its size and risk profile.[36] Under the NAIC regime, the formula for minimum risk-based capital for P&C insurers takes into account asset risk, credit risk, underwriting risk, and other relevant risks.

6 ANALYZING LIFE AND HEALTH INSURANCE COMPANIES

f describe key ratios and other factors to consider in analyzing an insurance company.

Insurance companies provide protection against adverse events. Insurance companies earn revenues from **premiums** (amounts paid by the purchaser of insurance products) and from investment income earned on the **float** (amounts collected as premium and not yet paid out as benefits). Insurance companies are typically categorized as property and casualty (P&C) or life and health (L&H). The products of the two types of insurance companies differ in contract duration and variability of claims. P&C insurers' policies are usually short term, and the final cost will usually be known within a year of occurrence of an insured event, whereas L&H insurers' policies are usually longer term. P&C insurers' claims are more variable and "lumpier" because they arise from accidents and other unpredictable events, whereas L&H insurers' claims are more predictable because they correlate closely with relatively stable actuarially based mortality rates when applied to large populations.

For both types of insurance companies, important areas for analysis include business profile, earnings characteristics, investment returns, liquidity, and capitalization. In addition, for P&C companies, analysis of reserves and the combined ratio, an indicator of overall underwriting profitability, are important.

Some countries, including, for example, the United States, require insurance companies to prepare financial reports according to statutory accounting rules, which differ from US GAAP and IFRS, and have a greater focus on solvency. This section discusses analysis based on US GAAP and IFRS financial reports. A discussion of P&C insurers is followed by a discussion of L&H insurers.

6.1 Life and Health Insurance Companies

Life and health insurance companies generate revenue from collecting premiums by selling life and health insurance policies—and for many firms, by providing investment products and services. Investment income is the other primary source of revenues.

6.1.1 *Operations: Products and Distribution*

The types of life insurance products vary widely, with some solely providing a benefit upon the death of the insured and others providing a savings vehicle. In the simplest types of life insurance, a premium is paid for coverage and when the insured dies, the beneficiary receives payment. For example, a term life policy provides a benefit if the insured dies within the fixed term of the contract but expires without value if the insured is still living at the end of the term. In other types of life insurance, the policy both provides a benefit upon the death of the insured and serves as a savings vehicle. Life insurance companies may also offer such investment products as annuities, with fixed payments or variable payments linked to market returns.

36 See www.naic.org/cipr_topics/topic_risk_based_capital.htm.

Health-related insurance products vary primarily by the type of coverage. Some products cover specific medical expenses and treatments, and others provide income payments if the policyholder is injured or becomes ill.

L&H companies sell their products either directly to consumers via electronic media or through agents. The agents may be either employees of the company, exclusive agents, or independent agents. Distribution via independent agents is more expensive for the insurance company but offers the benefits of minimizing fixed costs and increasing flexibility to pursue growth opportunities.[37]

It is helpful to understand the source of a company's revenue and any changes over time. Diversification reduces risks. L&H companies can be diversified across revenue sources, product offerings, geographic coverage, distribution channels, and investment assets.

EXAMPLE 7

Revenue Diversification

Exhibits 32 and 33 present selected income statement information for Aegon N.V. and MetLife, Inc., respectively.

Exhibit 32 Selected Consolidated Income Statement Information: Aegon N.V.

(In EUR millions)	2016	2015	2014	2013	2012
Amounts based upon IFRS					
Premium income	23,453	22,925	19,864	19,939	19,049
Investment income	7,788	8,525	8,148	7,909	8,413
Fees, commissions, other	2,414	2,452	2,145	1,957	1,865
Total revenues	33,655	33,902	30,157	29,805	29,327

Exhibit 33 Selected Income Statement Information: MetLife, Inc.

Years Ended 31 December	2016	2015	2014	2013	2012
(In $ millions)					
Premiums	$39,153	$38,545	$39,067	$37,674	$37,975
Investment income, including derivatives gains	13,358	19,916	22,273	19,154	19,713

(continued)

37 D. Nissim, "Analysis and Valuation of Insurance Companies," Columbia Business School Center for Excellence in Accounting and Security Analysis (November 2010).

Exhibit 33 (Continued)

Years Ended 31 December	2016	2015	2014	2013	2012
Universal life and investment-type product policy fees, and other	10,965	11,490	11,976	11,371	10,462
Total revenues	$63,476	$69,951	$73,316	$68,199	$68,150

Notes: To create comparability in this illustration, the above exhibit combines certain line items from MetLife's income statement. The company's audited financial statements should be used for purposes other than this example.

1 Based on the data for 2016 in Exhibits 32 and 33, compare the companies' diversification across revenue sources.

2 Based on the data in Exhibits 32 and 33, describe the trends in each company's diversification across revenue sources, with specific reference to premium income.

Solutions Exhibit

Aegon N.V. Data in Exhibit 32

As percentage of total revenues	2016	2015	2014	2013	2012
Premium income	69.7%	67.6%	65.9%	66.9%	65.0%
Investment income	23.1%	25.1%	27.0%	26.5%	28.7%
Fees, commissions, other	7.2%	7.2%	7.1%	6.6%	6.4%

YOY percent change	2016	2015	2014	2013
Premium income	2.3%	15.4%	–0.4%	4.7%
Investment income	–8.6%	4.6%	3.0%	–6.0%
Fees, commissions, other	–1.5%	14.3%	9.6%	4.9%

MetLife, Inc., Data in Exhibit 33

As percentage of total revenues	2016	2015	2014	2013	2012
Premiums	61.7%	55.1%	53.3%	55.2%	55.7%
Investment income, including derivatives gains	21.0%	28.5%	30.4%	28.1%	28.9%
Universal life and investment-type product policy fees, and other	17.3%	16.4%	16.3%	16.7%	15.4%
YOY percent change					
Premiums	1.6%	–1.3%	3.7%	–0.8%	

(Continued)

As percentage of total revenues	2016	2015	2014	2013	2012
Investment income, including derivatives gains	−32.9%	−10.6%	16.3%	−2.8%	
Universal life and investment-type product policy fees, and other	−4.6%	−4.1%	5.3%	8.7%	

Solution to 1:

MetLife appears to have greater diversification across revenue sources because it generates only about 62% of total revenues from premiums, compared to nearly 70% for Aegon. It should be noted that premium income can be a more stable source of revenue, and thus greater diversification of revenues should be considered along with potentially greater variability in revenues.

Solution to 2:

For both companies, the percentage of total revenues earned from premiums is greater in 2016 than in any of the previous four years. For Aegon, the increase in the proportion of revenue from premiums resulted in part from significant growth in premium income (15.4% in 2015) as well as a decline in investment income (−8.6%) in 2016. For MetLife, the increase in the proportion of revenue from premiums resulted primarily from the decline in investment income in 2015 and 2016 (−10.6% and −32.9%, respectively).

6.1.2 *Earnings Characteristics*

The major components of L&H insurers' expenses are for benefit payments to policyholders under life insurance, other types of insurance policies, annuity contracts, and other types of contracts. Some types of insurance products that accumulate a cash value include provisions for the policyholder to cancel the contract before its contractual maturity and receive the accumulated cash value. Such early cancellation is known as a contract surrender. Contract surrenders may result in additional expenses for L&H insurers.

Similar to P&C insurers, L&H insures' earnings reflect a number of accounting items that require a significant amount of judgement and estimates. L&H companies must estimate future policyholder benefits and claims based on actuarial assumptions (e.g., about life expectancy). The amounts expensed in a given period are affected by both policyholder benefits actually paid and interest on the estimated liability for future policyholder benefit. As another example of the importance of estimates, L&H companies capitalize the costs of acquiring new and renewal insurance business, which are then amortized on the basis of actual and expected future profits from that business. Another area where accounting judgement can significantly affect L&H companies' earnings—securities valuation—is discussed below in the section on investment returns.

Some general profitability measures can be applied to L&H companies, such as, for example, return on assets (ROA), return on equity (ROE), growth and volatility of capital, and book value per share. Other common profitability measures include pre- and post-tax operating margin (operating profit as a percentage of total revenues) and pre- and post-tax operating return on assets and return on equity.[38] However, most analysis goes beyond these general measures because of the complexity of L&H companies' earnings. Given the possibility of operational distortion and the importance of accounting estimates to L&H companies' reported earnings, a variety of earnings metrics specific to the insurance sector are helpful in providing a good understanding of performance. For example, the profitability ratios used by A.M. Best include (1) total benefits paid as a percentage of net premiums written and deposits and (2) commissions and expenses incurred as a percentage of net premiums written and deposits.[39]

Exhibit 34 shows return on average equity and pretax operating return on average equity for the US L&H sector and MetLife, Inc. In 2011, MetLife had a higher return on average equity than the industry average and a similar pretax operating return on average equity. After 2011, MetLife has not performed as well as the industry on these two measures. Further investigation into causes of the differences between MetLife and the industry and into the reason why the pretax operating return on average equity and return on average equity were similar for MetLife in 2014 and 2015 is needed.

Exhibit 34 Return on Equity—US L&H Sector and MetLife, Inc.

	2011	2012	2013	2014	2015	2016
US L&H Sector Return on Average Equity	4.70%	12.60%	12.90%	11.00%	11.20%	na*
US L&H Sector Pretax Operating Return on Average Equity	9.10%	18.70%	19.10%	14.30%	15.10%	na
MetLife, Inc., Return on Average Equity (Source: 10-K)	12.20%	2.00%	5.40%	9.40%	7.50%	1.00%
MetLife, Inc., Pretax Operating Return on Average Equity (Calculated)	9.00%	9.30%	9.90%	9.80%	7.80%	7.50%

* not available
Source for Sector Data: "Annual Report on the Insurance Industry," Federal Insurance Office (September 2016).

L&H companies' earnings can also be distorted by the accounting treatment of certain items. For example, mismatches between the valuation approach for assets and liabilities can introduce distortion when interest rate changes occur. In some cases, significant distortions to reported earnings have occurred because companies' assets are reported on the basis of current market values whereas liabilities are reported at fixed historical costs, which reflect assumptions in place at the time the liabilities were booked.[40]

38 "Annual Report on the Insurance Industry," Federal Insurance Office, US Department of the Treasury (September 2016), available at www.treasury.gov.

39 A.M. Best is a widely known rating agency for insurance companies. "Best's Credit Rating Methodology: Global Life and Non-Life Insurance Edition" (28 April 2016) is available at www3.ambest.com/ambv/ratingmethodology/openpdf.aspx?ubcr=1&ri=1011.

40 See, for example, Alistair Gray, "MetLife Loss Raises Accounting 'Noise' Concerns," *Financial Times* (16 February 2017).

6.1.3 *Investment Returns*

Investment returns are an important source of income for L&H companies. Key aspects in evaluating L&H companies' investment activities include diversification, investment performance, and interest rate risk. Liquidity of the portfolio is also relevant for L&H companies and is discussed in the following section.

Investment diversification begins with an assessment of allocation across asset classes and an evaluation of how the allocation corresponds to the insurer's liabilities to policyholders. Compared to P&C companies, L&H companies' relative predictability of claims generally allows them to more often seek the higher returns offered by riskier investments. However, higher-yielding assets, such as equity or real estate investments, experience greater fluctuations in valuation than investments in debt. The insurance industry has also faced investment return challenges from the low-interest rate environment of the last 10 years. It has been harder to earn an adequate risk-adjusted return on financial assets because the low interest rates have limited the available opportunities. Overall, asset concentrations by type, maturity, low credit quality, industry, or geographic location or within single issuers can be a concern, particularly to rating agencies.[41]

Investment performance of L&H companies, as with any investment portfolio's performance, can be measured broadly as the amount of investment income divided by the amount of invested assets (cash and investments). The measure can use investment income plus realized gains (losses) with and without unrealized capital gains (losses). In addition, a common metric for evaluating interest rate risk of L&H companies is the comparison of the duration of the company's assets with the duration of its liabilities.

EXAMPLE 8

Investment Portfolio

Exhibit 35 presents information on the investment portfolio of AIA Group. AIA's portfolio of financial investments constitutes 82% of its total assets (and 84% including investment properties).

Exhibit 35 AIA Group Limited Investment Portfolio

	30-Nov-16		30-Nov-15	
	US$m	% Total	US$m	% Total
Loans and deposits	7,062	4.6%	7,211	5.1%
Debt securities	113,618	73.3%	104,640	73.3%
Equity securities	30,211	19.5%	27,159	19.0%
Derivative financial instruments	107	0.1%	73	0.1%
Total financial investments	150,998	97.5%	139,083	97.4%
Investment property	3,910	2.5%	3,659	2.6%
Total	154,908	100.0%	142,742	100.0%

41 Standard & Poor's, "Standard & Poor's Insurance Ratings Criteria: Life Edition" (2004): www.lifecriteria.standardandpoors.com. Note that Standard & Poor's makes ongoing updates to its ratings criteria.

AIA Group Limited Investment Income

	30-Nov-16
Investment Returns	**US$m**
Interest income	$5,290
Dividend income	654
Rental income	140
Investment income	6,084
Gains and losses	1,471
Total investment return	$7,555

Of the $1,471 million in gains and losses, approximately $127 million was related to debt securities.

1 Based on the information in Exhibit 35, describe AIA's investment allocation in 2016 and changes from the prior year.
2 Based on the information in Exhibit 35, estimate the return on average fixed-income assets. (For the purposes of this question, consider loans and deposits and debt securities as a single class of assets—namely, fixed-income assets.)

Solution to 1:

The portfolio, which is mainly invested in debt securities, shows a very small shift from loans and deposits to equity securities in 2016.

Solution to 2:

The return (in $ millions) can be estimated as Investment income on fixed-income securities divided by Average investment in fixed-income securities.

The Investment income on fixed-income securities equals Interest income plus Gains on debt securities = $5,290 + $127 = $5,417.

The average amounts invested in loans and deposits and debt securities was [($7,062 + $113,618) + ($7,211 +$104,640)]/2 = $232,531/2 = $116,265.5.

Therefore, the estimated return on the fixed-income investments was 4.7% (calculated as $5,417/$116,265.5).

6.1.4 *Liquidity*

An L&H company's requirements for liquidity are driven by its liabilities to creditors and, primarily, its liabilities to policyholders, including both benefits and policy surrenders. Historically, liquidity was less important to life insurers because of the long-term nature of traditional life insurance products; however, liquidity has become more important to life insurers as new products have been introduced.[42] An L&H company's sources of liquidity include its operating cash flow and the liquidity of its investment assets. An analysis of liquidity includes a review of the overall liquidity of the investment portfolio. Such investments as non-investment-grade bonds and equity real estate are typically less liquid than investment-grade fixed-income investments.[43]

42 "Insurance Regulatory Information Systems (IRIS) Manual: IRIS Ratios Manual for Property/Casualty, Life/Accident & Health, and Fraternal—2016 Edition," National Association of Insurance Commissioners (2016): www.naic.org/prod_serv/UIR-ZB-16_UIR_2016.pdf.

43 Standard & Poor's Liquidity Model for US and Canadian Life Insurers.

In general, liquidity measures compare the amount of the company's more liquid assets, such as cash and marketable securities, to the amount of its near-term liabilities. Other liquidity measures—for example, the liquidity model used by Standard & Poor's—compare the amount of the company's assets (individually adjusted for assumptions about ready convertibility to cash) with the amount of the company's obligations (individually adjusted for assumptions about potential for withdrawals).[44] The adjusted amounts are calculated under both normal market conditions and stress. The typical "current ratio" is not directly applicable to L&H companies because their balance sheets often do not include the classifications "current" and "non-current."

6.1.5 *Capitalization*

As noted with P&C insurers, L&H companies are not guided by a global risk-based capital standard. Various jurisdictions do, however, have standards specifying the amount of capital an insurer must have based on its risk profile. If an insurer's capital falls below the minimum requirement, generally, a supervisory authority intervenes.

Differences between the P&C and L&H businesses are reflected in differences in the risk-based capital requirement. For example, because L&H claims are considered more predictable than those of P&C insurers, L&H insurers do not need as high an equity cushion and can have lower capital requirements.[45] Another difference between the factors considered in establishing minimum capital requirements for L&H companies is that many life insurance products create material exposure to interest rate risk. Accordingly, the calculation of risk-based capital for an L&H company incorporates interest rate risk.[46]

SUMMARY

- Financial institutions' systemic importance results in heavy regulation of their activities.
- Systemic risk refers to the risk of impairment in some part of the financial system that then has the potential to spread throughout other parts of the financial system and thereby to negatively affect the entire economy.
- The Basel Committee, a standing committee of the Bank for International Settlements, includes representatives from central banks and bank supervisors from around the world.
- The Basel Committee's international regulatory framework for banks includes minimum capital requirements, minimum liquidity requirements, and stable funding requirements.
- Among the international organizations that focus on financial stability are the Financial Stability Board, the International Association of Insurance Supervisors, the International Association of Deposit Insurers, and the International Organization of Securities Commissions.
- Another distinctive feature of financial institutions (compared to manufacturing or merchandising companies) is that their productive assets are predominantly financial assets, such as loans and securities, creating greater direct exposures

44 "Annual Report on the Insurance Industry," Federal Insurance Office, US Department of the Treasury (September 2016).
45 Nissim, "Analysis and Valuation of Insurance Companies."
46 See www.naic.org/cipr_topics/topic_risk_based_capital.htm.

to a variety of risks, such as credit risk, liquidity risk, market risk, and interest rate risk. In general, the values of their assets are relatively close to fair market values.

- A widely used approach to analyzing a bank, CAMELS, considers a bank's **C**apital adequacy, **A**sset quality, **M**anagement capabilities, **E**arnings sufficiency, **L**iquidity position, and **S**ensitivity to market risk.
- "**C**apital adequacy," described in terms of the proportion of the bank's assets that is funded with capital, indicates that a bank has enough capital to absorb potential losses without severely damaging its financial position.
- "**A**sset quality" includes the concept of quality of the bank's assets—credit quality and diversification—and the concept of overall sound risk management.
- "**M**anagement capabilities" refers to the bank management's ability to identify and exploit appropriate business opportunities and to simultaneously manage associated risks.
- "**E**arnings" refers to the bank's return on capital relative to cost of capital and also includes the concept of earnings quality.
- "**L**iquidity" refers to the amount of liquid assets held by the bank relative to its near-term expected cash flows. Under Basel III, liquidity also refers to the stability of the bank's funding sources.
- "**S**ensitivity to market risk" pertains to how adverse changes in markets (including interest rate, exchange rate, equity, and commodity markets) could affect the bank's earnings and capital position.
- In addition to the CAMELS components, important attributes deserving analysts' attention include government support, the banking entity's mission, corporate culture and competitive environment, off-balance-sheet items, segment information, currency exposure, and risk disclosures.
- Insurance companies are typically categorized as property and casualty (P&C) or life and health (L&H).
- Insurance companies earn revenues from premiums (amounts paid by the purchaser of insurance products) and from investment income earned on the float (amounts collected as premiums and not yet paid out as benefits).
- P&C insurers' policies are usually short term, and the final cost will usually be known within a year of a covered event, whereas L&H insurers' policies are usually longer term. P&C insurers' claims are more variable, whereas L&H insurers' claims are more predictable.
- For both types of insurance companies, important areas for analysis include business profile, earnings characteristics, investment returns, liquidity, and capitalization. In addition, analysis of P&C companies' profitability includes analysis of loss reserves and the combined ratio.

PRACTICE PROBLEMS

The following information relates to questions 1–7

Viktoria Smith is a recently hired junior analyst at Aries Investments. Smith and her supervisor, Ingrid Johansson, meet to discuss some of the firm's investments in banks and insurance companies.

Johansson asks Smith to explain why the evaluation of banks is different from the evaluation of non-financial companies. Smith tells Johansson the following:

Statement 1 As intermediaries, banks are more likely to be systemically important than non-financial companies.

Statement 2 The assets of banks mostly consist of deposits, which are exposed to different risks than the tangible assets of non-financial companies.

Smith and Johansson also discuss key aspects of financial regulations, particularly the framework of Basel III. Johansson tells Smith:

> "Basel III specifies the minimum percentage of its risk-weighted assets that a bank must fund with equity. This requirement of Basel III prevents a bank from assuming so much financial leverage that it is unable to withstand loan losses or asset write-downs."

Johansson tells Smith that she uses the CAMELS approach to evaluate banks, even though it has some limitations. To evaluate P&C insurance companies, Johansson tells Smith that she places emphasis on the efficiency of spending on obtaining new premiums. Johansson and Smith discuss differences between P&C and L&H insurance companies. Smith notes the following differences:

Difference 1: L&H insurers' claims are more predictable than P&C insurers' claims.

Difference 2: P&C insurers' policies are usually short term, whereas L&H insurers' policies are usually longer term.

Difference 3: Relative to L&H insurers, P&C insurers often have lower capital requirements and can also seek higher returns offered by riskier investments.

Johansson asks Smith to review key performance ratios for three P&C insurers in which Aries is invested. The ratios are presented in Exhibit 1.

Exhibit 1 Key Performance Ratios for Selected P&C Insurers

	Insurer A	Insurer B	Insurer C
Loss and loss adjustment expense ratio	68.8%	65.9%	64.1%
Underwriting expense ratio	33.7%	37.8%	32.9%
Combined ratio	102.5%	103.7%	97.0%

© 2019 CFA Institute. All rights reserved.

Johansson also asks Smith to review key performance ratios for ABC Bank, a bank in which Aries is invested. The ratios are presented in Exhibit 2.

Exhibit 2 Key Performance Ratios for ABC Bank*

	2017	2016	2015
Common equity Tier 1 capital ratio	10.7%	11.5%	12.1%
Tier 1 capital ratio	11.5%	12.6%	13.4%
Total capital ratio	14.9%	14.8%	14.9%
Liquidity coverage ratio	123.6%	121.4%	119.1%
Net stable funding ratio	114.9%	113.2%	112.7%
Total trading VaR (all market risk factors)	$11	$13	$15
Total trading and credit portfolio VaR	$15	$18	$21

* *Note*: VaR amounts are in millions and are based on a 99% confidence interval and a single-day holding period.

1 Which of Smith's statements regarding banks is correct?
   A Only Statement 1
   B Only Statement 2
   C Both Statement 1 and Statement 2

2 The aspect of the Basel III framework that Johansson describes to Smith relates to minimum:
   A capital requirements.
   B liquidity requirements.
   C amounts of stable funding requirements.

3 One limitation of the approach used by Johansson to evaluate banks is that it fails to address a bank's:
   A sensitivity to market risk.
   B management capabilities.
   C competitive environment.

4 The best indicator of the operations of a P&C insurance company emphasized by Johansson when evaluating P&C insurance companies is the:
   A combined ratio.
   B underwriting loss ratio.
   C underwriting expense ratio.

5 Which of the differences between P&C insurers and L&H insurers noted by Smith is *incorrect*?
   A Difference 1
   B Difference 2
   C Difference 3

6 Based on Exhibit 1, Smith should conclude that the insurer with the most efficient underwriting operation is:
   A Insurer A.
   B Insurer B.
   C Insurer C.

7 Based on Exhibit 2, Smith and Johansson should conclude that over the past three years, ABC Bank's:

A liquidity position has declined.

B capital adequacy has improved.

C sensitivity to market risk has improved.

The following information relates to questions 8–14

Ivan Paulinic, an analyst at a large wealth management firm, meets with his supervisor to discuss adding financial institution equity securities to client portfolios. Paulinic focuses on Vermillion Insurance (Vermillion), a property and casualty company, and Cobalt Life Insurance (Cobalt). To evaluate Vermillion further, Paulinic compiles the information presented in Exhibit 1.

Exhibit 1 Select Financial Ratios for Vermillion Insurance

Ratio	2017	2016
Loss and loss adjustment expense	59.1%	61.3%
Underwriting expense	36.3%	35.8%
Combined	95.4%	97.1%
Dividend	2.8%	2.6%

In addition to the insurance companies, Paulinic gathers data on three national banks that meet initial selection criteria but require further review. This information is shown in Exhibits 2, 3, and 4.

Exhibit 2 Select Balance Sheet Data for National Banks—Trading: Contribution to Total Revenues

Bank	2017	2013	2009	2005
N-bank	4.2%	7.0%	10.1%	8.9%
R-bank	8.3%	9.1%	17.0%	7.9%
T-bank	5.0%	5.0%	11.9%	6.8%

Focusing on N-bank and T-bank, Paulinic prepares the following data.

Exhibit 3 2017 Select Data for N-bank and T-bank

	N-bank		T-bank	
	2017	2016	2017	2016
Average daily trading VaR ($ millions)	11.3	12.6	21.4	20.5
Annual trading revenue/average daily trading VaR	160×	134×	80×	80×

Paulinic investigates R-bank's risk management practices with respect to the use of credit derivatives to enhance earnings, following the 2008 financial crisis. Exhibit 4 displays R-bank's exposure over the last decade to credit derivatives not classified as hedges.

Exhibit 4 R-bank's Exposure to Freestanding Credit Derivatives

Credit Derivative Balances	2017	2012	2007
Notional amount ($ billions)	13.4	15.5	305.1

All of the national banks under consideration primarily make long-term loans and source a significant portion of their funding from retail deposits. Paulinic and the rest of the research team note that the central bank is unwinding a long period of monetary easing as evidenced by two recent increases in the overnight funding rate. Paulinic informs his supervisor that:

Statement 1 Given the recently reported stronger-than-anticipated macro-economic data, there is an imminent risk that the yield curve will invert.

Statement 2 N-bank is very active in the 30-day reverse repurchase agreement market during times when the bank experiences significant increases in retail deposits.

8 Paulinic's analysis of the two insurance companies *most likely* indicates that:

- **A** Cobalt has more-predictable claims than Vermillion.
- **B** Cobalt has a higher capital requirement than Vermillion.
- **C** Vermillion's calculated risk-based capital is more sensitive than Cobalt's to interest rate risk.

9 Based only on the information in Exhibit 1, in 2017 Vermillion *most likely*:

- **A** experienced a decrease in overall efficiency.
- **B** improved its ability to estimate insured risks.
- **C** was more efficient in obtaining new premiums.

10 Based only on Exhibit 2, which of the following statements is correct?

- **A** The quality of earnings for R-bank was the highest in 2009.
- **B** Relative to the other banks, N-bank has the highest quality of earnings in 2017.
- **C** Trading represented a sustainable revenue source for T-bank between 2005 and 2013.

11 Based only on Exhibit 3, Paulinic should conclude that:

- **A** trading activities are riskier at T-bank than N-bank.
- **B** trading revenue per unit of risk has improved more at N-bank than T-bank.
- **C** compared with duration, the metric used is a better measure of interest rate risk.

12 Based only on Exhibit 4, R-bank's use of credit derivatives since 2007 *most likely*:

- **A** increased posted collateral.
- **B** decreased the volatility of earnings from trading activities.
- **C** indicates consistent correlations among the relevant risks taken.

13 Based on Statement 1, the net interest margin for the three banks' *most likely* will:

- **A** decrease.
- **B** remain unchanged.
- **C** increase.

14 Based on Statement 2, the financial ratio *most* directly affected is the:

- **A** Tier 2 capital ratio.
- **B** net stable funding ratio.
- **C** liquidity coverage ratio.

The following information relates to questions 15–20

Judith Yoo is a financial sector analyst writing an industry report. In the report, Yoo discusses the relative global systemic risk across industries, referencing Industry A (international property and casualty insurance), Industry B (credit unions), and Industry C (global commercial banks).

Part of Yoo's analysis focuses on Company XYZ, a global commercial bank, and its CAMELS rating, risk management practices, and performance. First, Yoo considers the firm's capital adequacy as measured by the key capital ratios (common equity Tier 1 capital, total Tier 1 capital, and total capital) in Exhibit 1.

Exhibit 1 Company XYZ: Excerpt from Annual Report Disclosure

At 31 December	2017	2016	2015
Regulatory capital	**$m**	**$m**	**$m**
Common equity Tier 1 capital	146,424	142,367	137,100
Additional Tier 1 capital	22,639	20,443	17,600
Tier 2 capital	22,456	27,564	38,200
Total regulatory capital	191,519	190,374	192,900
Risk-weighted assets (RWAs) by risk type			
Credit risk	960,763	989,639	968,600

(continued)

Exhibit 1 (Continued)

At 31 December	2017	2016	2015
Regulatory capital	**$m**	**$m**	**$m**
Market risk	44,100	36,910	49,600
Operational risk	293,825	256,300	224,300
Total RWAs	1,298,688	1,282,849	1,242,500

Yoo turns her attention to Company XYZ's asset quality using the information in Exhibit 2.

Exhibit 2 Company XYZ: Asset Composition

At 31 December	2017	2016	2015
	$m	**$m**	**$m**
Total liquid assets	361,164	354,056	356,255
Investments	434,256	367,158	332,461
Consumer loans	456,957	450,576	447,493
Commercial loans	499,647	452,983	403,058
Goodwill	26,693	26,529	25,705
Other assets	151,737	144,210	121,780
Total assets	1,930,454	1,795,512	1,686,752

To assess Company XYZ's risk management practices, Yoo reviews the consumer loan credit quality profile in Exhibit 3 and the loan loss analysis in Exhibit 4.

Exhibit 3 Company XYZ: Consumer Loan Profile by Credit Quality

At 31 December	2017	2016	2015
	$m	**$m**	**$m**
Strong credit quality	338,948	327,345	320,340
Good credit quality	52,649	54,515	54,050
Satisfactory credit quality	51,124	55,311	56,409
Substandard credit quality	23,696	24,893	27,525
Past due but not impaired	2,823	2,314	2,058
Impaired	8,804	9,345	10,235
Total gross amount	478,044	473,723	470,617
Impairment allowances	–5,500	–4,500	–4,000
Total	472,544	469,223	466,617

Exhibit 4 Company XYZ: Loan Loss Analysis Data			
At 31 December	2017	2016	2015
	$m	$m	$m
Consumer loans			
Allowance for loan losses	11,000	11,500	13,000
Provision for loan losses	3,000	2,000	1,300
Charge-offs	3,759	3,643	4,007
Recoveries	1,299	1,138	1,106
Net charge-offs	2,460	2,505	2,901
Commercial loans			
Allowance for loan losses	1,540	1,012	169
Provision for loan losses	1,100	442	95
Charge-offs	1,488	811	717
Recoveries	428	424	673
Net charge-offs	1,060	387	44

Finally, Yoo notes the following supplementary information from Company XYZ's annual report:

- Competition in the commercial loan space has become increasingly fierce, leading XYZ managers to pursue higher-risk strategies to increase market share.
- The net benefit plan obligation has steadily decreased during the last three years.
- Company XYZ awards above-average equity-based compensation to its top managers.

15 Which of the following industries *most likely* has the highest level of global systemic risk?

A Industry A

B Industry B

C Industry C

16 Based on Exhibit 1, Company XYZ's capital adequacy over the last three years, as measured by the three key capital ratios, signals conditions that are:

A mixed.

B declining.

C improving.

17 Based only on Exhibit 2, asset composition from 2015 to 2017 indicates:

A declining liquidity.

B increasing risk based on the proportion of total loans to total assets.

C decreasing risk based on the proportion of investments to total assets.

18 Based on Exhibit 3, the trend in impairment allowances is reflective of the changes in:

A impaired assets.

B strong credit quality assets.

C past due but not impaired assets.

19 Based on Exhibit 4, a loan loss analysis for the last three years indicates that:

A Company XYZ has become less conservative in its provisioning for consumer loans.

B the provision for commercial loan losses has trailed the actual net charge-off experience.

C the cushion between the allowance and the net commercial loan charge-offs has declined.

20 Which of the following supplemental factors is consistent with a favorable assessment of Company XYZ's financial outlook?

A Competitive environment

B Net benefit plan obligation

C Equity-based compensation policy

SOLUTIONS

1 A is correct. Banks are more likely to be systemically important than non-financial companies because, as intermediaries, they create financial linkages across all types of entities, including households, banks, corporates, and governments. The network of linkages across entities means that the failure of one bank will negatively affect other financial and non-financial entities (a phenomenon known as financial contagion). The larger the bank and the more widespread its network of linkages, the greater its potential impact on the entire financial system. The assets of banks are predominantly financial assets, such as loans and securities (not deposits, which represent most of a bank's liabilities). Compared to the tangible assets of non-financial companies, financial assets create direct exposure to a different set of risks, including credit risks, liquidity risks, market risks, and interest rate risks.

2 A is correct. Basel III specifies the minimum percentage of its risk-weighted assets that a bank must fund with equity capital. This minimum funding requirement prevents a bank from assuming so much financial leverage that it is unable to withstand loan losses or asset write-downs.

3 C is correct. The approach used by Johansson to evaluate banks, the CAMELS approach, has six components: (1) capital adequacy, (2) asset quality, (3) management capabilities, (4) earnings sufficiency, (5) liquidity position, and (6) sensitivity to market risk. While the CAMELS approach to evaluating a bank is fairly comprehensive, some attributes of a bank are not addressed by this method. One such attribute is a bank's competitive environment. A bank's competitive position relative to its peers may affect how it allocates capital and assesses risks.

4 C is correct. The underwriting expense ratio is an indicator of the efficiency of money spent on obtaining new premiums. The underwriting loss ratio is an indicator of the quality of a company's underwriting activities—the degree of success an underwriter has achieved in estimating the risks insured. The combined ratio, a measure of the overall underwriting profitability and efficiency of an underwriting operation, is the sum of these two ratios.

5 C is correct. The products of the two types of insurance companies, P&C and L&H, differ in contract duration and claim variability. P&C insurers' policies are usually short term, and the final cost will usually be known within a year of the occurrence of an insured event, while L&H insurers' policies are usually longer term. P&C insurers' claims are more variable and "lumpier" because they arise from accidents and other less predictable events, while L&H insurers' claims are more predictable because they correlate closely with relatively stable, actuarially based mortality rates applied to large populations. The relative predictability of L&H insurers' claims generally allows these companies to have lower capital requirements and to seek higher returns than P&C insurers.

6 C is correct. The combined ratio, which is the sum of the underwriting expense ratio and the loss and loss adjustment expense ratio, is a measure of the efficiency of an underwriting operation. A combined ratio of less than 100% is considered efficient; a combined ratio greater than 100% indicates an underwriting loss. Insurer C is the only insurer that has a combined ratio less than 100%.

7 C is correct. Over the past three years, there has been a downward trend in the two VaR measures—total trading VaR (all market risk factors) and total trading and credit portfolio VaR. This trend indicates an improvement in ABC Bank's sensitivity, or a reduction in its exposure, to market risk. The two liquidity

measures—the liquidity coverage ratio and the net stable funding ratio—have increased over the past three years, indicating an improvement in ABC Bank's liquidity position. Trends in the three capital adequacy measures—common equity Tier 1 capital ratio, Tier 1 capital ratio, and total capital ratio—indicate a decline in ABC Bank's capital adequacy. While the total capital ratio has remained fairly constant over the past three years, the common equity Tier 1 capital ratio and the Tier 1 capital ratio have declined. This trend suggests that ABC Bank has moved toward using more Tier 2 capital and less Tier 1 capital, indicating an overall decline in capital adequacy.

8 A is correct. Claims associated with life and health insurance companies (Cobalt) are more predicable than those for property and casualty insurance companies (Vermillion). Property and casualty insurers' claims are more variable and "lumpier" because they arise from accidents and other unpredictable events, whereas life and health insurers' claims are more predictable because they correlate closely with relatively stable actuarially based mortality rates when applied to large populations.

9 B is correct. The loss and loss adjustment expense ratio decreased from 61.3% to 59.1% between 2016 and 2017. This ratio is calculated as follows: (Loss Expense + Loss Adjustment Expense)/Net Premiums Earned. The loss and loss adjustment expense ratio indicates the degree of success an underwriter has achieved in estimating the risks insured. A lower ratio indicates greater success in estimating insured risks.

10 B is correct. The quality of earnings is directly related to the level of sustainable sources of income. Trading income tends to be volatile and not necessarily sustainable. Higher-quality income would be net interest income and fee-based service income. Because N-bank's 2017 trading revenue contribution is the lowest relative to other banks, its quality of earnings would be considered the best of the three banks.

11 B is correct. Trading revenue per unit of risk can be represented by the ratio of annual trading revenue to average daily trading value at risk (VaR) and represents a measure of reward-to-risk. The trading revenue per unit of risk improved at N-bank (from 134× to 160×) between 2016 and 2017, and there was no change at T-bank (80×). VaR can be used for gauging trends in intra-company risk taking.

12 B is correct. Exhibit 4 indicates that exposure to free-standing credit derivatives dramatically declined from a peak during the global financial crisis in 2008. If a derivatives contract is classified as freestanding, changes in its fair value are reported as income or expense in the income statement at each reporting period. The immediate recognition of a gain or loss in earnings, instead of reporting it in other comprehensive income, can lead to unexpected volatility of earnings and missed earnings targets. As a result, earnings volatility from the use of credit derivatives most likely decreased.

13 A is correct. A bank's net interest revenue represents the difference between interest earned on loans and other interest-bearing assets and the level of interest paid on deposits and other interest-bearing liabilities. Banks typically borrow money for shorter terms (retail deposits) and lend to customers for longer periods (mortgages and car loans). If the yield curve unexpectedly inverts, the short-term funding costs will increase and the net interest margin will most likely decrease (not remain unchanged or increase).

14 C is correct. Reverse repurchase agreements represent collateralized loans between a bank and a borrower. A reverse repo with a 30-day maturity is a highly liquid asset and thus would directly affect the liquidity coverage ratio (LCR). LCR evaluates short-term liquidity and represents the percentage of a bank's expected cash outflows in relation to highly liquid assets.

15 C is correct. Industry C, representing global commercial banks, most likely has the highest level of global systemic risk because global commercial banks have the highest proportion of cross-border business. Unlike banks, the overall insurance market (of which Industry A is a subset) has a smaller proportion of cross-border business, and insurance companies' foreign branches are generally required to hold assets in a jurisdiction that are adequate to cover the related policy liabilities in that jurisdiction. As an international property and casualty (P&C) insurer, Company A provides protection against adverse events related to autos, homes, or commercial activities; many of these events have local, rather than international, impact. Industry B, credit unions, most likely has the lowest level of global systemic risk. Credit unions are depository institutions that function like banks and offer many of the same services, but they are owned by their members rather than being publicly traded as many banks are.

16 A is correct. Company XYZ's key capital adequacy ratios show mixed conditions. The ratios are calculated as follows:

$$\text{Common Equity Tier 1 Capital Ratio} = \frac{\text{Total Common Equity Tier 1 Capital}}{\text{Total Risk-Weighted Assets}}$$

$$\text{2015 Common Equity Tier 1 Capital Ratio} = \frac{137{,}100}{1{,}242{,}500} = 11.0\%$$

$$\text{2016 Common Equity Tier 1 Capital Ratio} = \frac{142{,}367}{1{,}282{,}849} = 11.1\%$$

$$\text{2017 Common Equity Tier 1 Capital Ratio} = \frac{146{,}424}{1{,}298{,}688} = 11.3\%$$

$$\text{Tier 1 Ratio} = \frac{\text{Common Equity Tier 1 Capital} + \text{Additional Tier 1 Capital}}{\text{Total Risk-Weighted Assets}}$$

$$\text{2015 Tier 1 Ratio} = \frac{137{,}100 + 17{,}600}{1{,}242{,}500} = 12.5\%$$

$$\text{2016 Tier 1 Ratio} = \frac{142{,}367 + 20{,}443}{1{,}282{,}849} = 12.7\%$$

$$\text{2017 Tier 1 Ratio} = \frac{146{,}424 + 22{,}639}{1{,}298{,}688} = 13.0\%$$

$$\text{Total Capital Ratio} = \frac{\text{Total Capital}}{\text{Total Risk-Weighted Assets}}$$

$$\text{2015 Total Capital Ratio} = \frac{192{,}900}{1{,}242{,}500} = 15.5\%$$

$$\text{2016 Total Capital Ratio} = \frac{190{,}374}{1{,}282{,}849} = 14.8\%$$

$$2017 \text{ Total Capital Ratio} = \frac{191{,}519}{1{,}298{,}688} = 14.7\%$$

	2017	2016	2015
Common equity Tier 1 capital ratio	11.3%	11.1%	11.0%
Tier 1 capital ratio	13.0%	12.7%	12.5%
Total capital ratio	14.7%	14.8%	15.5%

The common equity Tier 1 capital ratio and the Tier 1 capital ratio both strengthened from 2015 to 2017, but the total capital ratio weakened during that same period, signaling mixed conditions.

17 A is correct. Company XYZ's liquid assets as a percentage of total assets declined each year since 2015, indicating declining liquidity.

	2017		2016		2015	
	$m	% of Total Assets	$m	% of Total Assets	$m	% of Total Assets
Total liquid assets	361,164	18.7%	354,056	19.7%	356,255	21.1%
Investments	434,256	22.5%	367,158	20.4%	332,461	19.7%
Loans						
Consumer loans	456,957		450,576		447,493	
Commercial loans	499,647		452,983		403,058	
Total loans	956,604	49.6%	903,559	50.3%	850,551	50.4%
Goodwill	26,693	1.4%	26,529	1.5%	25,705	1.5%
Other assets	151,737	7.9%	144,210	8.0%	121,780	7.2%
Total assets	1,930,454	100%	1,795,512	100%	1,686,752	100%

18 C is correct. Impairment allowances have increased proportionately to the increases in the amount of past due but not impaired assets, which may be in anticipation of these past due assets becoming impaired. Impaired assets have decreased each year while strong credit quality assets have increased each year, which suggests lowering impairment allowances as a result of improving credit quality of these financial instruments.

At 31 December	2017	2016	2015
	$m	$m	$m
Strong credit quality	338,948	327,345	320,340
Good credit quality	52,649	54,515	54,050
Satisfactory credit quality	51,124	55,311	56,409
Substandard credit quality	23,696	24,893	27,525
Past due but not impaired	2,823	2,314	2,058
Impaired	8,804	9,345	10,235
Total gross amount	478,044	473,723	470,617
Impairment allowances	–5,500	–4,500	–4,000
Total	472,544	469,223	466,617
YoY change in impaired assets	–5.8%	–8.7%	
YoY change in strong credit quality assets	3.5%	2.2%	

At 31 December	2017	2016	2015
	$m	$m	$m
YoY change in past due but not impaired assets	22.0%	12.4%	
YoY change in impairment allowances	22.2%	12.5%	

Note: YoY = year-over-year

2015 to 2016 change in impaired assets: $\left(\frac{9,345}{10,235}\right) - 1 = -8.7\%$

2015 to 2016 change in strong credit quality assets: $\left(\frac{327,345}{320,340}\right) - 1 = 2.2\%$

2015 to 2016 change in past due but not impaired assets: $\left(\frac{2,314}{2,058}\right) - 1 = 12.4\%$

2015 to 2016 change in impairment allowances: $\left(\frac{-4,500}{-4,000}\right) - 1 = 12.5\%$

2016 to 2017 change in impaired assets: $\left(\frac{8,804}{9,345}\right) - 1 = -5.8\%$

2016 to 2017 change in strong credit quality assets: $\left(\frac{338,948}{327,345}\right) - 1 = 3.5\%$

2016 to 2017 change in past due but not impaired assets: $\left(\frac{2,823}{2.314}\right) - 1 = 22.0\%$

2016 to 2017 change in impairment allowances: $\left(\frac{-5,500}{-4,500}\right) - 1 = 22.2\%$

19 C is correct. The allowance for loan losses to net commercial loan charge-offs has been declining during the last three years, which indicates that the cushion between the allowance and the net commercial loan charge-offs has deteriorated.

2015 Consumer: $\frac{\text{Allowance for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{13,000}{2,901} = 4.48$

2016 Consumer: $\frac{\text{Allowance for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{11,500}{2,505} = 4.59$

2017 Consumer: $\frac{\text{Allowance for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{11,000}{2,460} = 4.47$

2015 Commercial: $\frac{\text{Allowance for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{169}{44} = 3.84$

$$\text{2016 Commercial: } \frac{\text{Allowance for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{1{,}012}{387} = 2.61$$

$$\text{2017 Commercial: } \frac{\text{Allowance for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{1{,}540}{1{,}060} = 1.45$$

$$\text{2015 Consumer: } \frac{\text{Provision for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{1{,}300}{2{,}901} = 0.45$$

$$\text{2016 Consumer: } \frac{\text{Provision for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{2{,}000}{2{,}505} = 0.80$$

$$\text{2017 Consumer: } \frac{\text{Provision for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{3{,}000}{2{,}460} = 1.22$$

$$\text{2015 Commercial: } \frac{\text{Provision for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{95}{44} = 2.16$$

$$\text{2016 Commercial: } \frac{\text{Provision for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{442}{387} = 1.14$$

$$\text{2017 Commercial: } \frac{\text{Provision for Loan Losses}}{\text{Net Loan Charge-Offs}} = \frac{1{,}100}{1{,}060} = 1.04$$

	2017	2016	2015
	$m	$m	$m
Consumer loans			
Allowance for loan losses	11,000	11,500	13,000
Provision for loan losses	3,000	2,000	1,300
Charge-offs	3,759	3,643	4,007
Recoveries	1,299	1,138	1,106
Net charge-offs	2,460	2,505	2,901
Commercial loans			
Allowance for loan losses	1,540	1,012	169
Provision for loan losses	1,100	442	95
Charge-offs	1,488	811	717
Recoveries	428	424	673
Net charge-offs	1,060	387	44
Allowance for loan losses to net loan charge-offs: consumer	4.47	4.59	4.48
Allowance for loan losses to net loan charge-offs: commercial	1.45	2.61	3.84

	2017	2016	2015
	$m	$m	$m
Provision for loan losses to net loan charge-offs: consumer	1.22	0.80	0.45
Provision for loan losses to net loan charge-offs: commercial	1.04	1.14	2.16

20 B is correct. The net benefit plan obligation has steadily decreased during the last three years, which indicates a lower degree of risk posed by the benefit plan.

FINANCIAL STATEMENT ANALYSIS

STUDY SESSION

5

Financial Statement Analysis (2)

This study session focuses on evaluating financial reporting quality and applying financial analysis techniques to investment decisions. A conceptual framework for assessing the quality of a company's financial reports, including the quality of earnings, is provided. Indicators of low-quality reporting, including quality of earnings, cash flow, and balance sheet are examined. The session concludes with mini cases, which demonstrate the value in applying financial statement analysis to inform practical investment decisions.

READING ASSIGNMENTS

Reading 13	Evaluating Quality of Financial Reports by Jack T. Ciesielski, CPA, CFA, Elaine Henry, PhD, CFA, and Thomas I. Selling, PhD, CPA
Reading 14	Integration of Financial Statement Analysis Techniques by Jack T. Ciesielski, CPA, CFA

Note: Changes in accounting standards as well as new rulings and/or pronouncements issued after the publication of the readings on financial reporting and analysis may cause some of the information in these readings to become dated. Candidates are *not* responsible for anything that occurs after the readings were published. In addition, candidates are expected to be familiar with the analytical frameworks contained in the readings, as well as the implications of alternative accounting methods for financial analysis and valuation discussed in the readings. Candidates are also responsible for the content of accounting standards, but not for the actual reference numbers. Finally, candidates should be aware that certain ratios may be defined and calculated differently. When alternative ratio definitions exist and no specific definition is given, candidates should use the ratio definitions emphasized in the readings.

© 2021 CFA Institute. All rights reserved.

READING

13

Evaluating Quality of Financial Reports

by Jack T. Ciesielski, CPA, CFA, Elaine Henry, PhD, CFA, and Thomas I. Selling, PhD, CPA

Jack T. Ciesielski, CPA, CFA, is at R.G. Associates, Inc., former publisher of The Analyst's Accounting Observer (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA). Thomas I. Selling, PhD, CPA, is at the Cox School of Business, Southern Methodist University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** demonstrate the use of a conceptual framework for assessing the quality of a company's financial reports;
☐	**b.** explain potential problems that affect the quality of financial reports;
☐	**c.** describe how to evaluate the quality of a company's financial reports;
☐	**d.** evaluate the quality of a company's financial reports;
☐	**e.** describe the concept of sustainable (persistent) earnings;
☐	**f.** describe indicators of earnings quality;
☐	**g.** explain mean reversion in earnings and how the accruals component of earnings affects the speed of mean reversion;
☐	**h.** evaluate the earnings quality of a company;
☐	**i.** describe indicators of cash flow quality;
☐	**j.** evaluate the cash flow quality of a company;
☐	**k.** describe indicators of balance sheet quality;
☐	**l.** evaluate the balance sheet quality of a company;
☐	**m.** describe sources of information about risk.

© 2019 CFA Institute. All rights reserved.

1 INTRODUCTION

The ability to assess the quality of reported financial information can be a valuable skill. An analyst or investor who can recognize high-quality financial reporting can have greater confidence in analysis based on those financial reports and the resulting investment decisions. Similarly, an analyst or investor who can recognize poor financial reporting quality early—before deficiencies become widely known—is more likely to make profitable investment decisions or to reduce or even avoid losses.

An example of early recognition of an ultimate financial disaster is James Chanos's short position in Enron in November 2000 (Chanos 2002)—more than a year before Enron filed for bankruptcy protection (in December 2001). Despite Enron's high profile and reputation,[1] Chanos had a negative view of Enron based on both quantitative and qualitative factors. Chanos noted that Enron's return on capital was both lower than comparable companies' return on capital and lower than the company's own cost of capital. Qualitative factors contributing to Chanos's view included the company's aggressive revenue recognition policy, its complex and difficult-to-understand disclosures on related-party transactions, and one-time earnings-boosting gains. Later events that substantiated Chanos's perspective included sales of the company's stock by insiders and the resignation of senior executives.

Another example of early recognition of eventual financial troubles is June 2001 reports by analyst Enitan Adebonojo. These reports highlighted questionable accounting by Royal Ahold, a European food retailer. The questionable accounting included "claiming profits of acquired firms as 'organic growth,' booking capital gains from sale-and-leaseback deals as profit, and keeping billions in debt off its balance sheet."[2] In 2003, Royal Ahold announced that it had significantly overstated its profits in the prior two years. The CEO and CFO resigned, various regulators announced investigations, and Royal Ahold's market value dropped significantly.

This reading focuses on reporting quality and the interrelated attribute of results quality. *Reporting quality* pertains to the information disclosed in financial reports. High-quality reporting provides decision-useful information—information that is relevant and faithfully represents the economic reality of the company's activities during the reporting period and the company's financial condition at the end of the period. A separate, but interrelated, attribute of quality is *results* or *earnings quality*, which pertains to the earnings and cash generated by the company's actual economic activities and the resulting financial condition relative to expectations of current and future financial performance. Note that the term "earnings quality" is more commonly used in practice than "results quality," so throughout this reading, earnings quality is used broadly to encompass the quality of earnings, cash flow, and/or balance sheet items.

High-quality earnings reflect an adequate level of return on investment and are derived from activities that a company will likely be able to sustain in the future. Thus, high-quality earnings increase the value of a company more than low-quality earnings. When reported earnings are described as being high quality, it means that the company's underlying economic performance was good (i.e., value enhancing), and it also implies that the company had high reporting quality (i.e., that the information that the company calculated and disclosed was a good reflection of the economic reality).

Earnings can be termed "low quality" either because the reported information properly represents genuinely bad performance or because the reported information misrepresents economic reality. In theory, a company could have low-quality earnings while simultaneously having high reporting quality. Consider a company with

1 In October 2000, Enron was named in the top 25 on *Fortune* magazine's list of the World's Most Admired Companies.

2 "Ahold: Europe's Enron," *The Economist*, (27 February 2003).

low-quality earnings—for example, one whose only source of earnings in a period is a one-off settlement of a lawsuit without which the company would have reported huge losses. The company could nonetheless have high reporting quality if it calculated its results properly and provided decision-useful information. Although it is theoretically possible that a company could have low-quality earnings while simultaneously having high reporting quality, experiencing poor financial performance can motivate the company's management to misreport.

This reading begins in Sections 2–5 with a description of a conceptual framework for and potential problems with financial reporting quality. This is followed in Sections 6 and 7 with a discussion of how to evaluate financial reporting quality. Sections 8–16 focus on the quality of reported earnings, cash flows, and balance sheets, respectively. Sections 17–19 covers sources of information about risk. A summary and practice problems in the CFA Institute item set format complete the reading.

QUALITY OF FINANCIAL REPORTS: CONCEPTUAL FRAMEWORK

2

a demonstrate the use of a conceptual framework for assessing the quality of a company's financial reports;

This section reviews a conceptual framework for assessing the quality of financial reports and then outlines potential problems that affect the quality of financial reports.

2.1 Conceptual Framework for Assessing the Quality of Financial Reports

As indicated in the introduction, financial reporting quality and results or earnings quality are related attributes of quality. Exhibit 1 illustrates this relationship and its implications. Low financial reporting quality can make it difficult or impossible to assess a company's results, and as a result, it is difficult to make investment and other decisions, such as lending and extending credit to the company.

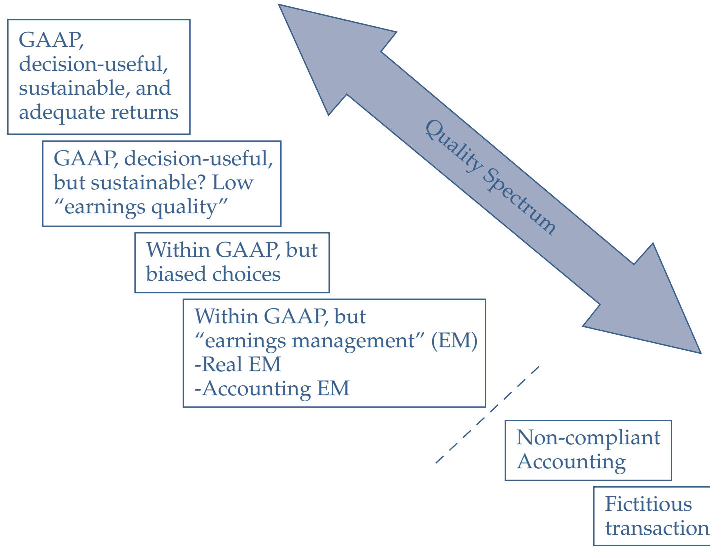
Exhibit 1 Relationships between Financial Reporting Quality and Earnings Quality

		Financial Reporting Quality: Low	Financial Reporting Quality: High
Earnings (Results) Quality	High	LOW financial reporting quality impedes assessment of earnings quality and impedes valuation.	HIGH financial reporting quality enables assessment. HIGH earnings quality increases company value.
	Low		HIGH financial reporting quality enables assessment. LOW earnings quality decreases company value.

Financial reporting quality varies across companies. Financial reports can range from those that contain relevant and faithfully representational information to those that contain information that is pure fabrication. Earnings (results) quality can range from high and sustainable to low and unsustainable. The presence of high-quality financial reporting is a necessary condition for enabling investors to evaluate results quality. High-quality financial reporting alone is an insufficient condition to ensure the presence of high-quality results, but the existence of high-quality financial reporting allows the investor to make such an assessment.

Combining the two aspects of quality—financial reporting and earnings—the overall quality of financial reports from a user perspective can be thought of as spanning a continuum from the highest to the lowest. Exhibit 2 presents a spectrum that provides a basis for evaluating better versus poorer quality reports.

Exhibit 2 Quality Spectrum of Financial Reports



Essentially, the analyst needs to consider two basic questions:

1 Are the financial reports GAAP-compliant and decision-useful?

2 Are the results (earnings) of high quality? In other words, do they provide an adequate level of return, and are they sustainable?

These two questions provide a basic conceptual framework to assess the quality of a company's financial reports and to locate the company's financial reports along the quality spectrum. At the top of the spectrum, labeled in Exhibit 2 as "GAAP, decision-useful, sustainable, and adequate returns" are high-quality reports that provide decision-useful information about high-quality earnings. "GAAP" refers generically to the generally accepted accounting principles or the accepted accounting standards of the jurisdiction under which the company reports. Examples of GAAP are International Financial Reporting Standards (IFRS), US GAAP, and other home-country accounting standards. *Decision-useful* information embodies the characteristics of relevance and faithful representation.[3] High-quality earnings provide an *adequate level of return* on investment (i.e., a return equal to or in excess of the cost of capital) and are sustainable. *Sustainable* indicates that the earnings are derived from activities that a company will likely be able to sustain in the future. Sustainable earnings that provide a high return on investment contribute to a higher valuation of a company and its securities.

Any deviation from the highest point on the quality spectrum can be assessed in terms of the two-question conceptual framework. For example, a company that provides GAAP-compliant, decision-useful information about low-quality earnings (they can be of low quality because they do not provide an adequate level of return and/or they are not sustainable) would appear lower on the quality spectrum. Even lower on the spectrum would be companies that provide GAAP-compliant information, which is less decision-useful because of biased choices.

Biased accounting choices result in financial reports that do not faithfully represent economic phenomena. Biased choices can be made not only in the context of reported amounts but also in the context of how information is presented. For example, companies can disclose information transparently and in a manner that facilitates analysis, or they can disclose information in a manner that aims to obscure unfavorable information and/or to emphasize favorable information.

The problem with bias in accounting choices, as with other deficiencies in financial reporting quality, is that it impedes an investor's ability to correctly assess a company's past performance, to accurately forecast future performance, and thus to appropriately value the company. Choices are deemed to be "aggressive" if they increase the company's reported performance and financial position in the current period. Aggressive choices may decrease the company's reported performance and financial position in later periods. In contrast, choices are deemed to be "conservative" if they decrease the company's reported performance and financial position in the current period. Conservative choices may increase the company's reported performance and financial position in later periods.

Another type of bias is "earnings management." An example of this bias is earnings "smoothing" to understate earnings volatility relative to the volatility if earnings were faithfully represented. Earnings volatility is decreased by understating earnings in periods when a company's operations are performing well and overstating in periods when the company's operations are struggling.

The next levels down on the spectrum mark a departure from GAAP. Financial reports that depart from GAAP can generally be considered low quality; they are of poor financial reporting quality and cannot be relied on to assess earnings quality. The lowest-quality financial reports portray fictitious transactions or omit actual transactions; such financial reports are fabrications.

---

3 These characteristics are from the *Conceptual Framework for Financial Reporting* (IASB 2010). The characteristics of decision-useful information are identical under IFRS and US GAAP. Relevant information is defined as information that can affect a decision and encompasses the notion of materiality. Faithful representation of economic events is complete, neutral, and free from error. The *Framework* also identifies enhancing characteristics of useful information: comparability, verifiability, timeliness, and understandability. High-quality information results when necessary trade-offs among these characteristics are made in an unbiased, skillful manner.

3 POTENTIAL PROBLEMS THAT AFFECT THE QUALITY OF FINANCIAL REPORTS AND REPORTED AMOUNTS AND TIMING OF RECOGNITION

b explain potential problems that affect the quality of financial reports;

The basic choices that give rise to potential problems with quality of financial reports include reported amounts and timing of recognition and classification. Remember that even GAAP-compliant financial reports can diverge from economic reality if GAAP allows for biased choices. In addition to GAAP-compliant choices, a financial statement preparer may choose to present fraudulent reports. This choice represents a divergence from GAAP and economic reality.

3.1 Reported Amounts and Timing of Recognition

The choice of the reported amount and timing of recognition may focus on a single financial statement element (assets, liabilities, owners' equity, revenue and gains [income], or expenses and losses). However, this choice may affect other elements and more than one financial statement because financial statements are interrelated.[4] It is useful to think of the impact of accounting choices in terms of the basic accounting equation (Assets = Liabilities + Equity). This equation can be restated as Assets – Liabilities = Equity, which is also equivalent to Net Assets = Equity. Choices related to income statement elements will affect the balance sheet through equity, and if equity is affected, then another balance sheet element(s) has to be affected or the balance sheet will not balance.

Following are some examples of choices—accounting choices that comply with GAAP, accounting choices that depart from GAAP, and operating choices—and their effects in the current period:

- Aggressive, premature, and fictitious revenue recognition results in overstated income and thus overstated equity. Assets, usually accounts receivable, are also overstated.
- Conservative revenue recognition, such as deferred recognition of revenue, results in understated net income, understated equity, and understated assets.
- Omission and delayed recognition of expenses results in understated expenses and overstated income, overstated equity, overstated assets, and/or understated liabilities. An understatement of bad debt expense results in overstated accounts receivable. Understated depreciation or amortization expense results in the overstatement of the related long-lived asset. Understated interest, taxes, or other expenses result in the understatement of the related liability: accrued interest payable, taxes payable, or other payable.
- Understatement of contingent liabilities is associated with overstated equity resulting from understated expenses and overstated income or overstated other comprehensive income.

4 Depending on management's motivation, poor-quality financial reports may either over-state or understate results. Fraudulent financial reports almost always overstate results.

- Overstatement of financial assets and understatement of financial liabilities, reported at fair value, are associated with overstated equity resulting from overstated unrealized gains or understated unrealized losses.
- Cash flow from operations may be increased by deferring payments on payables, accelerating payments from customers, deferring purchases of inventory, and deferring other expenditures related to operations, such as maintenance and research.

Example 1 describes events and choices at Satyam Computer Services Limited, which resulted in the issuance of fraudulent reports.

EXAMPLE 1

Fictitious Reports

Satyam Computer Services Limited

Satyam Computer Services Limited, an Indian information technology company, was founded in 1987 and grew rapidly by providing business process outsourcing (BPO) on a global basis. In 2007, its CEO, Ramalinga Raju, was named "Entrepreneur of the Year" by Ernst & Young, and in 2008, the World Council for Corporate Governance recognized the company for "global excellence in corporate accountability." In 2009, the CEO submitted a letter of resignation that outlined a massive financial fraud at the company. The company's decline was so rapid and significant that it came to be referred to as "India's Enron."

In late 2008, the World Bank terminated its relationship with the company after finding that Satyam gave kickbacks to bank staff and billed for services that were not provided. These initial revelations of wrongdoing had the effect of putting the company under increased scrutiny. Among other misconduct, the CEO eventually admitted that he created fictitious bank statements to inflate cash and to show interest income. The CEO also created fake salary accounts and took the money paid to those "employees." The company's head of internal auditing created fictitious customer accounts and invoices to inflate revenues.[5]

The external auditors did not independently verify much of the information provided by the company. Even when bank confirmations, which were sent to them directly as opposed to indirectly through Satyam, contained significantly different balances than those reported by Satyam, they did not follow up.

1 Based on the information provided, characterize Satyam's financial reports, with reference to the quality spectrum of financial reports.
2 Explain each of the following misconducts with reference to the basic accounting equation:
   A Transactions with World Bank
   B Fictitious interest income
   C CEO's embezzlement
   D Fictitious revenue
3 Based on the information provided, what documents were falsified to support the misconducts listed in Question 2?

5 See Bhasin (2012) for more information.

Solution to 1:

Based on the information provided, Satyam's financial reports were of the lowest quality. They clearly are at the bottom of the quality spectrum of financial reports: reports based on fictitious information.

Solution to 2:

The effects on the basic accounting equation of the different acts of misconduct are as follows:

A Upon billing for fictitious services, the company would increase an asset, such as accounts receivable, and a revenue account, such as service revenues. The kickbacks to the customer's staff, if recorded, would increase an expense account, such as commissions paid, and increase a liability, such as commissions payable, or decrease an asset, such as cash. The net effect of this misconduct is the overstatement of income, net assets, and equity.

B Fictitious interest income would result in overstated income; overstated assets, such as cash and interest receivable; and overstated equity. These overstatements were hidden by falsifying revenue and cash balances.

C The embezzlement by creating fictitious employees would increase an expense account, such as wages and salaries, and decrease the asset, cash. The resulting understatement of income and equity was offset by a real but fraudulent decrease in cash, which was hidden by falsifying revenue and cash balances.

D Fictitious revenues would result in overstated revenues and income; overstated assets, such as cash and accounts receivable; and overstated equity.

Solution to 3:

Based on the information provided, the documents that were falsified include

- invoices to the World Bank for services that were not provided,
- bank statements,
- employee records, and
- customer accounts and invoices.

The falsified documents were intended to mislead the external auditors.

An astute reader of financial statements may have identified a potential problem at Satyam by comparing the growth in revenue with the growth in assets on its balance sheet, such as short-term and long-term trade receivables and unbilled revenue. Long-term trade receivables and unbilled revenue accounts may have raised questions. Also, there was an account separate from cash, investments in bank deposits, which may have raised questions. However, fraudulent reports that are well constructed can be very challenging to identify.

4 CLASSIFICATION

b explain potential problems that affect the quality of financial reports;

Choices with respect to reported amounts and timing of recognition typically affect more than one financial element, financial statement, and financial period. Classification choices typically affect one financial statement and relate to how an item is classified within a particular financial statement. The balance sheet, the statement of comprehensive income, or the cash flow statement may be the primary focus of the choice.

With respect to the balance sheet, the concern may be to make the balance sheet ratios more attractive or to hide an issue. For example, a company may focus on accounts receivable because it wants to hide liquidity or revenue collection issues. Choices include removing the accounts receivable from the balance sheet by selling them externally or transferring them to a controlled entity, converting them to notes receivable, or reclassifying them within the balance sheet, such as by reporting them as long-term receivables. Although these amounts remain on the balance sheet as receivables of some sort, a result of their reclassification is a lower accounts receivable balance. This could imply to investors that a collection has taken place and also might favorably skew receivables measures, such as days' sales outstanding and receivables turnover.

In the 2003 Merck Annual Report, Merck & Co. reclassified a portion of its inventory to "Other assets," a long-term asset. This reclassification affects the balance sheet and financial ratios as demonstrated in Example 2.

EXAMPLE 2

Balance Sheet Reclassifications

Merck & Co., Inc. and Subsidiaries

In the 2002 Annual Report, inventory was reported at $3,411.8 million. In the 2003 Annual Report, the 2002 inventory value was reported at $2,964.3 million and $447.5 million of inventory was included in other assets. This information was contained in Note 6 to the financial statements, reproduced in Exhibit 3.

Exhibit 3 Note 6 to Consolidated Financial Statements

6. Inventories

Inventories at December 31 consisted of:

($ in millions)	2003	2002
Finished goods	$552.5	$1,262.3
Raw materials and work in process	2,309.8	2,073.8
Supplies	90.5	75.7
Total (approximate current cost)	$2,952.8	$3,411.8
Reduction to LIFO cost	—	—
	$2,952.8	$3,411.8
Recognized as:		
Inventories	$2,554.7	$2,964.3
Other assets	398.1	447.5

Inventories valued under the LIFO method comprised approximately 51% and 39% of inventories at December 31, 2003 and 2002, respectively. Amounts recognized as Other assets consist of inventories held

(continued)

Exhibit 3 (Continued)

in preparation for product launches and not expected to be sold within one year. The reduction in finished goods is primarily attributable to the spin-off of Medco Health in 2003.

1 The reclassification of a portion of inventory to other assets will *most likely* result in the days of inventory on hand:

A decreasing.

B staying the same.

C increasing.

2 As a result of the reclassification of a portion of inventory to other assets, the current ratio will *most likely*:

A decrease.

B stay the same.

C increase.

Solution to 1:

A is correct. The number of days of inventory on hand calculated using the reported inventory number will most likely decrease because the amount of inventory relative to cost of goods sold will decrease.

Solution to 2:

A is correct. The current ratio will decrease because current assets will decrease and current liabilities will stay the same.

From Exhibit 3, notice that the reclassification is described in the sentence, "Amounts recognized as Other assets consist of inventories held in preparation for product launches and not expected to be sold within one year." The reasoning behind the reclassification's explanation is logical: Current assets include assets to be consumed or converted into cash in a company's operating cycle, which is usually one year. The inventory items associated with product launches beyond one year are more appropriately classified as "other assets." Yet, the change in classification poses analytical problems. Inventory turnover is a key indicator of efficiency in managing inventory levels and is calculated as cost of sales divided by average inventory. Although the inventory turnover can be calculated for 2003, it cannot be calculated on a consistent basis for 2002, or any year before then, because the amount of inventory that would have been classified as "other assets" in those periods is not disclosed. An investor has to recognize that a time-series comparison of Merck's inventory turnover is going to produce an inconsistent history because of the lack of consistent information.

The classification of revenues between operating and non-operating may help the user to determine sustainability of a company's earnings, but the classification has potential for misuse by a company. The classification of revenues as being derived from core, continuing operations could mislead financial statement users into considering inflated amounts of income as being sustainable. Similarly, the classification of expenses as non-operating could mislead financial statement users into considering inflated amounts of income as being sustainable. In non-GAAP metrics reported outside of the financial statements, the classification of income-reducing items as non-recurring could also mislead financial statement users into considering inflated amounts of income as being sustainable.

Classifications that result in an item being reported in other comprehensive income rather than on the income statement can affect analysis and comparison. For example, if two otherwise identical companies classify investments differently, net income may differ because the change in value of the investments may flow through net income for one company and through other comprehensive income for the other company.

Classification issues also arise specifically with the statement of cash flows for which management may have incentives to maximize the amount of cash flows that are classified as "operating." Management may be motivated to classify activities, such as the sale of long-term assets, as operating activities rather than investing activities. Operating activities are part of the day-to-day functioning of a company, such as selling inventory or providing services. For most companies, the sale of property or other long-term assets are not operating activities, and including them in operating activities overstates the company's ability to generate cash from its operations. Management may capitalize rather than expense operating expenditures. As a result, the outflow may be classified as an investing activity rather than an operating activity.

Exhibit 4 presents a selection of potential issues, possible actions, and warning signs of possible deviations from high-quality financial reports, some of which will be specifically discussed in later sections of this reading. The warning signs may be visible in the financial statements themselves, in the notes to the financial statements, or in ratios calculated by the analyst that are assessed over time or compared with those of peer companies. Frequently, the chosen actions bias net income upward. However, a new management or management of a company in financial difficulty may be motivated to bias current income downward to enhance future periods.

Exhibit 4 Accounting Warning Signs

Potential Issues	Possible Actions/Choices	Warning Signs
■ Overstatement or non-sustainability of operating income and/or net income ● Overstated or accelerated revenue recognition ● Understated expenses ● Misclassification of revenue, gains, expenses, or losses	■ Contingent sales with right of return, "channel stuffing" (the practice of inducing customers to order products they would otherwise not order or order at a later date through generous terms), "bill and hold" sales (encouraging customers to order goods and retain them on seller's premises) ■ Fictitious (fraudulent) revenue ■ Capitalizing expenditures as assets ■ Classifying non-operating income or gains as part of operations ■ Classifying ordinary expenses as non-recurring or non-operating ■ Reporting gains through net income and losses through other comprehensive income	■ Growth in revenue higher than that of industry or peers ■ Increases in discounts to and returns from customers ■ Higher growth rate in receivables than revenue ■ Large proportion of revenue in final quarter of year for a non-seasonal business ■ Cash flow from operations is much lower than operating income ■ Inconsistency over time in the items included in operating revenues and operating expenses ■ Increases in operating margin ■ Aggressive accounting assumptions, such as long, depreciable lives ■ Losses in non-operating income or other comprehensive income and gains in operating income or net income ■ Compensation largely tied to financial results
■ Misstatement of balance sheet items (may affect income statement) ● Over- or understatement of assets ● Over- or understatement of liabilities ● Misclassification of assets and/or liabilities	■ Choice of models and model inputs to measure fair value ■ Classification from current to non-current ■ Over- or understating reserves and allowances ■ Understating identifiable assets and overstating goodwill	■ Models and model inputs that bias fair value measures ■ Inconsistency in model inputs when measuring fair value of assets compared with that of liabilities ■ Typical current assets, such as accounts receivable and inventory, included in non-current assets ■ Allowances and reserves that fluctuate over time or are not comparable with peers ■ High goodwill value relative to total assets ■ Use of special purpose vehicles ■ Large changes in deferred tax assets and liabilities ■ Significant off-balance-sheet liabilities

Exhibit 4 (Continued)

Potential Issues	Possible Actions/Choices	Warning Signs
■ Overstatement of cash flow from operations	■ Managing activities to affect cash flow from operations ■ Misclassifying cash flows to positively affect cash flow from operations	■ Increase in accounts payable and decrease in accounts receivable and inventory ■ Capitalized expenditures in investing activities ■ Sales and leaseback ■ Increases in bank overdrafts

5 QUALITY ISSUES AND MERGERS AND ACQUISITIONS & FINANCIAL REPORTING THAT DIVERGES FROM ECONOMIC REALITY DESPITE COMPLIANCE WITH ACCOUNTING RULES

b explain potential problems that affect the quality of financial reports;

Quality issues with respect to financial reports often arise in connection with mergers and acquisitions. Mergers and acquisitions provide opportunities and motivations to manage financial results. For accounting purposes, the business combination is accounted for using the acquisition method, and one company is identified as the acquirer. The financial results of the combined companies are reported on a consolidated basis.

Companies with faltering cash-generating ability may be motivated to acquire other companies to increase cash flow from operations. The acquisition will be reported in the investing cash flows if paid in cash, or not even appear on the cash flow statement if paid for with equity. The consolidated cash flow from operations will include the cash flow of the acquired company, effectively concealing the acquirer's own cash flow problems. Such an acquisition can provide a one-time boost to cash from operations that may or may not be sustainable. There are no required post-acquisition "with and without acquisitions" disclosures, making it impossible for investors to reliably assess whether or not the acquirer's cash flow problems are worsening.

A potential acquisition may create an incentive for a company to report using aggressive choices or even misreport. For example, an acquirer's managers may be motivated to make choices to increase earnings to make an acquisition on more favorable terms. Evidence indicates that acquirers making an acquisition for stock may manipulate their reported earnings prior to the acquisition to inflate the value of shares being used to pay for the acquisition (Erickson and Wang 1999). Similarly, the target company's managers may be motivated to make choices to increase earnings to secure a more favorable price for their company. As another example, the acquiring managers may try to manipulate earnings upward after an acquisition if they want to positively influence investors' opinion of the acquisition.[6]

In other cases, misreporting can be an incentive to make an acquisition. Acquisitions complicate a company's financial statements and thus can conceal previous accounting misstatements. Some evidence indicates that companies engaged in intentional

6 Findings consistent with this possibility are presented in Bens, Goodman, and Neamtiu (2012).

misreporting (specifically, companies that were subsequently accused of accounting fraud by the US SEC) are more likely than non-misreporting companies to make an acquisition. They are also more likely to acquire a company that would reduce the comparability and consistency of their financial statements, such as by targeting companies that have less public information and less similar operations (Erickson, Heitzman, and Zhang 2012).

There are also opportunities to make choices that affect the initial consolidated balance sheet and consolidated income statements in the future. When a business combination occurs, the acquirer must measure and recognize identifiable assets acquired and liabilities assumed at their fair values as of the acquisition date. These may include assets and liabilities that the acquired company had not previously recognized as assets and liabilities in its financial statements. For example, identifiable intangible assets that the acquired company developed internally and some contingent liabilities would be recognized by the acquirer. The excess of the purchase price over the recognized value of the identified assets acquired and liabilities assumed is reported as goodwill. Unlike other long-lived assets, goodwill is not amortized; however, it is subject to impairment testing. Because goodwill is not amortized, unless appropriate impairment charges are recorded, the capitalized goodwill amount continues indefinitely.

The default accounting treatment for goodwill—no future amortization expense—provides an incentive to acquirers to understate the value of amortizable intangibles when recording an acquisition. Being a residual amount, more of the value of an acquisition will thus be classified as goodwill, with its future earnings-friendly accounting treatment. That bias may result in postponement of the recognition of an uneconomic acquisition until impairment charges on the goodwill are recorded, which may be long after the acquisition. Managements may be willing to take this chance because they may be able to convince analysts and investors that a goodwill impairment charge is a non-recurring, non-cash charge—something that many will overlook. Nevertheless, the presence of goodwill should make an investor more inquisitive about a company's record in recognizing impairments and should also motivate an investor to evaluate a company's impairment testing process for goodwill. Fair value measurement, except in the case of assets and liabilities with quoted prices in active markets for identical assets or liabilities, presents an opportunity for the acquirer's management to exercise judgment and affect reported values. For example, they could understate fair value of assets to avoid future charges to expense. Understating the fair value of assets will result in a higher goodwill amount. In the absence of impairment of goodwill, there will be no charges associated with the goodwill. Many analysts question whether reported goodwill reflects economic reality.

5.1 Financial Reporting that Diverges from Economic Reality Despite Compliance with Accounting Rules

Certain accounting standards may give rise to financial reporting that an analyst may find less useful because he or she does not view it as reflective of economic reality. Examples 3 and 4 illustrate these types of situations. When possible, an analyst should adjust the reported information to better reflect his or her view of economic reality. If an adjustment is not possible because the relevant data are not disclosed, an analyst can instead make a qualitative assessment of the effect.

Example 3 describes one of the earlier cases of creative consolidation accounting that raised the need for an in-depth consideration of consolidation accounting and the related issue of control. Many entities are governed by the votes of shareholders under which the majority rules. However, exceptions may exist and both US GAAP and IFRS have endeavored to create regimes under which consolidation is required when it is appropriate to depict economic substance.

EXAMPLE 3

Treatment of Variable Interest (Special Purpose) Entities

SEC enforcement action regarding the financial statements of Digilog, Inc.

In order to develop and introduce a new product, Digilog created a separate business entity, DBS, that was capitalized with $10 million of convertible debt issued to Digilog. Upon conversion, Digilog would end up owning nearly 100% of DBS. Initially, owners' equity of DBS consisted of a few thousand dollars of common stock issued to DBS's manager.

During the first two years of DBS's operations, Digilog did not consolidate DBS; it argued that DBS was controlled by its manager, who owned 100% of the outstanding common shares. Even though DBS generated substantial losses over its first two years of existence, Digilog reported interest income on its investment in the convertible debt. After two years, when DBS started to generate profits, Digilog exercised its conversion option and consolidated from that point forward.

Although DBS had been set up as an "independent" corporation, the SEC took the position that the contractual and operating relationships between the two companies were such that they should have been viewed as constituting a single enterprise for financial reporting purposes. The defendants in the enforcement action, Digilog's auditors, consented to a settlement. The settlement included the opinion by the SEC that consolidation would have provided a user of the financial statements with the most meaningful presentation in accordance with GAAP—even though no specific GAAP at that time directly addressed Digilog's "creative" accounting solution.

Eventually, after many more years of debate, and in the wake of the Enron scandal, which also involved abuse of subsequent consolidation rules, the concept of a "variable interest entity" (VIE) was created. A key aspect is control for consolidation purposes; even in the absence of voting control, consolidation is necessary if the investor has the ability to exert influence on the financial and operating policy of the entity and is exposed, or has rights, to variable returns from its investment in the entity. Although the term VIE is not employed by IFRS, its provisions are similar.

Given the facts above and the consolidation rules for a variable interest entity, Digilog is *most likely* to try to argue that it does not need to consolidate DBS because:

A Digilog does not have voting control.

B Digilog's interest income from DBS is not variable.

C DBS's manager has operational and financial control.

Solution:

C is correct. Digilog is most likely to assert that operational and financial control rest with DBS's manager. However, the assertion is not likely to be accepted because the manager's investment is a few thousand dollars compared with $10 million by Digilog. Simply not having voting control is not sufficient to avoid consolidation. Digilog is exposed to variable returns because of possible losses and the convertibility option.

Example 4 considers asset impairments and restructuring charges and their implications.

EXAMPLE 4

Asset Impairments and Restructuring Charges

Two related topics that almost always require special consideration on the part of analysts are asset impairments and restructuring charges. Asset impairments are write-downs of assets required when circumstances indicate that the carrying amount of an asset is excessive compared with the expected future benefits.

The term "restructuring charge" is used under IFRS to indicate a sale or termination of a line of business, closure of business locations, changes in management structure, and/or a fundamental reorganization. All of these events could also give rise to the recognition of a liability (e.g., a commitment to make employee severance payments or to make a payment to settle a lease).

On 25 April 2013, Fuji Electric Co., Ltd, a Japanese company reporting under the GAAP of its home country, announced an impairment loss on land, buildings, structures, and leased assets employed in its "solar cell and module business" in the amount of ¥6.5 billion (Fuji Electric 2013). The entire loss was recorded in its 2012 fiscal year (ending 31 March). Assets and net income were reduced by ¥6.5 billion.

Elan Corporation, plc, a biotechnology company headquartered in Ireland, reported US$42.4 million in restructuring and other costs incurred during fiscal year 2012 related to its decision to close a research facility in San Francisco, with the loss of around 200 jobs, and to shift much of its operations back to Ireland because of changing business conditions. Some of these costs were associated with the obligation to make current and deferred employee severance payments (Leuty 2012).[7]

Recognizing an impairment loss and restructuring charges in a single period, although consistent with most GAAP, is *most likely* to overstate:

A prior periods' net incomes.

B current period's net income.

C future periods' net incomes.

Solution:

A is correct. The impairment and the restructuring were likely the result of past activities and should be taken into account when evaluating past net incomes. The current period's net income, unless the impairment or restructuring is expected to be repeated, is understated. Future period net income may be overstated if reversals occur, but such behavior is not likely. Charging the entire impairment loss and restructuring charge in the current period are examples of conservative accounting principles.

An analyst would likely consider it probable that the events giving rise to Fuji Electric's impairment loss (evidently, declining activity and future prospects for its solar business) had actually occurred over a longer period than that single year. Similarly, an analyst might view the restructuring charge at Elan as relating to previous periods.

When faced with a restructuring charge, an impairment charge, or a combination of the two, an analyst should consider whether similar events occur regularly enough such that they should be factored into estimates of permanent earnings, or whether they should be regarded as one-off items that provide little information about the future earnings of the remaining activities of the company. If it is the former, then the

7 See also Elan Corporation, plc, Form 20-F, filed 12 February 2013.

analyst should attempt to "normalize" earnings by essentially spreading the current restructuring/impairment charge(s) over past periods as well as the current period. If an item is truly one-off—say, the financial effects of a natural disaster—then the analyst is justified in "normalizing" earnings by excluding the item from earnings. This process will require a significant amount of judgment, best informed by knowledge of the underlying facts and circumstances.

Items that are commonly encountered by analysts include the following:

- Revisions to ongoing estimates, such as the remaining economic lives of assets, may lead an analyst to question whether an earlier change in estimate would have been more appropriate.
- Sudden increases to allowances and reserves could call into question whether the prior estimates resulted in overstatement of prior periods' earnings instead of an unbiased picture of economic reality.
- Large accruals for losses (e.g., environmental or litigation-related liabilities) suggest that prior periods' earnings may have been overstated because of the failure to accrue losses earlier.

Management may use items such as reserves and allowances to manage or smooth earnings. The application of accounting standards illustrated in Examples 3 and 4 results in financial statements that may not reflect economic reality. Accounting standards may result in some economic assets and liabilities not being reflected in the financial statements. An example is research and development (R&D) expense. Accounting standards do not permit the capitalization of expenditures for R&D expense, yet R&D produces assets that, in turn, produce future benefits. Accounting standards prohibit R&D's capitalization because of the difficulty in assessing which expenditures will actually produce future benefits and which expenditures will produce nothing. Accounting standards may also result in some information being reported in other comprehensive income rather than through net income. For example, classifying marketable securities as "available for sale" will result in their changes in fair value being reported in other comprehensive income. Contrast that reporting result against that for marketable securities classified as "trading": Their changes in fair value are reported in net income.

No basis of accounting can be expected to recognize all of the economic assets and liabilities for an entity. Consequently, figuring out what *is not* reported can be challenging. One frequently encountered example of an unrecognized asset is a company's sales order backlog. Under most GAAP, revenue is not recognized (and an asset is not created) until services have been performed and other criteria have been met. However, in certain industries, particularly large-scale manufacturing, such as airplane manufacturing, the order backlog can be a significant unrecognized asset. When the amount of backlog is significant, it is typically discussed in the management commentary, and an analyst can use this information to adjust reported amounts and to prepare forecasts.

Another dilemma for analysts is judging whether an item presented in other comprehensive income (OCI) should be included in their analysis as net income. Examples of items presented in OCI include the following:

- unrealized holding gains and losses on certain investments in equity securities,
- unrealized holding gains (and subsequent losses) on items of property and equipment for which the "revaluation option" is elected (IFRS only),
- effects on owners' equity resulting from the translation of the foreign currency-denominated financial statements of a foreign operation to the reporting currency of the consolidated entity,

- certain changes to net pension liability or asset, and
- gains and losses on derivative financial instruments (and certain foreign currency-denominated non-derivative financial instruments) accounted for as a hedge of future cash flows.

When an analyst decides that a significant item presented in OCI should be included in net income, the analyst can adjust reported and forecasted amounts accordingly.

6 EVALUATING THE QUALITY OF FINANCIAL REPORTS: GENERAL STEPS

c describe how to evaluate the quality of a company's financial reports;

Prior to beginning any financial analysis, an analyst should clarify the purpose and context and clearly understand the following:

- What is the purpose of the analysis? What questions will this analysis answer?
- What level of detail will be needed to accomplish this purpose?
- What data are available for the analysis?
- What are the factors or relationships that will influence the analysis?
- What are the analytical limitations, and will these limitations potentially impair the analysis?

In the context of evaluating the quality of financial reports, an analyst is attempting to answer two basic questions:

1. Are the financial reports GAAP-compliant and decision-useful?
2. Are the results (earnings) of high quality? Do they provide an adequate level of return, and are they sustainable?

General steps, which fit within the general framework just mentioned, are discussed first. Following these steps may help an analyst evaluate the quality of financial reports (answering the two basic questions). Then, quantitative tools for evaluating the quality of financial reports are discussed.

6.1 General Steps to Evaluate the Quality of Financial Reports

It is important to note that the steps presented here are meant to serve as a general guideline only. An analyst may choose to add steps, emphasize or deemphasize steps, or alter the order of the steps. Companies are unique, and variation in specific analytical projects will require specific approaches.

1. Develop an understanding of the company and its industry. Understanding the economic activities of a company provides a basis for understanding why particular accounting principles may be appropriate and why particular financial metrics matter. Understanding the accounting principles used by a company *and* its competitors provides a basis for understanding what constitutes the norm—and to assess whether a company's treatment is appropriate.
2. Learn about management. Evaluate whether the company's management has any particular incentives to misreport. Review disclosures about compensation and insider transactions, especially insiders' sales of the company's stock. Review the disclosures concerning related-party transactions.

3 Identify significant accounting areas, especially those in which management judgment or an unusual accounting rule is a significant determinant of reported financial performance.

4 Make comparisons:

   A Compare the company's financial statements and significant disclosures in the current year's report with the financial statements and significant disclosures in the prior year's report. Are there major differences in line items or in key disclosures, such as risk disclosures, segment disclosures, classification of specific expense, or revenue items? Are the reasons for the changes apparent?

   B Compare the company's accounting policies with those of its closest competitors. Are there significant differences? If so, what is the directional effect of the differences?

   C Using ratio analysis, compare the company's performance with that of its closest competitors.

5 Check for warnings signs of possible issues with the quality of the financial reports. For example,

   - declining receivables turnover could suggest that some revenues are fictitious or recorded prematurely or that the allowance for doubtful accounts is insufficient;
   - declining inventory turnover could suggest obsolescence problems that should be recognized; and
   - net income greater than cash provided by operations could suggest that aggressive accrual accounting policies have shifted current expenses to later periods.

6 For firms operating in multiple segments by geography or product—particularly multinational firms—consider whether inventory, sales, and expenses have been shifted to make it appear that a company is positively exposed to a geographic region or product segment that the investment community considers to be a desirable growth area. An analyst may suspect that this shift is occurring if the segment is showing strong performance while the consolidated results remain static or worsen.

7 Use appropriate quantitative tools to assess the likelihood of misreporting.

The first six steps listed describe a qualitative approach to evaluating the quality of financial reports. In addition to the qualitative approach, quantitative tools have been developed to help in evaluating financial reports.

QUANTITATIVE TOOLS TO ASSESS THE LIKELIHOOD OF MISREPORTING 7

c describe how to evaluate the quality of a company's financial reports;

d evaluate the quality of a company's financial reports;

This section describes some tools for assessing the likelihood of misreporting (Step 7 above). If the likelihood of misreporting appears high, an analyst should take special care in analyzing, including qualitatively analyzing, the financial reports of the company.

7.1 Beneish Model

Messod D. Beneish and colleagues conducted studies to identify quantitative indicators of earnings manipulation and to develop a model to assess the likelihood of misreporting (Beneish 1999; Beneish, Lee, and Nichols 2013). The following is the Beneish model and its variables. After the description of each variable, an intuitive explanation of why it is included is given.

The probability of manipulation (M-score) is estimated using a probit model:[8]

$$M\text{-score} = -4.84 + 0.920\ (\text{DSR}) + 0.528\ (\text{GMI}) + 0.404\ (\text{AQI}) + 0.892\ (\text{SGI}) + 0.115\ (\text{DEPI}) - 0.172\ (\text{SGAI}) + 4.679\ (\text{Accruals}) - 0.327\ (\text{LEVI})$$

where

M-score = Score indicating probability of earnings manipulation

DSR (days sales receivable index) = $(\text{Receivables}_t/\text{Sales}_t)/(\text{Receivables}_{t-1}/\text{Sales}_{t-1})$.

Changes in the relationship between receivables and sales could indicate inappropriate revenue recognition.

GMI (gross margin index) = $\text{Gross margin}_{t-1}/\text{Gross margin}_t$.

Deterioration in margins could predispose companies to manipulate earnings.

AQI (asset quality index) = $[1 - (\text{PPE}_t + \text{CA}_t)/\text{TA}_t]/[1 - (\text{PPE}_{t-1} + \text{CA}_{t-1})/\text{TA}_{t-1}]$, where PPE is property, plant, and equipment; CA is current assets; and TA is total assets.

Change in the percentage of assets other than in PPE and CA could indicate excessive expenditure capitalization.

SGI (sales growth index) = $\text{Sales}_t/\text{Sales}_{t-1}$.

Managing the perception of continuing growth and capital needs from actual growth could predispose companies to manipulate sales and earnings.

DEPI (depreciation index) = $\text{Depreciation rate}_{t-1}/\text{Depreciation rate}_t$, where Depreciation rate = Depreciation/(Depreciation + PPE).

Declining depreciation rates could indicate understated depreciation as a means of manipulating earnings.

SGAI (sales, general, and administrative expenses index) = $(\text{SGA}_t\ /\text{Sales}_t)/(\text{SGA}_{t-1}/\text{Sales}_{t-1})$.

An increase in fixed SGA expenses suggests decreasing administrative and marketing efficiency, which could predispose companies to manipulate earnings.

Accruals = (Income before extraordinary items[9] − Cash from operations)/Total assets.

Higher accruals can indicate earnings manipulation.

LEVI (leverage index) = $\text{Leverage}_t/\text{Leverage}_{t-1}$, where Leverage is calculated as the ratio of debt to assets.

Increasing leverage could predispose companies to manipulate earnings.

---

8 Variables that are statistically significant in the empirical results of Beneish (1999) include the days sales receivable index, gross margin index, asset quality index, sales growth index, and accruals.

9 US GAAP for fiscal periods beginning after December 15, 2015, will no longer include the concept of extraordinary items.

The *M*-score in the Beneish model is a normally distributed random variable with a mean of 0 and a standard deviation of 1.0. Consequently, the probability of earnings manipulation indicated by the model can be calculated by using the cumulative probabilities for a standard normal distribution or the NORMSDIST function in Excel. For example, *M*-scores of –1.49 and –1.78 indicate that the probability of earnings manipulation is 6.8% and 3.8%, respectively. Higher *M*-scores (i.e., less negative numbers) indicate an increased probability of earnings manipulation. The probability is given by the amount in the left side of the distribution.

The use of the *M*-score to classify companies as potential manipulators depends on the relative cost of Type I errors (incorrectly classifying a manipulator company as a non-manipulator) and Type II errors (incorrectly classifying a non-manipulator as a manipulator). The cutoff value for classification minimizes the cost of misclassification. Beneish considered that the likely relevant cutoff for investors is a probability of earnings manipulation of 3.8% (an *M*-score exceeding –1.78).[10] Example 5 shows an application of the Beneish model.

EXAMPLE 5

Application of the Beneish Model

Exhibit 5 presents the variables and Beneish's *M*-Score for XYZ Corporation (a hypothetical company).

Exhibit 5 XYZ Corporation *M*-Score

	Value of Variable	Coefficient from Beneish Model	Calculations
DSR	1.300	0.920	1.196
GMI	1.100	0.528	0.581
AQI	0.800	0.404	0.323
SGI	1.100	0.892	0.981
DEPI	1.100	0.115	0.127
SGAI	0.600	–0.172	–0.103
Accruals	0.150	4.679	0.702
LEVI	0.600	–0.327	–0.196
Intercept			–4.840
M-score			–1.231
Probability of manipulation			10.93%

1 Would the results of the Beneish model lead an analyst, using a –1.78 *M*-score as the cutoff, to flag XYZ as a likely manipulator?

2 The values of DSR, GMI, SGI, and DEPI are all greater than one. In the Beneish model, what does this indicate for each variable?

10 See Beneish (1999) for an explanation and derivation of the cutoff values. Beneish et al. (2013) use an *M*-score exceeding –1.78 as the cutoff value.

Solution to 1:

Yes, the model could be expected to lead an analyst to flag XYZ as a likely manipulator. The *M*-score is higher than the cutoff of −1.78, indicating a higher-than-acceptable probability of manipulation. For XYZ Corporation, the model estimates the probability of manipulation as 10.93%. Although the classification of companies as manipulators depends on the relative cost of Type I errors and Type II errors, the value of 10.93% greatly exceeds the cutoff of 3.8% that Beneish identified as the relevant cutoff.

Solution to 2:

Indications are as follows:

A The value greater than one for DSR indicates that receivables as a percentage of sales have increased; this change may be an indicator of inappropriate revenue recognition. XYZ may have shipped goods prematurely and recognized revenues belonging in later periods. Alternatively, it may be caused by customers with deteriorating credit-paying ability—still a problem for the analyst of XYZ.

B The value greater than one for GMI indicates that gross margins were higher last year; deteriorating margins could predispose companies to manipulate earnings.

C The value greater than one for SGI indicates positive sales growth relative to the previous year. Companies could be predisposed to manipulate earnings to manage perceptions of continuing growth and also to obtain capital needed to support growth.

D The value greater than one for DEPI indicates that the depreciation rate was higher in the prior year; a declining depreciation rate can indicate manipulated earnings.

7.2 Other Quantitative Models

Researchers have examined numerous factors that contribute to assessing the probability that a company is engaged in accounting manipulation. Variables that have been found useful for detecting misstatement include accruals quality; deferred taxes; auditor change; market-to-book value; whether the company is publicly listed and traded; growth rate differences between financial and non-financial variables, such as number of patents, employees, and products; and aspects of corporate governance and incentive compensation.[11]

7.3 Limitations of Quantitative Models

Accounting is a partial representation of economic reality. Consequently, financial models based on accounting numbers are only capable of establishing associations between variables. The underlying cause and effect can only be determined by a deeper analysis of actions themselves—perhaps through interviews, surveys, or investigations by financial regulators with enforcement powers.

An additional concern is that earnings manipulators are just as aware as analysts of the power of quantitative models to screen for possible cases of earnings manipulation. It is not surprising to learn, therefore, that Beneish et al.'s 2013 study found

11 A summary of research on predicting accounting misstatement is provided in Dechow, Ge, Larson, and Sloan (2011).

that the predictive power of the Beneish model is declining over time. Undoubtedly, many managers have learned to test the detectability of earnings manipulation tactics by using the model to anticipate analysts' perceptions. Thus, as useful as the Beneish model may be, the search for more powerful analytical tools continues. It is necessary for analysts to use qualitative, not just quantitative, means to assess quality.

EARNINGS QUALITY INDICATORS AND RECURRING EARNINGS 8

f describe indicators of earnings quality;

This section first discusses indicators of earnings quality and then describes how to evaluate the earnings quality of a company. Analytical tools related to identifying very poor earnings/results quality, such as quantitative approaches to assessing the probability of bankruptcy, are also discussed.

8.1 Indicators of Earnings Quality

In general, the term "earnings quality" can be used to encompass earnings, cash flow, and balance sheet quality. This section, however, focuses specifically on earnings quality. High earnings quality is often considered to be evidenced by earnings that are sustainable and represent returns equal to or in excess of the company's cost of capital.[12] High-quality earnings increase the value of the company more than low-quality earnings, and the term "high-quality earnings" assumes that reporting quality is high. In contrast, low-quality earnings are insufficient to cover the company's cost of capital and/or are derived from non-recurring, one-off activities. In addition, the term "low-quality earnings" can also be used when the reported information does not provide a useful indication of the company's performance.

A variety of alternatives have been used as indicators of earnings quality: recurring earnings, earnings persistence and related measures of accruals, beating benchmarks, and after-the-fact confirmations of poor-quality earnings, such as enforcement actions and restatements.

8.1.1 *Recurring Earnings*

When using a company's current and prior earnings as an input to forecast future earnings (for example, for use in an earnings-based valuation), an analyst focuses on the earnings that are expected to recur in the future. For example, earnings from subsidiaries that have been selected for disposal, which must be separately identified as "discontinued operations," are typically excluded from forecasting models. A wide range of other types of items may be non-recurring—for example, one-off asset sales, one-off litigation settlements, or one-off tax settlements. Reported earnings that contain a high proportion of non-recurring items are less likely to be sustainable and are thus considered lower quality.

Enron, an energy distribution company and a company famous for misreporting, presented non-recurring items, among other reporting issues, in such a way that they created an illusion of a solidly performing company. Example 6 shows aspects of Enron's reporting.

12 The residual income model of valuation is most closely linked to this concept of high earnings quality.

EXAMPLE 6

Non-Recurring Items

Enron Corp.

Exhibit 6 Excerpts from Enron and Subsidiaries Consolidated Income Statement, Year-Ended 31 December

(In millions, except per share amounts)	2000	1999	1998
Total revenues	$100,789	$40,112	$31,260
Total costs and expenses	98,836	39,310	29,882
Operating income	$1,953	$802	$1,378
Other income and deductions			
Equity in earnings of unconsolidated equity affiliates	$87	$309	$97
Gains on sales of non-merchant assets	146	541	56
Gain on the issuance of stock by TNPC, Inc.	121	0	0
Interest income	212	162	88
Other income, net	–37	181	–37
Income before interest, minority interests, and income taxes	$2,482	$1,995	$1,582

1 How does the trend in Enron's operating income compare with the trend in its income after other income and deductions (i.e., Income before interest, minority interests, and income taxes)?

2 What items appear to be non-recurring as opposed to being a result of routine operations? How significant are these items?

3 The Enron testimony of short seller James Chanos before US Congress referred to "a number of one-time gains that boosted Enron's earnings" as one of the items that "strengthened our conviction that the market was mispricing Enron's stock" (Chanos 2002). What does Chanos's statement indicate about how Enron's earnings information was being used in valuation?

Solution to 1:

Enron's operating income varied dramatically from year to year, declining from 1998 to 1999 and then more than doubling in 2000. In contrast, Enron's income before interest, minority interests, and income taxes shows a smooth, upward trend with significant increases each year. The increases were 24% and 26% for 2000 and 1999 relative to 1999 and 1998, respectively.

Solution to 2:

Items that appear to be non-recurring are gains on sales of non-merchant assets and the gain on the issuance of stock by TNPC. Although gains from sales of non-merchant assets do recur in each year, this type of activity is not a part of Enron's energy distribution operations. In addition, two other non-operating items—the amount of equity in earnings from unconsolidated subsidiaries and the amount of other income—are highly variable. Two aspects of these items are significant. First, the smooth, upward trend in Enron's income is the direct result of these items. Second, these items collectively represent a significant percentage of the company's income before interest, minority interests, and income taxes, particularly in 1999 when these items represent 52% of the total: ($309 + $541+ $181)/ $1,995 = $1,031/$1,995.

Solution to 3:

Chanos's statement suggests that at least some market participants were mistakenly using Enron's reported income as an input to earnings-based valuation, without adjusting for non-recurring items.

Although evaluating non-recurring items for inclusion in operating metrics is important for making appropriate historical comparisons and for developing appropriate inputs in valuation, another aspect of non-recurring items merits mention. Because classification of items as non-recurring is a subjective decision, classification decisions can provide an opportunity to inflate the amount potentially identified by a user of the income statement as repeatable earnings—those earnings expected from the company's business operations, which investors label as "recurring" or "core" earnings. In the absence of special or one-time items (such as restructuring charges, employee separation costs, goodwill impairment charges, or gains on disposals of assets), operating income is representative of these kinds of earnings. So-called classification shifting, which does not affect total net income, can inflate the amount reported as recurring or core earnings. This could be accomplished by re-classifying normal expenses to special items or by shifting operating expenses to income-decreasing discontinued operations. Anecdotal evidence of classification shifting exists (see Exhibit 7), but the evidence only emerges after the fact.[13] From an analyst's perspective, after-the-fact evidence of earnings management is not particularly useful for anticipating issues with earnings quality. Although it may not be possible to identify whether a company might be engaging in classification shifting, an analyst should nonetheless give special attention to income-decreasing special items, particularly if the company is reporting unusually high operating earnings for the period or if the classification of the item enabled the company to meet or beat forecasts for operating earnings.

Exhibit 7 Anecdotal Evidence of Classification Shifting

- Borden, a food and chemicals company: The SEC determined that the company had classified $146 million of operating expenses as part of a special item (restructuring charges) when the expenses should have been included in selling, general, and administrative expenses (Hwang 1994).
- AmeriServe Food Distribution Inc., which declared bankruptcy only four months after completing a $200 million junk bond issuance: A bankruptcy court–appointed examiner found that the company's financial statements "classified substantial operating expenses… as restructuring charges," which "masked the company's serious financial underperformance and delayed recognition by all parties of the severity of the problems faced by the company (Sherer 2000)."

(continued)

13 Archival evidence of classification shifting is presented in McVay (2006). McVay first models "expected core earnings" and then documents a relationship between reported-minus-expected core earnings and the number of special items. But in any given year, a company's management could attribute the unexpectedly high core earnings to economic improvements related to the special items; therefore, only the *ex post* evidence that unexpectedly high core earnings tend to reverse in the following year is suggestive of earnings management through classification shifting.

Exhibit 7 (Continued)

- Waste Management, which, in 1998, issued the then-largest restatement in SEC history: The enforcement documentation indicates that the company had improperly inflated operating income by netting non-operating gains from the sale of investments and discontinued operations against unrelated operating expenses (SEC 2001b).
- IBM: Revised disclosures, prompted by SEC scrutiny and analysts' requests, showed that the company had classified intellectual property income as an offset to selling, general, and administrative expenses. This classification resulted in an understatement of operating expenses and thus an overstatement of core earnings by $1.5 billion and $1.7 billion in 2001 and 2000, respectively (Bulkeley 2002).

Companies understand that investors differentiate between recurring and non-recurring items. Therefore, in addition to presenting components of income on the face of the income statement, many companies voluntarily disclose additional information to facilitate the differentiation between recurring and non-recurring items. Specifically, companies may disclose both total income and so-called *pro forma* income (or adjusted income, also referred to as non-GAAP measures, or non-IFRS measures if IFRS is applicable) that has been adjusted to exclude non-recurring items. Disclosures of *pro forma* income must be accompanied by a reconciliation between *pro forma* income and reported income. It is important to be aware, however, that determination of whether an item is non-recurring involves judgment, and some companies' managers may be motivated to consider an item non-recurring if it improves a performance metric relevant to investors. For example, Groupon, an online discount provider, included in its original initial public offering (IPO) filing a *pro forma* (i.e., non-GAAP) measure of operating income that excluded online marketing costs. The SEC determined that the measure was misleading and subsequently required the company to eliminate that measure as reported. Overall, although voluntarily disclosed adjustments to reported income can be informative, an analyst should review the information to ensure that excluded items are truly non-recurring.[14]

9 EARNINGS PERSISTENCE AND RELATED MEASURES OF ACCRUALS

e describe the concept of sustainable (persistent) earnings;

One property of high earnings quality is earnings persistence—that is, sustainability of earnings excluding items that are obviously non-recurring and persistence of growth in those earnings. The assumption is that, for equity valuation models involving earnings forecasts, more persistent earnings are more useful inputs. Persistence can be expressed as the coefficient on current earnings in a simple model:[15]

$$\text{Earnings}_{t+1} = \alpha + \beta_1 \text{Earnings}_t + \varepsilon$$

A higher coefficient (β_1) represents more persistent earnings.

14 A survey of non-GAAP earnings in the S&P 500 is presented in Ciesielski and Henry (2017). In the article, the authors provide key prescriptions in evaluating non-GAAP earnings disclosure.

15 Descriptions of certain indicators in this section follow Dechow, Ge, and Schrand (2010).

Earnings can be viewed as being composed of a cash component and an accruals component. The accrual component arises from accounting rules that reflect revenue in the period earned and expenses in the period incurred—not at the time of cash movement. For example, a sale of goods on account results in accounting income in the period the sale is made. If the cash collection occurs in a subsequent period, the difference between reported net income and cash collected constitutes an accrual. When earnings are decomposed into a cash component and an accruals component, research has shown that the cash component is more persistent (Sloan 1996). In the following model, the coefficient on cash flow (β_1) has been shown to be higher than the coefficient on accruals (β_2), indicating that the cash flow component of earnings is more persistent:

$$\text{Earnings}_{t+1} = \alpha + \beta_1\text{Cash flow}_t + \beta_2\text{Accruals}_t + \varepsilon$$

Because of the greater persistence of the cash component, indicators of earnings quality evolved to measure the relative size of the accruals component of earnings. Earnings with a larger component of accruals would be less persistent and thus of lower quality.

An important distinction is between accruals that arise from normal transactions in the period (called "non-discretionary") and accruals that result from transactions or accounting choices outside the normal, which are possibly made with the intent to distort reported earnings (called "discretionary accruals"). Outlier discretionary accruals are an indicator of possibly manipulated—and thus low-quality—earnings. One common approach to identifying abnormal accruals is first to model companies' normal accruals and then to determine outliers. A company's normal accruals are modeled as a function of economic factors, such as growth in credit sales and the amount of depreciable assets. Growth in credit sales would be expected to result in accounts receivable growth, and depreciable assets would be associated with the amount of depreciation. To apply this approach, total accruals are regressed on the factors expected to give rise to normal accruals, and the residual of the regression would be considered a proxy for abnormal accruals.

This approach was pioneered by academics and subsequently adopted in practice.[16] The SEC describes its approach to modeling abnormal accruals:

> Our Accounting Quality Model extends the traditional approach [often based on the popular Jones Model or the Modified Jones Model] by allowing discretionary accrual factors to be a part of the estimation. Specifically, we take filings information across all registrants and estimate total accruals as a function of a large set of factors that are proxies for discretionary and non-discretionary components…. Discretionary accruals are calculated from the model estimates and then used to screen firms that appear to be managing earnings most aggressively. (Lewis 2012)

One simplified approach to screening for abnormal accruals is to compare the magnitude of total accruals across companies. To make a relevant comparison, the accruals would be scaled—for example, by average assets or by average net operating income. Under this approach, high amounts of accruals are an indicator of possibly manipulated and thus low-quality earnings.

A more dramatic signal of questionable earnings quality is when a company reports positive net income but negative operating cash flows. This situation is illustrated in Example 7.

16 See Jones (1991) and Dechow, Sloan, and Sweeney (1995). These seminal academic papers produced the Jones Model and the Modified Jones Model.

EXAMPLE 7

Discrepancy between Net Income and Operating Cash Flows

Allou Health & Beauty Care, Inc.

Allou Health & Beauty Care, Inc. was a manufacturer and distributor of hair and skin care products. Exhibit 8 presents excerpts from the company's financial statements from 2000 to 2002. Following the periods reported in these statements, Allou's warehouses were destroyed by fire, for which the management was found to be responsible. Allou was subsequently shown to have fraudulently inflated the amount of its sales and inventories in those years.

Exhibit 8 Illustration of Fraudulent Reporting in which Reported Net Income Significantly Exceeded Reported Operating Cash Flow, Annual Data 10-K for Allou Health & Beauty Care, Inc., and Subsidiaries

Years ended 31 March	2002	2001	2000
Excerpt from Income Statement			
Revenues, net	$564,151,260	$548,146,953	$421,046,773
Costs of revenue	500,890,588	482,590,356	367,963,675
Gross profit	$63,260,672	$65,556,597	$53,083,098
	⋮	⋮	⋮
Income from operations	27,276,779	28,490,063	22,256,558
	⋮	⋮	⋮
Income from continuing operations*	$6,589,658	$2,458,367	$7,043,548
Excerpt from Statement of Cash Flows			
Cash flows from operating activities:			
Net income from continuing operations	$6,589,658	$2,458,367	$7,043,548
Adjustments to reconcile net income to net cash used in operating activities:			
[Portions omitted]	⋮	⋮	⋮
Decrease (increase) in operating assets:			
Accounts receivable	(24,076,150)	(9,725,776)	(25,691,508)
Inventories	(9,074,118)	(12,644,519)	(40,834,355)
Net cash used in operating activities	$(17,397,230)	$(34,195,838)	$(27,137,652)

* The difference between income from operations and income from continuing operations included deductions for interest expense and provision for income taxes in each year and for a $5,642,678 loss on impairment of investments in 2001.

Referring to Exhibit 8, answer the following questions:

1 Based on the income statement data, evaluate Allou's performance over the period shown.

2 Compare Allou's income from continuing operations and cash flows from operating activities.

3 Interpret the amounts shown as adjustments to reconcile income from continuing operations to net cash used in operating activities.

Solution to 1:

Based on the income statement, the following aspects of Allou's performance are notable. Revenues grew in each of the past three years, albeit more slowly in the latest year shown. The company's gross margin declined somewhat over the past three years but has been fairly stable. Similarly, the company's operating margin declined somewhat over the past three years but has been fairly stable at around 5%. The company's income from continuing operations was sharply lower in 2001 as a result of an impairment loss. The company showed positive net income in each year. Overall, the company showed positive net income in each year, and its performance appears to be reasonably stable based on the income statement data.

Note: Gross margin is gross profit divided by revenues. For example, for 2002, \$63,260,672 divided by \$564,151,260 is 11.2%. The ratios for 2001 and 2000 are 12.0% and 12.6%, respectively.

Operating margin is income from operations divided by revenues. For example, for 2002, \$27,276,779 divided by \$564,151,260 is 4.8%. The ratios for 2001 and 2000 are 5.2% and 5.3%, respectively.

Solution to 2:

Allou reported positive income from continuing operations but negative cash from operating activities in each of the three years shown. Persistent negative cash from operating activities is not sustainable for a going concern.

Solution to 3:

The excerpt from Allou's Statement of Cash Flows shows that accounts receivable and inventories increased each year. This increase can account for most of the difference between the company's income from continuing operations and net cash used in operating activities. The company seems to be accumulating inventory and not collecting on its receivables.

Note: The statement of cash flows, prepared using the indirect method, adjusts net income to derive cash from operating activities. An increase in current assets is subtracted from the net income number to derive the cash from operating activities.

Similar to Allou, the quarterly data for Enron shown in Exhibit 9 shows positive net income but negative cash from operating activities in quarters that were subsequently shown to have been misreported.

Exhibit 9

Quarterly Data 10-Q: Enron and Subsidiaries

Three months ended 31 March ($ millions)	2001	2000
Net income	425	338
Net cash used in operating activities	(464)	(457)

Annual Data 10-K: Enron and Subsidiaries

Year ended 31 December ($ millions)	2000	1999	1998
Net income	979	893	703
Net cash provided by operating activities	4,779	1,228	1,640

An analyst might also question why net cash provided by operating activities was more than double that of net income in 1998, almost 50% greater than net income in 1999, and almost five times net income in 2000.

Although sizable accruals (roughly, net income minus operating cash flow) can indicate possibly manipulated and thus low-quality earnings, it is not necessarily the case that fraudulently reporting companies will have such a profile. For example, as shown in Exhibit 9, Enron's annual operating cash flows exceeded net income in all three years during which fraudulent financial reporting was subsequently revealed. Some of the fraudulent transactions undertaken by Enron were specifically aimed at generating operating cash flow. It is advisable for investors to explore and understand why the differences exist. The company's ability to generate cash from operations ultimately affects investment and financing within the company.

Similarly, as shown in Exhibit 10, WorldCom showed cash from operating activities in excess of net income in each of the three years shown, although the company was subsequently found to have issued fraudulent reports. WorldCom's most significant fraudulent reporting was improperly capitalizing (instead of expensing) certain costs. Because capital expenditures are shown as investing cash outflows rather than operating cash outflows, the company's fraudulent reporting had the impact of inflating operating cash flows.

Exhibit 10 Example of Fraudulent Reporting in which Reported Net Income Did Not Significantly Exceed Reported Operating Cash Flow, WorldCom Inc. and Subsidiaries ($ millions)

For the years ended 31 December	1999	2000	2001
Net income (loss)	$4,013	$4,153	$1,501
Net cash provided by operating activities	11,005	7,666	7,994

In summary, although accrual measures (i.e., differences between net income and operating cash flows) can serve as indicators of earnings quality, they cannot be used in isolation or applied mechanically. WorldCom shows how comparing cash-basis measures, such as cash provided by operating activities, with net income may provide a false sense of confidence about net income. Net income is calculated using subjective estimates, such as expected life of long-term assets, that can be easily manipulated. In each year shown in Exhibit 10, the cash provided by operations exceeded net income (earnings), suggesting that the earnings were of high quality; an analyst looking at this without considering the investing activities would have felt a false sense of security in the reported net income.

MEAN REVERSION IN EARNINGS, BEATING BENCHMARKS AND EXTERNAL INDICATORS OF POOR-QUALITY EARNINGS

10

g explain mean reversion in earnings and how the accruals component of earnings affects the speed of mean reversion;

A key analyst responsibility is to forecast earnings for the purpose of valuation in making investment decisions. The accuracy and credibility of earnings forecasts should increase when a company's earnings stream possesses a high degree of persistence. As already discussed, earnings can be viewed as being composed of a cash flow element plus an accruals element. Sustainable, persistent earnings are driven by the cash flow element of earnings, whereas the accruals element adds information about the company's performance. At the same time, the accruals component can detract from the stability and persistence of earnings because of the estimation process involved in calculating them.

Academic research has shown empirically what we already know intuitively: Nothing lasts forever. Extreme levels of earnings, both high and low, tend to revert to normal levels over time. This phenomenon is known as "mean reversion in earnings" and is a natural attribute of competitive markets. A company experiencing poor earnings performance will shut down or minimize its losing operations and replace inferior managers with ones capable of executing an improved strategy, resulting in improved earnings. At the other extreme, a company experiencing abnormally high profits will attract competition unless the barriers to entry are insurmountable. New competitors may reduce their prices to gain a foothold in an existing company's markets, thereby reducing the existing company's profits over time. Whether a company is experiencing abnormally high or low earnings, the net effect over time is that a return to the mean should be anticipated.

Nissim and Penman (2001) demonstrated that the mean reversion principle exists across a wide variety of accounting-based measures. In a time-series study encompassing companies listed on the New York Stock Exchange and the American Stock Exchange between 1963 and 1999, they tracked such measures as residual income, residual operating income, return on common equity, return on net operating assets, growth in common equity, core sales profit margins, and others. Beginning with data from 1964, they sorted the companies into 10 equal portfolios based on their ranking for a given measure and tracked the median values in each portfolio in each of the next five-year periods. At the end of each fifth year, the portfolios were re-sorted. The process was extended through 1994, yielding means of portfolio medians over seven rankings. The findings were similar across the metrics, showing a clear reversion to the mean over time.

For example, looking at the pattern for return on net operating assets (RNOA),[17] they found that the range of observed RNOAs was between 35% and –5% at the start of the observations but had compressed to a range of 22% to 7% by the end of the study. Their work illustrates the point that extremely strong or weak performance cannot be sustained forever. They also found that the RNOAs of the portfolios that were not outliers in either direction in Year 1—outperformance or underperformance—did not stray over time, staying constant or nearly so over the entire observation period.

The lesson for analysts is clear: One cannot simply extrapolate either very high or very low earnings into the future and expect to construct useful forecasts. In order to be useful, analysts' forecasts need to take into account normalized earnings over the relevant valuation time frame. As discussed, earnings are the sum of cash flows and accruals, and they will be more sustainable and persistent when the cash flow component dominates earnings. If earnings have a significant accruals component, it may hasten the earnings' reversion to the mean, even more so when the accrual elements are outliers relative to the normal amount of accruals in a company's earnings. In constructing their forecasts of future earnings, analysts need to develop a realistic cash flow model and realistic estimates of accruals as well.

10.1 Beating Benchmarks

Announcements of earnings that meet or exceed benchmarks, such as analysts' consensus forecasts, typically result in share price increases. However, meeting or beating benchmarks is not necessarily an indicator of high-quality earnings. In fact, exactly meeting or only narrowly beating benchmarks has been proposed as an indicator of earnings manipulation and thus low-quality earnings. Academic research has documented a statistically large clustering slightly above zero of actual benchmark differences, and this clustering has been interpreted by some as evidence of earnings management.[18] There is, however, disagreement about whether exactly meeting or only narrowly beating is an indicator of earnings manipulation.[19] Nonetheless, a company that consistently reports earnings that exactly meet or only narrowly beat benchmarks can raise questions about its earnings quality.

10.2 External Indicators of Poor-Quality Earnings

Two external indicators of poor-quality earnings are enforcement actions by regulatory authorities and restatements of previously issued financial statements. From an analyst's perspective, recognizing poor earnings quality is generally more valuable if it can be done before deficiencies become widely known and confirmed. Therefore, the external indicators of poor earnings quality are relatively less useful to an analyst. Nonetheless, even though it might be better to recognize poor earnings quality early, an analyst should be alert to external indicators and be prepared to re-evaluate decisions.

---

17 Nissim and Penman define return on net operating assets as Operating income_t/Net operating assets_{t-1}. Net operating assets are operating assets (those assets used in operations) net of operating liabilities (those generated by operations).

18 See Brown and Caylor (2005); Burgstahler and Dichev (1997); and Degeorge, Patel, and Zeckhauser (1999).

19 See Dechow, Richardson, and Tuna (2003).

EVALUATING THE EARNINGS QUALITY OF A COMPANY - REVENUE RECOGNITION CASE: SUNBEAM CORPORATION

11

h evaluate the earnings quality of a company;

The aim of analyzing earnings is to understand the persistence and sustainability of earnings. If earnings do not represent the financial realities faced by a company, then any forecast of earnings based on flawed reporting will also be flawed. Choices and estimates abound in financial reporting; and with those choices and estimates, the temptations for managers to improve their companies' performance by creative accounting are enormous. All too often, companies that appear to be extraordinary performers turn out to be quite ordinary or worse once their choice of accounting methods, including fraudulent choices, is uncovered by a regulator.

To avoid repeating the mistakes of the past, it may be helpful for analysts to learn how managers have used accounting techniques to enhance their companies' reported performance. Some cases provide useful lessons. In a study of 227 enforcement cases brought between 1997 and 2002, the SEC found that the most common accounting misrepresentation occurred in the area of revenue recognition (SEC 2003). Revenue is the largest single figure on the income statement and arguably the most important. Its sheer size and its effect on earnings, along with discretion in revenue recognition policies, have made it the most likely account to be intentionally misstated. For those reasons, investors should always thoroughly and skeptically analyze revenues. Too often, however, the chief concerns of analysts center on the quantitative aspects of revenues. They may ponder the growth of revenues and whether growth came from acquisitions or organically, but they rarely focus on the quality of revenues in the same way. A focus on the quality of revenues, including specifically on how it was generated, will serve analysts well. For example, was it generated by offering discounts or through bill-and-hold sales?

11.1 Revenue Recognition Case: Sunbeam Corporation

Premature/Fraudulent Revenue Recognition

Sunbeam Corporation was a consumer goods company focused on the production and sale of household appliances and outdoor products. In the mid- to late 1990s, it appeared that its new CEO, "Chainsaw Al" Dunlap, had engineered a turnaround at Sunbeam. He claimed to have done this through cutting costs and increasing revenues. The reality was different. Had more analysts performed basic but rigorous analysis of the financial statements in the earlier phases of Sunbeam's misreporting, they might have been more skeptical of the results produced by Chainsaw Al. Sunbeam engaged in numerous sales transactions that inflated revenues. Among them were the following:

- Sunbeam included one-time disposals of product lines in sales for the first quarter of 1997 without indicating that such non-recurring sales were included in revenues.
- At the end of the first quarter of 1997 (March), Sunbeam booked revenue and income from a sale of barbecue grills to a wholesaler. The wholesaler held the merchandise over the quarter's end without accepting ownership risks. The wholesaler could return the goods if it desired, and Sunbeam would pick up the cost of shipment both ways. All of the grills were returned to Sunbeam in the third quarter of 1997.

- Sunbeam induced customers to order more goods than they would normally through offers of discounts and other incentives. Often, the customers also had return rights on their purchases. This induced ordering had the effect of inflating current results by pulling future sales into the present. This practice is sometimes referred to as "channel stuffing." This policy was not disclosed by Sunbeam, which routinely made use of channel-stuffing practices at the end of 1997 and the beginning of 1998.
- Sunbeam engaged in bill-and-hold revenue practices. In a bill-and-hold transaction, revenue is recognized when the invoice is issued while the goods remain on the premises of the seller. These are unusual transactions, and the accounting requirements for them are very strict: The buyer must request such treatment, have a genuine business purpose for the request, and must accept ownership risks. Other criteria for justifying the use of this revenue recognition practice include the seller's past experience with bill-and-hold transactions, in which buyers took possession of the goods and the transactions were not reversed.

There was no real business purpose to the channel stuffing and bill-and-hold transactions at Sunbeam other than for the seller to accelerate revenue and for the buyers to take advantage of such eagerness without any risks on their part. In the words of the SEC, "these transactions were little more than projected orders disguised as sales" (SEC 2001a). Sunbeam did not make such transactions clear to analysts, and many of its disclosures from the fourth quarter of 1996 to the middle of 1998 were inadequate. Still, its methods of inflating revenue left indicators in the financial statements that should have alerted analysts to the low quality of its earnings and revenue reporting.

If customers are induced into buying goods they do not yet need through favorable payment terms or given substantial leeway in returning such goods to the seller, days' sales outstanding (DSO) may increase and returns may also increase. Furthermore, increases in revenue may exceed past increases and the increases of the industry and/or peers. Problems with and changes in collection, expressed through accounts receivable metrics, can give an analyst clues about the aggressiveness of the seller in making sales targets. Exhibit 11 contains relevant annual data on Sunbeam's sales and receivables from 1995 (before the misreporting occurred) through 1997 (when earnings management reached its peak level in the fourth quarter).

Exhibit 11 Information on Sunbeam's Sales and Receivables, 1995–1997

($ millions)	1995	1996	1997
Total revenue	$1,016.9	$984.2	$1,168.2
Change from prior year	—	–3.2%	18.7%
Gross accounts receivable	$216.2	$213.4	$295.6
Change from prior year	—	–1.3%	38.5%
Receivables/revenue	21.3%	21.7%	25.3%
Change in receivables/revenue	0.7%	0.4%	3.6%
Days' sales outstanding	77.6	79.1	92.4
Accounts receivable turnover	4.7	4.6	4.0

Source: Based on information in original company 10-K filings.

What can an analyst learn from the information in Exhibit 11?

- Although revenues dipped 3.2% in 1996, the year the misreporting began, they increased significantly in 1997 as Sunbeam's various revenue "enhancement" programs were implemented. The important factor to notice—the one that should have given an analyst insight into the quality of the revenues—is the simultaneous, and much greater, increase in the accounts receivable balance. Receivables increasing faster than revenues suggests that a company may be pulling future sales into current periods by offering favorable discounts or generous return policies. As it turned out, Sunbeam offered all of these inducements.
- The percentage relationship of receivables to revenue is another way of looking at the relationship between sales and the time it takes a company to collect cash from its customers. An increasing percentage of receivables to revenues means that a lesser percentage of sales has been collected. The decrease in collection on sales may indicate that customers' abilities to repay have deteriorated. It may also indicate that the seller created period-end sales by shipping goods that were not wanted by customers; the shipment would produce documentation, which serves as evidence of a sale. Receivables and revenue would increase by the same absolute amount, which would increase the percentage of receivables to revenue. Customers would return the goods to the seller in the following accounting period. The same thing would happen in the event of totally fictitious revenues. Revenues from a non-existent customer would simultaneously increase receivables by the same amount. An increase in the relationship between revenue and receivables provides analysts with a clue that collections on sales have declined or that there is a possible issue with revenue recognition.
- The number of days sales outstanding [Accounts receivable/(Revenues/365)] increased each year, indicating that the receivables were not being paid on a timely basis—or even that the revenues may not have been genuine in the first place. DSO figures increasing over time indicate that there are problems, either with collection or revenue recognition. The accounts receivable turnover (365/DSO) tells the same story in a different way: It is the number of times the receivables converted into cash each year, and the figure decreased each year. A trend of slower cash collections, as exhibited by Sunbeam, shows increasingly inefficient cash collections at best and should alert an analyst to the possibility of questionable sales or revenue recognition practices.
- The accounts receivable showed poor quality. In 1997, it increased 38.5% over the previous year, while revenues gained 18.7%. The simple fact that receivables growth greatly outstripped the revenue growth suggests receivables collection problems. Furthermore, analysts who paid attention to the notes might have found even more tiles to fit into the mosaic of accounting manipulations. According to a note in the 10-K titled "Accounts Receivable Securitization Facility," in December 1997 Sunbeam had entered into an arrangement for the sale of accounts receivable. The note said that "At December 28, 1997, the Company had received approximately $59 million from the sale of trade accounts receivable." Those receivables were not included in the year-end accounts receivable balance. As the *pro forma* column in Exhibit 12 shows, the accounts receivable would have shown an increase of 66.1% instead of 38.5%; the percentage of receivables to sales would have ballooned to 30.4%, and the days' sales outstanding would have been an attention-getting 110.8 days. Had this receivables sale not occurred, and the receivables been that large, perhaps analysts would have noticed a problem sooner. Careful attention to the notes might have alerted them to how this transaction improved the appearance of the financial statements and ratios.

Exhibit 12 Information on Sunbeam's Sales and Receivables, 1995–1997, and *Pro Forma* Information, 1997

($ millions)	1995	1996	1997	1997 *Pro Forma*
Total revenue	$1,016.9	$984.2	$1,168.2	*$1,168.2*
Change from prior year	—	−3.2%	18.7%	*18.7%*
Gross accounts receivable	$216.2	$213.4	$295.6	*$354.6*
Change from prior year	—	−1.3%	38.5%	*66.1%*
Receivables/revenue	21.3%	21.7%	25.3%	*30.4%*
Change in receivables/revenue	0.7%	0.4%	3.6%	*8.7%*
Days' sales outstanding	77.7	79.2	92.3	*110.8*
Accounts receivable turnover	4.7	4.6	4.0	*3.2*

Source: Based on information in original company 10-K filings.

Analysts observing the trend in days' sales outstanding would have been rightly suspicious of Sunbeam's revenue recognition practices, even if they were observing the days' sales outstanding simply in terms of Sunbeam's own history. If they took the analysis slightly further, they would have been even more suspicious. Exhibit 13 compares Sunbeam's DSO and accounts receivable turnover with those of an industry median based on the numbers from a group of other consumer products companies—Harman International, Jarden, Leggett & Platt, Mohawk Industries, Newell Rubbermaid, and Tupperware Brands.

Exhibit 13 Comparison of Sunbeam and Industry Median, 1995–1997

Sunbeam	1995	1996	1997
Days sales outstanding	77.7	79.2	92.3
Accounts receivable turnover	4.7	4.6	4.0
Industry median			
Days sales outstanding	44.6	46.7	50.4
Accounts receivable turnover	8.2	7.8	7.3
Sunbeam's underperformance relative to median			
Days sales outstanding	33.0	32.5	41.9
Accounts receivable turnover	(3.5)	(3.2)	(3.3)

Source: Based on information in company 10-K filings.

There was yet another clue that should have aroused suspicion in the analyst community. In the December 1997 annual report, the revenue recognition note had been expanded from the previous year's note:

> The Company recognizes revenues from product sales principally at the time of shipment to customers. *In limited circumstances, at the customer's request the Company may sell seasonal product on a bill and hold basis provided that the goods are completed, packaged and ready for shipment, such goods are segregated and the risks of ownership and legal title have passed to the customer.* ***The amount of such bill and hold sales at December 29, 1997 was approximately 3% of consolidated revenues.*** [Italics and emphasis added.]

Not only did Sunbeam hint at the fact that its revenue recognition policies included a method that was of questionable quality, a clue was dropped as to the degree to which it affected operations. That 3% figure may seem small, but the disclosure should have aroused suspicion in the mind of a thorough analyst. As shown in Exhibit 14, working through the numbers with some reasonable assumptions about the gross profit on the sales (28.3%) and the applicable tax rate (35%), an analyst would have seen that the bill-and-hold sales were significant to the bottom line.

Exhibit 14 Effect of Sunbeam's Bill-and-Hold Sales on Net Income ($ millions)

1997 revenue	$1,168.18
Bill-and-hold sales from note	3.0%
Bill-and-hold sales in 1997	$35.05
Gross profit margin	28.3%
Gross profit contribution	$9.92
After-tax earnings contribution	$6.45
Total earnings from continuing operations	$109.42
Earnings attributable to bill-and-hold sales	5.9%

An analyst questioning the genuineness of bill-and-hold sales and performing a simple test of the degree of exposure to their effects might have been disturbed to estimate that nearly 6% of net income depended on such transactions. This knowledge might have dissuaded an analyst from a favorable view of Sunbeam.

REVENUE RECOGNITION CASE: MICROSTRATEGY, INC. 12

h evaluate the earnings quality of a company;

12.1 Multiple-Element Contracts

MicroStrategy, Inc. was a fast-growing software and information services company that went public in 1998. After going public, the company engaged in more complex revenue transactions than it had previously. Its revenue stream increasingly involved less outright sales of software and began tilting more to transactions containing multiple deliverables, including obligations to provide services.

Product revenue is usually recognized immediately, depending on the delivery terms and acceptance by customers, whereas service revenue is recognized as the services are provided. The relevant accounting standards for multiple-deliverable arrangements at the time permitted recognition of revenue on a software delivery only if the software sale could be separated from the service portion of the contract and only if the service revenues were in fact accounted for separately.

Analysts studying MicroStrategy's financial statements should have understood the effects of such accounting conventions on the company's revenues. MicroStrategy's revenue recognition policy in the accounting policies note of its 1998 10-K stated that the standards' requirements were, in fact, its practice:

> Revenue from product licensing arrangements is generally recognized after execution of a licensing agreement and shipment of the product, provided that no significant Company obligations remain and the resulting receivable is deemed collectible by management… Services revenue, which includes training and consulting, is recognized at the time the service is performed. The Company defers and recognizes maintenance revenue ratably over the terms of the contract period, ranging from 12 to 36 months. (p. 49)

MicroStrategy took advantage of the ambiguity present in such arrangements, however, to mischaracterize service revenues and recognize them earlier than they should have as part of the software sale. For example, in the fourth quarter of 1998, MicroStrategy entered into a $4.5 million transaction with a customer for software licenses and a broad array of consulting services. Most of the software licenses acquired by the customer were intended to be used in applications that MicroStrategy would develop in the future, yet the company recognized all of the $4.5 million as software revenue (SEC 2000).

Similarly, in the fourth quarter of 1999, MicroStrategy entered into a multiple-deliverable arrangement with another customer that included the provision for extensive services. Again, the company improperly allocated the elements of the contract, skewing them toward an earlier-recognized software element and improperly recognizing $14.1 million of product revenue in the quarter, which was material.

How could analysts have recognized this pattern of behavior? Without in-depth knowledge of the contracts, it is not possible to approve or disapprove of the revenue allocation with certainty. The company still left a trail that could have aroused the suspicion of analysts, had they been familiar with MicroStrategy's stated revenue recognition policy.

Exhibit 15 shows the mix of revenues for 1996, 1997, and 1998 based on the income statement in MicroStrategy's 1998 10-K:

Exhibit 15 MicroStrategy's Mix of Licenses and Support Revenues, 1996–1998 ($ millions)

	1996	1997	1998
Licenses	$15,873	$36,601	$72,721
Support	6,730	16,956	33,709
Total	$22,603	$53,557	$106,430
Licenses	70.2%	68.3%	68.3%
Support	29.8	31.7	31.7
Total	100.0%	100.0%	100.0%

Between 1996 and 1997, the proportion of support revenues to total revenues increased slightly. It flattened out in 1998, which was the first year known to have mischaracterization between the support revenues and the software revenues. With perfect hindsight, had the $4.5 million of consulting services not been recognized at all, overall revenues would have been $101.930 million and support revenues would have been 33.1% of the total revenues. What could have alerted analysts that something was amiss, if they could not examine actual contracts?

Looking at the quarterly mix of revenues might have aroused analyst suspicions. Exhibit 16 shows the peculiar ebb and flow of revenues attributable to support services revenues.

Exhibit 16 MicroStrategy's Revenue Mix by Quarters, 1Q1998–4Q1999

Quarter	Licenses	Support
1Q98	71.8%	28.2%
2Q98	68.3	31.7
3Q98	62.7	37.3
4Q98	70.7	29.3
1Q99	64.6	35.4
2Q99	68.1	31.9
3Q99	70.1	29.9
4Q99	73.2	26.8

The support services revenue climbed in the first three quarters of 1998 and dropped sharply in the fourth quarter—the one in which the company characterized the $4.5 million of revenues that should have been deferred as software license revenue. Subsequently, the proportion rose again and then continued a downward trend, most sharply in the fourth quarter of 1999 when the company again mischaracterized $14.1 million of revenue as software license revenue.

There is no logical reason that the proportion of revenues from licensing and support services should vary significantly from quarter to quarter. The changes should arouse suspicions and generate questions to ask management. Management's answers, and the soundness of the logic embedded in them, might have made investors more comfortable or more skeptical.

If an analyst knows that a company has a policy of recognizing revenues for contracts with elements of multiple-deliverable arrangements—something apparent from a study of the accounting policy note—then the analyst should consider the risk that misallocation of revenue can occur. Observing trends and investigating deviations from observed trends become important habits for an analyst to practice in order to isolate exceptions. Although a study of revenue trends may not pinpoint a manipulated revenue transaction, it should be sufficient to raise doubts about the propriety of the accounting for transactions.

Enhancing the recognition of revenue is a way for managers to increase earnings, yet it can leave indicators that can be detected by analysts vigilant enough to look for them. Exhibit 17 provides a summary of how to assess the quality of revenues.

Exhibit 17 Summary: Looking for Quality in Revenues

Start with the basics

The first step should be to fully understand the revenue recognition policies as stated in the most recent annual report. Without context for the way revenue is recognized, an analyst will not understand the risks involved in the proper reporting of revenue. For instance, analysts should determine the following:

- What are the shipping terms?
- What rights of return does a customer have: limited or extensive?
- Do rebates affect revenues, and if so, how are they accounted for? What estimates are involved?
- Are there multiple deliverables to customers for one arrangement? If so, is revenue deferred until some elements are delivered late in the contract? If there are multiple deliverables, do deferred revenues appear on the balance sheet?

Age matters

A study of DSO can reveal much about their quality. Receivables do not improve with age. Analysts should seek reasons for exceptions appearing when they

- Compare the trend in DSOs or receivables turnover over a relevant time frame.
- Compare the DSO of one company with the DSOs of similar competitors over similar time frames.

Is it cash or accrual?

A high percentage of accounts receivable to revenues might mean nothing, but it might also mean that channel-stuffing has taken place, portending high future returns of inventory or decreased demand for product in the future. Analysts should

- Compare the percentage of accounts receivable to revenues over a relevant time frame.
- Compare the company's percentage of accounts receivable to revenues with that of competitors or industry measures over similar time frames.

Compare with the real world when possible

If a company reports non-financial data on a routine basis, try relating revenues to those data to determine whether trends in the revenue make sense. Examples include

- Airlines reporting extensive information about miles flown and capacity, enabling an analyst to relate increases in revenues to an increase in miles flown or capacity.
- Retailers reporting square footage used and number of stores open.
- Companies across all industries reporting employee head counts.

As always, analysts should compare any relevant revenue-per-unit measure with that of relevant competitors or industry measures.

Exhibit 17 (Continued)

Revenue trends and composition

Trend analysis, over time and in comparison with competitors, can prompt analysts to ask questions of managers, or it can simply evoke discomfort with the overall revenue quality. Some relationships to examine include

- The relationships between the kinds of revenue recognized. For example, how much is attributable to product sales or licenses, and how much is attributable to services? Have the relationships changed over time, and if so, why?
- The relationship between overall revenue and accounts receivable. Do changes in overall revenues make sense when compared with changes in accounts receivable?

Relationships

Does the company transact business with entities owned by senior officers or shareholders? This is a particularly sensitive area if the manager/shareholder-owned entities are private and there are revenues recognized from the private entity by a publicly owned company; it could be a dumping ground for obsolete or damaged inventory while inflating revenues.

Overstating revenues is not the only way to enhance earnings; according to the SEC study of enforcement cases brought between 1997 and 2002, the next most common financial misreporting was improper expense recognition (SEC 2003). Improper expense recognition typically involves understating expenses and has the same overstating effects on earnings as improper revenue recognition. Understating expenses also leaves indicators in the financial statements for the vigilant analyst to find and assess.

COST CAPITALIZATION CASE: WORLDCOM CORP. 13

h evaluate the earnings quality of a company;

13.1 Property/Capital Expenditures Analysis

WorldCom was a major global communications company, providing phone and internet services to both the business and consumer markets. It became a major player in the 1990s, largely through acquisitions. To keep delivering the earnings expected by analysts, the company engaged in the improper capitalization of operating expenses known as "line costs." These costs were fees paid by WorldCom to third-party telecommunications network providers for the right to use their networks, and the proper accounting treatment for them is to classify them as an operating expense. This improper treatment began in 1999 and continued through the first quarter of 2002. The company declared bankruptcy in July 2002; restatements of financial reports ensued.

The company was audited by Arthur Andersen, who had access to the company's records. According to the findings of the special committee that headed the investigation of the failure (Beresford, Katzenbach, and Rogers 2003), Arthur Andersen failed to identify the misclassification of line costs, among other things, because

> Andersen concluded—mistakenly in this case—that, year after year, the risk of fraud was minimal and thus it never devised sufficient auditing procedures to address this risk. Although it conducted a controls-based audit—relying on WorldCom's internal controls—it failed to recognize the nature and extent of senior management's top-side adjustments through reserve reversals with little or no support, highly questionable revenue items, and entries capitalizing line costs. Andersen did not conduct tests to corroborate the information it received in many areas. It assumed incorrectly that the absence of variances in the financial statements and schedules—in a highly volatile business environment—indicated there was no cause for heightened scrutiny. Andersen conducted only very limited auditing procedures in many areas where we found accounting irregularities. Even so, Andersen still had several chances to uncover problems we identify in this Report. (p. 230–231)

If auditors failed to detect fraud, could analysts really be expected to do better? Analysts may not have been able to pinpoint what was going on at WorldCom, all the way down to the under-reported line costs, but if they had focused on the company's balance sheet, they certainly could have been suspicious that all was not right. If they were looking for out-of-line relationships between accounts—something that the auditors would be expected to do—they might have uncovered questionable relationships that, if unsatisfactorily explained, should have led them to shun securities issued by WorldCom.

For an operating expense to be under-reported, an offsetting increase in the balance of another account must exist. A simple scan of an annual time-series common-size balance sheet, such as is shown in Exhibit 18, might identify the possibility that capitalization is being used to avoid expense recognition. An analyst might not have known that line costs were being under-reported, but simply looking at the time series in Exhibit 18 would have shown that something unusual was going on in gross property, plant, and equipment. The fraud began in 1999, and gross property, plant, and equipment had been 30% and 31% of total assets, respectively, in the two prior years. In 1999, property, plant, and equipment became a much more significant 37% of total assets and increased to 45% in 2000 and 47% in 2001. The company had not changed strategy or anything else to justify such an increase.

Exhibit 18 Common Size Asset Portion of Balance Sheet for WorldCom, 1997–2001

	1997	1998	1999	2000	2001
Cash and equivalents	0%	2%	1%	1%	1%
Net receivables	5	6	6	7	5
Inventories	0	0	0	0	0
Other current assets	2	4	4	2	2
Total current assets	7%	12%	11%	10%	8%
Gross property, plant, and equipment	***30%***	***31%***	***37%***	***45%***	***47%***
Accumulated depreciation	3%	2%	5%	7%	9%
Net property, plant, and equipment	27%	29%	32%	38%	38%
Equity investments	NA	NA	NA	NA	1
Other investments	0	0	0	2	1

Exhibit 18 (Continued)

	1997	1998	1999	2000	2001
Intangibles	61	54	52	47	49
Other assets	5	5	5	3	3
Total Assets	100%	100%	100%	100%	100%

Note: NA is not available.
Source: Based on information from Standard & Poor's Research Insight database.

A curious analyst in 1999 might not have *specifically* determined that line costs were being understated, but the buildup of costs in property, plant, and equipment should have at least made the analyst suspicious that expenses were under-reported somewhere in the income statement.

Capitalizing costs is not the only possible way of understating expenses. Exhibit 19 provides a summary of how to assess the quality of expense recognition, including some things to consider.

Exhibit 19 Summary: Looking for Quality in Expense Recognition

Start with the basics

The first step should be to fully understand the cost capitalization policies as stated in the most recent annual report. Without context for the costs stored on the balance sheet, analysts will not be able to comprehend practice exceptions they may encounter. Examples of policies that should be understood include the following:

- What costs are capitalized in inventory? How is obsolescence accounted for? Are there reserves established for obsolescence that might be artificially raised or lowered?
- What are the depreciation policies, including depreciable lives? How do they compare with competitors' policies? Have they changed from prior years?

Trend analysis

Trend analysis, over time and in comparison with competitors, can lead to questions the analyst can ask managers, or it can simply evoke discomfort with overall earnings quality because of issues with expenses. Some relationships to examine include the following:

- Each quarter, non-current asset accounts should be examined for quarter-to-quarter and year-to-year changes to see whether there are any unusual increases in costs. If present, they might indicate that improper capitalization of costs has occurred.
- Profit margins—gross and operating—are often observed by analysts in the examination of quarterly earnings. They are not often related to changes in the balance sheet, but they should be. If unusual buildups of non-current assets have occurred and the profit margins are improving or staying constant, it could mean that improper cost capitalization is taking place. Recall WorldCom and its improper capitalization of "line costs": Profitability was maintained by capitalizing costs that should have been

(continued)

Exhibit 19 (Continued)

expensed. Also, the overall industry environment should be considered: Are margins stable while balance sheet accounts are growing and the industry is slumping?

- Turnover ratio for total assets; property, plant, and equipment; and other assets should be computed (with revenues divided by the asset classification). Does a trend in the ratios indicate a slowing in turnover? Decreasing revenues might mean that the assets are used to make a product with declining demand and portend future asset write-downs. Steady or rising revenues and decreasing turnover might indicate improper cost capitalization.
- Compute the depreciation (or amortization) expense compared to the relevant asset base. Is it decreasing or increasing over time without a good reason? How does it compare with that of competitors?
- Compare the relationship of capital expenditures with gross property, plant, and equipment over time. Is the proportion of capital expenditures relative to total property, plant, and equipment increasing significantly over time? If so, it may indicate that the company is capitalizing costs more aggressively to prevent their recognition as current expenses.

Relationships

Does the company transact business with entities owned by senior officers or shareholders? This is a particularly sensitive area if the manager/shareholder-owned entities are private. Dealings between a public company and the manager-owned entity might take place at prices that are unfavorable for the public company in order to transfer wealth from the public company to the manager-owned entity. Such inappropriate transfers of wealth can also occur through excessive compensation, direct loans, or guarantees. These practices are often referred to as "tunneling" (Johnson, LaPorta, Shleifer, and Lopez-de-Silanes 2000).

In some cases, sham dealings between the manager-owned entity and the public company might be falsely reported to improve reported profits of the public company and thus enrich the managers whose compensation is performance based. In a different type of transaction, the manager-owned entity could transfer resources to the public company to ensure its economic viability and thus preserve the option to misappropriate or to participate in profits in the future. These practices are often referred to as "propping" (Friedman, Johnson, and Mitton 2003).

Assessing earnings quality should be an established practice for all analysts. Earnings quality should not automatically be accepted as "high quality" until accounting problems emerge and it is too late. Analysts should consider the quality of earnings before assigning value to the growth in earnings. In many cases, high reported earnings growth, which turned out to be fraudulent, preceded bankruptcy.

BANKRUPTCY PREDICTION MODELS: ALTMAN MODEL, DEVELOPMENTS IN BANKRUPTCY PREDICTION MODELS

14

h evaluate the earnings quality of a company;

j evaluate the cash flow quality of a company;

k describe indicators of balance sheet quality;

l evaluate the balance sheet quality of a company;

Bankruptcy prediction models address more than just the quality of a company's earnings and include aspects of cash flow and the balance sheet as well.[20] Various approaches have been used to quantify the likelihood that a company will default on its debt and/or declare bankruptcy.

14.1 Altman Model

A well-known and early model to assess the probability of bankruptcy is the Altman model (Altman 1968). The model is built on research that used ratio analysis to identify likely failures. An important contribution of the Altman model is that it provided a way to incorporate numerous financial ratios into a single model to predict bankruptcy. The model overcame a limitation of viewing ratios independently (e.g., viewing a company with poor profitability and/or solvency position as potentially bankrupt without considering the company's strong liquidity position).

Using discriminant analysis, Altman developed a model to discriminate between two groups: bankrupt and non-bankrupt companies. Altman's Z-score is calculated as follows:

Z-score = 1.2 (Net working capital/Total assets) + 1.4 (Retained earnings/Total assets) + 3.3 (EBIT/Total assets) + 0.6 (Market value of equity/Book value of liabilities) + 1.0 (Sales/Total assets)

The ratios in the model reflect liquidity, profitability, leverage, and activity. The first ratio—net working capital/total assets—is a measure of short-term liquidity risk. The second ratio—retained earnings/total assets—reflects accumulated profitability and relative age because retained earnings accumulate over time. The third ratio—EBIT (earnings before interest and taxes)/total assets, which is a variant of return on assets (ROA)—measures profitability. The fourth ratio—market value of equity/book value of liabilities—is a form of leverage ratio; it is expressed as equity/debt, so a higher number indicates greater solvency. The fifth ratio—sales/total assets—indicates the company's ability to generate sales and is an activity ratio.

Note that Altman's discriminant function shown in his original article (1968) was

$$Z\text{-score} = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$$

with each of the X variables corresponding to the ratios just described. Altman (2000) explains that "due to the original computer format arrangement, variables X_1 through X_4 must be calculated as absolute percentage values. For instance, the company whose net working capital to total assets (X_1) is 10% should be included as 10.0% and not 0.10. Only variable X_5 (sales to total assets) should be expressed in a different manner: that

20 Recall that the term "earnings quality" is used broadly to encompass the quality of earnings, cash flow, and/or balance sheet items.

is, a S/TA [sales/total assets] ratio of 200 percent should be included as 2.0" (p. 14). For this reason, the *Z*-score model is often expressed as shown in the first equation of this section.

The interpretation of the score is that a higher *Z*-score is better. In Altman's application of the model to a sample of manufacturing companies that had experienced losses, scores of less than 1.81 indicated a high probability of bankruptcy, scores greater than 3.00 indicated a low probability of bankruptcy, and scores between 1.81 and 3.00 were not clear indicators.

14.2 Developments in Bankruptcy Prediction Models

Subsequent research addressed various shortcomings in the Altman prediction model. One shortcoming is the single-period, static nature of the Altman model; it uses only one set of financial measures, taken at a single point in time. Shumway (2001) addressed this shortcoming by using a hazard model, which incorporates all available years of data to calculate each company's bankruptcy risk at each point in time.

Another shortcoming of the Altman model (and other accounting-based bankruptcy prediction models) is that financial statements measure past performance and incorporate the going-concern assumption. The reported values on a company's balance sheet assume that the company is a going concern rather than one that might be failing. An alternative is to use market-based bankruptcy prediction models. For example, market-based prediction models building on Merton's concept of equity as a call option on the company's assets infer the default probability from the company's equity value, amount of debt, equity returns, and equity volatility (Kealhofer 2003). Credit default swap data and corporate bond data can also be used to derive default probabilities. Other research indicates that the most effective bankruptcy prediction models include both accounting-based data and market-based data as predictor variables. For example, Bharath and Shumway (2008) model default probability based on market value of equity, face value of debt, equity volatility, stock returns relative to market returns over the previous year, and the ratio of net income to total assets to identify companies likely to default.

15 CASH FLOW QUALITY

i. describe indicators of cash flow quality;

j. evaluate the cash flow quality of a company;

Cash flow statements are free of some of the discretion embedded in the financial statements based on accrual accounting. As a result, analysts may place a great deal of importance and reliance on the cash flow statement. However, there are opportunities for management to affect the cash flow statement.

15.1 Indicators of Cash Flow Quality

Operating cash flow (OCF) is the cash flow component that is generally most important for assessing a company's performance and valuing a company or its securities. Therefore, discussions of cash flow quality typically focus on OCF.

Similar to the term "earnings quality," when reported cash flows are described as being of high quality, it means that the company's underlying economic performance was good (i.e., value enhancing) and it also implies that the company had high reporting quality (i.e., that the information calculated and disclosed by the company was a reasonable reflection of economic reality). Cash flow can be described as "low quality" either because the reported information correctly represents bad economic performance (poor results quality) or because the reported information misrepresents economic reality (poor reporting quality).

From an economic perspective, the corporate life cycle and industry profile affect cash flow and must be considered when analyzing the statement of cash flows. For example, a start-up company might be expected to have negative operating and investing cash flows, which would be funded from borrowing or from equity issuance (i.e., financing cash flows). In contrast, an established company would typically have positive operating cash flow from which it would fund necessary investments and returns to providers of capital (i.e., dividends, share repurchases, or debt repayments—all of which are financing cash flows).

In general, for established companies, high-quality cash flow would typically have most or all of the following characteristics:

- Positive OCF
- OCF derived from sustainable sources
- OCF adequate to cover capital expenditures, dividends, and debt repayments
- OCF with relatively low volatility (relative to industry participants)

As always, high quality requires not only high results quality, as in the previous list, but also high reporting quality. The reported cash flows should be relevant and faithfully represent the economic reality of the company's activities. For example, classifying a financing inflow as an operating inflow would misrepresent the economic reality.

From the perspective of cash flow reporting quality, OCF is generally viewed as being less easily manipulated than operating or net income. Large differences between earnings and OCF or increases in such differences can be an indication of earnings manipulation. The statement of cash flows can be used to highlight areas of potential earnings manipulation.

Even though OCF is viewed as being less subject to manipulation than earnings, the importance of OCF may create incentives for managers to manipulate the amounts reported. Therefore, quality issues with cash flow reporting can exist. One issue that arises with regard to cash flow reporting quality is timing. For example, by selling receivables to a third party and/or by delaying paying its payables, a company can boost OCF. An increase in such activities would be reflected as a decrease in the company's days' sales outstanding and an increase in the company's days of payables. Thus, an analyst can potentially detect management choices to decrease current assets or increase current liabilities, choices that will increase OCF, by looking at asset utilization (activity) ratios, changes in balance sheet accounts, and disclosures in notes to the financial statements. Another issue that arises with regard to cash flow reporting quality is related to classification of cash flows: Management may try to shift positive cash flow items from investing or financing activities to operating activities to inflate operating cash flows.

15.2 Evaluating Cash Flow Quality

Because OCF is viewed as being less subject to manipulation than earnings, the statement of cash flows can be used to identify areas of potential earnings manipulation. The financial fraud at Satyam Computer Services, an Indian information technology company, was described earlier in this reading. In that case, the use of a computer model based on accruals may have failed to detect the fraud. A *New York Times* article (Kahn 2009) provides anecdotal evidence:

> In September, [an analyst] used a computer model to examine India's 500 largest public companies for signs of accounting manipulation. He found that more than 20 percent of them were potentially engaged in aggressive accounting, but Satyam was not on the list. This is because the automated screens that analysts … use to pick up signs of fraud begin by searching for large discrepancies between reported earnings and cash flow. In Satyam's case, the cash seemed to keep pace with profits.

In other words, a computer model that screened for companies with operating cash flow persistently lower than earnings would not have identified Satyam as a potential problem because its reported operating cash flow was relatively close to reported profits.

It may be helpful to examine pertinent indicators using a more qualitative approach. Exhibit 20 presents an excerpt from the statement of cash flows for Satyam for the quarter ended 30 June 2008.

Exhibit 20 Excerpt from Satyam's IFRS Consolidated Interim Cash Flow Statement (All amounts $ millions except per share data and as otherwise stated.)

	Quarter ended 30 June 2008 (unaudited)	Quarter ended 30 June 2007 (unaudited)	Year ended 31 March 2008 (audited)
Profit before income tax	**143.1**	**107.1**	**474.3**
Adjustments for			
Share-based payment expense	4.3	5.9	23.0
Financial costs	1.3	0.8	7.0
Finance income	(16.2)	(16.4)	(67.4)
Depreciation and amortisation	11.5	9.3	40.3
(Gain)/loss on sale of premises and equipment	0.1	0.1	0.6
Changes in value of preference shares designated at fair value through profit or loss	0.0	0.0	(1.6)
Gain/(loss) on foreign exchange forward and option contracts	53.0	(21.1)	(7.4)
Share of (profits)/losses of joint ventures, net of taxes	(0.1)	0.0	(0.1)
	197.0	**85.7**	**468.7**
Movements in working capital			
— Trade and other receivables	(81.4)	(64.9)	(184.3)
— Unbilled revenue	(23.5)	(6.0)	(39.9)
— Trade and other payables	34.1	2.2	48.8
— Unearned revenue	5.8	2.4	11.4
— Other liabilities	(6.3)	30.3	61.2
— Retirement benefit obligations	3.7	1.3	17.8
Cash generated from operations	**129.4**	**51.0**	**383.7**
Income taxes paid	–3.8	–9.8	–49.4
Net cash provided by operating activities	**125.6**	**41.2**	**334.3**

Source: Based on information from Satyam's Form 6-K, filed 25 July 2008.

One item of note on this statement of cash flows is the $53 million non-cash item labeled "Gain/(loss) on foreign exchange forward and options contracts" (i.e., derivative instruments) in the quarter ended 30 June 2008. The item appears to be shown as a gain based on the labeling; however, it would not be correct to add back a gain in this calculation of operating cash flow because it is already included in profit before tax. When the company was asked about this item in the quarterly conference call with analysts, no answer was readily available. Instead, the company's manager said that he would "get back to" the questioner. The fact that the company's senior executives could not explain the reason for an item that represented almost 40% of the total pre-tax profit for the quarter ($53/$143.1 = 37%) is clearly a signal of potential problems. Refer to Exhibit 21 for an excerpt from the conference call.

Exhibit 21 Excerpt from Conference Call regarding Quarterly Results of Satyam, 18 July 2008

George Price, analyst at Stifel Nicolaus:	One question which is on the cash flow statement. You had a—you had $53 million in unrealized gain on derivative financial instruments in the quarter and it's a line item that just, on quick check, I don't think we've seen in past quarters. Can you comment on exactly what that is? … On the comparison periods, there were more modest losses. What drove that large benefit? How should we think about timing of cash flow maybe over the next couple quarters? Any one-time issues like that?
Srinivas Vadlamani:	I—can you repeat that, please?
George Price:	Srinivas, there's was a $53 million unrealized gain in the cash flow statement, and I'm just wondering if you could explain that in a little bit more detail…. The magnitude is a little surprising.
Srinivas Vadlamani:	No, let me—let me check on that. I'll get back to you.

Another item of note on the statement of cash flows is the steady growth in receivables. Analysts examine a company's ratios, such as days' sales outstanding. Exhibit 22 presents selected annual data for Satyam. The large jump in days' sales outstanding from 2006 to 2007 could cause concern. Furthermore, the management commentary in the company's Form 20-F indicated that "Net accounts receivable… increased… primarily as a result of an increase in our revenues and increase in collection period." An increase in the collection period of receivables raises questions about the creditworthiness of the company's customers, about the efficiency of the company's collection efforts, and about the quality of the revenue recognized.

Exhibit 22 Selected Annual Data on Accounts Receivable for Satyam, 2005–2008

($ millions)	2008	2007	2006	2005
Total revenue	$2,138.1	$1,461.4	$1,096.3	$793.6
% Change from previous year	*46.3%*	*33.3%*	*38.1%*	
Gross accounts receivable	$539.1	$386.9	$238.1	$178.3
% Change from previous year	*39.3%*	*62.5%*	*33.5%*	

(continued)

Exhibit 22 (Continued)

($ millions)	2008	2007	2006	2005
Allowance for doubtful debts	$31.0	$22.8	$19.1	$17.5
% Change from previous year	*36.0%*	*19.4%*	*9.1%*	
Gross receivables/revenue	25.21%	26.47%	21.72%	22.47%
Change in receivables/revenue	*–4.8%*	*21.9%*	*–3.3%*	
Days' sales outstanding	92.0	96.6	79.3	82.0
Accounts receivable turnover	4.0	3.8	4.6	4.5

Source: Based on data from Satyam's 20-F filings.

A signal of problems related to cash, which would not have appeared on the statement of cash flows, was the purported use of the company's cash. Satyam reported increasing amounts invested in current accounts. On a conference call excerpted in Exhibit 23, an analyst asked for a specific reason why such large amounts would be held in non-interest-bearing accounts. Instead of providing a reason, the company officer instead stated that the amounts would be transferred to higher-earning accounts soon.

Exhibit 23 Excerpt from Conference Call regarding Quarterly Results for Satyam, 17 October 2008

Kawaljeet Saluja, analyst at Kotak Institutional Equities:	Hi, my questions are for Srinivas. Srinivas, any specific reason why you have $500m parked in current accounts which are not [gaining] any interest?
Srinivas Vadlamani:	No, that is basically—as on the quarter ending, but there is a statement to that [inaudible] to the deposit accounts. We have [inaudible] deposits now.
Kawaljeet Saluja:	But, Srinivas, if I look at the deposit accounts for the last four quarters, that number has remained absolutely flat. And most of the incremental cash that is parked in current accounts and this is not something which is this quarter changed. Would you highlight some of the reasons for it?
Srinivas Vadlamani:	No, basically, what will happen is these amounts will be basically in different countries. And then we will be bringing them to India based on the need. So we will be—basically, some of them are in overnight deposits and all that. So, now we have placing them into normal current deposits. So, next quarter onwards, we will see that as part of the deposits.

In Satyam CEO's January 2009 letter of resignation, he confessed that "the Balance Sheet carries as of September 30, 2008 [i]nflated (non-existent) cash and bank balances of Rs. 5,040 crore[21] (as against Rs. 5,361 crore reflected in the books)...."[22] In other words, of the amount shown as cash on the company's balance sheet, more than 90% was non-existent. It is suggested that some of the cash balances had existed but had been "siphoned off to a web of companies controlled by Mr. Raju and his family." (Kahn 2009)

Overall, the Satyam example illustrates how the statement of cash flows can suggest potential areas of misreporting. In Satyam's case, two items that raised questions were a large non-cash gain on derivatives and an increase in days' sales outstanding. Potential areas of misreporting can then be investigated by reference to the company's other financial reports. The following example illustrates how the statement of cash flows can highlight earnings manipulation and also illustrates how the cash flow information corresponds to information gleaned from analysis of the company's earnings.

Example 8 covers the application of cash flow evaluation to determine quality of earnings.

EXAMPLE 8

Sunbeam Statement of Cash Flows

As noted in the previous section, Sunbeam engaged in various improper accounting practices. Refer to the excerpt from Sunbeam's statement of cash flows in Exhibit 24 to answer the following questions:

1. One of the ways that Sunbeam misreported its financial statements was improperly inflating and subsequently reversing restructuring charges. How do these items appear on the statement of cash flows?
2. Another aspect of Sunbeam's misreporting was improper revenue recognition. What items on the statement of cash flow would primarily be affected by that practice?

Exhibit 24 Excerpt from Sunbeam's Consolidated Statement of Cash Flows, 1995–1997 ($ thousands)

Fiscal Years Ended	28 Dec. 1997	29 Dec. 1996	31 Dec. 1995
Operating Activities:			
Net earnings (loss)	109,415	(228,262)	50,511
Adjustments to reconcile net earnings (loss) to net cash provided by (used in) operating activities:			
Depreciation and amortization	38,577	47,429	44,174
Restructuring, impairment, and other costs	—	154,869	—
Other non-cash special charges	—	128,800	—
Loss on sale of discontinued operations, net of taxes	13,713	32,430	—
Deferred income taxes	57,783	(77,828)	25,146
Increase (decrease) in cash from changes in working capital:			

(continued)

21 Crore is used in India to denote 10,000,000.

22 From Mr. B. Ramalinga Raju's resignation letter attached to Form 6-K that was filed with the SEC on 7 January 2009.

Exhibit 24 (Continued)

Fiscal Years Ended	28 Dec. 1997	29 Dec. 1996	31 Dec. 1995
Receivables, net	(84,576)	(13,829)	(4,499)
Inventories	(100,810)	(11,651)	(4,874)
Account payable	(1,585)	14,735	9,245
Restructuring accrual	(43,378)	—	—
Prepaid expenses and other current assets and liabilities	(9,004)	2,737	(8,821)
Income taxes payable	52,844	(21,942)	(18,452)
Payment of other long-term and non-operating liabilities	(14,682)	(27,089)	(21,719)
Other, net	(26,546)	13,764	10,805
Net cash provided by (used in) operating activities	(8,249)	14,163	81,516

Note: The reason that an increase in sales is shown as a negative number on the statement of cash flows prepared using the indirect method is to reverse any sales reported in income for which cash has not yet been received.

Solution to 1:

Sunbeam's statement of cash flows is prepared using the indirect method (i.e., the operating section shows a reconciliation between reported net income and operating cash flow). This reconciliation highlights that the amount of non-cash charges recorded in 1996 for restructuring, impairment, and other costs totaled about $284 million ($154.869 million + $128.8 million). In the following year, the reversal of the restructuring accrual was $43 million. By inflating and subsequently reversing restructuring charges, the company's income would misleadingly portray significant improvements in performance following the arrival of its new CEO in mid-1996.

Solution to 2:

The items on the statement of cash flows that would primarily be affected by improper revenue recognition include net income, receivables, and inventories. Net income and receivables would be overstated. The statement of cash flows, in which an increase in receivables is shown as a negative number, highlights the continued growth of receivables. In addition, Sunbeam's practice of recording sales that lacked economic substance—because the purchaser held the goods over the end of an accounting period but subsequently returned all the goods—is highlighted in the substantial increase in inventory in 1997.

An issue that arises with regard to cash flow reporting quality is classification shifting: shifting positive cash flow items from investing or financing to inflate operating cash flows. A shift in classification does not change the total amount of cash flow, but it can affect investors' evaluation of a company's cash flows and investors' expectations for future cash flows.

Flexibility in classification exists within accounting standards. For example, IFRS permits companies to classify interest paid either as operating or as financing. IFRS also permits companies to classify interest and dividends received as operating or as investing. In contrast, US GAAP requires that interest paid, interest received, and dividends received all be classified as operating cash flows. Thus, an analyst comparing an IFRS-reporting company to a US GAAP-reporting company would want to ensure comparable classification of interest and dividends and would adjust the reported amounts, if necessary. In addition, an analyst examining an IFRS-reporting company should be alert to any year-to-year changes in classification of interest and dividends. For example, consider an IFRS-reporting company that changed its classification of interest paid from operating to financing. All else equal, the company's operating cash flow would appear higher than the prior period even if no other activities occurred in the period.

As another example of the flexibility permitted by accounting standards, cash flows from non-trading securities are classified as investing cash flows, whereas cash flows from trading securities are typically classified as operating cash flows. However, each company decides what constitutes trading and non-trading activities, depending on how it manages its securities holdings. This discretion creates an opportunity for managers to shift cash flows from one classification to another.

Example 9 illustrates a shift of cash flows from investing to operating.

EXAMPLE 9

Classification of Cash Flows

Nautica Enterprises[23]

An excerpt from the statement of cash flows from the fiscal 2000 annual report of Nautica Enterprises, an apparel manufacturer, is shown as Exhibit 25. An excerpt from the statement of cash flows from the company's fiscal 2001 annual report is shown in Exhibit 26. Use these two excerpts to answer the questions below.

Exhibit 25 Excerpt from Nautica Enterprises' Consolidated Statement of Cash Flow from Annual Report, filed 27 May 2000 (amounts in thousands)

	Year ended 4 March 2000
Cash flows from operating activities	
Net earnings	$46,163
Adjustments to reconcile net earnings to net cash provided by operating activities, net of assets and liabilities acquired	
Minority interest in net loss of consolidated subsidiary	—
Deferred income taxes	(1,035)
Depreciation and amortization	17,072
Provision for bad debts	1,424
Changes in operating assets and liabilities	
Accounts receivable	(6,562)
Inventories	(3,667)
Prepaid expenses and other current assets	(20)
Other assets	(2,686)
Accounts payable: trade	(548)

(continued)

23 Example adapted from Mulford and Comiskey (2005).

Exhibit 25 (Continued)

	Year ended 4 March 2000
Accrued expenses and other current liabilities	9,086
Income taxes payable	3,458
Net cash provided by operating activities	62,685
Cash flows from investing activities	
Purchase of property, plant, and equipment	(33,289)
Acquisitions, net of cash acquired	—
Sale (purchase) of short-term investments	21,116
Payments to register trademark	(277)
Net cash used in investing activities	(12,450)

Exhibit 26 Excerpt from Nautica Enterprises' Consolidated Statements of Cash Flows from Annual Report, filed 29 May 2001 (amounts in thousands)

	Year Ended 3 March 2001	Year Ended 4 March 2000
Cash flows from operating activities		
Net earnings	46,103	46,163
Adjustments to reconcile net earnings to net cash provided by operating activities, net of assets and liabilities acquired		
Minority interest in net loss of consolidated subsidiary	—	—
Deferred income taxes	(2,478)	(1,035)
Depreciation and amortization	22,968	17,072
Provision for bad debts	1,451	1,424
Changes in operating assets and liabilities		
Short-term investments	28,445	21,116
Accounts receivable	(17,935)	(768)
Inventories	(24,142)	(3,667)
Prepaid expenses and other current assets	(2,024)	(20)
Other assets	(36)	(2,686)
Accounts payable: trade	14,833	(548)
Accrued expenses and other current liabilities	7,054	3,292
Income taxes payable	3,779	3,458
Net cash provided by operating activities	78,018	83,801

Exhibit 26 (Continued)

	Year Ended 3 March 2001	Year Ended 4 March 2000
Cash flows from investing activities		
Purchase of property, plant, and equipment	(41,712)	(33,289)
Acquisitions, net of cash acquired	—	—
Purchase of short-term investments	—	—
Payments to register trademark	(199)	(277)
Net cash used in investing activities	(41,911)	(33,566)

1. What amount does Nautica report as operating cash flow for the year ended 4 March 2000 in Exhibit 25? What amount does Nautica report as operating cash flow for the same year in Exhibit 26?
2. Exhibit 25 shows that the company had investing cash flows of $21,116 thousand from the sale of short-term investments for the year ended 4 March 2000. Where does this amount appear in Exhibit 26?
3. As actually reported (Exhibit 26), how did the company's operating cash flow for fiscal year 2001 compare with that for 2000? If Nautica had not changed the classification of its short-term investing activities, how would the company's operating cash flows for fiscal year 2001 have compared with that for 2000?

Solution to 1:

In Exhibit 25, Nautica reports operating cash flow for the year ended 4 March 2000 of $62,685 thousand. In Exhibit 26, Nautica reports operating cash flow for the same year of $83,801 thousand.

Solution to 2:

The $21,116 thousand (i.e., the difference between the amounts of operating cash flow reported in Exhibits 25 and 26) that appears in Exhibit 25 as investing cash flows from the sale of short-term investments for the year ended 4 March 2000 has been reclassified. In Exhibit 26, this amount appears under changes in operating assets and liabilities (i.e., as a component of operating cash flow).

Solution to 3:

As reported in Exhibit 26, the company's cash flows declined by 7% from fiscal year 2000 to fiscal year 2001 (= 78,018/83,801 − 1 = −7%). If Nautica had not changed the classification of its short-term investing activities, the company's operating cash flows for fiscal year 2001 would have been $49,573 thousand (=78,018 − 28,445), and would have shown a decline of 21% from fiscal year 2000 to fiscal year 2001 (= 49,573/62,685 − 1 = −21%).

An analyst could have identified Nautica's classification shift by comparing the statement of cash flows for 2000 in the fiscal year 2000 annual report with the statement in the fiscal year 2001 annual report. In general, comparisons of period-to-period reports issued by a company can be useful in assessing financial reporting quality.

If a company restates prior years' financial statements (because of an error), recasts prior years' financial statements (because of a change in accounting policy), omits some information that was previously voluntarily disclosed, or adds some item, such as a new risk disclosure that was not previously disclosed, an analyst should aim to understand the reasons for the changes.

16 BALANCE SHEET QUALITY

k describe indicators of balance sheet quality;

l evaluate the balance sheet quality of a company;

With regard to the balance sheet, high financial *reporting* quality is indicated by completeness, unbiased measurement, and clear presentation. High financial *results* quality (i.e., a strong balance sheet) is indicated by an optimal amount of leverage, adequate liquidity, and economically successful asset allocation. Balance sheet strength is assessed using ratio analysis, including common-size financial statements, which is covered by the financial statement analysis readings. There are no absolute values for ratio analysis that indicate adequate financial strength; such analysis must be undertaken in the context of a firm's earnings and cash flow outlook, coupled with an understanding of the environment in which the firm operates. In this section, the focus is on high financial reporting quality.

An important aspect of financial reporting quality for the balance sheet is *completeness*. Significant amounts of off-balance-sheet obligations could be a concern for an analyst because exclusion of these obligations could understate the company's leverage. One common source of off-balance-sheet obligation is purchase contracts, which may be structured as take-or-pay contracts. Analysts typically adjust reported financial statement information by constructively capitalizing, where material, purchase obligations. Constructive capitalization means that the analyst estimates the amount of the obligation as the present value of future purchase obligation payments and then adds the amount of the obligation to the company's reported assets and liabilities.

The use of unconsolidated joint ventures or equity-method investees may reflect off-balance-sheet liabilities. In addition, certain profitability ratios (return on sales, also called "net profit margin") may be overstated because the parent company's consolidated financial statements include its share of the investee's profits but not its share of the investee's sales. If disclosures are adequate, an analyst can adjust the reported amounts to better reflect the combined amounts of sales, assets, and liabilities. A company operating with numerous or material unconsolidated subsidiaries for which ownership levels approach 50% could be a warning sign of accounting issues. Understanding why a company structures its operations in such a manner—industry practice or need for strategic alliances in certain businesses or geographies—can allay concerns.

Another important aspect of financial reporting quality for the balance sheet is *unbiased measurement*. Unbiased measurement is particularly important for assets and liabilities for which valuation is subjective. The following list presents several examples:

- As previously discussed, understatement of impairment charges for inventory; plant, property, and equipment; or other assets not only results in overstated profits on the income statement but also results in overstatement of the assets on the balance sheet. A company with substantial amounts of reported goodwill but with a market value of equity less than the book value of shareholders' equity may indicate that appropriate goodwill impairments have not been taken.

- Similarly, understatement of valuation allowance for deferred tax assets would understate tax expenses and overstate the value of the assets on the balance sheet. (Overstatement would have the opposite effect.) Significant, unexplainable variations in the valuation account can signal biased measurement.
- A company's investments in the debt or equity securities of another company would ideally be based on observable market data. For some investments, no observable market data exist and the valuation must be based solely on management estimates. The balance sheet of a company with a substantial portion of its assets valued using non-observable inputs likely warrants closer scrutiny.
- A company's pension liabilities require various estimates, such as the discount rate at which future obligations are present valued. If pension obligations exist, the level and changes for the discount rate should be examined.

Example 10 shows a company with overstated goodwill.

EXAMPLE 10

Goodwill

Sealed Air Corporation

In August 2012, a *Wall Street Journal* article listed six companies that were carrying more goodwill on their balance sheets than the companies' market values (Thurm 2012). At the top of the list was Sealed Air Corporation, a company operating in the packaging and containers industry. Exhibit 27 presents an excerpt from the company's income statement for the following year, and Exhibit 28 presents an excerpt from the company's balance sheet.

Exhibit 27 Sealed Air Corporation and Subsidiaries Consolidated Statements of Operations ($ millions, except per share amounts)

Year ended 31 December	2012	2011	2010
Net sales	$7,648.1	$5,550.9	$4,490.1
Cost of sales	5,103.8	3,950.6	3,237.3
Gross profit	2,544.3	1,600.3	1,252.8
Marketing, administrative, and development expenses	1,785.2	1,014.4	699.0
Amortization expense of intangible assets acquired	134.0	39.5	11.2
Impairment of goodwill and other intangible assets	1,892.3	—	—
Costs related to the acquisition and integration of Diversey	7.4	64.8	—
Restructuring and other charges	142.5	52.2	7.6
Operating (loss) profit	(1,417.1)	429.4	535.0
Interest expense	(384.7)	(216.6)	(161.6)
Loss on debt redemption	(36.9)	—	(38.5)
Impairment of equity method investment	(23.5)	—	—
Foreign currency exchange (losses) gains related to Venezuelan subsidiaries	(0.4)	(0.3)	5.5
Net gains on sale (other-than-temporary impairment) of available-for-sale securities	—	—	5.9
Other expense, net	(9.4)	(14.5)	(2.9)
(Loss) earnings from continuing operations before income tax provision	(1,872.0)	198.0	343.4

(continued)

Exhibit 27 (Continued)

Year ended 31 December	2012	2011	2010
Income tax (benefit) provision	(261.9)	59.5	87.5
Net (loss) earnings from continuing operations	(1,610.1)	138.5	255.9
Net earnings from discontinued operations	20.9	10.6	—
Net gain on sale of discontinued operations	178.9	—	—
Net (loss) earnings available to common stockholders	$(1,410.3)	$149.1	$255.9

Exhibit 28 Excerpt from Sealed Air Corporation and Subsidiaries Consolidated Balance Sheets ($ millions, except share data)

Year Ended 31 December	2012	2011
ASSETS		
Current assets		
Cash and cash equivalents	$679.6	$703.6
Receivables, net of allowance for doubtful accounts of $25.9 in 2012 and $16.2 in 2011	1,326.0	1,314.2
Inventories	736.4	777.5
Deferred tax assets	393.0	156.2
Assets held for sale	—	279.0
Prepaid expenses and other current assets	87.4	119.7
Total current assets	$3,222.4	$3,350.2
Property and equipment, net	$1,212.8	$1,269.2
Goodwill	3,191.4	4,209.6
Intangible assets, net	1,139.7	2,035.7
Non-current deferred tax assets	255.8	112.3
Other assets, net	415.1	455.0
Total assets	$9,437.2	$11,432.0

1 Sealed Air Corporation's financial statements indicate that the number of common shares issued and outstanding in 2011 was 192,062,185. The price per share of Sealed Air Corporation's common stock was around $18 per share in December 2011 and around $14 in August 2012; the *Wall Street Journal* article (Thurm 2012) was written in 2012. What was the company's market value?

2 How did the amount of goodwill as of 31 December 2011 compare with the company's market value?

3 Why did the *Wall Street Journal* article state that goodwill in excess of the company's market value is "a potential clue to future write-offs"?

4 Based on the information in Exhibit 28, does the *Wall Street Journal* article statement appear to be correct?

Solution to 1:

Sealed Air Corporation's market cap was about \$3,457 million (= 192,062,185 shares × \$18 per share) in December 2011 and around \$2,689 million (= 192,062,185 shares × \$14 per share) when the *Wall Street Journal* article was written in August 2012.

Solution to 2:

The amount of goodwill on Sealed Air Corporation's balance sheet as of 31 December 2011 was \$4,209.6 million. The amount of goodwill exceeded the company's market value. (Also note that goodwill and other intangible assets represented about 55% of Sealed Air Corporation's total assets as of 31 December 2011.)

Solution to 3:

If the market capitalization exactly equaled the reported amount of goodwill, the value implicitly assigned to all the company's other assets would equal zero. In this case, because the market capitalization is less than the reported amount of goodwill, the value implicitly attributed to all the company's other assets is less than zero. This suggests that the amount of goodwill on the balance sheet is overvalued, so a future write-off is likely.

Solution to 4:

Yes, based on the information in Exhibit 28, the *Wall Street Journal* article statement appears correct. In the fiscal year ending 31 December 2012 after the article, Sealed Air Corporation recorded impairment of goodwill and other intangible assets of \$1,892.3 million.

Finally, *clear presentation* is also important for financial reporting quality for the balance sheet. Although accounting standards specify many aspects of what appears on the balance sheet, companies have discretion, for example, in determining which line items should be shown separately and which should be aggregated into a single total. For items shown as a single total, an analyst can usually consult the notes for information about the components. For example, in consulting the inventory note, an analyst may learn that inventory is carried on a last-in, first-out basis and that, consequently, in an inflationary environment, the inventory is carried on the balance sheet at a cost that is significantly lower than its current cost. This information would provide the analyst with comfort that the inventory is unlikely to be overstated.

SOURCES OF INFORMATION ABOUT RISK AND LIMITED USEFULNESS OF AUDITOR'S REPORT 17

m describe sources of information about risk.

A company's financial statements can provide useful indicators of financial, operating, or other risk. For example, high leverage ratios (or, similarly, low coverage ratios) derived from financial statement data can signal financial risk. As described in a previous section, analytical models that incorporate various financial data can signal bankruptcy risk, and others can predict reporting risks (i.e., the risk of a company misreporting). Operating risks can be indicated by financial data, such as highly variable operating cash flows or negative trends in profit margins. Additional information about risk can be obtained from sources other than the financial statements.

An audit opinion(s) covering financial statements (and internal controls over financial reporting, where required) can provide some information about reporting risk. However, the content of an audit opinion is unlikely to be a timely source of information about risk. A related item that is potentially a signal of problems (and thus potentially represents information about risk) is a discretionary change in auditor. For example, Allou Health & Beauty Care, discussed in Example 7, had a different auditor for 2000, 2001, and 2002.

The notes are an integral part of the financial statements. They typically contain information that is useful in understanding a company's risk. Beyond the information about risk that can be derived from a company's financial statements and notes, various other disclosures can provide information about financial, operating, reporting, or other risks. An important source of information is the management commentary, which provides management's assessment of the important risks faced by the company. Although risk-related disclosures in the management commentary sometimes overlap with disclosures contained in the financial statement notes or elsewhere in regulatory filings, the commentary should reveal the management perspective, and its content often differs from the note disclosures.

Other required disclosures that are specific to an event, such as capital raising, non-timely filing of financial reports, management changes, or mergers and acquisitions, can provide important information relevant to assessing risk. Finally, the financial press, including online media, if used judiciously, can be a useful source of information about risk.

17.1 Limited Usefulness of Auditor's Opinion as a Source of Information about Risk

An auditor's opinion is unlikely to be an analyst's first source of information about a company's risk.[24] For financial statements, a clean audit opinion states that the financial statements present the information fairly and in conformity with the relevant accounting principles. For internal controls, a clean audit opinion states that the company maintained effective internal controls over financial reporting. A negative or going-concern audit opinion on financial statements or a report indicating an internal control weakness would clearly be a warning sign for an analyst. However, an audit opinion relates to historical information and would, therefore, typically not provide information on a timely enough basis to be a useful source of information about risk.

For example, Eastman Kodak Company filed for bankruptcy on 19 January 2012. The audit opinion for fiscal 2011 (dated 28 February 2012) is shown in Exhibit 29. The opinion is identical to the company's audit opinion for the prior fiscal year except for two differences: (1) the years have been updated, and (2) the paragraph highlighted in bold has been added. The added paragraph states that the financial statements were prepared under the "going-concern" assumption; the company has subsequently declared bankruptcy, which raises doubt about the company's ability to continue as a going concern; and the financial statements have not been adjusted to reflect the bankruptcy. An analyst would have learned about Eastman Kodak's bankruptcy on 19 January, so the audit opinion is not useful as a source of that information. In addition, the audit opinion addresses financial statements that had not been adjusted to reflect the bankruptcy, which would limit usefulness to an analyst.

24 Regulators globally are considering changes to increase the usefulness of audit reports. For example, the Financial Reporting Council in the UK requires auditors to include more information in their reports on risks identified during the audit and on how the concept of materiality was applied.

Exhibit 29 Post-Bankruptcy Audit Opinion for Eastman Kodak

Report of Independent Registered Public Accounting Firm

To the Board of Directors and Shareholders of Eastman Kodak Company:

In our opinion, the consolidated financial statements listed in the index appearing under Item 15(a)(1) present fairly, in all material respects, the financial position of Eastman Kodak Company and its subsidiaries at December 31, 2011 and 2010, and the results of their operations and their cash flows for each of the three years in the period ended December 31, 2011 in conformity with accounting principles generally accepted in the United States of America. In addition, in our opinion, the financial statement schedule listed in the index appearing under Item 15(a)(2) presents fairly, in all material respects, the information set forth therein when read in conjunction with the related consolidated financial statements. Also in our opinion, the Company maintained, in all material respects, effective internal control over financial reporting as of December 31, 2011, based on criteria established in *Internal Control - Integrated Framework* issued by the Committee of Sponsoring Organizations of the Treadway Commission (COSO). The Company's management is responsible for these financial statements and financial statement schedule, for maintaining effective internal control over financial reporting and for its assessment of the effectiveness of internal control over financial reporting, included in Management's Report on Internal Control over Financial Reporting appearing under Item 9A. Our responsibility is to express opinions on these financial statements, on the financial statement schedule, and on the Company's internal control over financial reporting based on our integrated audits. We conducted our audits in accordance with the standards of the Public Company Accounting Oversight Board (United States). Those standards require that we plan and perform the audits to obtain reasonable assurance about whether the financial statements are free of material misstatement and whether effective internal control over financial reporting was maintained in all material respects. Our audits of the financial statements included examining, on a test basis, evidence supporting the amounts and disclosures in the financial statements, assessing the accounting principles used and significant estimates made by management, and evaluating the overall financial statement presentation. Our audit of internal control over financial reporting included obtaining an understanding of internal control over financial reporting, assessing the risk that a material weakness exists, and testing and evaluating the design and operating effectiveness of internal control based on the assessed risk. Our audits also included performing such other procedures as we considered necessary in the circumstances. We believe that our audits provide a reasonable basis for our opinions.

The accompanying financial statements have been prepared assuming that the Company will continue as a going concern. As more fully discussed in Note 1 to the financial statements, on January 19, 2012, the Company and its US subsidiaries filed voluntary petitions for relief under chapter 11 of the United States Bankruptcy Code. Uncertainties inherent in the bankruptcy process raise substantial doubt about the Company's ability to continue as a going concern. Management's plans in regard to these matters are also described in Note 1. The accompanying financial statements do not include any adjustments that might result from the outcome of this uncertainty.

A company's internal control over financial reporting is a process designed to provide reasonable assurance regarding the reliability of financial reporting and the preparation of financial statements for external purposes in accordance with generally accepted accounting principles. A company's internal control over

(continued)

Exhibit 29 (Continued)

financial reporting includes those policies and procedures that (i) pertain to the maintenance of records that, in reasonable detail, accurately and fairly reflect the transactions and dispositions of the assets of the company; (ii) provide reasonable assurance that transactions are recorded as necessary to permit preparation of financial statements in accordance with generally accepted accounting principles, and that receipts and expenditures of the company are being made only in accordance with authorizations of management and directors of the company; and (iii) provide reasonable assurance regarding prevention or timely detection of unauthorized acquisition, use, or disposition of the company's assets that could have a material effect on the financial statements.

Because of its inherent limitations, internal control over financial reporting may not prevent or detect misstatements. Also, projections of any evaluation of effectiveness to future periods are subject to the risk that controls may become inadequate because of changes in conditions, or that the degree of compliance with the policies or procedures may deteriorate.

/s/ PricewaterhouseCoopers LLP

PricewaterhouseCoopers LLP
Rochester, New York
February 28, 2012

Note: Bold-face type is added for emphasis.

In the case of Kodak, an analyst would not have obtained very useful information about risk from the auditor's report. Other sources of information—financial and market data—would have provided clear and timely indications of the company's financial difficulty.

Groupon provides another example of the timing of availability of information about risk in external auditors' reports. Exhibit 30 presents a timeline of events related to the company's material weakness in internal controls. Note that no negative external auditor opinion appeared before or during the time frame in which the weakness existed. No external opinion was required for the first annual filing, and the weakness had been remedied by the second annual filing.

Exhibit 30 Material Weaknesses in Internal Controls at Groupon

November 2011:	The company goes public (initial public offering)
March 2012:	The company revises financial results and discloses that management concluded there was a "material weakness" in internal controls over financial reporting, as of 31 December. Shares fall 17%. (Because of an exemption for newly public companies, no external auditor opinion on the effectiveness of internal controls was required.)
May 2012:	In its first-quarter filing, the company discloses that it is "taking steps" to correct the weaknesses but cannot provide assurance that internal controls will be considered effective by the end of the year.
August 2012:	Second-quarter filing includes a disclosure similar to that in first-quarter filing.

Exhibit 30 (Continued)

November 2012:	Third-quarter filing includes a disclosure similar to that in first-quarter filing.
February 2013:	Full-year filing indicates that the company "concluded that we have remediated the previously identified material weakness as of December 31, 2012." (As required for public companies, the filing includes Groupon's first external auditor opinion on the effectiveness of internal controls. The company received a clean opinion.)

In the case of Groupon, an analyst would not have obtained any useful information from the auditor's report. Other data would have given more useful indicators of the company's reporting difficulties. For example, the company was required to change its revenue recognition policy and to restate the amount of revenue reported in its IPO filing—clearly a sign of reporting difficulties. Another item of information providing a signal of likely reporting difficulties was the company's extensive number of acquisitions and explosive growth. Groupon's reported revenues for 2009 were more than 300 times the amount of 2008 reported revenues, and 2010 reported revenues were 23 times larger than 2009 revenues. As described in an August 2011 accounting blog (Catanach and Ketz 2011):

> It is absolutely ludicrous to think that Groupon is anywhere close to having an effective set of internal controls over financial reporting having done 17 acquisitions in a little over a year. When a company expands to 45 countries, grows merchants from 212 to 78,466, and expands its employee base from 37 to 9,625 in only two years, there is little doubt that internal controls are not working somewhere.

The growth data, particularly coupled with disclosures in the IPO filing about management inexperience, are a warning sign of potential reporting risks. These reporting risks were observable many months before the company disclosed its internal control weakness, and the control weaknesses did not appear in an audit opinion.

Although the content of an audit opinion is unlikely to provide timely information about risk, a change in the auditor—and especially multiple changes in the auditor—can signal possible reporting problems. For example, one of the largest feeder funds for Bernie Madoff (the perpetrator of a multi-billion-dollar Ponzi scheme) had three different auditors for the three years from 2004 to 2006, a fact highlighted in testimony as a huge warning sign indicating "auditor shopping."[25] Similarly, the use of an auditor whose capabilities seem inadequate for the complexity of the company can indicate risk. For example, the accounting/auditing firm that audited Madoff's $50 billion operation consisted of three people (two principals and a secretary). The small size of the auditing firm relative to the size of Madoff's operations should have caused serious concern for any potential investor. In general, it is important to understand the relationship between the auditor and the firm. Any questions about the auditor's independence would be a cause for concern—for example, if the auditor and company management are particularly close or if the company represents a substantial portion of the auditing firm's revenue.

25 From the testimony of Harry Markopolos, CFA, given before the US House of Representatives Committee on Financial Services, 4 February 2009.

18 RISK-RELATED DISCLOSURES IN THE NOTES

m describe sources of information about risk.

The notes, an integral part of the financial statements, typically contain information that is useful in understanding a company's risk. For example, both IFRS and US GAAP require specific disclosures about risks related to contingent obligations, pension and post-employment benefits, and financial instrument risks.

Disclosures about contingent obligations include a description of the obligation, estimated amounts, timing of required payments, and related uncertainties.[26] Exhibit 31 shows excerpts from two of Royal Dutch Shell's financial statement notes disclosing information about provisions and contingencies. The year-to-year changes in management's estimated costs for items such as future decommissioning and restoration could have implications for risk evaluation. The disclosure also emphasizes the uncertain timing and amounts.

Exhibit 31 Disclosures about Contingent Obligations, Excerpt from Royal Dutch Shell's Note 19 and Note 25

19 Decommissioning and Other Provisions

	Current		Non-Current		Total	
	31 Dec 2012	31 Dec 2011	31 Dec 2012	31 Dec 2011	31 Dec 2012	31 Dec 2011
Decommissioning and restoration	1,356	894	14,715	13,072	16,071	13,966
Environmental	366	357	1,032	1,078	1,398	1,435
Redundancy	228	406	275	297	503	703
Litigation	390	256	307	330	697	586
Other	881	1,195	1,106	854	1,987	2,049
Total	3,221	3,108	17,435	15,631	20,656	18,739

The timing and amounts settled in respect of these provisions are uncertain and dependent on various factors that are not always within management's control. Additional provisions are stated net of reversals of provisions recognised in previous periods.

Of the decommissioning and restoration provision at December 31, 2012, an estimated $4,666 million is expected to be utilised within one to five years, $3,483 million within six to ten years, and the remainder in later periods.

Reviews of estimated decommissioning and restoration costs are carried out annually, which in 2012 resulted in an increase of $1,586 million …

26 Contingent losses are recognized (i.e., reported on the financial statements) when it is probable the loss will occur and the amount can be reasonably estimated. Contingencies are disclosed (but not recognized) when the occurrence of a loss is less than probable but greater than remote and/or the amount cannot be reliably estimated. The concepts are similar under IFRS and US GAAP despite differences in terminology. IFRS makes a distinction between "provisions," which are recognized as liabilities because they meet the definition of a liability, and "contingent liabilities," which are disclosed but not recognized.

Exhibit 31 (Continued)

25 Legal Proceedings and Other Contingencies

Groundwater contamination

Shell Oil Company (including subsidiaries and affiliates, referred to collectively as SOC), along with numerous other defendants, has been sued by public and quasi-public water purveyors, as well as governmental entities. The plaintiffs allege responsibility for groundwater contamination caused by releases of gasoline containing oxygenate additives. Most of these suits assert various theories of liability, including product liability, and seek to recover actual damages, including clean-up costs. Some assert claims for punitive damages. Fewer than 10 of these cases remain. On the basis of court rulings in SOC's favour in certain cases claiming damages from threats of contamination, the claims asserted in remaining matters, and Shell's track record with regard to amounts paid to resolve varying claims, the management of Shell currently does not believe that the outcome of the remaining oxygenate-related litigation pending, as at December 31, 2012, will have a material impact on Shell.

Nigerian claims

Shell subsidiaries and associates operating in Nigeria are parties to various environmental and contractual disputes. These disputes are at different stages in litigation, including at the appellate stage, where judgments have been rendered against Shell. If taken at face value, the aggregate amount of these judgments could be seen as material. The management of Shell, however, believes that these matters will ultimately be resolved in a manner favourable to Shell. While no assurance can be provided as to the ultimate outcome of any litigation, these matters are not expected to have a material effect on Shell.

Other

In the ordinary course of business, Shell subsidiaries are subject to a number of other loss contingencies arising from litigation and claims brought by governmental and private parties. The operations and earnings of Shell subsidiaries continue, from time to time, to be affected to varying degrees by political, legislative, fiscal and regulatory developments, including those relating to the protection of the environment and indigenous groups, in the countries in which they operate. The industries in which Shell subsidiaries are engaged are also subject to physical risks of various types. The nature and frequency of these developments and events, as well as their effect on future operations and earnings, are unpredictable.

Disclosures about pensions and post-employment benefits include information relevant to actuarial risks that could result in actual benefits differing from the reported obligations based on estimated benefits or investment risks that could result in actual assets differing from reported amounts based on estimates.

Disclosures about financial instruments include information about risks, such as credit risk, liquidity risk, and market risks that arise from the company's financial instruments, and how they have been managed.

EXAMPLE 11

Use of Disclosures

Use the excerpts from Royal Dutch Shell's note disclosing information about financial instruments in Exhibit 32 to answer the following questions:

1. Does Shell appear to take a centralized or decentralized approach to managing interest rate risk?
2. For the year ended 31 December 2012, Shell reported pre-tax income of $50,289 million. How significant is Shell's exposure to a 1% increase in interest rates?
3. For the year ended 31 December 2012, what would be the impact on Shell's pre-tax income of a 10% appreciation of the Australian dollar against the US dollar?

Exhibit 32 Disclosures about Financial Instruments, Excerpt from Royal Dutch Shell's Note 21

21 Financial Instruments and Other Derivative Contracts

A – Risks

In the normal course of business, financial instruments of various kinds are used for the purposes of managing exposure to interest rate, currency and commodity price movements.

....

Interest rate risk

Most debt is raised from central borrowing programmes. Interest rate swaps and currency swaps have been entered into to effectively convert most centrally issued debt to floating rate linked to dollar Libor (London Inter-Bank Offer Rate), reflecting Shell's policy to have debt principally denominated in dollars and to maintain a largely floating interest rate exposure profile. Consequently, Shell is exposed predominantly to dollar Libor interest rate movements. The financing of most subsidiaries is also structured on a floating-rate basis and, except in special cases, further interest rate risk management is discouraged.

On the basis of the floating rate net debt position at December 31, 2012, and assuming other factors (principally foreign exchange rates and commodity prices) remained constant and that no further interest rate management action were taken, an increase in interest rates of 1% would decrease pre-tax income by $27 million (2011: $146 million).

Foreign exchange risk

Many of the markets in which Shell operates are priced, directly or indirectly, in dollars. As a result, the functional currency of most Upstream companies and those with significant cross-border business is the dollar. For Downstream companies, the local currency is typically the functional currency. Consequently, Shell is exposed to varying levels of foreign exchange risk when it enters into transactions that are not denominated in the companies' functional currencies, when foreign currency monetary

Exhibit 32 (Continued)

assets and liabilities are translated at the reporting date and as a result of holding net investments in operations that are not dollar-functional. The main currencies to which Shell is exposed are sterling, the Canadian dollar, euro and Australian dollar. Each company has treasury policies in place that are designed to measure and manage its foreign exchange exposures by reference to its functional currency.

Exchange rate gains and losses arise in the normal course of business from the recognition of receivables and payables and other monetary items in currencies other than individual companies' functional currency. Currency exchange risk may also arise in connection with capital expenditure. For major projects, an assessment is made at the final investment decision stage whether to hedge any resulting exposure.

Hedging of net investments in foreign operations or of income that arises in foreign operations that are non-dollar functional is not undertaken.

Assuming other factors (principally interest rates and commodity prices) remained constant and that no further foreign exchange risk management action were taken, a 10% appreciation against the dollar at December 31 of the main currencies to which Shell is exposed would have the following pre-tax effects:

	Increase (decrease) in income		Increase in net assets	
$ millions	**2012**	**2011**	**2012**	**2011**
10% appreciation against the dollar of:				
Sterling	(185)	(58)	1,214	1,042
Canadian dollar	131	(360)	1,384	1,364
Euro	30	458	1,883	1,768
Australian dollar	246	153	142	120

The above sensitivity information is calculated by reference to carrying amounts of assets and liabilities at December 31 only. The pre-tax effect on income arises in connection with monetary balances denominated in currencies other than the relevant entity's functional currency; the pre-tax effect on net assets arises principally from the translation of assets and liabilities of entities that are not dollar-functional.

Solution to 1:

Shell appears to take a centralized approach to managing interest rate risk based on its statements that most debt is raised centrally and that interest rate swaps and currency swaps have been used to convert most interest rate exposure to dollar Libor. In addition, Shell states that apart from structuring subsidiary financing on a floating-rate basis, it discourages subsidiary's further interest rate risk management.

Solution to 2:

For the year ended 31 December 2012, Shell's exposure to a 1% increase in interest rates is relatively insignificant. An increase in interest rates of 1% would decrease pre-tax income by $27 million, which is less than 0.1% of Shell's 2012 reported pre-tax income of $50,289 million.

Solution to 3:

The impact on Shell's pre-tax income of a 10% appreciation of the Australian dollar against the US dollar would be an increase of $246 million, which is about 0.5% of Shell's 2012 reported pre-tax income of $50,289 million.

These disclosures, along with expectations about future market conditions, can help an analyst assess whether the company's exposures to interest rate risk and foreign exchange risks pose a significant threat to the company's future performance.

19 MANAGEMENT COMMENTARY (MD&A), OTHER REQUIRED DISCLOSURES, FINANCIAL PRESS

m describe sources of information about risk.

The IFRS Practice Statement, *Management Commentary*, issued in December 2010, is a non-binding framework for commentary related to financial statements prepared in accordance with IFRS. One purpose of the commentary is to help users of the financial reports in understanding the company's risk exposures, approach to managing risks, and effectiveness of risk management. The practice statement includes five elements that should be contained in the commentary: (1) nature of the business; (2) objectives and strategies; (3) resources, risks, and relationships; (4) results and prospects; and (5) performance measures and indicators. The section on risks can be particularly useful (IFRS 2010).

> Management should disclose its principal strategic, commercial, operational, and financial risks, which are those that may significantly affect the entity's strategies and progress of the entity's value. The description of the principal risks facing the entity should cover both exposures to negative consequences and potential opportunities.... The principal risks and uncertainties can constitute either a significant external or internal risk to the entity. (p. 13)

Public US companies are required to include an MD&A as Item 7 of Form 10-K. The MD&A disclosures include information about (1) liquidity, (2) capital resources, (3) results of operations, (4) off-balance-sheet arrangements, and (5) contractual arrangements. Information about off-balance-sheet arrangements and contractual arrangements can enable an analyst to anticipate future impact on cash flow. Companies are required to present quantitative and qualitative information about the company's exposure to market risks as Item 7A of the 10-K. This disclosure should enable analysts to understand the impact of fluctuations in interest rates, foreign exchange, and commodity prices.[27]

The IFRS Practice Statement states specifically that companies should present only the principal risks and not list all possible risks and uncertainties. Similarly, the SEC Division of Corporation Finance's internal reference document, *Financial Reporting Manual*, states, "MD&A should not consist of generic or boilerplate disclosure. Rather, it should reflect the facts and circumstances specific to each individual registrant" (p. 296). In practice, disclosures do not always reflect the intent. One challenge faced by analysts is identifying important risks and distinguishing between risks that are generic and thus relevant to all companies and risks that are more specific to an individual company.

This challenge is illustrated by an excerpt from the "Key Risks and Uncertainties" section of Autonomy Corporation's 2010 Annual Report, its last annual report before it was acquired by Hewlett-Packard Company (HP) for $11.1 billion in 2011.[28] As shown in Exhibit 33, Autonomy's risk disclosures contain many items that are arguably generic, such as the inability to maintain the competitive value of its technology, loss of key executives, and continued unfavorable economic conditions. These types of risks would be faced by any technology company. This significant amount of generic commentary (two pages) could potentially distract a reader whose aim was to identify the specific and important risks faced by the company.

Exhibit 33 Autonomy Corporation, Key Risks and Uncertainties

Risk	Description	Impact/Sensitivity	Mitigation/Comment
Technology	Business depends on our core technology, and our strategy concentrates on developing and marketing software based on our proprietary technology.	Since substantially all of revenues derive from licensing our core technology, if unable to maintain and enhance the competitive value of our core technology, our business will be adversely affected.	Continue to invest heavily in research and development to maintain competitive advantage. Monitor market to maintain competitiveness. Apply core technology to new and additional vertical market applications.
Competition	Technology which significantly competes with our technology.	Could render our products out of date and could result in rapid loss of market share.	Invest heavily in new product development to ensure that we have products at various stages of the product life cycle.

(continued)

27 Although not part of the MD&A, disclosures about risk factors relevant to the company's securities are also required as Item 1A of Form 10-K.

28 HP subsequently took a multi-billion-dollar write-down on its investment, which it attributed to misreporting by Autonomy Corporation, stating that "the majority of this impairment charge is linked to serious accounting improprieties, disclosure failures and outright misrepresentations at Autonomy Corporation plc that occurred prior to HP's acquisition of Autonomy and the associated impact of those improprieties, failures and misrepresentations on the expected future financial performance of the Autonomy business over the long-term" (HP earnings announcement, 20 November 2012). Of course, HP's due diligence prior to purchasing the company would have gone far beyond the published financial reports; HP would have had access to all of the company's internal reporting as well.

Exhibit 33 (Continued)

Risk	Description	Impact/Sensitivity	Mitigation/Comment
Variability and visibility	There may be fluctuations in results due to quarterly reporting, and variability in results due to late-in-the-quarter purchasing cycles common in the software industry.	Although quarter-to-quarter results may not be meaningful due to the short periods, negative sentiment may arise based on interpretation of results. Due to late purchasing cycles common in the software industry, variability in closure rates could become exaggerated resulting in a negative effect on operations.	Close management of sales pipelines on a quarterly basis to improve visibility in results expectations. Close monitoring of macro and micro economic conditions to understand variability in closure rates. Annual and quarterly target setting to enable results achievement.
Margins	Expenditures increasing without a commensurate increase in revenues, and rapid changes in market conditions.	If increased expenses are not accompanied by increased revenues, we could experience decreased margins or operating losses.	Close monitoring by management of revenue and cost forecasts. Adjustment to expenditures in the event of anticipated revenue shortfalls.
Average selling prices	The average selling prices of our products could decrease rapidly.	May negatively impact revenues and gross margins.	Monitor market prices on an ongoing basis. Pricing responsibility at a senior level of management for deviations from standard.
Market conditions	The continuation of unfavourable economic and market conditions.	Could result in a rapid deterioration of operating results.	Regular monitoring of economic conditions. Adjustments to costs and product offerings to anticipate and match market conditions.
Resellers	Our ability to expand sales through indirect sellers and our general reliance on sales of our products by third parties.	Inability to recruit and retain resellers who can successfully penetrate their markets could adversely affect our business.	Invest in training resources for resellers. Close monitoring of reseller sales cycles. Investment in direct sales channel.
Management	The continued service of our executive directors.	The loss of any key member of management may affect the leadership of the company.	Establish succession plan. Maintain effective management training programme. Attract and retain senior personnel.
Hiring	The hiring and retention of qualified personnel.	Without the appropriate quality and quantity of skills throughout the organisation, it would be difficult to execute the business plans and grow.	Use of external recruiters and internal bonuses. Rigorous talent management plans and reviews. Provide competitive compensation packages. Ensure that work is challenging and rewarding.
Product errors	Errors or defects in our products.	Could negatively affect our revenues and the market acceptance of our products and increase our costs.	Invest in quality control programmes. Monitor integrity and effectiveness of software. Solicit and act on customer feedback.
Acquisitions	Problems encountered in connection with potential acquisitions.	We may not successfully overcome problems in connection with potential acquisitions, which could lead to a deterioration in our results.	Carefully evaluate transactions. Conduct thorough due diligence on all targets. Carefully plan for post-acquisition integration.
IP infringement	Claims by others that we infringe on their intellectual property rights.	If our technology infringed on other parties' intellectual property rights, we could be exposed to costs and injunctive relief.	Monitor market developments closely to identify potential violations of our patents, and by the company, and take action where necessary. Maintain a significant number of patents to support our business and protect competitive advantage.

Exhibit 33 (Continued)

Risk	Description	Impact/Sensitivity	Mitigation/Comment
Growth	Our ability to effectively manage our growth.	Expansion places demands on management, engineering, support, operations, legal, accounting, sales and marketing personnel, and other resources. Failure to manage effectively will impact business and financial results	Recruitment and retention of key personnel. Investment in corporate infrastructure, including support, operations, legal, and accounting personnel. Focus on internal controls.
International risks	Additional operational and financial risks as we continue to expand our international operations.	Exposure to movements in exchange rates and lack of familiarity with local laws could lead to infractions.	Pricing of contracts in US dollars to the extent possible to minimise exchange risk. Retention of local staff and local advisors, reporting to headquarters, to manage risk.
Security breaches	Any breach of our security measures and unauthorised access to a customer's or our data.	Could result in significant legal liability and negative publicity.	Establish and maintain strict security standards. Test security standards on a regular basis.

Source: Section from Autonomy Corporation's 2010 Annual Report.

19.1 Other Required Disclosures

Other required disclosures that are specific to an event, such as capital raising, non-timely filing of financial reports, management changes, or mergers and acquisitions, can provide important information relevant to assessing risk. In the United States, public companies would report such events to the SEC in a Form 8-K (and NT—"notification of inability to timely file"—when appropriate). Delays in filing are often the result of accounting difficulties. Such accounting difficulties could be internal disagreement on an accounting principle or estimate, the lack of adequate financial staff, or the discovery of an accounting fraud that requires further examination. In general, an NT filing is highly likely to signal problems with financial reporting quality.

For public companies in Europe, the Committee of European Securities Regulators (CESR)[29] has published guidance concerning the types of inside information that must be disclosed on an ad hoc basis to the market. Examples of such information include changes in control; changes in management and supervisory boards; mergers, splits, and spinoffs; legal disputes; and new licenses, patents, and registered trademarks. Companies use the disclosure mechanisms specified by their relevant national authorities to make such disclosures. For example, in the United Kingdom, a company would release an announcement to the market via an approved regulatory information service.

In these cases, an examination of the information announced would be necessary to determine whether reporting quality would be affected. For example, an announcement of the sudden resignation of a company's most senior financial officer or external auditor would clearly be a warning sign of potential problems with financial reporting quality. As another example, an announcement of a legal dispute related to one of the company's important assets or products would warrant attention because it could negatively affect the company's future earnings. Announcements of mergers and acquisitions, although they might indicate future positive developments for the company, could also indicate changes in the company's risk profile, particularly during the transaction.

29 CESR has been replaced by the European Securities and Markets Authority (ESMA).

19.2 Financial Press as a Source of Information about Risk

The financial press can be a useful source of information about risk when, for example, a financial reporter uncovers financial reporting issues that had not previously been recognized. For example, a *Wall Street Journal* financial reporter, Jonathan Weil (2000), was one of the first people to identify problems with the accounting at Enron (and other companies that were using "gain-on-sale" accounting, an aggressive policy allowing immediate revenue recognition on long-term contracts). Indeed, the well-known investor James (Jim) Chanos cites an article by Weil as the catalyst of his investigation of Enron (Chanos 2002).

It is important to emphasize that even if an initial idea comes from a news article, further investigation is essential—first, by using definitive sources (i.e., regulatory filings) to confirm any accounting and financial disclosures and, second, by seeking supporting information from other sources, where available. For example, although a financial press article was the initial source of information for Chanos, the first step in his research was to analyze Enron's annual SEC filings (Form 10-K and 10-Q). In addition, Chanos obtained information about insider stock sales, the company's business strategy and tactics, and stock analysts' perspectives.

It is also important—and likely will become increasingly important as electronic media via the internet expands—to consider the source of any particular news article. Information reported by a well-known financial news provider is more likely to be factual than information from less-established sources. Similarly, stories or blogs written by financial journalists are more likely to be unbiased than those written by individuals with a related service or product to sell.

SUMMARY

Assessing the quality of financial reports—both reporting quality and results quality—is an important analytical skill.

- The quality of financial reporting can be thought of as spanning a continuum from the highest quality to the lowest.
- Potential problems that affect the quality of financial reporting broadly include revenue and expense recognition on the income statement; classification on the statement of cash flows; and the recognition, classification, and measurement of assets and liabilities on the balance sheet.
- Typical steps involved in evaluating financial reporting quality include an understanding of the company's business and industry in which the company is operating; comparison of the financial statements in the current period and the previous period to identify any significant differences in line items; an evaluation of the company's accounting policies, especially any unusual revenue and expense recognition compared with those of other companies in the same industry; financial ratio analysis; examination of the statement of cash flows with particular focus on differences between net income and operating cash flows; perusal of risk disclosures; and review of management compensation and insider transactions.
- High-quality earnings increase the value of the company more than low-quality earnings, and the term "high-quality earnings" assumes that reporting quality is high.

- Low-quality earnings are insufficient to cover the company's cost of capital and/or are derived from non-recurring, one-off activities. In addition, the term "low-quality earnings" can be used when the reported information does not provide a useful indication of the company's performance.
- Various alternatives have been used as indicators of earnings quality: recurring earnings, earnings persistence and related measures of accruals, beating benchmarks, and after-the-fact confirmations of poor-quality earnings, such as enforcement actions and restatements.
- Earnings that have a significant accrual component are less persistent and thus may revert to the mean more quickly.
- A company that consistently reports earnings that exactly meet or only narrowly beat benchmarks can raise questions about its earnings quality.
- Cases of accounting malfeasance have commonly involved issues with revenue recognition, such as premature recognition of revenues or the recognition of fraudulent revenues.
- Cases of accounting malfeasance have involved misrepresentation of expenditures as assets rather than as expenses or misrepresentation of the timing or amount of expenses.
- Bankruptcy prediction models, used in assessing financial results quality, quantify the likelihood that a company will default on its debt and/or declare bankruptcy.
- Similar to the term "earnings quality," when reported cash flows are described as being high quality, it means that the company's underlying economic performance was satisfactory in terms of increasing the value of the firm, and it also implies that the company had high reporting quality (i.e., that the information calculated and disclosed by the company was a good reflection of economic reality). Cash flow can be described as "low quality" either because the reported information properly represents genuinely bad economic performance or because the reported information misrepresents economic reality.
- For the balance sheet, high financial *reporting* quality is indicated by completeness, unbiased measurement, and clear presentation.
- A balance sheet with significant amounts of off-balance-sheet debt would lack the completeness aspect of financial reporting quality.
- Unbiased measurement is a particularly important aspect of financial reporting quality for assets and liabilities for which valuation is subjective.
- A company's financial statements can provide useful indicators of financial or operating risk.
- The management commentary (also referred to as the management discussion and analysis, or MD&A) can give users of the financial statements information that is helpful in assessing the company's risk exposures and approaches to managing risk.
- Required disclosures regarding, for example, changes in senior management or inability to make a timely filing of required financial reports can be a warning sign of problems with financial reporting quality.
- The financial press can be a useful source of information about risk when, for example, a financial reporter uncovers financial reporting issues that had not previously been recognized. An analyst should undertake additional investigation of any issue identified.

REFERENCES

"Ahold: Europe's Enron." 2003. *The Economist* (27 February).

Altman, Edward I. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance*, vol. 23, no. 4 (September):589–609.

Altman, Edward I. 2000. "Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta® Models." Working paper (July).

Beneish, Messod D. 1999. "The Detection of Earnings Manipulation." *Financial Analysts Journal*, vol. 55, no. 5 (September/October):24–36.

Beneish, Messod D., Charles M.C. Lee, and D. Craig Nichols. 2013. "Earnings Manipulation and Expected Returns." *Financial Analysts Journal*, vol. 69, no. 2 (March/April):57–82.

Bens, Daniel A., Theodore H. Goodman, and Monica Neamtiu. 2012. "Does Investment-Related Pressure Lead to Misreporting? An Analysis of Reporting Following M&A Transactions." *Accounting Review*, vol. 87, no. 3 (May):839–865.

Beresford, Dennis R., Nicholas deB. Katzenbach, and C.B. Rogers, Jr. 2003. "Report of Investigation by the Special Investigative Committee of the Board of Directors of WorldCom, Inc." (31 March): www.sec.gov/Archives/edgar/data/723527/000093176303001862/dex991.htm.

Bharath, Sreedhar T., and Tyler Shumway. 2008. "Forecasting Default with the Merton Distance to Default Model." *Review of Financial Studies*, vol. 21, no. 3 (May):1339–1369.

Bhasin, Madan. 2012. "Corporate Accounting Frauds: A Case Study of Satyam Computers Limited." *International Journal of Contemporary Business Studies*, vol. 3, no. 10 (October):16–42.

Brown, Lawrence D., and Marcus L. Caylor. 2005. "A Temporal Analysis of Quarterly Earnings Thresholds: Propensities and Valuation Consequences." *Accounting Review*, vol. 80, no. 2 (April):423–440.

Bulkeley, W. 2002. "Questioning the Books: IBM Annual Report Shows Stronger Core Earnings." *Wall Street Journal* (12 March).

Burgstahler, D., and Ilia Dichev. 1997. "Earnings Management to Avoid Earnings Decreases and Losses." *Journal of Accounting and Economics*, vol. 24, no. 1 (December):99–126.

Catanach, Anthony H., and J. Edward Ketz. 2011. "Trust No One, Particularly Not Groupon's Accountants," Grumpy Old Accountants (August): http://blogs.smeal.psu.edu/grumpyoldaccountants.

Chanos, James. 2002. "Anyone Could Have Seen Enron Coming: Prepared Witness Testimony Given Feb. 6, 2002 to the House Committee on Energy and Commerce," *Wall $treet Week with FORTUNE* (http://www.pbs.org/wsw/opinion/chanostestimony.html).

Ciesielski, Jack T., and Elaine Henry. 2017. "Accounting's Tower of Babel: Key Considerations in Assessing Non-GAAP Earnings." *Financial Analysts Journal*, vol. 73, no. 2:34–50.

Dechow, Patricia M., Richard G. Sloan, and Amy P. Sweeney. 1995. "Detecting Earnings Management." *Accounting Review*, vol. 70, no. 2 (April):193–225.

Dechow, Patricia M., Scott A. Richardson, and Irem Tuna. 2003. "Why Are Earnings Kinky? An Examination of the Earnings Management Explanation." *Review of Accounting Studies*, vol. 8, no. 2–3 (June):355–384.

Dechow, Patricia M., Weili Ge, and Catherine Schrand. 2010. "Understanding Earnings Quality: A Review of the Proxies, Their Determinants and Their Consequences." *Journal of Accounting and Economics*, vol. 50, no. 2–3 (December):344–401.

Dechow, Patricia, Seili Ge, Chad Larson, and Richard Sloan. 2011. "Predicting Material Accounting Misstatements." *Contemporary Accounting Research*, vol. 28, no. 1 (Spring):17–82.

Degeorge, François, Jayendu Patel, and Richard Zeckhauser. 1999. "Earnings Management to Exceed Thresholds." *Journal of Business*, vol. 72, no. 1 (January):1–33.

Fuji Electric. 2013. "Announcement of Impairment Losses on Noncurrent Assets (Extraordinary Losses)." memo (25 April): www.fujielectric.com/company/news/box/doc/130425_evaluation.pdf.

Erickson, Merle, and Shiing-wu Wang. 1999. "Earnings Management by Acquiring Firms in Stock for Stock Mergers." *Journal of Accounting and Economics*, vol. 27, no. 2 (April):149–176.

Erickson, M., S. Heitzman, and X.F. Zhang. 2012. "The Effect of Financial Misreporting on Corporate Mergers and Acquisitions." Working paper.

Friedman, Eric, Simon Johnson, and Todd Mitton. 2003. "Propping and Tunneling." *Journal of Comparative Economics*, vol. 31, no. 4 (December):732–750.

Hwang, S.L. 1994. "Borden to Reverse, Reclassify 40% of 1992 Charge." *Wall Street Journal* (22 March).

IASB. 2010. *Conceptual Framework for Financial Reporting 2010*. International Accounting Standards Board (September).

IFRS. 2010. *Management Commentary, A Framework for Presentation*. IFRS Practice Statement (December).

Johnson, S., R. LaPorta, A. Shleifer, and F. Lopez-de-Silanes. 2000. "Tunneling." *American Economic Review*, vol. 90, no. 2 (May):22–27.

Jones, Jennifer J. 1991. "Earnings Management during Import Relief Investigations." *Journal of Accounting Research*, vol. 29, no. 2 (Autumn):193–228.

Kahn, Jeremy. 2009. "In India, Clues Unfold to a Fraud's Framework." *New York Times* (26 January).

Kealhofer, Stephen. 2003. "Quantifying Credit Risk I: Default Prediction." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):30–44.

Leuty, Ron. 2012. "Elan Will Shutter South S.F. Center as It Shifts R&D to New Company." *San Francisco Business Times* (5 October 2012): www.bizjournals.com/sanfrancisco/blog/biotech/2012/09/elan-neotope-onclave-alzheimers.html?page=all.

Lewis, Craig M. 2012. "Risk Modeling at the SEC: The Accounting Quality Model," Speech given at the Financial Executives International Committee on Finance and Information Technology (13 December): www.sec.gov/news/speech/2012/spch121312cml.htm.

McVay, Sarah E. 2006. "Earnings Management Using Classification Shifting: An Examination of Core Earnings and Special Items." *Accounting Review*, vol. 81, no. 3 (May):501–532.

Mulford, Charles W., and Eugene E. Comiskey. 2005. *Creative Cash Flow Reporting: Uncovering Sustainable Financial Performance*. Hoboken, NJ: John Wiley & Sons.

Nissim, Doron, and Stephen H. Penman. 2001. "Ratio Analysis and Equity Valuation: From Research to Practice." *Review of Accounting Studies*, vol. 6, no. 1 (March):109–154.

SEC. 2000. "Accounting and Auditing Enforcement, Release No. 1350." US Securities and Exchange Commission (14 December): www.sec.gov/litigation/admin/34-43724.htm.

SEC. 2001a "Accounting and Auditing Enforcement, Release No. 1393." US Securities and Exchange Commission (15 May): www.sec.gov/litigation/admin/33-7976.htm.

SEC. 2001b. "Accounting and Auditing Enforcement, Release No. 1405." US Securities and Exchange Commission (19 June): www.sec.gov/litigation/admin/34-44444.htm.

SEC. 2003. "Report Pursuant to Section 704 of the Sarbanes-Oxley Act of 2002" US Securities and Exchange Commission (24 January): www.sec.gov/news/studies/sox704report.pdf.

Sherer, P. 2000. "AmeriServe Examination Finds Financial Woes." *Wall Street Journal* (3 July).

Shumway, Tyler. 2001. "Forecasting Bankruptcy More Accurately: A Simple Hazard Model." *Journal of Business*, vol. 74, no. 1 (January):101–124.

Sloan, Richard G. 1996. "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?" *Accounting Review*, vol. 71, no. 3 (July):289–315.

Thurm, Scott. 2012. "Buyers Beware: The Goodwill Games." *Wall Street Journal* (12 August).

Weil, Jonathan. 2000. "Energy Traders Cite Gains, But Some Math is Missing," *Wall Street Journal* (20 September).

PRACTICE PROBLEMS

The following information relates to Questions 1–4

Mike Martinez is an equity analyst who has been asked to analyze Stellar, Inc. by his supervisor, Dominic Anderson. Stellar exhibited strong earnings growth last year; however, Anderson is skeptical about the sustainability of the company's earnings. He wants Martinez to focus on Stellar's financial reporting quality and earnings quality.

After conducting a thorough review of the company's financial statements, Martinez concludes the following:

Conclusion 1	Although Stellar's financial statements adhere to generally accepted accounting principles (GAAP), Stellar understates earnings in periods when the company is performing well and overstates earnings in periods when the company is struggling.
Conclusion 2	Stellar most likely understated the value of amortizable intangibles when recording the acquisition of Solar, Inc. last year. No goodwill impairment charges have been taken since the acquisition.
Conclusion 3	Over time, the accruals component of Stellar's earnings is large relative to the cash component.
Conclusion 4	Stellar reported an unusually sharp decline in accounts receivable in the current year, and an increase in long-term trade receivables.

1 Based on Martinez's conclusions, Stellar's financial statements are *best* categorized as:

- A non-GAAP compliant.
- B GAAP compliant, but with earnings management.
- C GAAP compliant and decision useful, with sustainable and adequate returns.

2 Based on Conclusion 2, after the acquisition of Solar, Stellar's earnings are *most likely*:

- A understated.
- B fairly stated.
- C overstated.

3 In his follow-up analysis relating to Conclusion 3, Martinez should focus on Stellar's:

- A total accruals.
- B discretionary accruals.
- C non-discretionary accruals.

4 What will be the impact on Stellar in the current year if Martinez's belief in Conclusion 4 is correct? Compared with the previous year, Stellar's:

- A current ratio will increase.

© 2018 CFA Institute. All rights reserved.

B days sales outstanding (DSO) will decrease.

C accounts receivable turnover will decrease.

The following information relates to questions 5–11

Ioana Matei is a senior portfolio manager for an international wealth management firm. She directs research analyst Teresa Pereira to investigate the earnings quality of Miland Communications and Globales, Inc.

Pereira first reviews industry data and the financial reports of Miland Communications for the past few years. Pereira then makes the following three statements about Miland:

Statement 1 Miland shortened the depreciable lives for capital assets.

Statement 2 Revenue growth has been higher than that of industry peers.

Statement 3 Discounts to customers and returns from customers have decreased.

Pereira also observes that Miland has experienced increasing inventory turnover, increasing receivables turnover, and net income greater than cash flow from operations. She estimates the following regression model to assess Miland's earnings persistence:

$$\text{Earnings}_{t+1} = \alpha + \beta_1\text{Cash flow}_t + \beta_2\text{Accruals}_t + \varepsilon$$

Pereira and Matei discuss quantitative models such as the Beneish model, used to assess the likelihood of misreporting. Pereira makes the following two statements to Matei:

Statement 4 An advantage of using quantitative models is that they can determine cause and effect between model variables.

Statement 5 A disadvantage of using quantitative models is that their predictive power declines over time because many managers have learned to test the detectability of manipulation tactics by using the model.

Pereira concludes her investigation of Miland by examining the company's reported pre-tax income of $5.4 billion last year. This amount includes $1.2 billion of acquisition and divestiture-related expenses, $0.5 billion of restructuring expenses, and $1.1 billion of other non-operating expenses. Pereira determines that the acquisition and divestiture-related expenses as well as restructuring expenses are non-recurring expenses, but other expenses are recurring expenses.

Matei then asks Pereira to review last year's financial statements for Globales, Inc. and assess the effect of two possible misstatements. Upon doing so, Pereira judges that Globales improperly recognized EUR50 million of revenue and improperly capitalized EUR100 million of its cost of revenue. She then estimates the effect of these two misstatements on net income, assuming a tax rate of 25%.

Pereira compares Globales, Inc.'s financial statements with those of an industry competitor. Both firms have similar, above-average returns on equity (ROE), although Globales has a higher cash flow component of earnings. Pereira applies the mean reversion principle in her forecasts of the two firms' future ROE.

5 Which of Pereira's statements describes an accounting warning sign of potential overstatement or non-sustainability of operating and/or net income?

A Statement 1

B Statement 2

C Statement 3

6 Which of Pereira's statements about Miland Communications is *most likely* a warning sign of potential earnings manipulation?

A The trend in inventory turnover

B The trend in receivables turnover

C The amount of net income relative to cash flow from operations

7 Based on the regression model used by Pereira, earnings persistence for Miland would be highest if:

A β_1 is less than 0.

B β_1 is greater than β_2.

C β_2 is greater than β_1.

8 Which of Pereira's statements regarding the use of quantitative models to assess the likelihood of misreporting is correct?

A Only Statement 4

B Only Statement 5

C Both Statement 4 and Statement 5

9 Based on Pereira's determination of recurring and non-recurring expenses for Miland, the company's recurring or core pre-tax earnings last year is *closest* to:

A $4.3 billion.

B $4.8 billion.

C $7.1 billion.

10 After adjusting the Globales, Inc. income statement for the two possible misstatements, the decline in net income is *closest* to:

A EUR37.5 million.

B EUR112.5 million.

C EUR150.0 million.

11 Pereira should forecast that the ROE for Globales is likely to decline:

A more slowly than that of the industry competitor.

B at the same rate as the industry competitor.

C more rapidly than that of the industry competitor.

The following information relates to questions 12–18

Emmitt Dodd is a portfolio manager for Upsilon Advisers. Dodd meets with Sonya Webster, the firm's analyst responsible for the machinery industry, to discuss three established companies: BIG Industrial, Construction Supply, and Dynamic Production. Webster provides Dodd with research notes for each company that reflect trends during the last three years:

BIG Industrial:

Note 1 Operating income has been much lower than operating cash flow (OCF).

Note 2 Accounts payable has increased, while accounts receivable and inventory have substantially decreased.

Note 3 Although OCF was positive, it was just sufficient to cover capital expenditures, dividends, and debt repayments.

Construction Supply:

Note 4 Operating margins have been relatively constant.

Note 5 The growth rate in revenue has exceeded the growth rate in receivables.

Note 6 OCF was stable and positive, close to its reported net income, and just sufficient to cover capital expenditures, dividends, and debt repayments.

Dynamic Production:

Note 7 OCF has been more volatile than that of other industry participants.

Note 8 OCF has fallen short of covering capital expenditures, dividends, and debt repayments.

Dodd asks Webster about the use of quantitative tools to assess the likelihood of misreporting. Webster tells Dodd she uses the Beneish model, and she presents the estimated *M*-scores for each company in Exhibit 1.

Exhibit 1 Beneish Model *M*-scores

Company	2017	2016	Change in *M*-score
BIG Industrial	−1.54	−1.82	0.28
Construction Supply	−2.60	−2.51	−0.09
Dynamic Production	−1.86	−1.12	−0.74

Webster tells Dodd that Dynamic Production was required to restate its 2016 financial statements as a result of its attempt to inflate sales revenue. Customers of Dynamic Production were encouraged to take excess product in 2016, and they were then allowed to return purchases in the subsequent period, without penalty.

Webster's industry analysis leads her to believe that innovations have caused some of the BIG Industrial's inventory to become obsolete. Webster expresses concern to Dodd that although the notes to the financial statements for BIG Industrial are informative about its inventory cost methods, its inventory is overstated.

The BIG Industrial income statement reflects a profitable 49% unconsolidated equity investment. Webster calculates the return on sales of BIG Industrial based on the reported income statement. Dodd notes that industry peers consolidate similar investments. Dodd asks Webster to use a comparable method of calculating the return on sales for BIG Industrial.

12 Which of Webster's notes about BIG Industrial provides an accounting warning sign of a potential reporting problem?

A Only Note 1

B Only Note 2

C Both Note 1 and Note 2

13 Do either of Webster's Notes 4 or 5 about Construction Supply describe an accounting warning sign of potential overstatement or non-sustainability of operating income?-

A No

B Yes, Note 4 provides a warning sign

C Yes, Note 5 provides a warning sign

14 Based on Webster's research notes, which company would *most likely* be described as having high-quality cash flow?

A BIG Industrial

B Construction Supply

C Dynamic Production

15 Based on the Beneish model results for 2017 in Exhibit 1, which company has the highest probability of being an earnings manipulator?

A BIG Industrial

B Construction Supply

C Dynamic Production

16 Based on the information related to its restatement, Dynamic Production reported poor operating cash flow quality in 2016 by understating:

A inventories.

B net income.

C trade receivables.

17 Webster's concern about BIG Industrial's inventory suggests poor reporting quality, *most likely* resulting from a lack of:

A completeness.

B clear presentation.

C unbiased measurement.

18 In response to Dodd's request, Webster's recalculated return on sales will *most likely*:

A decrease.

B remain the same.

C increase.

SOLUTIONS

1 B is correct. Stellar's financial statements are GAAP compliant (Conclusion 1) but cannot be relied upon to assess earnings quality. There is evidence of earnings management: understating and overstating earnings depending upon the results of the period (Conclusion 1), understated amortizable intangibles (Conclusion 2), and a high accruals component in the company's earnings (Conclusion 3).

2 C is correct. Martinez believes that Stellar most likely understated the value of amortizable intangibles when recording the acquisition of a rival company last year. Impairment charges have not been taken since the acquisition (Conclusion 2). Consequently, the company's earnings are likely to be overstated because amortization expense is understated. This understatement has not been offset by an impairment charge.

3 B is correct. Martinez concluded that the accruals component of Stellar's earnings was large relative to the cash component (Conclusion 3). Earnings with a larger component of accruals are typically less persistent and of lower quality. An important distinction is between accruals that arise from normal transactions in the period (called non-discretionary) and accruals that result from transactions or accounting choices outside the normal (called discretionary accruals). The discretionary accruals are possibly made with the intent to distort reported earnings. Outlier discretionary accruals are an indicator of possibly manipulated—and thus low quality earnings. Thus, Martinez is primarily focused on discretionary accruals, particularly outlier discretionary accruals (referred to as abnormal accruals).

4 B is correct. Because accounts receivable will be lower than reported in the past, Stellar's DSO [Accounts receivable/(Revenues/365)] will decrease. Stellar's accounts receivable turnover (365/days' sales outstanding) will increase with the lower DSO, giving the false impression of a faster turnover. The company's current ratio will decrease (current assets will decrease with no change in current liabilities).

5 B is correct. Higher growth in revenue than that of industry peers is an accounting warning sign of potential overstatement or non-sustainability of operating income. Shortening the depreciable lives of capital assets is a conservative change and not a warning sign. An increase (not a decrease) in discounts and returns would be a warning sign.

6 C is correct. Net income being greater than cash flow from operations is a warning sign that the firm may be using aggressive accrual accounting policies that shift current expenses to future periods. Decreasing, not increasing, inventory turnover could suggest inventory obsolescence problems that should be recognized. Decreasing, not increasing, receivables turnover could suggest that some revenues are fictitious or recorded prematurely or that the allowance for doubtful accounts is insufficient.

7 B is correct. When earnings are decomposed into a cash component and an accruals component, research has shown that the cash component is more persistent. A beta coefficient (β_1) on the cash flow variable that is larger than the beta coefficient (β_2) on the accruals variable indicates that the cash flow component of earnings is more persistent than the accruals component. This result provides evidence of earnings persistence.

8 B is correct. Earnings manipulators have learned to test the detectability of earnings manipulation tactics by using the model to anticipate analysts' perceptions. They can reduce their likelihood of detection; therefore, Statement 5 is correct. As a result, the predictive power of the Beneish model can decline over time. An additional limitation of using quantitative models is that they cannot determine cause and effect between model variables. Quantitative models establish only associations between variables, and Statement 4 is incorrect.

A is incorrect because quantitative models cannot determine cause and effect between model variables. They are capable only of establishing associations between variables. Therefore, Statement 4 is incorrect.

9 C is correct. Recurring or core pre-tax earnings would be \$7.1 billion, which is the company's reported pre-tax income of \$5.4 billion plus the \$1.2 billion of non-recurring (i.e., one-time) acquisitions and divestiture expenses plus the \$0.5 billion of non-recurring restructuring expenses.

10 B is correct. The correction of the revenue misstatement would result in lower revenue by EUR50 million, and the correction of the cost of revenue misstatement would result in higher cost of revenue by EUR100 million. The result is a reduction in pre-tax income of EUR150 million. Applying a tax rate of 25%, the reduction in net income would be $150 \times (1 - 0.25)$ = EUR112.5 million.

11 A is correct. Based on the principle of mean reversion, the high ROE for both firms should revert towards the mean. Globales has a higher cash flow component to its return than the peer firm, however, so its high return on common equity should persist longer than that of the peer firm. The peer firm has a higher accruals component, so it is likely to revert more quickly.

12 B is correct. Only Note 2 provides a warning sign. The combination of increases in accounts payable with substantial decreases in accounts receivable and inventory are an accounting warning sign that management may be overstating cash flow from operations. Note 1 does not necessarily provide a warning sign. Operating income being greater than operating cash flow is a warning sign of a potential reporting problem. In this case, however, BIG Industrial's operating income is lower than its operating cash flow.

13 A is correct. Neither Note 4 nor Note 5 provides an accounting warning sign of potential overstatement or non-sustainability of operating income.

Increases in operating margins can be a warning sign of potential overstatement or non-sustainability of operating and/or net income. In this case, however, operating margins for Construction Supply have been relatively constant during the last three years.

A growth rate in receivables exceeding the growth rate in revenue is an accounting warning sign of potential overstatement or non-sustainability of operating income. In this case, however, Construction Supply's revenue growth exceeds the growth rate in receivables.

14 B is correct. High-quality OCF means the performance is of high reporting quality and also of high results quality. For established companies, high-quality operating cash flow would typically be positive; be derived from sustainable sources; be adequate to cover capital expenditures, dividends, and debt repayments; and have relatively low volatility compared with industry peers. Construction Supply reported positive OCF during each of the last three years. The OCF appears to be derived from sustainable sources, because it compares closely with reported net income. Finally, OCF was adequate to cover capital expenditures, dividends, and debt repayments. Although the OCF for BIG Industrial has been positive and just sufficient to cover capital expenditures, dividends, and debt repayments, the increases in accounts payable and substantial decreases in accounts receivable and inventory during the last three years are an accounting warning sign that management may be overstating cash flow from operations. For Dynamic Production, OCF has been more volatile than other industry participants, and it has fallen short of covering capital expenditures, dividends, and debt repayments for the last three years. Both of these conditions are warning signs for Dynamic Production.

15 A is correct. Higher *M*-scores indicate an increased probability of earnings manipulation. The company with the highest *M*-score in 2017 is BIG Industrial, with an *M*-score of −1.54. Construction Supply has the lowest *M*-score at −2.60, and Dynamic Production also has a lower *M*-score at −1.86. The *M*-score for BIG Industrial is above the relevant cutoff of −1.78.

16 A is correct. The items primarily affected by improper revenue recognition include net income, receivables, and inventories. When revenues are overstated, net income and receivables will be overstated and inventories will be understated.

17 C is correct. Webster is concerned that innovations have made some of BIG Industrial's inventory obsolete. This scenario suggests impairment charges for inventory may be understated and that the inventory balance does not reflect unbiased measurement.

18 A is correct. The use of unconsolidated joint ventures or equity-method investees may reflect an overstated return on sales ratio, because the parent company's consolidated financial statements include its share of the investee's profits but not its share of the investee's sales. An analyst can adjust the reported amounts to better reflect the combined amounts of sales. Reported net income divided by the combined amount of sales will result in a decrease in the net profit margin.

READING

14

Integration of Financial Statement Analysis Techniques

by Jack T. Ciesielski, CPA, CFA

Jack T. Ciesielski, CPA, CFA, is at R.G. Associates, Inc., former publisher of The Analyst's Accounting Observer (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);
☐	**b.** identify financial reporting choices and biases that affect the quality and comparability of companies' financial statements and explain how such biases may affect financial decisions;
☐	**c.** evaluate the quality of a company's financial data and recommend appropriate adjustments to improve quality and comparability with similar companies, including adjustments for differences in accounting standards, methods, and assumptions;
☐	**d.** evaluate how a given change in accounting standards, methods, or assumptions affects financial statements and ratios;
☐	**e.** analyze and interpret how balance sheet modifications, earnings normalization, and cash flow statement related modifications affect a company's financial statements, financial ratios, and overall financial condition.

© 2019 CFA Institute. All rights reserved.

1 INTRODUCTION

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

It is important to keep in mind that financial analysis is a means to an end and not the end itself. Rather than try to apply every possible technique and tool to every situation, it is essential for the investor to consider and identify the proper type of analysis to apply in a given situation.

The primary reason for performing financial analysis is to help in making an economic decision. Before making such decisions as whether to lend to a particular long-term borrower or to invest a large sum in a common stock, venture capital vehicle, or private equity candidate, an investor or financial decision-maker wants to make sure that the probability of a successful outcome is on his or her side. Rather than leave outcomes to chance, a financial decision-maker should use financial analysis to identify and make more visible potential favorable and unfavorable outcomes.

The purpose of this reading is to provide examples of the effective use of financial analysis in decision making. The framework for the analysis is shown in Exhibit 1. The case study follows the basic framework shown in Exhibit 1.

Exhibit 1 A Financial Statement Analysis Framework

Phase	Sources of Information	Examples of Output
1 Define the purpose and context of the analysis.	■ The nature of the analyst's function, such as evaluating an equity or debt investment or issuing a credit rating ■ Communication with client or supervisor on needs and concerns ■ Institutional guidelines related to developing specific work product	■ Statement of the purpose or objective of the analysis ■ A list (written or unwritten) of specific questions to be answered by the analysis ■ Nature and content of report to be provided ■ Timetable and budgeted resources for completion
2 Collect input data.	■ Financial statements, other financial data, questionnaires, and industry/economic data ■ Discussions with management, suppliers, customers, and competitors ■ Company site visits (e.g., to production facilities or retail stores)	■ Organized financial statements ■ Financial data tables ■ Completed questionnaires, if applicable
3 Process input data, as required, into analytically useful data.	■ Data from the previous phase	■ Adjusted financial statements ■ Common-size statements ■ Ratios and graphs ■ Forecasts
4 Analyze/interpret the data.	■ Input data and processed data	■ Analytical results

Exhibit 1 (Continued)

Phase	Sources of Information	Examples of Output
5 Develop and communicate conclusions and recommendations (e.g., with an analysis report).	■ Analytical results and previous reports ■ Institutional guidelines for published reports	■ Analytical report answering questions posed in Phase 1 ■ Recommendation regarding the purpose of the analysis, such as whether to make an investment or grant credit
6 Follow-up.	■ Information gathered by periodically repeating above steps, as necessary, to determine whether changes to holdings or recommendations are necessary	■ Updated reports and recommendations

CASE STUDY 1: LONG -TERM EQUITY INVESTMENT: EARLY PHASES OF THE ANALYSIS

2

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

The portfolio manager for the food sector of a large public employee pension fund wants to take a long-term equity position in a publicly traded food company and has become interested in Nestlé S.A., a global company. In its 2014 annual report, Nestlé's management outlined its long-term objectives for organic growth, margin and earnings per share improvement, and capital efficiency. The management report indicated the following general strategic direction: "Our ambition is not just to be the leader but the industry reference for Nutrition, Health and Wellness. In recent years we have built on the strong foundations of our unrivalled food and beverage portfolio, exploring the benefits of nutrition's therapeutic role with Nestlé Health Science." Nestlé's stated objectives, including expansion of the company's mission into "nutrition's therapeutic role," captured the portfolio manager's attention: She became intrigued with Nestlé as an investment possibility. She asks an analyst to evaluate Nestlé for consideration as a large core holding. Before investing in the company, the portfolio manager has several concerns that she has conveyed to the analyst:

- What are Nestlé's sources of earnings growth? How sustainable is Nestlé's performance? Do the company's reported earnings represent its economic reality? And if Nestlé's performance is fairly reported, will it be sustainable for an extended period, such as 5 to 10 years, while the pension fund has the common stock as a core holding?
- In determining the quality of earnings over a long-term time frame, the portfolio manager wants to understand the relationship of earnings to cash flow.
- Having started out in the investment business as a lending officer, the portfolio manager wants to know how well Nestlé's balance sheet takes into account the company's full rights and obligations. She wants to know whether the capital structure of the company can support future operations and strategic plans. Even if the investor is primarily concerned with the earnings potential of a possible investee, the balance sheet matters. For example, if asset write-downs

or new legal liabilities decrease a company's financial position, it is difficult for a company to sustain profitability if it has to repair its balance sheet. Worse still for an investor: If "repairing the balance sheet" means the issuance of dilutive stock, it can be even more costly to existing investors.

The analyst develops a plan of analysis to address the portfolio manager's concerns by following the framework presented in Exhibit 1. Phases 3 and 4 will be the focus of most of the work.

Phase 1: Define a Purpose for the Analysis

The analyst states the purpose and context of the analysis as identifying the factors that have driven the company's financial success and assessing their sustainability. He also states the need to identify and understand the risks that may affect the sustainability of returns.

Phase 2: Collect Input Data

The analyst finds that Nestlé has an extensive collection of financial statements on its website. After gathering several years of annual reports, he is ready to begin processing the data.

3 PHASES 3 & 4: DUPONT ANALYSIS: ISOLATING "PURE NESTLE"

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

b identify financial reporting choices and biases that affect the quality and comparability of companies' financial statements and explain how such biases may affect financial decisions;

c evaluate the quality of a company's financial data and recommend appropriate adjustments to improve quality and comparability with similar companies, including adjustments for differences in accounting standards, methods, and assumptions;

Phase 3: Process Data and Phase 4: Analyze/Interpret the Processed Data

The analyst intends to accomplish his purpose stated in Phase 1 through a series of financial analyses, including

- a DuPont analysis;[1]
- an analysis of the composition of Nestlé's asset base;
- an analysis of Nestlé's capital structure;
- a study of the company's segments and the allocation of capital among them;
- an examination of the company's accruals in reporting as they affect earnings quality;
- a study of the company's cash flows and their adequacy for the company's continued operations and strategies; and
- a decomposition and analysis of the company's valuation.

While processing the input data consistent with the needs of these analyses, the analyst plans to simultaneously interpret and analyze the resulting data. In his view, Phases 3 and 4 of the framework are best considered jointly.

DuPont Analysis

The analyst decides to start the assessment of Nestlé with a DuPont analysis. The investment is expected to be in the company's common stock, and ultimately, the DuPont analysis separates the components affecting the return on common equity. Furthermore, the disaggregation of return on equity (ROE) components leads to more trails to follow in assessing the drivers of Nestlé's performance. The analyst also intends to investigate the quality of the earnings and underlying cash flows, as well as to understand the common shareholders' standing in the Nestlé capital structure.

One basic premise underlying all research and analysis is to constantly look beneath the level of information presented—to constantly search for meaningful insights through disaggregation of the presented information, whether it is a single line on a financial statement or within segments of an entire entity. This constant reduction of information into smaller components can reveal a company's earnings drivers; it can also highlight weaker operations being concealed by stronger ones in the aggregate. That premise of "seeking granularity" underlies DuPont analysis: By isolating the different components of ROE, it helps the analyst discover a company's strengths and allows the analyst to assess their sustainability.[2] Seeking granularity also helps the analyst find potential operational flaws and provides an opening for dialogue with management about possible problems.

The analyst begins to process the data gathered in Phase 2 in order to assemble the information required for the DuPont analysis. Exhibit 2 shows the last three years of income statements for Nestlé; Exhibit 3 shows the last four years of Nestlé balance sheets.

1 A reminder to the reader: This case study is an example, and starting the financial statement analysis with a DuPont analysis is not a mandate. Alternatively, another analyst might be more interested in the trends of various income and expense categories than in the sources of returns on shareholder equity as a financial statement analysis starting point. This analyst might have preferred starting with a time-series common-size income statement. The starting point depends on the perspective of the individual analyst.

2 ROE can be decomposed in a variety of ways:

ROE = Return on assets × Leverage

ROE = Net profit margin × Asset turnover × Leverage

ROE = EBIT margin × Tax burden × Interest burden × Asset turnover × Leverage

From his study of the income statement, the analyst notes that Nestlé has a significant amount of "income from associates and joint ventures" (hereafter referred to in the text as income from associates) in all three years. In 2014, this income amounted to CHF8,003 million, or 53.7%, of Nestlé's net income (referred to by Nestlé as "profit for the year"). The income from associates[3] is a pure net income figure, presented after taxes and with no related revenue in the income statement. Much of the income from associates relates to Nestlé's 23.4% stock ownership of L'Oréal, a cosmetics company.

In 2014, L'Oréal affected the amount of income from associates in a variety of ways. In 2014, Nestlé reduced its L'Oréal ownership by selling 48.5 million shares of its holding back to L'Oréal. In return, Nestlé gained full ownership of Galderma, a joint venture it had with L'Oréal. The partial disposal of L'Oréal shares resulted in a net gain of CHF4,569 million. Income from associates included a revaluation gain of CHF2,817 million from the increase in ownership of Galderma. Nestlé had owned 50% of Galderma, with L'Oréal holding the other 50%. When Nestlé bought the remaining ownership from L'Oréal, its original 50% ownership position was revalued at current fair value, which was based on the price paid. As of July 2014, Galderma became an affiliated company that was fully consolidated. Because of its L'Oréal stock ownership, Nestlé recognizes a share of L'Oréal's net income.

The share of results at other companies that Nestlé included in income from associates was CHF828 million in 2014.

The analyst wants to decompose the company's financial results as much as possible in order to identify any problem operations or to find hidden opportunities. Including the net investments and returns of associates with the full reported value of Nestlé's own assets and income would introduce noise into the analytical signals produced by the DuPont analysis. Unlike the "pure Nestlé" operations and resources, the returns earned by associates are not under the direct control of Nestlé's management. To avoid making incorrect inferences about the profitability of Nestlé's operations, the analyst wants to remove the effects of the investments in associates from the balance sheet and income statement. Otherwise, such DuPont analysis components as net profit margin and total asset turnover would combine the impact of pure Nestlé operations with that of the operations of associated companies: Conclusions about Nestlé-only business would be flawed because they would be based on commingled information.

Exhibit 2 Nestlé S.A. Income Statements, 2014–2012 (CHF millions)

	2014	2013	2012 (restated)[d]
Sales	91,612	92,158	89,721
Other revenue	253	215	210
Cost of goods sold	(47,553)	(48,111)	(47,500)
Distribution expenses	(8,217)	(8,156)	(8,017)
Marketing and administration expenses	(19,651)	(19,711)	(19,041)
Research and development costs	(1,628)	(1,503)	(1,413)
Other trading income	110	120	141
Other trading expenses[a]	(907)	(965)	(637)
Trading operating profit[b]	**14,019**	**14,047**	**13,464**
Other operating income	154	616	146

3 Associates are companies in which Nestlé has the power to exercise significant influence but does not exercise control. Associates and joint ventures are accounted for by the equity method.

Exhibit 2 (Continued)

	2014	2013	2012 (restated)[d]
Other operating expenses[c]	(3,268)	(1,595)	(222)
Operating profit (EBIT)	**10,905**	**13,068**	**13,388**
Financial income	135	219	120
Financial expense	(772)	(850)	(825)
Profit before taxes, associates, and joint ventures (EBT)	**10,268**	**12,437**	**12,683**
Taxes	(3,367)	(3,256)	(3,259)
Income from associates and joint ventures	8,003	1,264	1,253
Profit for the year	**14,904**	**10,445**	**10,677**
of which attributable to non-controlling interests	448	430	449
of which attributable to shareholders of the parent (net profit)	14,456	10,015	10,228
Earnings per share			
Basic earnings per share	4.54	3.14	3.21
Diluted earnings per share	4.52	3.13	3.20

Excerpted information from notes to the financial statements:	2014	2013	2012 (restated)
[a] Other trading expenses include:			
Restructuring costs	(257)	(274)	(88)
Impairment of PP&E	(136)	(109)	(74)
Impairment of intangible assets (other than goodwill)	(23)	(34)	—
Litigation and onerous contracts	(411)	(380)	(369)
Unusual charges contained within operating profit	(827)	(797)	(531)
[b] Expenses allocated by function:			
Depreciation of PP&E	(2,782)	(2,867)	(2,655)
Amortisation of intangible assets	(276)	(301)	(394)
	(3,058)	(3,168)	(3,049)
[c] Other operating expenses include:			
Impairment of goodwill	(1,908)	(114)	(14)

[d] The 2012 information came from the 2013 Annual Report; 2012 comparatives were restated by Nestlé following the implementation of IFRS 11 and IAS 19 revised, as described in Note 22.

Exhibit 3 Nestlé S.A. Balance Sheets, 2014–2011 (CHF millions)

	2014	2013	2012 (restated)[a]	2011 (revised)[b]
Assets				
Current assets				
Cash and cash equivalents	7,448	6,415	5,713	4,769
Short-term investments	1,433	638	3,583	3,013

(continued)

Exhibit 3 (Continued)

	2014	2013	2012 (restated)[a]	2011 (revised)[b]
Inventories	9,172	8,382	8,939	9,095
Trade and other receivables	13,459	12,206	13,048	12,991
Prepayments and accrued income	565	762	821	879
Derivative assets	400	230	576	722
Current income tax assets	908	1,151	972	1,053
Assets held for sale	576	282	368	16
Total current assets	**33,961**	**30,066**	**34,020**	**32,538**
Non-current assets				
Property, plant, and equipment (PP&E)	28,421	26,895	26,576	23,460
Goodwill	34,557	31,039	32,688	28,613
Intangible assets	19,800	12,673	13,018	8,785
Investments in associates and joint ventures	8,649	12,315	11,586	10,317
Financial assets	5,493	4,550	4,979	7,153
Employee benefits assets	383	537	84	127
Current income tax assets	128	124	27	39
Deferred tax assets	2,058	2,243	2,899	2,408
Total non-current assets	**99,489**	**90,376**	**91,857**	**80,902**
Total assets	**133,450**	**120,442**	**125,877**	**113,440**
Liabilities and equity				
Current liabilities				
Financial debt	8,810	11,380	18,408	15,945
Trade and other payables	17,437	16,072	14,627	13,544
Accruals and deferred income	3,759	3,185	3,078	2,780
Provisions	695	523	452	575
Derivative liabilities	757	381	423	632
Current income tax liabilities	1,264	1,276	1,608	1,379
Liabilities directly associated with assets held for sale	173	100	1	—
Total current liabilities	**32,895**	**32,917**	**38,597**	**34,855**
Non-current liabilities				
Financial debt	12,396	10,363	9,008	6,165
Employee benefits liabilities	8,081	6,279	8,360	6,912
Provisions	3,161	2,714	2,827	3,079
Deferred tax liabilities	3,191	2,643	2,240	1,974
Other payables	1,842	1,387	2,181	2,113
Total non-current liabilities	**28,671**	**23,386**	**24,616**	**20,243**
Total liabilities	**61,566**	**56,303**	**63,213**	**55,098**
Equity				
Share capital	322	322	322	330
Treasury shares	(3,918)	(2,196)	(2,078)	(6,722)
Translation reserve	(17,255)	(20,811)	(17,924)	(16,927)
Retained earnings and other reserves	90,981	85,260	80,687	80,184

Exhibit 3 (Continued)

	2014	2013	2012 (restated)[a]	2011 (revised)[b]
Total equity attributable to shareholders of the parent	**70,130**	**62,575**	**61,007**	**56,865**
Non-controlling interests	1,754	1,564	1,657	1,477
Total equity	**71,884**	**64,139**	**62,664**	**58,342**
Total liabilities and equity	**133,450**	**120,442**	**125,877**	**113,440**

[a] The 2012 information came from the 2013 Annual Report; 2012 comparatives were restated by Nestlé following the implementation of IFRS 11 and IAS 19 revised, as described in Note 22.
[b] The analyst revised the 2011 balance sheet from that reported in the 2012 Consolidated Financial Statements of the Nestlé Group.

To keep the DuPont analysis as logically consistent as possible throughout all the periods of study, the analyst revises the 2011 balance sheet (from that reported in the 2012 Consolidated Financial Statements of the Nestlé Group) for the effects of implementing IFRS 11 and IAS 19 revised. He identifies the 1 January 2012 adjustments from the 2013 financial statements and revises the 31 December 2011 year-end balances accordingly. The analyst's revisions to the as-reported 2011 balance sheet are shown in Exhibit 4.

Exhibit 4 Modifications to 2011 Balance Sheet (CHF millions)

	2011 (as reported)	Effects of IAS 19 (1)	Effects of IFRS 11 (2)	2011 (revised)
Assets				
Current assets				
Cash and cash equivalents	4,938	—	(169)	4,769
Short-term investments	3,050	—	(37)	3,013
Inventories	9,255	—	(160)	9,095
Trade and other receivables	13,340	—	(349)	12,991
Prepayments and accrued income	900	—	(21)	879
Derivative assets	731	—	(9)	722
Current income tax assets	1,094	—	(41)	1,053
Assets held for sale	16	—	—	16
Total current assets	**33,324**	—	**(786)**	**32,538**
Non-current assets				
Property, plant, and equipment	23,971	—	(511)	23,460
Goodwill	29,008	—	(395)	28,613
Intangible assets	9,356	—	(571)	8,785
Investments in associates and joint ventures	8,629	—	1,688	10,317
Financial assets	7,161	—	(8)	7,153
Employee benefits assets	127	—	—	127
Current income tax assets	39	—	—	39
Deferred tax assets	2,476	(5)	(63)	2,408
Total non-current assets	**80,767**	**(5)**	**140**	**80,902**

(continued)

Exhibit 4 (Continued)

	2011 (as reported)	Effects of IAS 19 (1)	Effects of IFRS 11 (2)	2011 (revised)
Total assets	**114,091**	**(5)**	**(646)**	**113,440**
Liabilities and equity				
Current liabilities				
Financial debt	16,100	—	(155)	15,945
Trade and other payables	13,584	—	(40)	13,544
Accruals and deferred income	2,909	—	(129)	2,780
Provisions	576	—	(1)	575
Derivative liabilities	646	—	(14)	632
Current income tax liabilities	1,417	—	(38)	1,379
Liabilities directly associated with assets held for sale	—	—	—	—
Total current liabilities	**35,232**	—	**(377)**	**34,855**
Non-current liabilities				
Financial debt	6,207	—	(42)	6,165
Employee benefits liabilities	7,105	(91)	(102)	6,912
Provisions	3,094	—	(15)	3,079
Deferred tax liabilities	2,060	18	(104)	1,974
Other payables	2,119	—	(6)	2,113
Total non-current liabilities	**20,585**	**(73)**	**(269)**	**20,243**
Total liabilities	**55,817**	**(73)**	**(646)**	**55,098**
Equity				
Share capital	330	—	—	330
Treasury shares	(6,722)	—	—	(6,722)
Translation reserve	(16,927)	—	—	(16,927)
Retained earnings and other reserves	80,116	68	—	80,184
Total equity attributable to shareholders of the parent	**56,797**	**68**	—	**56,865**
Non-controlling interests	1,477	—	—	1,477
Total equity	**58,274**	**68**	—	**58,342**
Total liabilities and equity	**114,091**	**(5)**	**(646)**	**113,440**

(1) IAS 19 Revised 2011—Employee Benefits was implemented in 2013, with comparative restatement made to 1 January 2012. This standard revised the calculation of benefit plan obligations. The 1 January 2012 adjustments were imposed on the 31 December 2011 balance sheet by the analyst, taken from Note 22 (Restatements and adjustments of 2012 comparatives) of the 2013 Annual Report.

(2) IFRS 11—Joint Arrangements was implemented in 2013, with comparative restatement made to 1 January 2012. Nestlé had used proportional consolidation for two of its joint arrangements (Cereal Partners Worldwide and Galderma), and the standard required that they be accounted for using the equity method of investments. The 1 January 2012 adjustments were imposed on the 31 December 2011 balance sheet by the analyst, taken from Note 22 (Restatements and adjustments of 2012 comparatives) of the 2013 Annual Report.

PHASES 3 & 4: DUPONT DECOMPOSITION

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

c evaluate the quality of a company's financial data and recommend appropriate adjustments to improve quality and comparability with similar companies, including adjustments for differences in accounting standards, methods, and assumptions;

The analyst considers what information he needs for a DuPont analysis. He extracts the data shown in Exhibit 5 from Exhibits 2 and 3:

Exhibit 5 Data Needed for DuPont Analysis (CHF millions)

	2014	2013	2012	2011
Income Statement Data:				
Sales	91,612	92,158	89,721	
Operating profit (EBIT)	10,905	13,068	13,388	
Profit before taxes, associates, and joint ventures (EBT)	10,268	12,437	12,683	
Profit for the year	14,904	10,445	10,677	
Income from associates and joint ventures	8,003	1,264	1,253	
Profit, excluding associates and joint ventures	6,901	9,181	9,424	
Balance Sheet Data:				
Total assets	133,450	120,442	125,877	113,440
Investments in associates and joint ventures	8,649	12,315	11,586	10,317
Total assets, excluding associates and joint ventures	124,801	108,127	114,291	103,123
Total equity	71,884	64,139	62,664	58,342

The five-way decomposition of ROE is expanded to isolate the effects of the investment in associates in Nestlé's asset base and earnings. The necessary modifications to the reported financial data to isolate these effects are shown in Exhibit 5. Subtracting income from associates from the net income (profit for the year) gives the profits generated by Nestlé's own asset base. Subtracting the amount of investment in associates from total assets results in a figure that more closely represents Nestlé's own asset base. With this information, the analyst can assess the profitability and returns of the largest and most relevant part of the entire Nestlé entity: the core Nestlé company.

Exhibit 6 shows the results of expanding the DuPont analysis. The net profit margin component and the asset turnover component require adjustments to remove the impact of the associates on the return on assets. To adjust the net profit margin component, the analyst subtracts the associates' income from the net income and divides the result by sales. For 2014, the Nestlé-only net profit margin was 7.53% (= Profit excluding income from associates/Sales = 6,901/91,612). To adjust the asset turnover, the analyst subtracts the investment in associates from total assets to arrive at the assets used by the core Nestlé company. Sales divided by the average of the beginning and ending assets (excluding investment in associates) gives the Nestlé-only asset turnover. For 2014, the Nestlé-only asset turnover was 0.787 {= 91,612/[(108,127 + 124,801)/2] = 91,612/116,464}. Including the investment in associates in total assets, the asset turnover was 0.722 {= 91,612/[(120,442 + 133,450)/2] = 91,612/126,946}. The difference between the asset turnover based on unadjusted financial statement amounts and the Nestlé-only asset turnover gives the effect on total asset turnover of the investment in associates: a decrease of 0.065 in 2014.

The net profit margin can be decomposed into three components: EBIT margin × Tax burden × Interest burden. The tax and interest burdens indicate what is left for the company after the effects of taxes and interest, respectively. To adjust the tax burden component, the analyst divides profit (excluding income from associates) by profit before taxes and income from associates (EBT). For 2014, the tax burden was 67.21% (= 6,901/10,268). The interest burden is calculated by dividing the profit before taxes, associates, and joint ventures (EBT) by operating profit (EBIT). For 2014, the interest burden was 94.16% (= 10,268/10,905). The EBIT margin is earnings before interest and taxes (operating profit) divided by revenue (sales). For 2014, the EBIT margin was 11.90% (= 10,905/91,612).

Multiplying the three components together yields the Nestlé-only net profit margin. In 2014, the Nestlé-only net profit margin was 7.53% (= 67.21% × 94.16% × 11.90%). Calculating the net profit margin without excluding income from associates gives 16.27% (= Net income/ Revenue = Profit for the year/Sales = 14,904/91,612), which is not representative of the Nestlé-only operations. Dividing the net profit margin by the net profit margin *without* the associates' income (16.27%/7.53% = 216.07%) quantifies the magnifying effect of the associates' income on Nestlé's own margins. The "Nestlé-only" entity earned 7.53% on every sale, but including the associates' income in net profit increases the net profit margins by 116.07% [(100.00% + 116.07%) × 7.53% = 16.27%]. A 16.27% level of profitability is not representative of what Nestlé's core operations can generate.

Exhibit 6 Expanded DuPont Analysis

	2014	2013	2012
Tax burden (excl. associates)	67.21%	73.82%	74.30%
× Interest burden	94.16%	95.17%	94.73%
× EBIT margin	11.90%	14.18%	14.92%
= Net profit margin (excl. associates)	7.53%	9.96%	10.50%
× Associates' effect on net profit margin	216.07%	113.76%	113.33%
= Net profit margin	**16.27%**	**11.33%**	**11.90%**
Total asset turnover (excl. associates)	0.787	0.829	0.825
Effect of associates' investments on turnover	(0.065)	(0.081)	(0.075)
× Total asset turnover	**0.722**	**0.748**	**0.750**
= Return on assets	11.75%	8.47%	8.93%
× Leverage	**1.87**	**1.94**	**1.98**
= Return on equity (ROE)	**21.97%**	**16.44%**	**17.67%**

Traditional ROE calculation (CHF millions):

Exhibit 6 (Continued)

	2014	2013	2012
Net income	14,904	10,445	10,677
÷ Average total equity	68,012	63,402	60,503
= ROE	**21.91%**	**16.47%**	**17.65%**

Note: Differences in ROE calculations because of rounding.

PHASES 3 & 4: ADJUSTING FOR UNUSUAL CHARGES

5

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

c evaluate the quality of a company's financial data and recommend appropriate adjustments to improve quality and comparability with similar companies, including adjustments for differences in accounting standards, methods, and assumptions;

d evaluate how a given change in accounting standards, methods, or assumptions affects financial statements and ratios;

e analyze and interpret how balance sheet modifications, earnings normalization, and cash flow statement related modifications affect a company's financial statements, financial ratios, and overall financial condition.

In 2012 and 2013, the net profit margin (including income from associates) was fairly stable at 11.90% and 11.33%, respectively. But it increased significantly in 2014—to 16.27%—as a result of the increase in income from associates attributable to the L'Oréal disposal and Galderma revaluation. The analyst, however, is interested in the ongoing operations of Nestlé, unaffected by such non-repeating types of gains. The net profit margin excluding income from associates shows a disturbing trend: It decreased each year in the 2012–2014 period. This finding prompts the analyst to try to identify a reason for the declining profitability of the Nestlé-only business. Searching the income statements and notes in the annual reports, he notices that Nestlé has recorded goodwill impairments over the period under study, with a particularly large one, CHF1,908 million, occurring in 2014. This impairment was related to Nestlé's acquisitions of ice cream and pizza businesses in the United States. He also notices that Nestlé has recorded provisions each year for restructuring activities, environmental liabilities, litigation reserves, and other activities. To see how much these events affected the Nestlé-only profitability, he constructs the table shown in Exhibit 7. He calls these events "unusual charges" for convenience of presentation.

Exhibit 7 Profitability Adjusted for Provisions and Impairment Charges (CHF millions)

	2014	2013	2012
Sales	91,612	92,158	89,721
Profit excluding income from associates (from Exhibit 5)	6,901	9,181	9,424
Impairment of goodwill	1,908	114	14
Total provisions for restructuring, environmental, litigation, and other (not tax-affected: assumed non-taxable in year of recognition)	920	862	618
Profit adjusted for unusual charges	9,729	10,157	10,056
Net profit margin: excl. associates, with all unusual charges incl.	7.53%	9.96%	10.50%
Net profit margin: excluding associates and unusual charges	10.62%	11.02%	11.21%
Profit margin consumed by unusual charges	3.09%	1.06%	0.71%

The analyst notices that the adjusted profits and the adjusted profit margins were more stable over the three-year period than the profits and profit margins excluding associates. However, the adjusted profits and profit margins and the profits and profit margins excluding associates decreased over the same period. Although the provisions and impairment charges potentially explain the significant decrease in the Nestlé-only profit margins, in particular from 2013 to 2014, the analyst decides *not* to adjust the remaining DuPont analysis to exclude these charges. They involve decisions by management, they recur regularly, and they affect the returns to shareholders. In assessing the company's prospects, he believes that these charges are important variables that should not to be ignored.

Returning to the DuPont analysis, he now realizes the significance of the associates' earnings to the entire Nestlé entity. The margin is greater in each year if the associates' earnings are included in net profit as opposed to looking at Nestlé alone. Consistently, the company's profit margins are smaller without the boost from associates' earnings. Asset turnover is consistently lower when assets include the investment in associates.

The adjustments thus far have isolated the operational aspects of Nestlé's performance and the assets that produced them from non-Nestlé operations. The financial leverage ratio has not been adjusted by the analyst in similar fashion to profit margin and asset turnover. The profit margin and asset turnover components of the DuPont analysis are relatively easy to consider when including or excluding associates: Both the Nestlé assets and the non-Nestlé assets produce a certain pre-tax return. Isolating those assets and their respective returns from each other makes it possible to see the contributions of each to the aggregate performance. It might be tempting to likewise adjust the financial leverage ratio by subtracting the investment in associates from total assets and equity, but the financial leverage component need not be adjusted. The analyst assumes that there will be no change in the Nestlé capital structure and that a similar blend of debt and equity in the company's capital structure finances the investment in associates' assets and the Nestlé-only assets.

From Exhibit 6, multiplying the three conventionally calculated components of ROE (net profit margin, total asset turnover, and leverage) yields the ROE when the effect of associates is included (top row of Exhibit 8). The ROE exhibits an overall increasing trend when examined without adjusting for investment in associates. The analyst wants to compare the ROE for Nestlé alone with the ROE including associates. Calculating the ROE on a Nestlé-only basis is done by multiplying the net profit margin excluding associates by the total asset turnover excluding associates by the financial leverage. For 2014, the Nestlé-only ROE was 11.08% (7.53% × 0.787 × 1.87 = 11.08%).

Exhibit 8 shows the ROE including and excluding the effects of associates. The difference between the two sets of ROE figures reveals the amount of ROE contribution from the associates. The trend in the ROE including associates, which shows a significant increase in 2014, is largely the result of the gains in 2014 from the transactions involving the investments in associates (exchange of L'Oréal shares for complete ownership of Galderma). Nestlé only shows a different trend: decreasing in each of the last two years.

Exhibit 8 ROE Performance Due to Investment in Associates

	2014 (%)	2013 (%)	2012 (%)
ROE including associates	21.97	16.44	17.67
Less Nestlé-only ROE	11.08	16.02	17.15
Associates' contribution to ROE	10.89	0.42	0.52

The analyst is particularly troubled by the sharp drop-off in the Nestlé-only ROE in 2014. He knows that there was an unusually large goodwill impairment charge in 2014, which may explain the sudden decrease. To see the role played by such unusual charges in the ROE trend, he reworks the Nestlé-only ROE figures on the basis of revised net profit margins (excluding associates and unusual charges) as shown in Exhibit 7. For 2014, the Nestlé-only ROE was 15.63% (10.62% × 0.787 × 1.87 = 15.63%). The results are shown in Exhibit 8A.

Exhibit 8A Nestlé-Only ROE, with Unusual Charges Removed from Pre-tax Margins

	2014	2013	2012
Nestlé-only ROE	15.63%	17.73%	18.31%

Absent the unusual charges, the magnitude of the Nestlé-only ROE improved significantly in all three years, but the trend remained on a downward slope. This trend is a genuine concern to the analyst; the investment in associates might provide incremental returns, but he believes the biggest part of the entire entity should be the most significant driver of returns.

Underscoring the significance of the investment in associates—and the deterioration of the Nestlé-only business—is the increasing spread between the as-reported and the Nestlé-only net profit margins in a with- and without-associates comparison (Exhibit 9). The profit margins include all the previously identified unusual charges because the analyst believes that they should not be excluded. They are real costs of doing business and seem to recur; they were actually incurred by the managers, who should be accountable for their stewardship of the shareholders' resources.

Exhibit 9 Net Profit Margin Spread

	2014	2013	2012
Consolidated net profit margin based on as-reported figures	16.27%	11.33%	11.90%
Nestlé-only profit margin	7.53%	9.96%	10.50%
Spread	8.74%	1.37%	1.40%

The analyst decides to focus on learning more about the drivers of Nestlé-only growth and revenues. He makes a note to himself to investigate the valuation aspects of the investment holdings later.

6 PHASES 3 & 4: ASSET BASE COMPOSITION

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

Asset Base Composition

The analyst examines the composition of the balance sheet over time, as shown in Exhibit 10.

Exhibit 10 Asset Composition as a Percentage of Total Assets

	2014 (%)	2013 (%)	2012 (%)	2011 (%)
Cash and equivalents	5.6	5.3	4.5	4.2
Short-term investments	1.1	0.5	2.8	2.7
Inventories	6.9	7.0	7.1	8.0
Trade and other receivables	10.1	10.1	10.4	11.5
Other current	1.8	2.0	2.2	2.4
Total current	**25.5**	**24.9**	**27.0**	**28.8**
Property, plant, and equipment, net	21.3	22.3	21.1	20.7
Goodwill	25.9	25.8	26.0	25.2
Intangible assets	14.8	10.5	10.3	7.7
Other non-current	12.5	16.4	15.6	17.7
Total	**100.0**	**99.9***	**100.0**	**100.1***

* Does not add to 100% because of rounding.

Although he expected significant investments in current assets, inventory, and physical plant assets—given that Nestlé is a food manufacturer and marketer—he is surprised to see so much investment in intangible assets, indicating that Nestlé's success may depend, in part, on successful acquisitions. Apparently, the company has been actively acquiring companies in the last four years. Goodwill and intangible assets, hallmarks of a growth-by-acquisition strategy, composed 40.7% of total assets in 2014; at the end of 2011, they amounted to 32.9% of total assets. The investing section of the statement of cash flows (Exhibit 11) shows that there have been acquisitions.

Exhibit 11 Nestlé Investing Activities, 2012–2014 (CHF millions)

	Total	2014	2013	2012
Capital expenditure	(14,115)	(3,914)	(4,928)	(5,273)
Expenditure on intangible assets	(1,236)	(509)	(402)	(325)
Acquisition of businesses	(13,223)	(1,986)	(321)	(10,916)
Disposal of businesses	884	321	421	142
Investments (net of divestments) in associates and joint ventures	3,851	3,958	(28)	(79)
Outflows from non-current treasury investments	(573)	(137)	(244)	(192)
Inflows from non-current treasury investments	4,460	255	2,644	1,561
Inflows/(outflows) from short-term treasury investments	115	(962)	400	677
Other investing activities	668	(98)	852	(86)
Cash flow from investing activities	(19,169)	(3,072)	(1,606)	(14,491)
Acquisitions' percentage of total investing activities	69.0%	64.6%	20.0%	75.3%

Except for a slowdown in acquisitions in 2013, Nestlé had been very active in devoting resources to acquisitions. For the full three-year span, 69.0% of the cash expenditures for investing activities were devoted to acquisitions. The largest single acquisition occurred in 2012, when Nestlé acquired the nutritional business of Wyeth for CHF10,846 million; this acquisition was 74.8% (= 10,846/14,491) of the cash used for investing activities in 2012.

PHASES 3 & 4: CAPITAL STRUCTURE ANALYSIS

7

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

Capital Structure Analysis

From the DuPont analysis, the analyst understands that Nestlé's overall financial leverage was rather stable over the last three years, which does not completely satisfy the analyst's curiosity regarding Nestlé's financing strategies. He knows that one shortcoming of financial leverage as a capital structure metric is that it says nothing about the nature, or riskiness, of the different financing instruments used by a company. For example, the financial burden imposed by bond debt is more onerous and bears more consequences in the event of default than do employee benefit plan obligations.

He decides to investigate Nestlé's capital structure more deeply by constructing a chart on a common-size basis, shown in Exhibit 12. The DuPont analysis indicated that the company's financial leverage remained within a narrow range over the last three years, from a low of 1.87 to a high of 1.98. A look at Exhibit 12, however, shows that Nestlé has been making its capital structure financially riskier over the last four years. Not only is the proportion of equity financing decreasing—from 74.2% in 2011 to 71.5% in 2014—but long-term financial liabilities have become a significantly greater part of the capital mix, increasing to 12.3% in 2014 from 7.8% in 2011. The "other long-term liabilities" (primarily employee benefit plan obligations and provisions) decreased from 17.9% in 2011 to 16.2% in 2014.

Exhibit 12 Percentages of Long-Term Capital Structure

	2014	2013	2012	2011
Long-term financial liabilities	12.3	11.8	10.3	7.8
Other long-term liabilities	16.2	14.9	17.9	17.9
Total equity	71.5	73.3	71.8	74.2
Total long-term capital	100.0	100.0	100.0	99.9*

* Does not add to 100% because of rounding.

Given the increased leverage in the long-term capital structure, the analyst wonders whether there have also been changes in the company's working capital accounts. He decides to examine Nestlé's liquidity situation. From the financial statements in Exhibits 2 and 3, he constructs the table shown in Exhibit 13.

Exhibit 13 Nestlé Working Capital Accounts and Ratios, 2011–2014

	2014	2013	2012	2011
Current ratio	1.03	0.91	0.88	0.93
Quick ratio	0.68	0.59	0.58	0.60
Defensive interval ratio*	106.6	91.9	110.0	110.5
Days sales outstanding (DSO)	51.1	50.0	53.0	54.7
Days on hand of inventory (DOH)	67.4	65.7	69.3	70.4
Number of days payables	(126.5)	(117.8)	(108.6)	(105.3)
Cash conversion cycle	(8.0)	(2.1)	13.7	19.8

* From Exhibit 2, for 2014: Daily cash expenditure = Expenses − Non-cash items = [Cost of goods sold + Distribution expenses + Marketing and administration expenses + R&D expenses − (Depreciation of PP&E + Amortisation of intangible assets) + Net trading expenses − (Impairment of PP&E and intangible assets) + (Net other operating expenses − Impairment of goodwill) + Net financial expenses]/365 = [47,553 + 8,217 + 19,651 + 1,628 − 3,058 + 797 − 159 + (3,114 − 1,908) + 637]/365 = 209.5. The defensive interval ratio is 22,340/209.5 = 106.6.

The analyst notices that the current and quick ratios improved slightly in 2014, after three years of relative stability. He also notices that the defensive interval ratio improved in 2014 after a significant decrease in 2013 from its prior levels. The improvements were modest; given the increase in long-term leverage, he was expecting more of a liquidity cushion in the working capital accounts. He found the cushion in that the speed of cash generation has been increasing: Since 2011, days' sales outstanding has decreased, as has days on hand of inventory, and the number of days payables has increased. In fact, the management of the working capital accounts has changed so much that Nestlé now has a negative eight days for its cash conversion cycle, mostly attributable to its steadily increasing delay in paying its vendors. In effect, Nestlé has been generating cash from its working capital accounts eight days before applying the cash to accounts payable.

PHASES 3 & 4: SEGMENT ANALYSIS: EARNINGS & CAPITAL

8

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

d evaluate how a given change in accounting standards, methods, or assumptions affects financial statements and ratios;

Segment Analysis and Capital Allocation

The DuPont analysis showed the declining profitability of Nestlé in its core operations, leading the analyst to subsequently learn more about the composition of the assets and to study the company's financing. He knows that asset turnover has been slowing at Nestlé and that the company has been looking to acquisitions for growth. But he still wonders about the health of the different businesses under the Nestlé umbrella and how effectively management has allocated capital to them. DuPont analysis does not provide answers to these kinds of questions, and he knows there is more information in the financial statements that might shed light on how management allocates capital internally as opposed to making acquisitions.

To understand any geopolitical investment risks, as well as the economies in which Nestlé operates, the analyst wants to know which geographic areas are of the greatest importance to the company. One issue the analyst identifies is that Nestlé reports segment information by management responsibility and geographic area (hereafter referred to as "segment"), not by segments based exclusively on geographic areas. From the segment information in Exhibit 14, he notes that the sales and operating profit of the European segment decreased in absolute terms and as a percentage of total business in 2014 compared with 2012. The decrease in profits has been consistent over the period. The sales of the Americas segment have also become a smaller contributor to the whole company's revenue base in the same period and, like the European segment, have decreased slightly since 2012. The Americas operating profit has decreased consistently since 2012, and like the European segment, the Americas contribution to total operating profit in 2014 is a smaller proportion than in 2012. The Asia, Oceania, and Africa segment repeated the pattern: lower sales and operating profit, with a decrease in both measures in each of the two years following 2012. The smallest segment, Nestlé Waters, was not a true geographic segment. It showed minor growth in revenues and operating profit between 2012 and 2014 and contributed essentially the same proportion of sales and operating profit in 2014 as it did in

2012. Nestlé Nutrition grew significantly during the period: It contributed 10.5% of revenues in 2014 (only 8.8% in 2012), and its operating profit contributed 14.2% of revenues in 2014 compared with 11.2% in 2012. The analyst remembers that Nestlé acquired the Wyeth Nutritionals business in 2012, which would explain the solid growth. "Other businesses," which is a collectively large group of disparate businesses, also increased in importance between 2012 and 2014, accounting for 15.2% of sales in 2014 (13.2% in 2012) and 18.9% of operating profit (15.3% in 2012). Both measures (sales and operating profit) grew in 2014, and the analyst attributes that growth to Nestlé's gaining full control of Galderma in 2014.

Exhibit 14 Sales and EBIT by Segment (CHF millions)

	2014		2013		2012		Year-to-Year % Change	
Sales	**Amount**	**% Total**	**Amount**	**% Total**	**Amount**	**% Total**	**2014**	**2013**
Europe	15,175	*16.6*	15,567	*16.9*	15,388	*17.2*	–2.5	1.2
Americas	27,277	*29.8*	28,358	*30.8*	28,613	*31.9*	–3.8	–0.9
Asia, Oceania, and Africa	18,272	*19.9*	18,851	*20.5*	18,875	*21.0*	–3.1	–0.1
Nestlé Waters	7,390	*8.1*	7,257	*7.9*	7,174	*8.0*	1.8	1.2
Nestlé Nutrition	9,614	*10.5*	9,826	*10.7*	7,858	*8.8*	–2.2	25.0
Other businesses[a]	13,884	*15.2*	12,299	*13.3*	11,813	*13.2*	12.9	4.1
	91,612	*100.0*	92,158	*100.0*	89,721	*100.0*		

	2014		2013		2012		Year-to-Year % Change	
Trading operating profit	**Amount**	**% Total**	**Amount**	**% Total**	**Amount**	**% Total**	**2014**	**2013**
Europe	2,327	16.6	2,331	16.6	2,363	17.6	–0.2	–1.4
Americas	5,117	36.5	5,162	36.7	5,346	39.7	–0.9	–3.4
Asia, Oceania, and Africa	3,408	24.3	3,562	25.4	3,579	26.6	–4.3	–0.5
Nestlé Waters	714	5.1	665	4.7	640	4.8	7.4	3.9
Nestlé Nutrition	1,997	14.2	1,961	14.0	1,509	11.2	1.8	30.0
Other businesses[a]	2,654	18.9	2,175	15.5	2,064	15.3	22.0	5.4
Unallocated items	(2,198)	–15.7	(1,809)	–12.9	(2,037)	–15.1	21.5	–11.2
	14,019	100.0	14,047	100.0	13,464	100.0		

[a] Group mainly includes Nespresso, Nestlé Professional, Nestlé Health Science, and Nestlé Skin Health.

For several reasons, the analyst is somewhat frustrated by the segment information presented by Nestlé. He would like to look at trends over more than just three years, but the change in accounting principles in 2013 (for IFRS 11) was not carried back in the segment information prior to 2012. That accounting change eliminated the proportional consolidation method of accounting for joint ventures and made the 2011 segment information non-comparable with the figures presented for 2012 and later. The earlier amounts included proportional amounts of sales and operating profits for the segments, and a comparison with later years would be flawed.

Another problem with the segment information is that it is not defined by category with fully geographic information or product information. The analyst notes that three geographically classified segments accounted for 66.3% of revenues in 2014 and 70.1% in 2012; the operating profit for the same three segments amounted to 77.4% in 2014 and 83.9% in 2012. Thus, these segments are declining in importance to Nestlé as a whole, whereas Nestlé Waters and Other businesses are increasing in size and importance. Yet, it would seem likely that both of these segments have geographically different operations as well, which are not being accounted for in the other three geographic segments. These segments are growing in relevance, and more information about them would be useful. For instance, the Other businesses segment includes a coffee product line, professional products, health care products, and skin care products. Together, they amount to almost 19% of operating profit, yet they seem unlikely to have similar distribution channels, profitability levels, and growth potential.

PHASES 3 & 4: SEGMENT ANALYSIS: CASH FLOW & CAPITAL

9

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

e analyze and interpret how balance sheet modifications, earnings normalization, and cash flow statement related modifications affect a company's financial statements, financial ratios, and overall financial condition.

The segment information is presented on the basis that management uses to make decisions. The analyst moves on with his segment analysis and evaluation of capital allocation, gathering the segment information shown in Exhibit 15 regarding Nestlé's capital expenditures and assets.

Exhibit 15 Asset and Capital Expenditure Segment Information (CHF millions)

	Assets*			Capital Expenditures		
	2014	2013	2012	2014	2013	2012
Europe	11,308	11,779	11,804	747	964	1,019
Americas	20,915	21,243	22,485	1,039	1,019	1,073
Asia, Oceania, and Africa	15,095	14,165	14,329	697	1,280	1,564
Nestlé Waters	6,202	6,046	6,369	308	377	407
Nestlé Nutrition	24,448	22,517	24,279	363	430	426
Other businesses[a]	21,345	9,564	9,081	573	642	550
	99,313	85,314	88,347	3,727	4,712	5,039

* Assets do not equal total assets on the balance sheet because of inter-segment assets and non-segment assets.
[a] Group mainly includes Nespresso, Nestlé Professional, Nestlé Health Science, and Nestlé Skin Health.

Using the information from Exhibit 14 to calculate EBIT margins, as well as the information about the asset and capital expenditure distribution from Exhibit 15, the analyst constructs the table in Exhibit 16, ranking by descending order of EBIT profitability in 2014.

Exhibit 16 EBIT Margin, Asset, and Capital Expenditure Proportions by Segment

	EBIT Margin %			% of Total Assets			% of Total Capital Expenditures		
	2014	2013	2012	2014	2013	2012	2014	2013	2012
Nestlé Nutrition	20.77	19.96	19.20	24.6	26.4	27.5	9.7	9.1	8.5
Other businesses[a]	19.12	17.68	17.47	21.5	11.2	10.3	15.4	13.6	10.9
Americas	18.76	18.20	18.68	21.1	24.9	25.5	27.9	21.6	21.3
Asia, Oceania, and Africa	18.65	18.90	18.96	15.2	16.6	16.2	18.7	27.2	31.0
Europe	15.33	14.97	15.36	11.4	13.8	13.4	20.0	20.5	20.2
Nestlé Waters	9.66	9.16	8.92	6.2	7.1	7.2	8.3	8.0	8.1
				100.0	100.0	100.1*	100.0	100.0	100.0

* Does not add to 100% because of rounding.
[a] Group mainly includes Nespresso, Nestlé Professional, Nestlé Health Science, and Nestlé Skin Health.

Although the segmentation is not purely geographic, the analyst can still make some judgments about the allocation of capital. On the premise that the largest investments in assets require a similar proportion of capital expenditures, he calculates ratios of the capital expenditure proportion to the total asset proportion for the last three years and compares them with the current EBIT profitability ranking. The resulting table is shown in Exhibit 17.

Exhibit 17 Ratio of Capital Expenditure Percentage to Total Asset Percentage Ranked by EBIT Margin

	EBIT Margin %	Ratio of Total Capital Expenditure % to Total Asset %		
	2014	2014	2013	2012
Nestlé Nutrition	20.77	0.39	0.34	0.31
Other businesses[a]	19.12	0.72	**1.21**	**1.06**
Americas	18.76	**1.32**	0.87	0.84
Asia, Oceania, and Africa	18.65	**1.23**	**1.64**	**1.91**
Europe	15.33	**1.75**	**1.49**	**1.51**
Nestlé Waters	9.66	**1.34**	**1.13**	**1.13**

[a] Group mainly includes Nespresso, Nestlé Professional, Nestlé Health Science, and Nestlé Skin Health.

A ratio of 1 indicates that the segment's proportion of capital expenditures is the same as its proportion of total assets. A ratio of *less than* 1 indicates that the segment is being allocated a lesser proportion of capital expenditures than its proportion of total assets; if a trend develops, the segment will become less significant over time. A ratio of *greater than* 1 indicates the company is growing the segment; the segment is receiving a "growth allocation" of capital spending. Comparing the ratio with the EBIT margin percentage gives the analyst an idea of whether the company is investing its capital in the most profitable segments. (In Exhibit 17, the ratios greater than 1 are bolded for ease of viewing.)

Equipped with these premises, the analyst is puzzled by the capital allocation taking place within Nestlé. The most profitable segment is Nestlé Nutrition, but over the last three years, it has received the lowest proportion of capital expenditures. The company has invested in the nutrition segment by acquisition, such as the Wyeth Nutritionals business in 2012. One would expect that a more substantial operation would require more capital expenditures on maintenance. The capital expenditures for the nutrition segment have increased only nominally since 2012.

The Other businesses segment is the next most profitable segment in EBIT margin terms. The analyst has difficulty understanding just why the profit margins are high in this segment because of the variety of businesses it contains. It appears that the company's managers are allocating capital to it in a significant way. Although it did not receive a "growth allocation" of capital expenditures in 2014, it received a growth allocation in the previous two years. The Americas segment and the Asia, Oceania, and Africa segment have similar EBIT margins, which are in the same range as those of the Nestlé Nutrition and Other businesses segments. Given their profitability levels and substantial operations, the analyst is encouraged to see that they are receiving "growth allocations" of capital spending.

Less encouraging, however, is the past and continuing significant allocation of capital spending to the European segment. Even more questionable is the high proportional allocation of capital spending to the Nestlé Waters segment, which has had the lowest profit margins. The analyst is uncomfortable with growth investments in such a low-return business but notes that the absolute levels of capital expenditures are the lowest of all the segments in each year.

In a worst-case scenario, if the company were to continue making growth allocations of capital toward the lowest-margined businesses, such as Europe and Nestlé Waters, the overall Nestlé-only returns might be affected negatively. As a result, Nestlé might become more dependent on its investment in associates to sustain performance.

The analyst knows that accrual performance measures, such as EBIT, can produce results that do not indicate an entity's ability to generate cash flow, and he wonders whether this limitation has any effect on Nestlé management's capital allocation decisions. He also knows that at the segment level, cash flow information is not publicly available. He decides to at least approximate cash flow by adding depreciation expense to operating profit and then relate the approximated cash flow to the average total assets of each segment. This approach provides an approximation of cash return relative to the continued investment in a particular segment.

The analyst combines the segment operating profit from Exhibit 14 and the segment depreciation and amortisation in Exhibit 18 to estimate the segment cash generation shown in Exhibit 18. Because he wants to eliminate the effects of any investment peaks or valleys, he also averages the total assets for each segment in Exhibit 18. The average total assets in 2012 include the 2011 total assets that were prepared on a pre–IFRS 11 basis, for which no adjustment is available. The analyst is aware of the irreconcilable difference but believes that the averaging of the two years' amounts will help dilute the difference. He notes that if any resulting measures based on 2011 data points appear to be outliers, he will dismiss them.

Exhibit 18 Segment Depreciation and Amortisation, Segment Cash Generation, and Average Assets (CHF millions)

	Depreciation and Amortisation			Segment Cash Generation			Average Total Assets*		
	2014	2013	2012	2014	2013	2012	2014	2013	2012
Europe	473	517	533	2,800	2,848	2,896	11,544	11,792	11,683
Americas	681	769	899	5,798	5,931	6,245	21,079	21,864	22,783
Asia, Oceania, and Africa	510	520	553	3,918	4,082	4,132	14,630	14,247	14,068
Nestlé Waters	403	442	491	1,117	1,107	1,131	6,124	6,208	6,486
Nestlé Nutrition	330	337	176	2,327	2,298	1,685	23,483	23,398	18,564
Other businesses[a]	525	437	295	3,179	2,612	2,359	15,455	9,323	10,009

* Average of total assets at beginning and end of the year.
[a] Group mainly includes Nespresso, Nestlé Professional, Nestlé Health Science, and Nestlé Skin Health.

In Exhibit 19, the analyst computes each segment's cash operating return on total assets and compares the results with the 2014 ranking of capital expenditures (Exhibit 17) as well as the EBIT margins. They are ranked in descending order of the ratio of capital expenditure percentage to percentage of total assets. The lighter shading indicates the highest EBIT margin and cash return on assets for each year, and the darker shading indicates the lowest EBIT margin and cash return on assets for each year.

Exhibit 19 Segment Cash Operating Return on Assets

	2014		Segment Cash Return on Average Total Assets		
	Capex %/ Asset %	EBIT %	2014 (%)	2013 (%)	2012 (%)
Europe	1.75	15.3	24.3	24.2	24.8
Nestlé Waters	1.34	9.7	18.2	17.8	17.4
Americas	1.32	18.8	27.5	27.1	27.4
Asia, Oceania, and Africa	1.23	18.7	26.8	28.7	29.4
Other businesses[a]	0.72	19.1	20.6	28.0	23.6
Nestlé Nutrition	0.39	20.8	9.9	9.8	9.1

[a] Group mainly includes Nespresso, Nestlé Professional, Nestlé Health Science, and Nestlé Skin Health.

PHASES 3 & 4: SEGMENT ANALYSIS BY PRODUCT GROUP

10

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

The analyst is surprised to see that the Nestlé Nutrition segment, which has the highest EBIT profit margin, consistently has the lowest cash return on total assets. When he looks at the segments with respect to EBIT margins, he is disappointed with the allocation of capital spending to Nestlé Nutrition, thinking that it is too low. When he looks at it using the cash return on total assets measure, the low allocation of spending makes much more sense. He is pleased to see that the segments with the highest cash return on total assets each year—the Americas and the Asia, Oceania, and Africa segments—are receiving growth allocations of capital spending. He is also encouraged that the European segment, though a poor performer with respect to EBIT margin, has cash returns on total assets that are competitive with the other segments and far ahead of Nestlé Waters and Nestlé Nutrition. Even Nestlé Waters, which had not appeared very attractive with respect to EBIT margin, is generating strong cash returns on total assets. The exercise restores the analyst's confidence that management is allocating capital in a rational manner. It makes sense to him that if management makes capital budgeting decisions on a cash flow basis, they should be evaluated on a cash flow basis also.

He decides to look at Nestlé's capital allocation process from a product group standpoint. The sales and EBIT information is shown in Exhibit 20. From the table, he notes that the Nutrition and Health Science product group is the only one with significant growth in either sales or EBIT, and that is the segment in which the company has been making its acquisitions in the last few years. He also notes that the EBIT margin for the Nutrition and Health Science product group has increased in each of the last two years, and although it is among the highest over the last three years, the Powdered and Liquid Beverages product group has consistently shown higher EBIT margins. The Powdered and Liquid Beverages product group EBIT margins far exceed the lowest-ranking EBIT margins of the Water product group.

Exhibit 20 Sales and EBIT Segment Information by Product Group (CHF millions)

	2014		2013		2012		Year-to-Year % Change	
Sales		**% Total**		**% Total**		**% Total**	**2014**	**2013**
Powdered and Liquid Beverages	20,302	22.2	20,495	22.2	20,248	22.6	–0.9	1.2
Water	6,875	7.5	6,773	7.3	6,747	7.5	1.5	0.4
Milk Products and Ice Cream	16,743	18.3	17,357	18.8	17,344	19.3	–3.5	0.1
Nutrition and Health Science	13,046	14.2	11,840	12.8	9,737	10.9	10.2	21.6
Prepared Dishes and Cooking Aids	13,538	14.8	14,171	15.4	14,394	16.0	–4.5	–1.5
Confectionery	9,769	10.7	10,283	11.2	10,441	11.6	–5.0	–1.5

(continued)

Exhibit 20 (Continued)

	2014		2013		2012		Year-to-Year % Change	
Sales		% Total		% Total		% Total	2014	2013
Pet Care	11,339	12.4	11,239	12.2	10,810	12.0	0.9	4.0
	91,612	100.0	92,158	100.0	89,721	100.0		

EBIT								
Powdered and Liquid Beverages	4,685	33.4	4,649	33.1	4,445	33.0	0.8	4.6
Water	710	5.1	678	4.8	636	4.7	4.7	6.6
Milk Products and Ice Cream	2,701	19.3	2,632	18.7	2,704	20.1	2.6	−2.7
Nutrition and Health Science	2,723	19.4	2,228	15.9	1,778	13.2	22.2	25.3
Prepared Dishes and Cooking Aids	1,808	12.9	1,876	13.4	2,029	15.1	−3.6	−7.5
Confectionery	1,344	9.6	1,630	11.6	1,765	13.1	−17.5	−7.6
Pet Care	2,246	16.0	2,163	15.4	2,144	15.9	3.8	0.9
Unallocated items	(2,198)	−15.7	(1,809)	−12.9	(2,037)	−15.1	21.5	−11.2
	14,019	100.0	14,047	100.0	13,464	100.0		

EBIT margin	2014	2013	2012
Powdered and Liquid Beverages	23.1%	22.7%	22.0%
Water	10.3%	10.0%	9.4%
Milk Products and Ice Cream	16.1%	15.2%	15.6%
Nutrition and Health Science	20.9%	18.8%	18.3%
Prepared Dishes and Cooking Aids	13.4%	13.2%	14.1%
Confectionery	13.8%	15.9%	16.9%
Pet Care	19.8%	19.2%	19.8%
Total	15.3%	15.2%	15.0%

Unfortunately for purposes of his analysis, Nestlé does not provide capital expenditure information by product group. Compared with the segment analysis he performed, the analyst's scope is more limited in examining product groups. All that can be done is to look at the return on assets with respect to EBIT rather than on a cash-generated basis. Nevertheless, the analyst decides to work with all the available information. To further examine capital allocation decisions, he gathers the asset information by product group from the financial statements, as shown in Exhibit 21. The reported total assets differ by segment and product group presentation because Nestlé reports its assets on an *average* basis for product groups and on a *year-end* basis for segments. A significant amount of assets is unallocated to segments, but there is no unallocated amount by product group. He calculates the EBIT return on assets as EBIT divided by average assets and determines the proportion of total average assets devoted to each product group. The highest EBIT percentage, EBIT return on assets, and percentage of total assets each year are lightly shaded, and the lowest are shaded darker.

Exhibit 21 Asset Segment Information by Product Group (CHF millions)

	Average Assets			EBIT %	EBIT Return on Assets			% Total Assets		
	2014	2013	2012	2014	2014	2013	2012	2014	2013	2012
Powdered and Liquid Beverages	11,599	11,044	10,844	23.1%	40.4%	42.1%	41.0%	11.6%	11.5%	12.4%
Water	5,928	6,209	6,442	10.3%	12.0%	10.9%	9.9%	6.0%	6.4%	7.4%
Milk Products and Ice Cream	14,387	14,805	14,995	16.1%	18.8%	17.8%	18.0%	14.4%	15.4%	17.1%
Nutrition and Health Science	32,245	28,699	19,469	20.9%	8.4%	7.8%	9.1%	32.4%	29.8%	22.2%
Prepared Dishes and Cooking Aids	13,220	13,289	13,479	13.4%	13.7%	14.1%	15.1%	13.3%	13.8%	15.4%
Confectionery	7,860	8,190	8,343	13.8%	17.1%	19.9%	21.2%	7.9%	8.5%	9.5%
Pet Care	14,344	14,064	13,996	19.8%	15.7%	15.4%	15.3%	14.4%	14.6%	16.0%
	99,583	96,300	87,568	15.3%	14.1%	14.6%	15.4%	100.0%	100.0%	100.0%

The analyst uses this information to make some important observations:

- The Nutrition and Health Science product group—which the company has been investing in over the last several years—has the lowest EBIT return on assets in each of the last three years and makes up the greatest portion of total assets.
- The EBIT return on assets for the Nutrition and Health Science product group is even lower than that of the Water product group, which has the lowest EBIT margin.
- The Nutrition and Health Science product group's EBIT return on assets is well below the total company's EBIT return on assets (8.4% versus 14.1% in 2014, 7.8% versus 14.6% in 2013, and 9.1% versus 15.4% in 2012).
- The Nutrition and Health Science product group drags down the overall return in each year as it becomes a bigger part of the whole.
- The EBIT return on assets is highest for the Powdered and Liquid Beverages product group, possibly because it might not need much in the way of assets or capital spending: It is one of the lesser components of total assets. Furthermore, it has the highest EBIT margin of all the product groups. Given the high EBIT margin, the high EBIT return on assets, and the low dedication of total assets, the analyst wonders whether the company is allocating capital among its product offerings effectively. It would make sense to devote as many resources as possible to where returns are best.
- He also wonders about management's capital allocation skills regarding acquisitions. The EBIT return on assets in the Nutrition and Health Science product group is weak, and the company has been making acquisitions in that group. He finds it troubling that Nestlé took a goodwill impairment charge of CHF1,908 million in 2014—something directly related to management's skill in making past acquisitions.

11 PHASES 3 & 4: ACCRUALS AND EARNINGS QUALITY

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

b identify financial reporting choices and biases that affect the quality and comparability of companies' financial statements and explain how such biases may affect financial decisions;

c evaluate the quality of a company's financial data and recommend appropriate adjustments to improve quality and comparability with similar companies, including adjustments for differences in accounting standards, methods, and assumptions;

At this point, the information reviewed by the analyst has not increased his enthusiasm for Nestlé's operating and capital allocation prowess. He considers a worst-case possibility: Could the company try to make up for weak operating performance by manipulating accounting inputs? He makes it a point to understand whether accruals play a role in the company's performance.

He decides to examine the balance-sheet-based accruals and the cash-flow-based accruals over the last few years. From the Nestlé financial statements, he assembles the information and intermediate calculations shown in Exhibit 22.

Exhibit 22 Selected Information from Balance Sheet and Statement of Cash Flows (CHF millions)

	2014	2013	2012	2011
Balance Sheet Accrual Info:				
Total assets	133,450	120,442	125,877	113,440
Cash and short-term investments	8,881	7,053	9,296	7,782
Operating assets (A)	124,569	113,389	116,581	105,658
Total liabilities	61,566	56,303	63,213	55,098
Long-term debt	12,396	10,363	9,008	6,165
Debt in current liabilities	8,810	11,380	18,408	15,945
Operating liabilities (B)	40,360	34,560	35,797	32,988
Net operating assets (NOA) [(A) – (B)]	84,209	78,829	80,784	72,670
Balance-sheet-based aggregate accruals (year-to-year change in NOA)	5,380	(1,955)	8,114	6,218
Average NOA	81,519	79,807	76,727	69,561
Statement of Cash Flows Accrual Info:				
Profit from continuing operations	14,904	10,445	10,677	
Operating cash flow	(14,700)	(14,992)	(15,668)	
Investing cash flow	3,072	1,606	14,491	
Cash-flow-based aggregate accruals	3,276	(2,941)	9,500	

The analyst calculates the balance-sheet-based and cash-flow-based accruals ratios, which are measures of financial reporting quality.[4] The ratios are calculated as follows:

Balance sheet accruals ratio for time t = $(NOA_t - NOA_{t-1})/[(NOA_t + NOA_{t-1})/2]$, and

Cash flow accruals ratio for time t = $[NI_t - (CFO_t + CFI_t)]/[(NOA_t + NOA_{t-1})/2]$,

where NI is net income, CFO is cash flow from operations, and CFI is cash flow from investing.

The accruals ratios for the last three years are shown in Exhibit 23.

Exhibit 23 Accruals Ratios (CHF millions)

	2014	2013	2012
Balance-sheet-based aggregate accruals (year-to-year change in NOA)	5,380	(1,955)	8,114
Divided by: Average NOA	81,519	79,807	76,727
Balance-sheet-based accruals ratio	**6.6%**	**−2.4%**	**10.6%**
Cash-flow-based aggregate accruals	3,276	(2,941)	9,500
Divided by: Average NOA	81,519	79,807	76,727
Cash-flow-based accruals ratio	**4.0%**	**−3.7%**	**12.4%**

The analyst notes that the absolute level of accruals on the balance sheet is minor relative to the size of the operating assets, on either an ending balance basis or an average basis. Similarly, the fluctuation in the balance-sheet-based accruals ratio is low. The analyst would have been more concerned if the absolute levels of the accruals ratio were high; even more worrisome would have been if they were consistently trending higher. That was not the case. The cash-flow-based accruals ratio exhibits a similar pattern. For the most recent two years, both ratios are lower than in 2012 and indicate that accruals are not a large factor in the financial results. The analyst still decides to examine the quality of Nestlé's cash flow and its relationship to net income.

PHASES 3 & 4: CASH FLOW RELATIONSHIPS

12

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

e analyze and interpret how balance sheet modifications, earnings normalization, and cash flow statement related modifications affect a company's financial statements, financial ratios, and overall financial condition.

4 If you are interested in subcomponents of accrual activity, simply focus on the relevant line item from the balance sheet. For example, looking at the change in net receivables over a fiscal period deflated by average NOA will give you a sense of the magnitude of accrued revenue attributable to net credit sales.

He begins his analysis with the compilation of Nestlé's statements of cash flows shown in Exhibit 24.

Exhibit 24 Nestlé's Statement of Cash Flows, 2012–2014 (CHF millions)

	2014	2013	2012
Operating activities			
Operating profit	10,905	13,068	13,388
Non-cash items of income and expense	6,323	4,352	3,217
Cash flow before changes in operating assets and liabilities	17,228	17,420	16,605
Decrease/(increase) in working capital	(114)	1,360	2,015
Variation of other operating assets and liabilities	85	(574)	(95)
Cash generated from operations	17,199	18,206	18,525
Net cash flows from treasury activities	(356)	(351)	(324)
Taxes paid	(2,859)	(3,520)	(3,118)
Dividends and interest from associates and joint ventures	716	657	585
Operating cash flow	**14,700**	**14,992**	**15,668**
Investing activities			
Capital expenditure	(3,914)	(4,928)	(5,273)
Expenditure on intangible assets	(509)	(402)	(325)
Acquisition of businesses	(1,986)	(321)	(10,916)
Disposal of businesses	321	421	142
Investments (net of divestments) in associates and joint ventures	3,958	(28)	(79)
Outflows from non-current treasury investments	(137)	(244)	(192)
Inflows from non-current treasury investments	255	2,644	1,561
Inflows/(outflows) from short-term treasury investments	(962)	400	677
Other investing activities	(98)	852	(86)
Cash flow from investing activities	**(3,072)**	**(1,606)**	**(14,491)**
Financing activities			
Dividends paid to shareholders of the parent	(6,863)	(6,552)	(6,213)
Dividends paid to non-controlling interests	(356)	(328)	(204)
Acquisition (net of disposal) of non-controlling interests	(49)	(337)	(165)
Purchase of treasury shares	(1,721)	(481)	(532)
Sale of treasury shares	104	60	1,199
Inflows from bonds and other non-current financial debt	2,202	3,814	5,226
Outflows from bonds and other non-current financial debt	(1,969)	(2,271)	(1,650)
Inflows/(outflows) from current financial debt	(1,985)	(6,063)	2,325
Cash flow from financing activities	**(10,637)**	**(12,158)**	**(14)**
Currency retranslations	42	(526)	(219)
Increase/(decrease) in cash and cash equivalents	**1,033**	**702**	**944**
Cash and cash equivalents at beginning of year	**6,415**	**5,713**	**4,769**
Cash and cash equivalents at end of year	**7,448**	**6,415**	**5,713**

The analyst's most pressing concerns include the following: Are Nestlé's operating earnings backed by cash flow? Are the accrual measures telling the whole story? Are the operating earnings the result of accounting choices? To convince himself of the representativeness of the Nestlé earnings, he first compares the cash generated by operations with the operating profit as shown in Exhibit 25. The amounts in Exhibit 25 are found in the cash flow statements in Exhibit 24.

Exhibit 25 Operating Cash Flow to Operating Profit, 2012–2014 (CHF millions)

	2014	2013	2012
Cash generated from operations	17,199	18,206	18,525
Operating profit	10,905	13,068	13,388
Cash generated from operations/Operating profit	**1.58**	**1.39**	**1.38**

The cash generated from operations is comparable to accrual basis operating income *but on a cash flow basis*. If the cash flow generated by operations was significantly or consistently less than operating profit, one would have reason to be suspicious about the quality of the operating profit. The analyst is encouraged by the fact that the cash generated from operations substantially exceeded the operating profit in each of the last three years.

Knowing that Nestlé has made a number of acquisitions, the analyst decides to examine the relationship between operating cash flow and total assets. *Cash flow* is a measure of the operational success of the company's investment projects: Successful investments generate cash rather than absorbing it. *Total assets* reflect the sum total of management's resource allocations over time. Cash generated by total assets indicates the kind of cash return that is generated by all investments. The relationship is shown in Exhibit 26.

Exhibit 26 Ratio of Operating Cash Flow to Total Assets, 2012–2014 (CHF millions)

	2014	2013	2012
Cash generated from operations	17,199	18,206	18,525
Average total assets	126,946	123,160	119,659
Cash return on total assets	13.5%	14.8%	15.5%

Again, the analyst finds himself concerned about the effectiveness of management's asset allocation decisions. Although the 13.5% cash return on total assets is a high return on investment, the trend is declining. The analyst thinks back to the 2014 goodwill impairment and the poor EBIT return on assets in the Nutrition and Health Science product group, in which acquisitions have been occurring lately.

Given the negative trend in asset returns, the analyst looks at Nestlé's liquidity and funding ability relative to cash flow. He decides to compare cash flow with reinvestment, debt, and debt-servicing capacity, as shown in Exhibit 27.

The analyst sees that reinvestment needs have been covered by cash flow by a factor of 3.89 in 2014, 3.42 in 2013, and 3.31 in 2012. Even better, the trend is improving.

He also sees that based on the relationship of cash flow to total debt, the company is not highly leveraged, with cash generated from operations at 78.3% of total debt at the end of 2014. The ratio is high enough to indicate that additional borrowing could be arranged should an investment opportunity arise. Furthermore, the analyst notes that Nestlé has the capacity to pay off its debt in approximately two years even while maintaining its current reinvestment policy [21,963/(17,199 – 4,423)].

Finally, the cash flow interest coverage ratio indicates more than satisfactory financial strength in the current year, with cash flow 33.2 times the interest paid. Like the ratio of cash flow to total debt, it indicates that the company has sufficient financial capacity to add more debt if there is an investment opportunity.

Exhibit 27 Ratio of Operating Cash Flow to Reinvestment, Debt, and Debt-Servicing Capacity, 2012–2014 (CHF millions)

	2014	2013	2012
*Cash flow to reinvestment:**			
Cash generated from operations	17,199	18,206	18,525
Capital expenditures	3,914	4,928	5,273
Expenditure on intangible assets	509	402	325
Total reinvestment spending	4,423	5,330	5,598
Ratio of cash flow to reinvestment	**3.89**	**3.42**	**3.31**
Cash flow to total debt:			
Cash generated from operations	17,199	18,206	18,525
Current debt (short-term financial liabilities)	8,810	11,380	18,408
Current derivative liabilities	757	381	423
Long-term debt (long-term financial liabilities)	12,396	10,363	9,008
Total debt	21,963	22,124	27,839
Ratio of cash flow to total debt	**78.3%**	**82.3%**	**66.5%**
Cash flow interest coverage:			
Cash generated from operations	17,199	18,206	18,525
Cash interest paid	518	505	559
Cash flow interest coverage	**33.2**	**36.1**	**33.1**

* Information is from Exhibit 24.

13 PHASES 3 & 4: DECOMPOSITION AND ANALYSIS OF THE COMPANY'S VALUATION

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

c evaluate the quality of a company's financial data and recommend appropriate adjustments to improve quality and comparability with similar companies, including adjustments for differences in accounting standards, methods, and assumptions;

At this point, the analyst believes he has obtained sufficient information about the company's sources of earnings and returns on shareholders' equity, its capital structure, the results of its capital allocation decisions, and its earnings quality. Before he makes his report to the portfolio manager, he wants to study the company's market valuation. During his reading of the annual reports, he noted that Nestlé has a significant equity position (23.4%) in L'Oréal (Paris exchange: OR), a French cosmetics company. L'Oréal is accounted for in the financial statements as an investment in associates because Nestlé's ownership position does not give it control. Although L'Oréal contributes to the earnings of Nestlé as a whole, it is also valued separately in the public markets, and its discrete valuations may be very different from its embedded Nestlé valuation. To determine the value that the market places solely on Nestlé operations, the analyst first removes the value of the L'Oréal holding from the Nestlé market value, as shown in Exhibit 28.

Exhibit 28 Nestlé Market Value without L'Oréal as of 31 December 2014 (Currency in millions, except share prices)

L'Oréal value:	
31 Dec 2014 share price	€139.30
Shares held by Nestlé (millions)	129.881
L'Oréal holding value	€18,092
31 Dec CHF/EUR rate	1.202
L'Oréal holding value	CHF21,747
Nestlé market value, with and without L'Oréal:	
Nestlé 29 Dec 2014 share price	CHF72.95
Shares outstanding (millions)	3,168.400
Nestlé market capitalization	CHF231,135
Value of L'Oréal holding	(21,747)
Implied value of Nestlé operations	CHF209,388
Pro rata market value:	
L'Oréal	9.4%
Nestlé	90.6%
	100.0%

The value of the L'Oréal holding is slightly less than 10% of the value of Nestlé's market capitalization. The analyst now wants to remove the earnings of L'Oréal from the earnings of the combined entity (Exhibit 29) to make a price-to-earnings comparison for Nestlé earnings alone. For L'Oréal, this comparison is simple: Nestlé discloses in its annual report that L'Oréal has contributed CHF934 million to current year earnings. After isolating the different earnings sources, the analyst prepares the table shown in Exhibit 30, which compares the different market values and price-to-earnings ratios.

Exhibit 29 Calculation of Nestlé Earnings without L'Oréal as of 31 December 2014 (CHF millions)

Calculation of Nestlé standalone earnings:	2014
Nestlé consolidated earnings	14,904
Less: L'Oréal earnings	(934)
Nestlé standalone earnings	13,970
Less: Non-controlling interests	(448)
Nestlé standalone earnings to shareholders	13,522

At the time of the analysis (early 2015), Nestlé's common stock traded at a price-to-earnings multiple of 16.0 based on its year-end market value of CHF231,135 million and trailing earnings (attributable to controlling interests) of CHF14,456 million: a discount of 20% to the price-to-earnings multiple of 19.9 for the S&P 500 Index at year-end 2014. Once the earnings and available market value of the L'Oréal holding are taken out of the price-to-earnings valuation, the shares of the "Nestlé-only" company are selling at a slightly higher discount: At 15.5 times earnings, the discount to the overall market's price-to-earnings multiple was a steeper 22%. At first, the analyst is surprised by Nestlé's discount to the market multiple, given that the company has consistently demonstrated meaningful cash flows and earnings and possesses low financial leverage. He considers whether the discount might be attributable to Nestlé's slipping core profitability. The analyst concludes that Nestlé shares may be discounted by the market because investors may be developing a skeptical attitude toward the company.

Exhibit 30 Comparison of Decomposed Nestlé Earnings and Price-to-Earnings Ratios

Earnings (CHF millions)	Market Value	Earnings (Group Shareholder Level)	Respective Price-to-Earnings Ratios
L'Oréal	21,747	934	23.3
Implied Nestlé-only	209,388	13,522	15.5
Actual earnings available to Nestlé parent company shareholders	231,135	14,456	16.0

Recap (%):	Market Value	Earnings
L'Oréal	9.4	6.5
Implied Nestlé-only	90.6	93.5
	100.0	100.0

At this point, the analyst believes that he has processed and analyzed the data sufficiently to pull together his findings and make his report to the portfolio manager.

PHASES 5 & 6: DEVELOP AND COMMUNICATE CONCLUSIONS AND RECOMMENDATIONS AND FOLLOW-UP

14

a demonstrate the use of a framework for the analysis of financial statements, given a particular problem, question, or purpose (e.g., valuing equity based on comparables, critiquing a credit rating, obtaining a comprehensive picture of financial leverage, evaluating the perspectives given in management's discussion of financial results);

Phase 5: Develop and Communicate Conclusions and Recommendations (e.g., with an Analysis Report)

As a result of the analyses performed, the analyst has gathered sufficient evidence regarding many of Nestlé's operational and financial characteristics and believes he is able to address the concerns initially expressed by the portfolio manager. Summary points he will cover in his report are divided into two classes: support for an investment in Nestlé shares and causes for concern.

Support for an Investment in Nestlé Shares

- Nestlé has the financial stability to fund growth in its existing operations and carry out its growth-by-acquisition strategy. The company's current liquidity and cash flows are more than adequate for future operating and investment purposes. The company has low leverage, and the capital structure is capable of supporting future operations and strategic plans.
- The operating cash flows have consistently exceeded the operating earnings. The ratio of operating cash to operating profit has been consistently favorable, providing confidence in the quality of the earnings. Measures comparing cash flow with reinvestment, debt, and debt-servicing capacity indicate strength in financial capacity.
- Decomposing earnings into Nestlé-only and L'Oréal and considering the respective price-to-earnings ratios, it appears that the implied Nestlé-only portion is undervalued. The implied Nestlé-only portion has a far lower price-to-earnings ratio than L'Oréal or the market. This finding should be considered an opportunity, given Nestlé's demonstrated cash flows and low financial leverage.

Causes for Concern

- Although Nestlé has significant, world-class brands and global reach, its core business has deteriorated in profitability in the last several years, as shown by the decomposition of the ROE. Even when taking into account the unusual items affecting profit margins, core operations still show decreases in profitability.
- The negative trend also shows in the cash returns on total assets. They have decreased each year since 2012.
- The acquisition activities in the Nutrition and Health Science product group do not appear to build on the company's traditional strengths. They do not seem to provide a remedy for the deterioration in the core profitability.

- The company's priorities in the allocation of capital in making acquisitions are of some concern. Although the Nutrition and Health Science product group and the Nestlé Nutrition segment show excellent EBIT margins, they rank very low in return on assets. This finding raises the question of whether management is overpaying for acquired companies.
- The company's write-down of goodwill from earlier acquisitions may signal ineffective allocation of capital. It is troubling that Nestlé has taken write-downs on previous acquisitions while actively making new ones.

The analyst concludes that Nestlé is not clearly a good investment opportunity *at this time* and recommends waiting to see whether a further discount makes it more attractive or the operations improve.

Phase 6: Follow-up

The portfolio manager is surprised by the analyst's findings and recommendations. The portfolio manager is convinced that the purchase of shares is justified because of the discount and because, in her opinion, Nestlé is experiencing only temporary issues. She commits the pension fund to a cautious, less-than-core investment holding of Nestlé common stock. The size of the holding is less than originally anticipated because, despite her enthusiasm for the company, the portfolio manager is troubled by the analyst's observations about the resource allocation within the company. She wants him to continually re-evaluate the holding. Unproductive capital spending may be a trigger for eliminating the holding. The analyst is asked to update his findings in the initial research report at each reporting period, emphasizing the quality measures expressed by the accruals tests and the cash flow support of earnings, with particular regard to return on assets.

SUMMARY

The case study demonstrates the use of a financial analysis framework in investment decision making. Although each analysis undertaken may have a different focus, purpose, and context that result in the application of different techniques and tools, the case demonstrates the use of a common financial statement analysis framework. The analyst starts with a global, summarized view of a company and its attributes and digs below the surface of the financial statements to find economic truths that are not apparent from a superficial review. In the case of Nestlé, the analyst applied disaggregation techniques to review the company's performance in terms of ROE and then successively examined the drivers of ROE in increasing detail to evaluate management's skills in capital allocation.

An economic decision is reached, which is consistent with the primary reason for performing financial analysis: to facilitate an economic decision.

PRACTICE PROBLEMS

The following information relates to Questions 1–7

Quentin Abay, CFA, is an analyst for a private equity firm interested in purchasing Bickchip Enterprises, a conglomerate. His first task is to determine the trends in ROE and the main drivers of the trends using DuPont analysis. To do so he gathers the data in Exhibit 1.

Exhibit 1 Selected Financial Data for Bickchip Enterprises (€ Thousands)

	2020	2019	2018
Revenue	72,448	66,487	55,781
Earnings before interest and tax	6,270	4,710	3,609
Earnings before tax	5,101	4,114	3,168
Net income	4,038	3,345	2,576
Asset turnover	0.79	0.76	0.68
Assets/Equity	3.09	3.38	3.43

After conducting the DuPont analysis, Abay believes that his firm could increase the ROE without operational changes. Further, Abay thinks that ROE could improve if the company divested segments that were generating the lowest returns on capital employed (total assets less non-interest-bearing liabilities). Segment EBIT margins in 2020 were 11 percent for Automation Equipment, 5 percent for Power and Industrial, and 8 percent for Medical Equipment. Other relevant segment information is presented in Exhibit 2.

Exhibit 2 Segment Data for Bickchip Enterprises (€ Thousands)

	Capital Employed			Capital Expenditures (Excluding Acquisitions)		
Operating Segments	**2020**	**2019**	**2018**	**2020**	**2019**	**2018**
Automation Equipment	10,705	6,384	5,647	700	743	616
Power and Industrial	15,805	13,195	12,100	900	849	634
Medical Equipment	22,870	22,985	22,587	908	824	749
	49,380	42,564	40,334	2,508	2,416	1,999

Abay is also concerned with earnings quality, so he intends to calculate Bickchip's cash-flow-based accruals ratio and the ratio of operating cash flow before interest and taxes to operating income. To do so, he prepares the information in Exhibit 3.

© 2019 CFA Institute. All rights reserved.

Exhibit 3 Earnings Quality Data for Bickchip Enterprises (€ Thousands)

	2020	2019	2018
Net income	4,038	3,345	2,576
Net cash flow provided by (used in) operating activity[a]	9,822	5,003	3,198
Net cash flow provided by (used in) investing activity	(10,068)	(4,315)	(5,052)
Net cash flow provided by (used in) financing activity[b]	(5,792)	1,540	(2,241)
Average net operating assets	43,192	45,373	40,421
[a] includes cash paid for taxes of:	(1,930)	(1,191)	(1,093)
[b] includes cash paid for interest of:	(1,169)	(596)	(441)

1 Over the three-year period presented in Exhibit 1, Bickchip's return on equity is *best* described as:

A stable.

B trending lower.

C trending higher.

2 Based on the DuPont analysis, Abay's belief regarding ROE is *most likely* based on:

A leverage.

B profit margins.

C asset turnover.

3 Based on Abay's criteria, the business segment *best* suited for divestiture is:

A medical equipment.

B power and industrial.

C automation equipment.

4 Bickchip's cash-flow-based accruals ratio in 2020 is *closest* to:

A 9.9%.

B 13.4%.

C 23.3%.

5 The cash-flow-based accruals ratios from 2018 to 2020 indicate:

A improving earnings quality.

B deteriorating earnings quality.

C no change in earnings quality.

6 The ratio of operating cash flow before interest and taxes to operating income for Bickchip for 2020 is *closest* to:

A 1.6.

B 1.9.

C 2.1.

7 Based on the ratios for operating cash flow before interest and taxes to operating income, Abay should conclude that:

A Bickchip's earnings are backed by cash flow.

B Bickchip's earnings are not backed by cash flow.

C Abay can draw no conclusion due to the changes in the ratios over time.

SOLUTIONS

1 C is correct. The ROE has been trending higher. ROE can be calculated by multiplying (net profit margin) × (asset turnover) × (financial leverage). Net profit margin is net income/sales. In 2018 the net profit margin was 2,576/55,781 = 4.6% and the ROE = 4.6% × 0.68 × 3.43 = 10.8%. Using the same method, ROE was 12.9 percent in 2019 and 13.6 percent in 2020.

2 A is correct. The DuPont analysis shows that profit margins and asset turnover have both increased over the last three years, but leverage has declined. The reduction in leverage offsets a portion of the improvement in profitability and turnover. Thus, ROE would have been higher if leverage had not decreased.

3 B is correct. The Power and Industrial segment has the lowest EBIT margins but uses about 31 percent of the capital employed. Further, Power and Industrial's proportion of the capital expenditures has increased from 32 percent to 36 percent over the three years. Its capital intensity only looked to get worse, as the segment's percentage of total capital expenditures was higher than its percentage of total capital in each of the three years. If Abay is considering divesting segments that do not earn sufficient returns on capital employed, this segment is most suitable.

4 A is correct. The cash-flow-based accruals ratio = [NI – (CFO + CFI)]/(Average NOA) = [4,038 – (9,822 – 10,068)]/43,192 = 9.9%.

5 A is correct. The cash-flow-based accruals ratio falls from 11.0 percent in 2018 to 5.9 percent in 2019, and then rises to 9.9 percent in 2020. However, the change over the three-year period is a net modest decline, indicating a slight improvement in earnings quality.

6 B is correct. Net cash flow provided by (used in) operating activity has to be adjusted for interest and taxes, as necessary, in order to be comparable to operating income (EBIT). Bickchip, reporting under IFRS, chose to classify interest expense as a financing cash flow so the only necessary adjustment is for taxes. The operating cash flow before interest and taxes = 9,822 + 1,930 = 11,752. Dividing this by EBIT of 6,270 yields 1.9.

7 A is correct. Operating cash flow before interest and taxes to operating income rises steadily (not erratically) from 1.2 to 1.3 to 1.9. The ratios over 1.0 and the trend indicate that earnings are supported by cash flow.

Glossary

Abandonment option The ability to terminate a project at some future time if the financial results are disappointing.

Abnormal earnings See *residual income*.

Abnormal return The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

Absolute convergence The idea that developing countries, regardless of their particular characteristics, will eventually catch up with the developed countries and match them in per capita output.

Absolute valuation model A model that specifies an asset's intrinsic value.

Absolute version of PPP An extension of the law of one price whereby the prices of goods and services will not differ internationally once exchange rates are considered.

Accounting estimates Estimates used in calculating the value of assets or liabilities and in the amount of revenue and expense to allocate to a period. Examples of accounting estimates include, among others, the useful lives of depreciable assets, the salvage value of depreciable assets, product returns, warranty costs, and the amount of uncollectible receivables.

Accumulated benefit obligation The actuarial present value of benefits (whether vested or non-vested) attributed, generally by the pension benefit formula, to employee service rendered before a specified date and based on employee service and compensation (if applicable) before that date. The accumulated benefit obligation differs from the projected benefit obligation in that it includes no assumption about future compensation levels.

Accuracy The percentage of correctly predicted classes out of total predictions. It is an overall performance metric in classification problems.

Acquirer The company in a merger or acquisition that is acquiring the target.

Acquiring company See *acquirer*.

Acquisition The purchase of some portion of one company by another; the purchase may be for assets, a definable segment of another entity, or the entire company.

Activation function A functional part of a neural network's node that transforms the total net input received into the final output of the node. The activation function operates like a light dimmer switch that decreases or increases the strength of the input.

Active factor risk The contribution to active risk squared resulting from the portfolio's different-than-benchmark exposures relative to factors specified in the risk model.

Active return The return on a portfolio minus the return on the portfolio's benchmark.

Active risk The standard deviation of active returns.

Active risk squared The variance of active returns; active risk raised to the second power.

Active share A measure of how similar a portfolio is to its benchmark. A manager who precisely replicates the benchmark will have an active share of zero; a manager with no holdings in common with the benchmark will have an active share of one.

Active specific risk The contribution to active risk squared resulting from the portfolio's active weights on individual assets as those weights interact with assets' residual risk.

Adjusted funds from operations (AFFO) Funds from operations adjusted to remove any non-cash rent reported under straight-line rent accounting and to subtract maintenance-type capital expenditures and leasing costs, including leasing agents' commissions and tenants' improvement allowances.

Adjusted present value As an approach to valuing a company, the sum of the value of the company, assuming no use of debt, and the net present value of any effects of debt on company value.

Adjusted R^2 A measure of goodness-of-fit of a regression that is adjusted for degrees of freedom and hence does not automatically increase when another independent variable is added to a regression.

Administrative regulations or administrative law Rules issued by government agencies or other regulators.

Advanced set An arrangement in which the reference interest rate is set at the time the money is deposited.

Advanced settled An arrangement in which a forward rate agreement (FRA) expires and settles at the same time, at the FRA expiration date.

Agency costs Costs associated with the conflict of interest present when a company is managed by non-owners. Agency costs result from the inherent conflicts of interest between managers and equity owners.

Agency costs of equity The smaller the stake managers have in the company, the less their share in bearing the cost of excessive perquisite consumption—consequently, the less their desire to give their best efforts in running the company.

Agency issues Conflicts of interest that arise when the agent in an agency relationship has goals and incentives that differ from the principal to whom the agent owes a fiduciary duty. Also called *agency problems* or *principal–agent problems*.

Agglomerative clustering A bottom-up hierarchical clustering method that begins with each observation being treated as its own cluster. The algorithm finds the two closest clusters, based on some measure of distance (similarity), and combines them into one new larger cluster. This process is repeated iteratively until all observations are clumped into a single large cluster.

Allowance for loan losses A balance sheet account; it is a contra asset account to loans.

Alpha The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

American Depositary Receipt A negotiable certificate issued by a depositary bank that represents ownership in a non-US company's deposited equity (i.e., equity held in custody by the depositary bank in the company's home market).

Analysis of variance (ANOVA) The analysis that breaks the total variability of a dataset (such as observations on the dependent variable in a regression) into components representing different sources of variation. With reference to regression, ANOVA provides the inputs for an *F*-test of

the significance of the regression as a whole, as well as the inputs for the coefficient of determination and the standard error of the estimate.

Application programming interface (API) A set of well-defined methods of communication between various software components and typically used for accessing external data.

Arbitrage (1) The simultaneous purchase of an undervalued asset or portfolio and sale of an overvalued but equivalent asset or portfolio in order to obtain a riskless profit on the price differential. Taking advantage of a market inefficiency in a risk-free manner. (2) The condition in a financial market in which equivalent assets or combinations of assets sell for two different prices, creating an opportunity to profit at no risk with no commitment of money. In a well-functioning financial market, few arbitrage opportunities are possible. (3) A risk-free operation that earns an expected positive net profit but requires no net investment of money.

Arbitrage-free models Term structure models that project future interest rate paths that emanate from the existing term structure. Resulting prices are based on a no-arbitrage condition.

Arbitrage-free valuation An approach to valuation that determines security values consistent with the absence of any opportunity to earn riskless profits without any net investment of money.

Arbitrage opportunity An opportunity to conduct an arbitrage; an opportunity to earn an expected positive net profit without risk and with no net investment of money.

Arbitrage portfolio The portfolio that exploits an arbitrage opportunity.

Ask price The price at which a trader will sell a specified quantity of a security. Also called *ask*, *offer price*, or *offer*.

Asset-based approach Approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities.

Asset-based valuation An approach to valuing natural resource companies that estimates company value on the basis of the market value of the natural resources the company controls.

Asset beta The unlevered beta; reflects the business risk of the assets; the asset's systematic risk.

Asset purchase An acquisition in which the acquirer purchases the target company's assets and payment is made directly to the target company.

Asymmetric information The differential of information between corporate insiders and outsiders regarding the company's performance and prospects. Managers typically have more information about the company's performance and prospects than owners and creditors.

At market contract When a forward contract is established, the forward price is negotiated so that the market value of the forward contract on the initiation date is zero.

Authorized participants (APs) A special group of institutional investors who are authorized by the ETF issuer to participate in the creation/redemption process. APs are large broker/dealers, often market makers.

Autocorrelations The correlations of a time series with its own past values.

Autoregressive model (AR) A time series regressed on its own past values in which the independent variable is a lagged value of the dependent variable.

Backtesting The process that approximates the real-life investment process, using historical data, to assess whether an investment strategy would have produced desirable results.

Backward integration A merger involving the purchase of a target ahead of the acquirer in the value or production chain; for example, to acquire a supplier.

Backward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving backward through the network's layers.

Backwardation A condition in futures markets in which the spot price exceeds the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is higher than the longer-term futures contract price.

Bag-of-words (BOW) A collection of a distinct set of tokens from all the texts in a sample dataset. BOW does not capture the position or sequence of words present in the text.

Bankruptcy A declaration provided for by a country's laws that typically involves the establishment of a legal procedure that forces creditors to defer their claims.

Barbell portfolio Fixed-income portfolio that combines short and long maturities.

Base error Model error due to randomness in the data.

Basic earnings per share (EPS) Net earnings available to common shareholders (i.e., net income minus preferred dividends) divided by the weighted average number of common shares outstanding during the period.

Basis The difference between the spot price and the futures price. As the maturity date of the futures contract nears, the basis converges toward zero.

Basis trade A trade based on the pricing of credit in the bond market versus the price of the same credit in the CDS market. To execute a basis trade, go long the "underpriced" credit and short the "overpriced" credit. A profit is realized as the implied credit prices converge.

Bear hug A tactic used by acquirers to circumvent target management's objections to a proposed merger by submitting the proposal directly to the target company's board of directors.

Bearish flattening Term structure shift in which short-term bond yields rise more than long-term bond yields, resulting in a flatter yield curve.

Benchmark value of the multiple In using the method of comparables, the value of a price multiple for the comparison asset; when we have comparison assets (a group), the mean or median value of the multiple for the group of assets.

Best ask The offer to sell with the lowest ask price. Also called *best offer* or *inside ask*.

Best bid The offer to buy with the highest bid price. Also called the *inside bid*.

Best offer See *best ask*.

Bias error Describes the degree to which a model fits the training data. Algorithms with erroneous assumptions produce high bias error with poor approximation, causing underfitting and high in-sample error.

Bid–ask spread The ask price minus the bid price.

Bid price The price at which a trader will buy a specified quantity of a security. Also called *bid*.

Bill-and-hold basis Sales on a bill-and-hold basis involve selling products but not delivering those products until a later date.

Blockage factor An illiquidity discount that occurs when an investor sells a large amount of stock relative to its trading volume (assuming it is not large enough to constitute a controlling ownership).

Bond indenture A legal contract specifying the terms of a bond issue.

Bond risk premium The expected excess return of a default-free long-term bond less that of an equivalent short-term bond.

Bond yield plus risk premium method An estimate of the cost of common equity that is produced by summing the before-tax cost of debt and a risk premium that captures the additional yield on a company's stock relative to its bonds. The additional yield is often estimated using historical spreads between bond yields and stock yields.

Bonding costs Costs borne by management to assure owners that they are working in the owners' best interest (e.g., implicit cost of non-compete agreements).

Bonus issue of shares *See stock dividend.*

Book value Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value of equity Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value per share The amount of book value (also called carrying value) of common equity per share of common stock, calculated by dividing the book value of shareholders' equity by the number of shares of common stock outstanding.

Bootstrap aggregating (or bagging) A technique whereby the original training dataset is used to generate *n* new training datasets or bags of data. Each new bag of data is generated by random sampling with replacement from the initial training set.

Bootstrapping The use of a forward substitution process to determine zero-coupon rates by using the par yields and solving for the zero-coupon rates one by one, from the shortest to longest maturities.

Bottom-up approach With respect to forecasting, an approach that usually begins at the level of the individual company or a unit within the company.

Breakup value The value derived using a sum-of-the-parts valuation.

Breusch–Pagan test A test for conditional heteroskedasticity in the error term of a regression.

Bullet portfolio A fixed-income portfolio concentrated in a single maturity.

Bullish flattening Term structure change in which the yield curve flattens in response to a greater decline in long-term rates than short-term rates.

Bullish steepening Term structure change in which short-term rates fall by more than long-term yields, resulting in a steeper term structure.

Buy-side analysts Analysts who work for investment management firms, trusts, bank trust departments, and similar institutions.

Buyback *See share repurchase.*

Callable bond Bond that includes an embedded call option that gives the issuer the right to redeem the bond issue prior to maturity, typically when interest rates have fallen or when the issuer's credit quality has improved.

Canceled shares Shares that were issued, subsequently repurchased by the company, and then retired (cannot be reissued).

Cannibalization Cannibalization occurs when an investment takes customers and sales away from another part of the company.

Capital charge The company's total cost of capital in money terms.

Capital deepening An increase in the capital-to-labor ratio.

Capital rationing A capital rationing environment assumes that the company has a fixed amount of funds to invest.

Capital structure The mix of debt and equity that a company uses to finance its business; a company's specific mixture of long-term financing.

Capitalization of earnings method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capitalization rate The divisor in the expression for the value of perpetuity. In the context of real estate, it is the divisor in the direct capitalization method of estimating value. The cap rate equals net operating income divided by value.

Capitalized cash flow method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity. Also called *capitalized cash flow model.*

Capitalized income method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capped floater Floating-rate bond with a cap provision that prevents the coupon rate from increasing above a specified maximum rate. It protects the issuer against rising interest rates.

Carry arbitrage model A no-arbitrage approach in which the underlying instrument is either bought or sold along with an opposite position in a forward contract.

Carry benefits Benefits that arise from owning certain underlyings; for example, dividends, foreign interest, and bond coupon payments.

Carry costs Costs that arise from owning certain underlyings. They are generally a function of the physical characteristics of the underlying asset and also the interest forgone on the funds tied up in the asset.

Cash available for distribution See *adjusted funds from operations.*

Cash-generating unit The smallest identifiable group of assets that generates cash inflows that are largely independent of the cash inflows of other assets or groups of assets.

Cash offering A merger or acquisition that is to be paid for with cash; the cash for the merger might come from the acquiring company's existing assets or from a debt issue.

Cash settlement A procedure used in certain derivative transactions that specifies that the long and short parties settle the derivative's difference in value between them by making a cash payment.

Catalyst An event or piece of information that causes the marketplace to re-evaluate the prospects of a company.

Categorical dependent variables An alternative term for qualitative dependent variables.

CDS spread A periodic premium paid by the buyer to the seller that serves as a return over a market reference rate required to protect against credit risk.

Ceiling analysis A systematic process of evaluating different components in the pipeline of model building. It helps to understand what part of the pipeline can potentially improve in performance by further tuning.

Centroid The center of a cluster formed using the *k*-means clustering algorithm.

Chain rule of forecasting A forecasting process in which the next period's value as predicted by the forecasting equation is substituted into the right-hand side of the equation to give a predicted value two periods ahead.

Cheapest-to-deliver The debt instrument that can be purchased and delivered at the lowest cost yet has the same seniority as the reference obligation.

Classification and regression tree A supervised machine learning technique that can be applied to predict either a categorical target variable, producing a classification tree, or a continuous target variable, producing a regression tree. CART is commonly applied to binary classification or regression.

Clean surplus relation The relationship between earnings, dividends, and book value in which ending book value is equal to the beginning book value plus earnings less dividends, apart from ownership transactions.

Club convergence The idea that only rich and middle-income countries sharing a set of favorable attributes (i.e., are members of the "club") will converge to the income level of the richest countries.

Cluster A subset of observations from a dataset such that all the observations within the same cluster are deemed "similar."

Clustering The sorting of observations into groups (clusters) such that observations in the same cluster are more similar to each other than they are to observations in other clusters.

Cobb–Douglas production function A function of the form $Y = K^{\alpha} L^{1-\alpha}$ relating output (Y) to labor (L) and capital (K) inputs.

Coefficient of determination The percentage of the variation of the dependent variable that is explained by the independent variable. Also referred to as the "*R*-squared" or "R^2."

Cointegrated Describes two time series that have a long-term financial or economic relationship such that they do not diverge from each other without bound in the long run.

Collateral return The component of the total return on a commodity futures position attributable to the yield for the bonds or cash used to maintain the futures position. Also called *collateral yield*.

Collection frequency (CF) The number of times a given word appears in the whole corpus (i.e., collection of sentences) divided by the total number of words in the corpus.

Commercial real estate properties Income-producing real estate properties; properties purchased with the intent to let, lease, or rent (in other words, produce income).

Commodity swap A type of swap involving the exchange of payments over multiple dates as determined by specified reference prices or indexes relating to commodities.

Common size statements Financial statements in which all elements (accounts) are stated as a percentage of a key figure, such as revenue for an income statement or total assets for a balance sheet.

Company fundamental factors Factors related to the company's internal performance, such as factors relating to earnings growth, earnings variability, earnings momentum, and financial leverage.

Company share-related factors Valuation measures and other factors related to share price or the trading characteristics of the shares, such as earnings yield, dividend yield, and book-to-market value.

Comparables Assets used as benchmarks when applying the method of comparables to value an asset. Also called *comps*, *guideline assets*, or *guideline companies*.

Competition laws A law that promotes or maintains market competition by regulating anti-competitive conduct. Known as "antitrust law" in the United States, "anti-monopoly law" in China and Russia, and often referred to as "trade practices law" in the United Kingdom and Australia.

Compiled financial statements Financial statements that are not accompanied by an auditor's opinion letter.

Complexity A term referring to the number of features, parameters, or branches in a model and to whether the model is linear or non-linear (non-linear is more complex).

Composite variable A variable that combines two or more variables that are statistically strongly related to each other.

Comprehensive income All changes in equity other than contributions by, and distributions to, owners; income under clean surplus accounting; includes all changes in equity during a period except those resulting from investments by owners and distributions to owners. Comprehensive income equals net income plus other comprehensive income.

Comps Assets used as benchmarks when applying the method of comparables to value an asset.

Concentrated ownership Ownership structure consisting of an individual shareholder or a group (controlling shareholders) with the ability to exercise control over the corporation.

Conditional convergence The idea that convergence of per capita income is conditional on the countries having the same savings rate, population growth rate, and production function.

Conditional heteroskedasticity Heteroskedasticity in the error variance that is correlated with the values of the independent variable(s) in the regression.

Conditional VaR (CVaR) The weighted average of all loss outcomes in the statistical (i.e., return) distribution that exceed the VaR loss. Thus, CVaR is a more comprehensive measure of tail loss than VaR is. Sometimes referred to as the *expected tail loss* or *expected shortfall*.

Confusion matrix A grid used for error analysis in classification problems, it presents values for four evaluation metrics including true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Conglomerate discount The discount possibly applied by the market to the stock of a company operating in multiple, unrelated businesses.

Conglomerate merger A merger involving companies that are in unrelated businesses.

Consolidation The combining of the results of operations of subsidiaries with the parent company to present financial statements as if they were a single economic unit. The assets, liabilities, revenues, and expenses of the subsidiaries are combined with those of the parent company, eliminating intercompany transactions.

Constant dividend payout ratio policy A policy in which a constant percentage of net income is paid out in dividends.

Constant returns to scale The condition that if all inputs into the production process are increased by a given percentage, then output rises by that same percentage.

Contango A condition in futures markets in which the spot price is lower than the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is lower than the longer-term futures contract price.

Contingent consideration Potential future payments to the seller that are contingent on the achievement of certain agreed-on occurrences.

Continuing earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *persistent earnings*, or *underlying earnings*.

Continuing residual income Residual income after the forecast horizon.

Continuing value The analyst's estimate of a stock's value at a particular point in the future.

Control premium An increment or premium to value associated with a controlling ownership interest in a company.

Convergence The property by which as expiration approaches, the price of a newly created forward or futures contract will approach the price of a spot transaction. At expiration, a forward or futures contract is equivalent to a spot transaction in the underlying.

Conversion period For a convertible bond, the period during which bondholders have the right to convert their bonds into shares.

Conversion price For a convertible bond, the price per share at which the bond can be converted into shares.

Conversion rate (or ratio) For a convertible bond, the number of shares of common stock that a bondholder receives from converting the bond into shares.

Conversion value For a convertible bond, the value of the bond if it is converted at the market price of the shares. Also called *parity value*.

Convertible bond Bond with an embedded conversion option that gives bondholders the right to convert their bonds into the issuer's common stock during a pre-determined period at a pre-determined price.

Convexity A measure of how interest rate sensitivity changes with a change in interest rates.

Core earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *persistent earnings*, or *underlying earnings*.

Core real estate investment style Investing in high-quality, well-leased, core property types with low leverage (no more than 30% of asset value) in the largest markets with strong, diversified economies. It is a conservative strategy designed to avoid real estate–specific risks, including leasing, development, and speculation in favor of steady returns. Hotel properties are excluded from the core categories because of the higher cash flow volatility resulting from single-night leases and the greater importance of property operations, brand, and marketing.

Corpus A collection of text data in any form, including list, matrix, or data table forms.

Cost approach An approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities. In the context of real estate, this approach estimates the value of a property based on what it would cost to buy the land and construct a new property on the site that has the same utility or functionality as the property being appraised.

Cost of carry model A model that relates the forward price of an asset to the spot price by considering the cost of carry (also referred to as future-spot parity model).

Cost of debt The cost of debt financing to a company, such as when it issues a bond or takes out a bank loan.

Cost of equity The required rate of return on common stock.

Covariance stationary Describes a time series when its expected value and variance are constant and finite in all periods and when its covariance with itself for a fixed number of periods in the past or future is constant and finite in all periods.

Covered bonds A senior debt obligation of a financial institution that gives recourse to the originator/issuer and a predetermined underlying collateral pool.

Covered interest rate parity The relationship among the spot exchange rate, the forward exchange rate, and the interest rates in two currencies that ensures that the return on a hedged (i.e., covered) foreign risk-free investment is the same as the return on a domestic risk-free investment. Also called *interest rate parity*.

Cox-Ingersoll-Ross model A general equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is directly related to the level of interest rates.

Creation basket The list of securities (and share amounts) the authorized participant (AP) must deliver to the ETF manager in exchange for ETF shares. The creation basket is published each business day.

Creation/redemption The process in which ETF shares are created or redeemed by authorized participants transacting with the ETF issuer.

Creation units Large blocks of ETF shares transacted between the authorized participant (AP) and the ETF manager that are usually but not always equal to 50,000 shares of the ETF.

Credit correlation The correlation of credit (or default) risks of the underlying single-name CDS contained in an index CDS.

Credit curve The credit spreads for a range of maturities of a company's debt.

Credit default swap A derivative contract between two parties in which the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit derivative A derivative instrument in which the underlying is a measure of the credit quality of a borrower.

Credit event The event that triggers a payment from the credit protection seller to the credit protection buyer.

Credit protection buyer One party to a credit default swap; the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit protection seller One party to a credit default swap; the seller makes a promise to pay compensation for credit losses resulting from the default.

Credit risk The risk that the borrower will not repay principal and interest. Also called *default risk*.

Credit valuation adjustment The value of the credit risk of a bond in present value terms.

Cross-validation A technique for estimating out-of-sample error directly by determining the error in validation samples.

Current exchange rate For accounting purposes, the spot exchange rate on the balance sheet date.

Current rate method Approach to translating foreign currency financial statements for consolidation in which all assets and liabilities are translated at the current exchange rate. The current rate method is the prevalent method of translation.

Curvature One of the three factors (the other two are level and steepness) that empirically explain most of the changes in the shape of the yield curve. A shock to the curvature factor affects mid-maturity interest rates, resulting in the term structure becoming either more or less hump-shaped.

Curve trade Buying a CDS of one maturity and selling a CDS on the same reference entity with a different maturity.

Cyclical businesses Businesses with high sensitivity to business- or industry-cycle influences.

Data mining The practice of determining a model by extensive searching through a dataset for statistically significant patterns.

Data preparation (cleansing) The process of examining, identifying, and mitigating (i.e., cleansing) errors in raw data.

Data snooping The subconscious or conscious manipulation of data in a way that produces a statistically significant result (i.e., the *p*-value is sufficiently small or the *t*-statistic sufficiently large to indicate statistical significance), such as by running multiple simulations and naively accepting the best result. Also known as p-hacking.

Data wrangling (preprocessing) This task performs transformations and critical processing steps on cleansed data to make the data ready for ML model training (i.e., preprocessing), and includes dealing with outliers, extracting useful variables from existing data points, and scaling the data.

"Dead-hand" provision A poison pill provision that allows for the redemption or cancellation of a poison pill provision only by a vote of continuing directors (generally directors who were on the target company's board prior to the takeover attempt).

Debt rating An objective measure of the quality and safety of a company's debt based upon an analysis of the company's ability to pay the promised cash flows. It includes an analysis of any indentures.

Deep learning Algorithms based on deep neural networks, ones with many hidden layers (more than two), that address highly complex tasks, such as image classification, face recognition, speech recognition, and natural language processing.

Deep neural networks Neural networks with many hidden layers—at least 2 but potentially more than 20—that have proven successful across a wide range of artificial intelligence applications.

Default risk See *credit risk*.

Defined benefit pension plans Plan in which the company promises to pay a certain annual amount (defined benefit) to the employee after retirement. The company bears the investment risk of the plan assets.

Defined contribution pension plans Individual accounts to which an employee and typically the employer makes contributions, generally on a tax-advantaged basis. The amounts of contributions are defined at the outset, but the future value of the benefit is unknown. The employee bears the investment risk of the plan assets.

Definitive merger agreement A contract signed by both parties to a merger that clarifies the details of the transaction, including the terms, warranties, conditions, termination details, and the rights of all parties.

Delay costs Implicit trading costs that arise from the inability to complete desired trades immediately. Also called *slippage.*

Delta The relationship between the option price and the underlying price, which reflects the sensitivity of the price of the option to changes in the price of the underlying. Delta is a good approximation of how an option price will change for a small change in the stock.

Dendrogram A type of tree diagram used for visualizing a hierarchical cluster analysis; it highlights the hierarchical relationships among the clusters.

Dependent variable The variable whose variation about its mean is to be explained by the regression; the left-side variable in a regression equation. Also referred to as the *explained variable*.

Depository Trust and Clearinghouse Corporation A US-headquartered entity providing post-trade clearing, settlement, and information services.

Descriptive statistics The study of how data can be summarized effectively.

Diluted earnings per share (Diluted EPS)Net income, minus preferred dividends, divided by the weighted average number of common shares outstanding considering all dilutive securities (e.g., convertible debt and options); the EPS that would result if all dilutive securities were converted into common shares.

Dilution A reduction in proportional ownership interest as a result of the issuance of new shares.

Dimension reduction A set of techniques for reducing the number of features in a dataset while retaining variation across observations to preserve the information contained in that variation.

Diminishing marginal productivity When each additional unit of an input, keeping the other inputs unchanged, increases output by a smaller increment.

Direct capitalization method In the context of real estate, this method estimates the value of an income-producing property based on the level and quality of its net operating income.

Discount To reduce the value of a future payment in allowance for how far away it is in time; to calculate the present value of some future amount. Also, the amount by which an instrument is priced below its face value.

Discount factor The present value or price of a risk-free single-unit payment when discounted using the appropriate spot rate.

Discount for lack of control An amount or percentage deducted from the pro rata share of 100% of the value of an equity interest in a business to reflect the absence of some or all of the powers of control.

Discount for lack of marketability An amount of percentage deducted from the value of an ownership interest to reflect the relative absence of marketability.

Discount function Discount factors for the range of all possible maturities. The spot curve can be derived from the discount function and vice versa.

Discount rate Any rate used in finding the present value of a future cash flow.

Discounted abnormal earnings model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Discounted cash flow (DCF) analysis In the context of merger analysis, an estimate of a target company's value found by discounting the company's expected future free cash flows to the present.

Discounted cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows. In the context of real estate, this method estimates the value of an income-producing property based on discounting future projected cash flows.

Discounted cash flow model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Discriminant analysis A multivariate classification technique used to discriminate between groups, such as companies that either will or will not become bankrupt during some time frame.

Dispersed ownership Ownership structure consisting of many shareholders, none of which has the ability to individually exercise control over the corporation.

Divestiture The sale, liquidation, or spin-off of a division or subsidiary.

Dividend A distribution paid to shareholders based on the number of shares owned.

Dividend coverage ratio The ratio of net income to dividends.

Dividend discount model (DDM) A present value model of stock value that views the intrinsic value of a stock as present value of the stock's expected future dividends.

Dividend displacement of earnings The concept that dividends paid now displace earnings in all future periods.

Dividend imputation tax system A taxation system that effectively assures corporate profits distributed as dividends are taxed just once and at the shareholder's tax rate.

Dividend index point A measure of the quantity of dividends attributable to a particular index.

Dividend payout ratio The ratio of cash dividends paid to earnings for a period.

Dividend policy The strategy a company follows with regard to the amount and timing of dividend payments.

Dividend rate The annualized amount of the most recent dividend.

Dividend yield Annual dividends per share divided by share price.

Divisive clustering A top-down hierarchical clustering method that starts with all observations belonging to a single large cluster. The observations are then divided into two clusters based on some measure of distance (similarity). The algorithm then progressively partitions the intermediate clusters into smaller ones until each cluster contains only one observation.

Document frequency (DF) The number of documents (texts) that contain a particular token divided by the total number of documents. It is the simplest feature selection method and often performs well when many thousands of tokens are present.

Document term matrix (DTM) A matrix where each row belongs to a document (or text file), and each column represents a token (or term). The number of rows is equal to the number of documents (or text files) in a sample text dataset. The number of columns is equal to the number of tokens from the BOW built using all the documents in the sample dataset. The cells typically contain the counts of the number of times a token is present in each document.

Dominance An arbitrage opportunity when a financial asset with a risk-free payoff in the future must have a positive price today.

Double taxation system Corporate earnings are taxed twice when paid out as dividends. First, corporate pretax earnings are taxed regardless of whether they will be distributed as dividends or retained at the corporate level. Second, dividends are taxed again at the individual shareholder level.

Downstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary) such that the investor company records a profit on its income statement. An example is a sale of inventory by the investor company to the associate or by a parent to a subsidiary company.

Dual-class shares Shares that grant one share class superior or even sole voting rights, whereas the other share class has inferior or no voting rights.

Due diligence Investigation and analysis in support of a recommendation; the failure to exercise due diligence may sometimes result in liability according to various securities laws.

Dummy variable A type of qualitative variable that takes on a value of 1 if a particular condition is true and 0 if that condition is false.

Duration A measure of the approximate sensitivity of a security to a change in interest rates (i.e., a measure of interest rate risk).

Dutch disease A situation in which currency appreciation driven by strong export demand for resources makes other segments of the economy (particularly manufacturing) globally uncompetitive.

Earnings surprise The difference between reported EPS and expected EPS. Also referred to as *unexpected earnings*.

Earnings yield EPS divided by price; the reciprocal of the P/E.

Economic profit See *residual income*.

Economic sectors Large industry groupings.

Economic value added (EVA®) A commercial implementation of the residual income concept; the computation of EVA® is the net operating profit after taxes minus the cost of capital, where these inputs are adjusted for a number of items.

Economies of scale A situation in which average costs per unit of good or service produced fall as volume rises. In reference to mergers, the savings achieved through the consolidation of operations and elimination of duplicate resources.

Edwards–Bell–Ohlson model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Effective convexity Sensitivity of duration to changes in interest rates.

Effective duration Sensitivity of the bond's price to a 100 bps parallel shift of the benchmark yield curve, assuming no change in the bond's credit spread.

Effective spread Two times the difference between the execution price and the midpoint of the market quote at the time an order is entered.

Eigenvalue A measure that gives the proportion of total variance in the initial dataset that is explained by each eigenvector.

Eigenvector A vector that defines new mutually uncorrelated composite variables that are linear combinations of the original features.

Embedded options Contingency provisions found in a bond's indenture or offering circular representing rights that enable their holders to take advantage of interest rate movements. They can be exercised by the issuer, by the bondholder, or automatically depending on the course of interest rates.

Ensemble learning A technique of combining the predictions from a collection of models to achieve a more accurate prediction.

Ensemble method The method of combining multiple learning algorithms, as in ensemble learning.

Enterprise value Total company value (the market value of debt, common equity, and preferred equity) minus the value of cash and investments.

Enterprise value multiple A valuation multiple that relates the total market value of all sources of a company's capital (net of cash) to a measure of fundamental value for the entire company (such as a pre-interest earnings measure).

Equilibrium The condition in which supply equals demand.

Equity carve-out A form of restructuring that involves the creation of a new legal entity and the sale of equity in it to outsiders.

Equity charge The estimated cost of equity capital in money terms.

Equity REITs REITs that own, operate, and/or selectively develop income-producing real estate.

Equity swap A swap transaction in which at least one cash flow is tied to the return on an equity portfolio position, often an equity index.

Error autocorrelations The autocorrelations of the error term.

Error term The difference between an observation and its expected value, where the expected value is based on the true underlying population relation between the dependent and independent variables. Also known simply as the *error*.

ESG integration An ESG investment approach that focuses on systematic consideration of material ESG factors in asset allocation, security selection, and portfolio construction decisions for the purpose of achieving the product's stated investment objectives.

Estimated parameters With reference to a regression analysis, the estimated values of the population intercept and population slope coefficients in a regression.

***Ex ante* tracking error** A measure of the degree to which the performance of a given investment portfolio might be expected to deviate from its benchmark; also known as *relative VaR*.

***Ex ante* version of PPP** The hypothesis that expected changes in the spot exchange rate are equal to expected differences in national inflation rates. An extension of relative purchasing power parity to expected future changes in the exchange rate.

Ex-dividend Trading ex-dividend refers to shares that no longer carry the right to the next dividend payment.

Ex-dividend date The first date that a share trades without (i.e., "ex") the right to receive the declared dividend for the period.

Excess earnings method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Exchange ratio The number of shares that target stockholders are to receive in exchange for each of their shares in the target company.

Exercise date The date when employees actually exercise stock options and convert them to stock.

Exercise value The value of an option if it were exercised. Also sometimes called *intrinsic value*.

Expanded CAPM An adaptation of the CAPM that adds to the CAPM a premium for small size and company-specific risk.

Expectations approach A procedure for obtaining the value of an option derived from discounting at the risk-free rate its expected future payoff based on risk neutral probabilities.

Expected exposure The projected amount of money an investor could lose if an event of default occurs, before factoring in possible recovery.

Expected holding-period return The expected total return on an asset over a stated holding period; for stocks, the sum of the expected dividend yield and the expected price appreciation over the holding period.

Expected shortfall See *conditional VaR*.

Expected tail loss See *conditional VaR*.

Exploratory data analysis (EDA) The preliminary step in data exploration, where graphs, charts, and other visualizations (heat maps and word clouds) as well as quantitative methods (descriptive statistics and central tendency measures) are used to observe and summarize data.

Exposure to foreign exchange risk The risk of a change in value of an asset or liability denominated in a foreign currency due to a change in exchange rates.

Extendible bond Bond with an embedded option that gives the bondholder the right to keep the bond for a number of years after maturity, possibly with a different coupon.

External growth Company growth in output or sales that is achieved by buying the necessary resources externally (i.e., achieved through mergers and acquisitions).

Extra dividend *See special dividend.*

F1 score The harmonic mean of precision and recall. F1 score is a more appropriate overall performance metric (than accuracy) when there is unequal class distribution in the dataset and it is necessary to measure the equilibrium of precision and recall.

Factor A common or underlying element with which several variables are correlated.

Factor betas An asset's sensitivity to a particular factor; a measure of the response of return to each unit of increase in a factor, holding all other factors constant.

Factor portfolio See *pure factor portfolio*.

Factor price The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors.

Factor risk premium The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors. Also called *factor price*.

Factor sensitivity See *factor betas*.

Failure to pay When a borrower does not make a scheduled payment of principal or interest on any outstanding obligations after a grace period.

Fair market value The market price of an asset or liability that trades regularly.

Fair value The amount at which an asset (or liability) could be bought (or incurred) or sold (or settled) in a current transaction between willing parties, that is, other than in a forced or liquidation sale. As defined in IFRS and US GAAP, it is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date.

Feature engineering A process of creating new features by changing or transforming existing features.
Feature selection A process whereby only pertinent features from the dataset are selected for model training. Selecting fewer features decreases model complexity and training time.
Features The independent variables (X's) in a labeled dataset.
Financial contagion A situation in which financial shocks spread from their place of origin to other locales. In essence, a faltering economy infects other, healthier economies.
Financial distress Heightened uncertainty regarding a company's ability to meet its various obligations because of lower or negative earnings.
Financial transaction A purchase involving a buyer having essentially no material synergies with the target (e.g., the purchase of a private company by a company in an unrelated industry or by a private equity firm would typically be a financial transaction).
First-differencing A transformation that subtracts the value of the time series in period $t - 1$ from its value in period t.
First-order serial correlation Correlation between adjacent observations in a time series.
Fitting curve A curve which shows in- and out-of-sample error rates (E_{in} and E_{out}) on the y-axis plotted against model complexity on the x-axis.
Fixed price tender offer Offer made by a company to repurchase a specific number of shares at a fixed price that is typically at a premium to the current market price.
Fixed-rate perpetual preferred stock Non-convertible, non-callable preferred stock with a specified dividend rate that has a claim on earnings senior to the claim of common stock, and no maturity date.
Flight to quality During times of market stress, investors sell higher-risk asset classes such as stocks and commodities in favor of default-risk-free government bonds.
Flip-in pill A poison pill takeover defense that dilutes an acquirer's ownership in a target by giving other existing target company shareholders the right to buy additional target company shares at a discount.
Flip-over pill A poison pill takeover defense that gives target company shareholders the right to purchase shares of the acquirer at a significant discount to the market price, which has the effect of causing dilution to all existing acquiring company shareholders.
Float Amounts collected as premium and not yet paid out as benefits.
Floored floater Floating-rate bond with a floor provision that prevents the coupon rate from decreasing below a specified minimum rate. It protects the investor against declining interest rates.
Flotation cost Fees charged to companies by investment bankers and other costs associated with raising new capital.
Forced conversion For a convertible bond, when the issuer calls the bond and forces bondholders to convert their bonds into shares, which typically happens when the underlying share price increases above the conversion price.
Foreign currency transactions Transactions that are denominated in a currency other than a company's functional currency.
Forward curve The term structure of forward rates for loans made on a specific initiation date.
Forward dividend yield A dividend yield based on the anticipated dividend during the next 12 months.
Forward integration A merger involving the purchase of a target that is farther along the value or production chain; for example, to acquire a distributor.
Forward P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.
Forward price The fixed price or rate at which the transaction, scheduled to occur at the expiration of a forward contract, will take place. This price is agreed to at the initiation date of the forward contract.
Forward pricing model The model that describes the valuation of forward contracts.
Forward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving forward through the network's layers.
Forward rate An interest rate determined today for a loan that will be initiated in a future period.
Forward rate agreement An over-the-counter forward contract in which the underlying is an interest rate on a deposit. A forward rate agreement (FRA) calls for one party to make a fixed interest payment and the other to make an interest payment at a rate to be determined at contract expiration.
Forward rate model The forward pricing model expressed in terms of spot and forward interest rates.
Forward rate parity The proposition that the forward exchange rate is an unbiased predictor of the future spot exchange rate.
Forward value The monetary value of an existing forward contract.
Franking credit A tax credit received by shareholders for the taxes that a corporation paid on its distributed earnings.
Free cash flow The actual cash that would be available to the company's investors after making all investments necessary to maintain the company as an ongoing enterprise (also referred to as free cash flow to the firm); the internally generated funds that can be distributed to the company's investors (e.g., shareholders and bondholders) without impairing the value of the company.
Free cash flow hypothesis The hypothesis that higher debt levels discipline managers by forcing them to make fixed debt service payments and by reducing the company's free cash flow.
Free cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows.
Free cash flow to equity The cash flow available to a company's common shareholders after all operating expenses, interest, and principal payments have been made and necessary investments in working and fixed capital have been made.
Free cash flow to equity model A model of stock valuation that views a stock's intrinsic value as the present value of expected future free cash flows to equity.
Free cash flow to the firm The cash flow available to the company's suppliers of capital after all operating expenses (including taxes) have been paid and necessary investments in working and fixed capital have been made.
Free cash flow to the firm model A model of stock valuation that views the value of a firm as the present value of expected future free cash flows to the firm.
Frequency analysis The process of quantifying how important tokens are in a sentence and in the corpus as a whole. It helps in filtering unnecessary tokens (or features).

Friendly transaction A potential business combination that is endorsed by the managers of both companies.

Functional currency The currency of the primary economic environment in which an entity operates.

Fundamental factor models A multifactor model in which the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices.

Fundamentals Economic characteristics of a business, such as profitability, financial strength, and risk.

Funds available for distribution (FAD) See *adjusted funds from operations*.

Funds from operations (FFO) Net income (computed in accordance with generally accepted accounting principles) *plus* (1) gains and losses from sales of properties and (2) depreciation and amortization.

Futures price The price at which the parties to a futures contract agree to exchange the underlying (or cash). In commodity markets, the price agreed on to deliver or receive a defined quantity (and often quality) of a commodity at a future date.

Futures value The monetary value of an existing futures contract.

FX carry trade An investment strategy that involves taking long positions in high-yield currencies and short positions in low-yield currencies.

Gamma A measure of how sensitive an option's delta is to a change in the underlying. The change in a given instrument's delta for a given small change in the underlying's value, holding everything else constant.

Generalize When a model retains its explanatory power when predicting out-of-sample (i.e., using new data).

Generalized least squares A regression estimation technique that addresses heteroskedasticity of the error term.

Going-concern assumption The assumption that the business will maintain its business activities into the foreseeable future.

Going-concern value A business's value under a going-concern assumption.

Goodwill An intangible asset that represents the excess of the purchase price of an acquired company over the value of the net identifiable assets acquired.

Grant date The day that stock options are granted to employees.

Green bond Bonds in which the proceeds are designated by issuers to fund a specific project or portfolio of projects that have environmental or climate benefits.

Greenmail The purchase of the accumulated shares of a hostile investor by a company that is targeted for takeover by that investor, usually at a substantial premium over market price.

Greenwashing The risk that a green bond's proceeds are not actually used for a beneficial environmental or climate-related project.

Grid search A method of systematically training a model by using various combinations of hyperparameter values, cross validating each model, and determining which combination of hyperparameter values ensures the best model performance.

Gross domestic product A money measure of the goods and services produced within a country's borders over a stated period.

Gross lease A lease under which the tenant pays a gross rent to the landlord, who is responsible for all operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Ground truth The known outcome (i.e., target variable) of each observation in a labelled dataset.

Growth accounting equation The production function written in the form of growth rates. For the basic Cobb–Douglas production function, it states that the growth rate of output equals the rate of technological change plus α multiplied by the growth rate of capital plus $(1 - \alpha)$ multiplied by the growth rate of labor.

Growth capital expenditures Capital expenditures needed for expansion.

Growth option The ability to make additional investments in a project at some future time if the financial results are strong. Also called *expansion option*.

Guideline assets Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline companies Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline public companies Public-company comparables for the company being valued.

Guideline public company method A variation of the market approach; establishes a value estimate based on the observed multiples from trading activity in the shares of public companies viewed as reasonably comparable to the subject private company.

Guideline transactions method A variation of the market approach; establishes a value estimate based on pricing multiples derived from the acquisition of control of entire public or private companies that were acquired.

Harmonic mean A type of weighted mean computed by averaging the reciprocals of the observations and then taking the reciprocal of that average.

Hazard rate The probability that an event will occur, given that it has not already occurred.

Hedonic index Unlike a repeat-sales index, a hedonic index does not require repeat sales of the same property. It requires only one sale. The way it controls for the fact that different properties are selling each quarter is to include variables in the regression that control for differences in the characteristics of the property, such as size, age, quality of construction, and location.

Heteroskedastic With reference to the error term of regression, having a variance that differs across observations.

Heteroskedasticity The property of having a nonconstant variance; refers to an error term with the property that its variance differs across observations.

Heteroskedasticity-consistent standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Hierarchical clustering An iterative unsupervised learning procedure used for building a hierarchy of clusters.

Highest and best use The concept that the best use of a vacant site is the use that would result in the highest value for the land. Presumably, the developer that could earn the highest risk-adjusted profit based on time, effort, construction and development cost, leasing, and exit value would be the one to pay the highest price for the land.

Historical exchange rates For accounting purposes, the exchange rates that existed when the assets and liabilities were initially recorded.

Historical scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Historical simulation A simulation method that uses past return data and a random number generator that picks observations from the historical series to simulate an asset's future returns.

Historical simulation method The application of historical price changes to the current portfolio.

Historical stress testing The process that tests how investment strategies would perform under some of the most negative (i.e., adverse) combinations of events and scenarios.

Ho–Lee model The first arbitrage-free term structure model. The model is calibrated to market data and uses a binomial lattice approach to generate a distribution of possible future interest rates.

Holding period return The return that an investor earns during a specified holding period; a synonym for total return.

Holdout samples Data samples that are not used to train a model.

Homoskedasticity The property of having a constant variance; refers to an error term that is constant across observations.

Horizontal merger A merger involving companies in the same line of business, usually as competitors.

Horizontal ownership Companies with mutual business interests (e.g., key customers or suppliers) that have cross-holding share arrangements with each other.

Hostile transaction An attempt to acquire a company against the wishes of the target's managers.

Human capital The accumulated knowledge and skill that workers acquire from education, training, or life experience.

Hybrid approach With respect to forecasting, an approach that combines elements of both top-down and bottom-up analyses.

Hyperparameter A parameter whose value must be set by the researcher before learning begins.

I-spreads Shortened form of "interpolated spreads" and a reference to a linearly interpolated yield.

Illiquidity discount A reduction or discount to value that reflects the lack of depth of trading or liquidity in that asset's market.

Impairment Diminishment in value as a result of carrying (book) value exceeding fair value and/or recoverable value.

Impairment of capital rule A legal restriction that dividends cannot exceed retained earnings.

Implementation shortfall The difference between the money return (or value) on a notional or paper portfolio and the actual portfolio return (or value).

Implied volatility The standard deviation that causes an option pricing model to give the current option price.

In-sample forecast errors The residuals from a fitted time-series model within the sample period used to fit the model.

iNAVs "Indicated" net asset values are intraday "fair value" estimates of an ETF share based on its creation basket.

Income approach A valuation approach that values an asset as the present discounted value of the income expected from it. In the context of real estate, this approach estimates the value of a property based on an expected rate of return. The estimated value is the present value of the expected future income from the property, including proceeds from resale at the end of a typical investment holding period.

Incremental VaR (IVaR) A measure of the incremental effect of an asset on the VaR of a portfolio by measuring the difference between the portfolio's VaR while including a specified asset and the portfolio's VaR with that asset eliminated.

Indenture A written contract between a lender and borrower that specifies the terms of the loan, such as interest rate, interest payment schedule, or maturity.

Independent board directors Directors with no material relationship with the company with regard to employment, ownership, or remuneration.

Independent regulators Regulators recognized and granted authority by a government body or agency. They are not government agencies per se and typically do not rely on government funding.

Independent variable A variable used to explain the dependent variable in a regression; a right-side variable in a regression equation. Also referred to as the *explanatory variable*.

Index CDS A type of credit default swap that involves a combination of borrowers.

Indicator variable A variable that takes on only one of two values, 0 or 1, based on a condition. In simple linear regression, the slope is the difference in the dependent variable for the two conditions. Also referred to as a *dummy variable*.

Industry structure An industry's underlying economic and technical characteristics.

Information gain A metric which quantifies the amount of information that the feature holds about the response. Information gain can be regarded as a form of non-linear correlation between Y and X.

Information ratio (IR) Mean active return divided by active risk; or alpha divided by the standard deviation of diversifiable risk.

Informational frictions Forces that restrict availability, quality, and/or flow of information and its use.

Inside ask See *best ask*.

Inside bid See *best bid*.

Inside spread The spread between the best bid price and the best ask price. Also called the *market bid-ask spread*, *inside bid-ask spread*, or *market spread*.

Insiders Corporate managers and board directors who are also shareholders of a company.

Inter-temporal rate of substitution The ratio of the marginal utility of consumption s periods in the future (the numerator) to the marginal utility of consumption today (the denominator).

Intercept The expected value of the dependent variable when the independent variable in a simple linear regression is equal to zero.

Interest rate risk The risk that interest rates will rise and therefore the market value of current portfolio holdings will fall so that their current yields to maturity then match comparable instruments in the marketplace.

Interlocking directorates Corporate structure in which individuals serve on the board of directors of multiple corporations.

Internal rate of return Abbreviated as IRR. Rate of return that discounts future cash flows from an investment to the exact amount of the investment; the discount rate that makes the present value of an investment's costs (outflows) equal to the present value of the investment's benefits (inflows).

International Fisher effect The proposition that nominal interest rate differentials across currencies are determined by expected inflation differentials.

Intrinsic value The value of an asset given a hypothetically complete understanding of the asset's investment characteristics; the value obtained if an option is exercised based on current conditions. The difference between the spot exchange rate and the strike price of a currency.

Inverse price ratio The reciprocal of a price multiple—for example, in the case of a P/E, the "earnings yield" E/P (where P is share price and E is earnings per share).

Investment value The value to a specific buyer, taking account of potential synergies based on the investor's requirements and expectations.

ISDA Master Agreement A standard or "master" agreement published by the International Swaps and Derivatives Association. The master agreement establishes the terms for each party involved in the transaction.

Judicial law Interpretations of courts.

Justified (fundamental) P/E The price-to-earnings ratio that is fair, warranted, or justified on the basis of forecasted fundamentals.

Justified price multiple The estimated fair value of the price multiple, usually based on forecasted fundamentals or comparables.

***K*-fold cross-validation** A technique in which data (excluding test sample and fresh data) are shuffled randomly and then are divided into k equal sub-samples, with $k - 1$ samples used as training samples and one sample, the kth, used as a validation sample.

***K*-means** A clustering algorithm that repeatedly partitions observations into a fixed number, k, of non-overlapping clusters.

***K*-nearest neighbor** A supervised learning technique that classifies a new observation by finding similarities ("nearness") between this new observation and the existing data.

Kalotay–Williams–Fabozzi (KWF) model An arbitrage-free term structure model that describes the dynamics of the log of the short rate and assumes constant drift, no mean reversion, and constant volatility.

Key rate durations Sensitivity of a bond's price to changes in specific maturities on the benchmark yield curve. Also called *partial durations*.

***k*th-order autocorrelation** The correlation between observations in a time series separated by k periods.

Labeled dataset A dataset that contains matched sets of observed inputs or features (X's) and the associated output or target (Y).

Labor force Everyone of working age (ages 16 to 64) who either is employed or is available for work but not working.

Labor force participation rate The percentage of the working age population that is in the labor force.

Labor productivity The quantity of real GDP produced by an hour of labor. More generally, output per unit of labor input.

Labor productivity growth accounting equation States that potential GDP growth equals the growth rate of the labor input plus the growth rate of labor productivity.

Lack of marketability discount An extra return to investors to compensate for lack of a public market or lack of marketability.

LASSO Least absolute shrinkage and selection operator is a type of penalized regression which involves minimizing the sum of the absolute values of the regression coefficients. LASSO can also be used for regularization in neural networks.

Latency The elapsed time between the occurrence of an event and a subsequent action that depends on that event.

Law of one price A principle that states that if two investments have the same or equivalent future cash flows regardless of what will happen in the future, then these two investments should have the same current price.

Leading dividend yield Forecasted dividends per share over the next year divided by current stock price.

Leading P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Learning curve A curve that plots the accuracy rate (= 1 – error rate) in the validation or test samples (i.e., out-of-sample) against the amount of data in the training sample, which is thus useful for describing under- and overfitting as a function of bias and variance errors.

Learning rate A parameter that affects the magnitude of adjustments in the weights in a neural network.

Level One of the three factors (the other two are steepness and curvature) that empirically explain most yield curve shape changes. A shock to the level factor changes the yield for all maturities by an almost identical amount.

Leveraged buyout A transaction whereby the target company management team converts the target to a privately held company by using heavy borrowing to finance the purchase of the target company's outstanding shares.

Leveraged recapitalization A post-offer takeover defense mechanism that involves the assumption of a large amount of debt that is then used to finance share repurchases. The effect is to dramatically change the company's capital structure while attempting to deliver a value to target shareholders in excess of a hostile bid.

Libor–OIS spread The difference between Libor and the overnight indexed swap rate.

Limit order book The book or list of limit orders to buy and sell that pertains to a security.

Lin-log model A regression model in which the independent variable is in logarithmic form.

Linear classifier A binary classifier that makes its classification decision based on a linear combination of the features of each data point.

Linear regression Regression that models the straight-line relationship between the dependent and independent variables. Also known as *least squares regression* and *ordinary least squares regression*.

Linear trend A trend in which the dependent variable changes at a constant rate with time.

Liquidating dividend A dividend that is a return of capital rather than a distribution from earnings or retained earnings.

Liquidation To sell the assets of a company, division, or subsidiary piecemeal, typically because of bankruptcy; the form of bankruptcy that allows for the orderly satisfaction of creditors' claims after which the company ceases to exist.

Liquidation value The value of a company if the company were dissolved and its assets sold individually.

Liquidity preference theory A term structure theory that asserts liquidity premiums exist to compensate investors for the added interest rate risk they face when lending long term.

Liquidity premium The premium or incrementally higher yield that investors demand for lending long term.

Local currency The currency of the country where a company is located.

Local expectations theory A term structure theory that contends the return for all bonds over short periods is the risk-free rate.

Log-lin model A regression model in which the dependent variable is in logarithmic form.

Log-linear model With reference to time-series models, a model in which the growth rate of the time series as a function of time is constant.

Log-log model A regression model in which both the dependent and independent variables are in logarithmic form. Also known as the *double-log model*.

Log-log regression model A regression that expresses the dependent and independent variables as natural logarithms.

Logistic regression (logit model) A qualitative-dependent-variable multiple regression model based on the logistic probability distribution.

Long/short credit trade A credit protection seller with respect to one entity combined with a credit protection buyer with respect to another entity.

Look-ahead bias The bias created by using information that was unknown or unavailable in the time periods over which backtesting is conducted, such as company earnings and macroeconomic indicator values.

Lookback period The time period used to gather a historical data set.

Loss given default The amount that will be lost if a default occurs.

Macroeconomic factor model A multifactor model in which the factors are surprises in macroeconomic variables that significantly explain equity returns.

Macroeconomic factors Factors related to the economy, such as the inflation rate, industrial production, or economic sector membership.

Maintenance capital expenditures Capital expenditures needed to maintain operations at the current level.

Majority shareholders Shareholders that own more than 50% of a corporation's shares.

Majority-vote classifier A classifier that assigns to a new data point the predicted label with the most votes (i.e., occurrences).

Managerialism theories Theories that posit that corporate executives are motivated to engage in mergers to maximize the size of their company rather than shareholder value (a form of agency cost).

Marginal VaR (MVaR) A measure of the effect of a small change in a position size on portfolio VaR.

Market approach Valuation approach that values an asset based on pricing multiples from sales of assets viewed as similar to the subject asset.

Market conversion premium per share For a convertible bond, the difference between the market conversion price and the underlying share price, which allows investors to identify the premium or discount payable when buying a convertible bond rather than the underlying common stock.

Market conversion premium ratio For a convertible bond, the market conversion premium per share expressed as a percentage of the current market price of the shares.

Market efficiency A finance perspective on capital markets that deals with the relationship of price to intrinsic value. The **traditional efficient markets formulation** asserts that an asset's price is the best available estimate of its intrinsic value. The **rational efficient markets formulation** asserts that investors should expect to be rewarded for the costs of information gathering and analysis by higher gross returns.

Market fragmentation Trading the same instrument in multiple venues.

Market impact The effect of the trade on transaction prices. Also called *price impact.*

Market timing Asset allocation in which the investment in the market is increased if one forecasts that the market will outperform T-bills.

Market value of invested capital The market value of debt and equity.

Mature growth rate The earnings growth rate in a company's mature phase; an earnings growth rate that can be sustained long term.

Maximum drawdown The worst cumulative loss ever sustained by an asset or portfolio. More specifically, maximum drawdown is the difference between an asset's or a portfolio's maximum cumulative return and its subsequent lowest cumulative return.

Mean reversion The tendency of a time series to fall when its level is above its mean and rise when its level is below its mean; a mean-reverting time series tends to return to its long-term mean.

Mean square error (MSE) The sum of squares error divided by the degrees of freedom, $n - k - 1$; in a simple linear regression, $n - k - 1 = n - 2$.

Mean square regression (MSR) The sum of squares regression divided by the number of independent variables k; in a simple linear regression, $k = 1$.

Merger The absorption of one company by another; two companies become one entity and one or both of the pre-merger companies ceases to exist as a separate entity.

Metadata Data that describes and gives information about other data.

Method based on forecasted fundamentals An approach to using price multiples that relates a price multiple to forecasts of fundamentals through a discounted cash flow model.

Method of comparables An approach to valuation that involves using a price multiple to evaluate whether an asset is relatively fairly valued, relatively undervalued, or relatively overvalued when compared to a benchmark value of the multiple.

Midquote price The average, or midpoint, of the prevailing bid and ask prices.

Minority interest The proportion of the ownership of a subsidiary not held by the parent (controlling) company.

Minority shareholders Shareholders that own less than 50% of a corporation's shares.

Mispricing Any departure of the market price of an asset from the asset's estimated intrinsic value.

Mixed offering A merger or acquisition that is to be paid for with cash, securities, or some combination of the two.

Model specification With reference to regression, the set of variables included in the regression and the regression equation's functional form.

Molodovsky effect The observation that P/Es tend to be high on depressed EPS at the bottom of a business cycle and tend to be low on unusually high EPS at the top of a business cycle.

Momentum indicators Valuation indicators that relate either price or a fundamental (such as earnings) to the time series of their own past values (or in some cases to their expected value).

Monetary assets and liabilities Assets and liabilities with value equal to the amount of currency contracted for, a fixed amount of currency. Examples are cash, accounts receivable, accounts payable, bonds payable, and mortgages payable. Inventory is not a monetary asset. Most liabilities are monetary.

Monetary/non-monetary method Approach to translating foreign currency financial statements for consolidation in which monetary assets and liabilities are translated at the current exchange rate. Non-monetary assets and liabilities are translated at historical exchange rates (the exchange rates that existed when the assets and liabilities were acquired).

Monetizing Unwinding a position to either capture a gain or realize a loss.

Monitoring costs Costs borne by owners to monitor the management of the company (e.g., board of director expenses).

Monte Carlo simulation A technique that uses the inverse transformation method for converting a randomly generated uniformly distributed number into a simulated value of a random variable of a desired distribution. Each key decision variable in a Monte Carlo simulation requires an assumed statistical distribution; this assumption facilitates incorporating non-normality, fat tails, and tail dependence as well as solving high-dimensionality problems.

Mortgages Loans with real estate serving as collateral for the loans.

Multicollinearity A regression assumption violation that occurs when two or more independent variables (or combinations of independent variables) are highly but not perfectly correlated with each other.

Multiple linear regression Linear regression involving two or more independent variables.

Multiple linear regression model A linear regression model with two or more independent variables.

Mutual information Measures how much information is contributed by a token to a class of texts. MI will be 0 if the token's distribution in all text classes is the same. MI approaches 1 as the token in any one class tends to occur more often in only that particular class of text.

Mutually exclusive projects Mutually exclusive projects compete directly with each other. For example, if Projects A and B are mutually exclusive, you can choose A or B, but you cannot choose both.

N-grams A representation of word sequences. The length of a sequence varies from 1 to n. When one word is used, it is a unigram; a two-word sequence is a bigram; and a 3-word sequence is a trigram; and so on.

***n*-Period moving average** The average of the current and immediately prior $n - 1$ values of a time series.

Naked credit default swap A position where the owner of the CDS does not have a position in the underlying credit.

Name entity recognition An algorithm that analyzes individual tokens and their surrounding semantics while referring to its dictionary to tag an object class to the token.

Negative serial correlation Serial correlation in which a positive error for one observation increases the chance of a negative error for another observation, and vice versa.

Net asset balance sheet exposure When assets translated at the current exchange rate are greater in amount than liabilities translated at the current exchange rate. Assets exposed to translation gains or losses exceed the exposed liabilities.

Net asset value The difference between assets and liabilities, all taken at current market values instead of accounting book values.

Net asset value per share Net asset value divided by the number of shares outstanding.

Net lease A lease under which the tenant pays a net rent to the landlord and an additional amount based on the tenant's pro rata share of the operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Net liability balance sheet exposure When liabilities translated at the current exchange rate are greater assets translated at the current exchange rate. Liabilities exposed to translation gains or losses exceed the exposed assets.

Net operating income Gross rental revenue minus operating costs but before deducting depreciation, corporate overhead, and interest expense. In the context of real estate, a measure of the income from the property after deducting operating expenses for such items as property taxes, insurance, maintenance, utilities, repairs, and insurance but before deducting any costs associated with financing and before deducting federal income taxes. It is similar to EBITDA in a financial reporting context.

Net regulatory burden The private costs of regulation less the private benefits of regulation.

Network externalities The impact that users of a good, a service, or a technology have on other users of that product; it can be positive (e.g., a critical mass of users makes a product more useful) or negative (e.g., congestion makes the product less useful).

Neural networks Highly flexible machine learning algorithms that have been successfully applied to a variety of supervised and unsupervised tasks characterized by nonlinearities and interactions among features.

No-arbitrage approach A procedure for obtaining the value of an option based on the creation of a portfolio that replicates the payoffs of the option and deriving the option value from the value of the replicating portfolio.

No-growth company A company without positive expected net present value projects.

No-growth value per share The value per share of a no-growth company, equal to the expected level amount of earnings divided by the stock's required rate of return.

Non-cash rent An amount equal to the difference between the average contractual rent over a lease term (the straight-line rent) and the cash rent actually paid during a period. This figure is one of the deductions made from FFO to calculate AFFO.

Non-convergence trap A situation in which a country remains relatively poor, or even falls further behind, because it fails to implement necessary institutional reforms and/or adopt leading technologies.

Non-monetary assets and liabilities Assets and liabilities that are not monetary assets and liabilities. Non-monetary assets include inventory, fixed assets, and intangibles, and non-monetary liabilities include deferred revenue.

Non-renewable resources Finite resources that are depleted once they are consumed; oil and coal are examples.

Non-residential properties Commercial real estate properties other than multi-family properties, farmland, and timberland.

Nonearning assets Cash and investments (specifically cash, cash equivalents, and short-term investments).

Nonstationarity With reference to a random variable, the property of having characteristics, such as mean and variance, that are not constant through time.

Normal EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normalized EPS*.

Normalized earnings The expected level of mid-cycle earnings for a company in the absence of any unusual or temporary factors that affect profitability (either positively or negatively).

Normalized EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normal EPS*.

Normalized P/E P/E based on normalized EPS data.

Notional amount The amount of protection being purchased in a CDS.

NTM P/E Next 12-month P/E: current market price divided by an estimated next 12-month EPS.

Off-the-run A series of securities or indexes that were issued/created prior to the most recently issued/created series.

On-the-run The most recently issued/created series of securities or indexes.

One hot encoding The process by which categorical variables are converted into binary form (0 or 1) for machine reading. It is one of the most common methods for handling categorical features in text data.

One-sided durations Effective durations when interest rates go up or down, which are better at capturing the interest rate sensitivity of bonds with embedded options that do not react symmetrically to positive and negative changes in interest rates of the same magnitude.

One-tier board Board structure consisting of a single board of directors, composed of executive (internal) and non-executive (external) directors.

Opportunity cost The value that investors forgo by choosing a particular course of action; the value of something in its best alternative use.

Optimal capital structure The capital structure at which the value of the company is maximized.

Option-adjusted spread (OAS) Constant spread that, when added to all the one-period forward rates on the interest rate tree, makes the arbitrage-free value of the bond equal to its market price.

Orderly liquidation value The estimated gross amount of money that could be realized from the liquidation sale of an asset or assets, given a reasonable amount of time to find a purchaser or purchasers.

Organic growth Company growth in output or sales that is achieved by making investments internally (i.e., excludes growth achieved through mergers and acquisitions).

Other comprehensive income Changes to equity that bypass (are not reported in) the income statement; the difference between comprehensive income and net income.

Other post-employment benefits Promises by the company to pay benefits in the future, such as life insurance premiums and all or part of health care insurance for its retirees.

Out-of-sample forecast errors The differences between actual and predicted values of time series outside the sample period used to fit the model.

Overfitting When a model fits the training data too well and so does not generalize well to new data.

Overnight indexed swap (OIS) rate An interest rate swap in which the periodic floating rate of the swap equals the geometric average of a daily unsecured overnight rate (or overnight index rate).

Pairs trading An approach to trading that uses pairs of closely related stocks, buying the relatively undervalued stock and selling short the relatively overvalued stock.

Par curve A hypothetical yield curve for coupon-paying Treasury securities that assumes all securities are priced at par.

Par swap A swap in which the fixed rate is set so that no money is exchanged at contract initiation.

Parametric method A method of estimating VaR that uses the historical mean, standard deviation, and correlation of security price movements to estimate the portfolio VaR. Generally assumes a normal distribution but can be adapted to non-normal distributions with the addition of skewness and kurtosis. Sometimes called the *variance–covariance method* or the *analytical method*.

Partial regression coefficients The slope coefficients in a multiple regression. Also called *partial slope coefficients*.

Partial slope coefficients The slope coefficients in a multiple regression. Also called *partial regression coefficients*.

Parts of speech An algorithm that uses language structure and dictionaries to tag every token in the text with a corresponding part of speech (i.e., noun, verb, adjective, proper noun, etc.).

Payout amount The loss given default times the notional.

Payout policy The principles by which a company distributes cash to common shareholders by means of cash dividends and/or share repurchases.

Payouts Cash dividends and the value of shares repurchased in any given year.

Pecking order theory The theory that managers consider how their actions might be interpreted by outsiders and thus order their preferences for various forms of corporate financing. Forms of financing that are least visible to outsiders (e.g., internally generated funds) are most preferable to managers and those that are most visible (e.g., equity) are least preferable.

PEG ratio The P/E-to-growth ratio, calculated as the stock's P/E divided by the expected earnings growth rate.

Penalized regression A regression that includes a constraint such that the regression coefficients are chosen to minimize the sum of squared residuals *plus* a penalty term that increases in size with the number of included features.

Pension obligation The present value of future benefits earned by employees for service provided to date.

Perfect capital markets Markets in which, by assumption, there are no taxes, transaction costs, or bankruptcy costs and in which all investors have equal ("symmetric") information.

Perpetuity A perpetual annuity, or a set of never-ending level sequential cash flows, with the first cash flow occurring one period from now.

Persistent earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *continuing earnings*, or *underlying earnings*.

Pet projects Projects in which influential managers want the corporation to invest. Often, unfortunately, pet projects are selected without undergoing normal capital budgeting analysis.

Physical settlement Involves actual delivery of the debt instrument in exchange for a payment by the credit protection seller of the notional amount of the contract.

Point-in-time data Data consisting of the exact information available to market participants as of a given point in time. Point-in-time data is used to address look-ahead bias.

Poison pill A pre-offer takeover defense mechanism that makes it prohibitively costly for an acquirer to take control of a target without the prior approval of the target's board of directors.

Poison puts A pre-offer takeover defense mechanism that gives target company bondholders the right to sell their bonds back to the target at a pre-specified redemption price, typically at or above par value; this defense increases the need for cash and raises the cost of the acquisition.

Portfolio balance approach A theory of exchange rate determination that emphasizes the portfolio investment decisions of global investors and the requirement that global investors willingly hold all outstanding securities denominated in each currency at prevailing prices and exchange rates.

Positive serial correlation Serial correlation in which a positive error for one observation increases the chance of a positive error for another observation; a negative error for one observation increases the chance of a negative error for another observation.

Potential GDP The maximum amount of output an economy can sustainably produce without inducing an increase in the inflation rate. The output level that corresponds to full employment with consistent wage and price expectations.

Precision In error analysis for classification problems it is ratio of correctly predicted positive classes to all predicted positive classes. Precision is useful in situations where the cost of false positives (FP), or Type I error, is high.

Preferred habitat theory A term structure theory that contends that investors have maturity preferences and require yield incentives before they will buy bonds outside of their preferred maturities.

Premise of value The status of a company in the sense of whether it is assumed to be a going concern or not.

Premium leg The series of payments the credit protection buyer promises to make to the credit protection seller.

Premiums Amounts paid by the purchaser of insurance products.

Present value model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Present value of growth opportunities The difference between the actual value per share and the no-growth value per share. Also called *value of growth*.

Presentation currency The currency in which financial statement amounts are presented.

Price improvement When trade execution prices are better than quoted prices.

Price momentum A valuation indicator based on past price movement.

Price multiples The ratio of a stock's market price to some measure of value per share.

Price-setting option The operational flexibility to adjust prices when demand varies from what is forecast. For example, when demand exceeds capacity, the company could benefit from the excess demand by increasing prices.

Price-to-earnings ratio (P/E) The ratio of share price to earnings per share.

Priced risk Risk for which investors demand compensation for bearing (e.g., equity risk, company-specific factors, macroeconomic factors).

Principal components analysis (PCA) An unsupervised ML technique used to transform highly correlated features of data into a few main, uncorrelated composite variables.

Principle of no arbitrage In well-functioning markets, prices will adjust until there are no arbitrage opportunities.

Prior transaction method A variation of the market approach; considers actual transactions in the stock of the subject private company.

Private market value The value derived using a sum-of-the-parts valuation.

Probability of default The probability that a bond issuer will not meet its contractual obligations on schedule.

Probability of survival The probability that a bond issuer will meet its contractual obligations on schedule.

Procedural law The body of law that focuses on the protection and enforcement of the substantive laws.

Production-flexibility option The operational flexibility to alter production when demand varies from forecast. For example, if demand is strong, a company may profit from employees working overtime or from adding additional shifts.

Project sequencing To defer the decision to invest in a future project until the outcome of some or all of a current project is known. Projects are sequenced through time, so that investing in a project creates the option to invest in future projects.

Projection error The vertical (perpendicular) distance between a data point and a given principal component.

Prospective P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Protection leg The contingent payment that the credit protection seller may have to make to the credit protection buyer.

Protection period Period during which a bond's issuer cannot call the bond.

Provision for loan losses An income statement expense account that increases the amount of the allowance for loan losses.

Proxy fight An attempt to take control of a company through a shareholder vote.

Proxy statement A public document that provides the material facts concerning matters on which shareholders will vote.

Prudential supervision Regulation and monitoring of the safety and soundness of financial institutions to promote financial stability, reduce system-wide risks, and protect customers of financial institutions.

Pruning A regularization technique used in CART to reduce the size of the classification or regression tree—by pruning, or removing, sections of the tree that provide little classifying power.

Purchasing power gain A gain in value caused by changes in price levels. Monetary liabilities experience purchasing power gains during periods of inflation.

Purchasing power loss A loss in value caused by changes in price levels. Monetary assets experience purchasing power loss during periods of inflation.

Purchasing power parity (PPP) The idea that exchange rates move to equalize the purchasing power of different currencies.

Pure expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *unbiased expectations theory*.

Pure factor portfolio A portfolio with sensitivity of 1 to the factor in question and a sensitivity of 0 to all other factors.

Putable bond Bond that includes an embedded put option, which gives the bondholder the right to put back the bonds to the issuer prior to maturity, typically when interest rates have risen and higher-yielding bonds are available.

Qualitative dependent variables Dummy variables used as dependent variables rather than as independent variables.

Quality of earnings analysis The investigation of issues relating to the accuracy of reported accounting results as reflections of economic performance. Quality of earnings analysis is broadly understood to include not only earnings management but also balance sheet management.

Random forest classifier A collection of a large number of decision trees trained via a bagging method.

Random walk A time series in which the value of the series in one period is the value of the series in the previous period plus an unpredictable random error.

Rational efficient markets formulation See *market efficiency.*

Readme files Text files provided with raw data that contain information related to a data file. They are useful for understanding the data and how they can be interpreted correctly.

Real estate investment trusts Tax-advantaged entities (companies or trusts) that own, operate, and—to a limited extent—develop income-producing real estate property.

Real estate operating companies Regular taxable real estate ownership companies that operate in the real estate industry in countries that do not have a tax-advantaged REIT regime in place or that are engage in real estate activities of a kind and to an extent that do not fit in their country's REIT framework.

Real interest rate parity The proposition that real interest rates will converge to the same level across different markets.

Real options Options that relate to investment decisions such as the option to time the start of a project, the option to adjust its scale, or the option to abandon a project that has begun.

Rebalance return A return from rebalancing the component weights of an index.

Recall Also known as *sensitivity*, in error analysis for classification problems it is the ratio of correctly predicted positive classes to all actual positive classes. Recall is useful in situations where the cost of false negatives (FN), or Type II error, is high.

Reconstitution When dealers recombine appropriate individual zero-coupon securities and reproduce an underlying coupon Treasury.

Recovery rate The percentage of the loss recovered.

Redemption basket The list of securities (and share amounts) the authorized participant (AP) receives when it redeems ETF shares back to the ETF manager. The redemption basket is published each business day.

Reference entity The borrower (debt issuer) covered by a single-name CDS.

Reference obligation A particular debt instrument issued by the borrower that is the designated instrument being covered.

Regime With reference to a time series, the underlying model generating the times series.

Regression analysis A tool for examining whether a variable is useful for explaining another variable.

Regression coefficients The intercept and slope coefficient(s) of a regression.

Regular expression (regex) A series of texts that contains characters in a particular order. Regex is used to search for patterns of interest in a given text.

Regularization A term that describes methods for reducing statistical variability in high-dimensional data estimation problems.

Regulatory arbitrage Entities identify and use some aspect of regulations that allows them to exploit differences in economic substance and regulatory interpretation or in foreign and domestic regulatory regimes to their (the entities') advantage.

Regulatory burden The costs of regulation for the regulated entity.

Regulatory capture Theory that regulation often arises to enhance the interests of the regulated.

Regulatory competition Regulators may compete to provide a regulatory environment designed to attract certain entities.

Reinforcement learning Machine learning in which a computer learns from interacting with itself or data generated by the same algorithm.

Relative-strength indicators Valuation indicators that compare a stock's performance during a period either to its own past performance or to the performance of some group of stocks.

Relative valuation models A model that specifies an asset's value relative to the value of another asset.

Relative VaR See *ex ante tracking error*.

Relative version of PPP The hypothesis that changes in (nominal) exchange rates over time are equal to national inflation rate differentials.

Renewable resources Resources that can be replenished, such as a forest.

Rental price of capital The cost per unit of time to rent a unit of capital.

Repeat sales index As the name implies, this type of index relies on repeat sales of the same property. In general, the idea supporting this type of index is that because it is the same property that sold twice, the change in value between the two sale dates indicates how market conditions have changed over time.

Replacement cost In the context of real estate, the value of a building assuming it was built today using current construction costs and standards.

Reporting unit For financial reporting under US GAAP, an operating segment or one level below an operating segment (referred to as a component).

Required rate of return The minimum rate of return required by an investor to invest in an asset, given the asset's riskiness.

Residential properties Properties that provide housing for individuals or families. Single-family properties may be owner-occupied or rental properties, whereas multi-family properties are rental properties even if the owner or manager occupies one of the units.

Residual The difference between an observation and its predicted value, where the predicted value is based on the estimated linear relation between the dependent and independent variables using sample data.

Residual autocorrelations The sample autocorrelations of the residuals.

Residual income Earnings for a given period, minus a deduction for common shareholders' opportunity cost in generating the earnings. Also called *economic profit* or *abnormal earnings*.

Residual income method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Residual income model (RIM) A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share. Also called *discounted abnormal earnings model* or *Edwards–Bell–Ohlson model*.

Residual loss Agency costs that are incurred despite adequate monitoring and bonding of management.

Restructuring Reorganizing the capital structure of a firm.

Return on capital employed Operating profit divided by capital employed (debt and equity capital).

Return on invested capital A measure of the after-tax profitability of the capital invested by the company's shareholders and debtholders.

Reverse carry arbitrage A strategy involving the short sale of the underlying and an offsetting opposite position in the derivative.

Reverse stock split A reduction in the number of shares outstanding with a corresponding increase in share price but no change to the company's underlying fundamentals.

Reverse stress testing A risk management approach in which the user identifies key risk exposures in the portfolio and subjects those exposures to extreme market movements.

Reviewed financial statements A type of non-audited financial statements; typically provide an opinion letter with representations and assurances by the reviewing accountant that are less than those in audited financial statements.

Rho The change in a given derivative instrument for a given small change in the risk-free interest rate, holding everything else constant. Rho measures the sensitivity of the option to the risk-free interest rate.

Risk budgeting The allocation of an asset owner's total risk appetite among groups or divisions (in the case of a trading organization) or among strategies and managers (in the case of an institutional or individual investor).

Risk decomposition The process of converting a set of holdings in a portfolio into a set of exposures to risk factors.

Risk factors Variables or characteristics with which individual asset returns are correlated. Sometimes referred to simply as *factors*.

Risk parity A portfolio allocation scheme that weights stocks or factors based on an equal risk contribution.

Robust standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Roll When an investor moves its investment position from an older series to the most current series.

Roll return The component of the return on a commodity futures contract attributable to rolling long futures positions forward through time. Also called *roll yield*.

Rolling down the yield curve A maturity trading strategy that involves buying bonds with a maturity longer than the intended investment horizon. Also called *riding the yield curve*.

Rolling windows A backtesting method that uses a rolling-window (or walk-forward) framework, rebalances the portfolio after each period, and then tracks performance over time. As new information arrives each period, the investment manager optimizes (revises and tunes) the model and readjusts stock positions.

Root mean squared error (RMSE) The square root of the average squared forecast error; used to compare the out-of-sample forecasting performance of forecasting models.

Sale-leaseback A situation in which a company sells the building it owns and occupies to a real estate investor and the company then signs a long-term lease with the buyer to continue to occupy the building. At the end of the lease, use of the property reverts to the landlord.

Sales comparison approach In the context of real estate, this approach estimates value based on what similar or comparable properties (comparables) transacted for in the current market.

Scaled earnings surprise Unexpected earnings divided by the standard deviation of analysts' earnings forecasts.

Scaling The process of adjusting the range of a feature by shifting and changing the scale of the data. Two of the most common ways of scaling are normalization and standardization.

Scatter plot A chart in which two variables are plotted along the axis and points on the chart represent pairs of the two variables. In regression, the dependent variable is plotted on the vertical axis and the independent variable is plotted along the horizontal axis. Also known as a scattergram and a *scatter diagram*.

Scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Scree plots A plot that shows the proportion of total variance in the data explained by each principal component.

Screening The application of a set of criteria to reduce a set of potential investments to a smaller set having certain desired characteristics.

Seasonality A characteristic of a time series in which the data experience regular and predictable periodic changes; for example, fan sales are highest during the summer months.

Secured overnight financing rate (SOFR) A daily volume-weighted index of rates on qualified cash borrowings collateralized by US Treasuries that is expected to replace Libor as a floating reference rate for swaps.

Securities offering A merger or acquisition in which target shareholders are to receive shares of the acquirer's common stock as compensation.

Security selection risk See *active specific risk*.

Segmented markets theory A term structure theory that contends yields are solely a function of the supply and demand for funds of a particular maturity.

Self-regulating organizations (SROs) Self-regulating bodies that are given recognition and authority, including enforcement power, by a government body or agency.

Self-regulatory bodies Private, non-governmental organizations that both represent and regulate their members. Some self-regulating organizations are also independent regulators.

Sell-side analysts Analysts who work at brokerages.

Sensitivity analysis A technique for exploring how a target variable (e.g., portfolio returns) and risk profiles are affected by changes in input variables (e.g., the distribution of asset or factor returns).

Sentence length The number of characters, including spaces, in a sentence.

Serially correlated With reference to regression errors, errors that are correlated across observations.

Service period For employee stock options, usually the period between the grant date and the vesting date.

Settled in arrears An arrangement in which the interest payment is made (i.e., settlement occurs) at the maturity of the underlying instrument.

Settlement In the case of a credit event, the process by which the two parties to a CDS contract satisfy their respective obligations.

Shaping risk The sensitivity of a bond's price to the changing shape of the yield curve.

Share repurchase A transaction in which a company buys back its own shares. Unlike stock dividends and stock splits, share repurchases use corporate cash.

Shareholder activism Strategies used by shareholders to attempt to compel a company to act in a desired manner.

Shareholders' equity Total assets minus total liabilities.

Shark repellents A pre-offer takeover defense mechanism involving the corporate charter (e.g., staggered boards of directors and supermajority provisions).

Simple linear regression (SLR) A regression that summarizes the relation between the dependent variable and a single independent variable.

Simulation A technique for exploring how a target variable (e.g. portfolio returns) would perform in a hypothetical environment specified by the user, rather than a historical setting.

Single-name CDS Credit default swap on one specific borrower.

Sinking fund bond A bond that requires the issuer to set aside funds over time to retire the bond issue, thus reducing credit risk.

Slope coefficient The coefficient of an independent variable that represents the average change in the dependent variable for a one-unit change in the independent variable.

Soft margin classification An adaptation in the support vector machine algorithm which adds a penalty to the objective function for observations in the training set that are misclassified.

Special dividend A dividend paid by a company that does not pay dividends on a regular schedule or a dividend that supplements regular cash dividends with an extra payment.

Spin-off A form of restructuring in which shareholders of a parent company receive a proportional number of shares in a new, separate entity; shareholders end up owning stock in two different companies where there used to be one.

Split-off A form of restructuring in which shareholders of the parent company are given shares in a newly created entity in exchange for their shares of the parent company.

Split-rate tax system In reference to corporate taxes, a split-rate system taxes earnings to be distributed as dividends at a different rate than earnings to be retained. Corporate profits distributed as dividends are taxed at a lower rate than those retained in the business.

Spot curve The term structure of spot rates for loans made today.

Spot price The current price of an asset or security. For commodities, the current price to deliver a physical commodity to a specific location or purchase and transport it away from a designated location.

Spot rate The interest rate that is determined today for a risk-free, single-unit payment at a specified future date.

Spot yield curve The term structure of spot rates for loans made today.

Stabilized NOI In the context of real estate, the expected NOI when a renovation is complete.

Stable dividend policy A policy in which regular dividends are paid that reflect long-run expected earnings. In contrast to a constant dividend payout ratio policy, a stable dividend policy does not reflect short-term volatility in earnings.

Standard error of the estimate A measure of the fit of a regression line, calculated as the square root of the mean square error. Also known as the *standard error of the regression* and the *root mean square error*.

Standard error of the forecast A measure of the uncertainty associated with a forecasted value of the dependent variable that depends on the standard error of the estimate, the variability of the independent variable, the deviation of the forecasted independent variable from the mean in the regression, and the number of observations.

Standard error of the slope coefficient The standard error of the slope, which in a simple linear regression is the ratio of the model's standard error of the estimate (s_e) to the square root of the variation of the independent variable.

Standardized beta With reference to fundamental factor models, the value of the attribute for an asset minus the average value of the attribute across all stocks, divided by the standard deviation of the attribute across all stocks.

Standardized unexpected earnings Unexpected earnings per share divided by the standard deviation of unexpected earnings per share over a specified prior time period.

Static trade-off theory of capital structure A theory pertaining to a company's optimal capital structure. The optimal level of debt is found at the point where additional debt would cause the costs of financial distress to increase by a greater amount than the benefit of the additional tax shield.

Statistical factor model A multifactor model in which statistical methods are applied to a set of historical returns to determine portfolios that best explain either historical return covariances or variances.

Statutes Laws enacted by legislative bodies.

Statutory merger A merger in which one company ceases to exist as an identifiable entity and all its assets and liabilities become part of a purchasing company.

Steady-state rate of growth The constant growth rate of output (or output per capita) that can or will be sustained indefinitely once it is reached. Key ratios, such as the capital–output ratio, are constant on the steady-state growth path.

Steepness The difference between long-term and short-term yields that constitutes one of the three factors (the other two are level and curvature) that empirically explain most of the changes in the shape of the yield curve.

Stock dividend A type of dividend in which a company distributes additional shares of its common stock to shareholders instead of cash.

Stock purchase An acquisition in which the acquirer gives the target company's shareholders some combination of cash and securities in exchange for shares of the target company's stock.

Stop-loss limit Constraint used in risk management that requires a reduction in the size of a portfolio, or its complete liquidation, when a loss of a particular size occurs in a specified period.

Straight bond An underlying option-free bond with a specified issuer, issue date, maturity date, principal amount and repayment structure, coupon rate and payment structure, and currency denomination.

Straight-line rent The average annual rent under a multi-year lease agreement that contains contractual increases in rent during the life of the lease.

Straight-line rent adjustment See *non-cash rent.*

Straight voting Voting structure in which shareholders are granted the right of one vote for each share owned.

Stranded assets Assets that are obsolete or not economically viable.

Strategic transaction A purchase involving a buyer that would benefit from certain synergies associated with owning the target firm.

Stress tests A risk management technique that assesses the portfolio's response to extreme market movements.

Stripping A dealer's ability to separate a bond's individual cash flows and trade them as zero-coupon securities.

Subsidiary merger A merger in which the company being purchased becomes a subsidiary of the purchaser.

Substantive law The body of law that focuses on the rights and responsibilities of entities and relationships among entities.

Succession event A change of corporate structure of the reference entity, such as through a merger, a divestiture, a spinoff, or any similar action, in which ultimate responsibility for the debt in question is unclear.

Sum of squares error (SSE) The sum of the squared deviations of (1) the value of the dependent variable and (2) the value of the dependent variable based on the estimated regression line. Also referred to as the *residual sum of squares.*

Sum of squares regression (SSR) The sum of the squared deviations of (1) the value of the dependent variable based on the estimated regression line and (2) the mean of the dependent variable.

Sum of squares total (SST) The sum of the squared deviations of the dependent variable from its mean; the variation of the dependent variable. Also referred to as the *total sum of squares.*

Sum-of-the-parts valuation A valuation that sums the estimated values of each of a company's businesses as if each business were an independent going concern.

Summation operator A functional part of a neural network's node that multiplies each input value received by a weight and sums the weighted values to form the total net input, which is then passed to the activation function.

Supernormal growth Above-average or abnormally high growth rate in earnings per share.

Supervised learning Machine learning where algorithms infer patterns between a set of inputs (the *X*'s) and the desired output (*Y*). The inferred pattern is then used to map a given input set into a predicted output.

Support vector machine A linear classifier that determines the hyperplane that optimally separates the observations into two sets of data points.

Survivorship bias The bias that results when data as of a given date reflects only those entities that have survived to that date. Entities can include any element of an index or list that is constituted through time: stocks, investment funds, etc. Survivorship bias is a form of look-ahead bias.

Sustainable growth rate The rate of dividend (and earnings) growth that can be sustained over time for a given level of return on equity, keeping the capital structure constant and without issuing additional common stock.

Swap curve The term structure of swap rates.

Swap rate The "price" that swap traders quote among one another. It is the rate at which the present value of all the expected floating-rate payments received over the life of the floating-rate bond equal the present value of all the expected fixed-rate payments made over the life of the fixed-rate bond.

Swap rate curve The term structure of swap rates.

Swap spread The difference between the fixed rate on an interest rate swap and the rate on a Treasury note with equivalent maturity; it reflects the general level of credit risk in the market.

Systematic risk Risk that affects the entire market or economy; it cannot be avoided and is inherent in the overall market. Systematic risk is also known as non-diversifiable or market risk.

Systemic risk The risk of failure of the financial system.

Tail risk The risk that losses in extreme events could be greater than would be expected for a portfolio of assets with a normal distribution.

Takeover A merger; the term may be applied to any transaction but is often used in reference to hostile transactions.

Takeover premium The amount by which the takeover price for each share of stock must exceed the current stock price in order to entice shareholders to relinquish control of the company to an acquirer.

Tangible book value per share Common shareholders' equity minus intangible assets reported on the balance sheet, divided by the number of shares outstanding.

Target In machine learning, the dependent variable (*Y*) in a labeled dataset; the company in a merger or acquisition that is being acquired.

Target capital structure A company's chosen proportions of debt and equity.

Target company See *target.*

Target payout ratio A strategic corporate goal representing the long-term proportion of earnings that the company intends to distribute to shareholders as dividends.

Taxable REIT subsidiaries Subsidiaries that pay income taxes on earnings from non-REIT-qualifying activities like merchant development or third-party property management.

Technical indicators Momentum indicators based on price.

TED spread A measure of perceived credit risk determined as the difference between Libor and the T-bill yield of matching maturity.

Temporal method A variation of the monetary/non-monetary translation method that requires not only monetary assets and liabilities, but also non-monetary assets and liabilities that are measured at their current value on the balance sheet date to be translated at the current exchange rate. Assets and liabilities are translated at rates consistent with the timing of their measurement value. This method is typically used when the functional currency is other than the local currency.

Tender offer A public offer whereby the acquirer invites target shareholders to submit ("tender") their shares in return for the proposed payment.

Term frequency (TF) Ratio of the number of times a given token occurs in all the texts in the dataset to the total number of tokens in the dataset.

Term premium The additional return required by lenders to invest in a bond to maturity net of the expected return from continually reinvesting at the short-term rate over that same time horizon.

Terminal price multiples The price multiple for a stock assumed to hold at a stated future time.

Terminal share price The share price at a particular point in the future.

Terminal value of the stock The analyst's estimate of a stock's value at a particular point in the future. Also called *continuing value of the stock*.

Test sample A data sample that is used to test a model's ability to predict well on new data.

Theta The change in a derivative instrument for a given small change in calendar time, holding everything else constant. Specifically, the theta calculation assumes nothing changes except calendar time. Theta also reflects the rate at which an option's time value decays.

Time series A set of observations on a variable's outcomes in different time periods.

Tobin's *q* The ratio of the market value of debt and equity to the replacement cost of total assets.

Token The equivalent of a word (or sometimes a character).

Tokenization The process of splitting a given text into separate tokens. Tokenization can be performed at the word or character level but is most commonly performed at word level.

Top-down approach With respect to forecasting, an approach that usually begins at the level of the overall economy. Forecasts are then made at more narrowly defined levels, such as sector, industry, and market for a specific product.

Total factor productivity (TFP) A multiplicative scale factor that reflects the general level of productivity or technology in the economy. Changes in total factor productivity generate proportional changes in output for any input combination.

Total invested capital The sum of market value of common equity, book value of preferred equity, and face value of debt.

Tracking error The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking risk*.

Tracking risk The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking error*.

Trailing dividend yield The reciprocal of current market price divided by the most recent annualized dividend.

Trailing P/E A stock's current market price divided by the most recent four quarters of EPS (or the most recent two semi-annual periods for companies that report interim data semi-annually). Also called *current P/E*.

Training sample A data sample that is used to train a model.

Tranche CDS A type of credit default swap that covers a combination of borrowers but only up to pre-specified levels of losses.

Transaction exposure The risk of a change in value between the transaction date and the settlement date of an asset of liability denominated in a foreign currency.

Treasury shares/stock Shares that were issued and subsequently repurchased by the company.

Trend A long-term pattern of movement in a particular direction.

Triangular arbitrage An arbitrage transaction involving three currencies that attempts to exploit inconsistencies among pairwise exchange rates.

Trimming Also called truncation, it is the process of removing extreme values and outliers from a dataset.

Triple-net leases Common leases in the United States and Canada that require each tenant to pay its share of the following three operating expenses: common area maintenance and repair expenses; property taxes; and building insurance costs. Also known as *NNN leases*.

Two-tier board Board structure consisting of a supervisory board that oversees a management board.

Unbiased expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *pure expectations theory*.

Unconditional heteroskedasticity Heteroskedasticity of the error term that is not correlated with the values of the independent variable(s) in the regression.

Uncovered interest rate parity The proposition that the expected return on an uncovered (i.e., unhedged) foreign currency (risk-free) investment should equal the return on a comparable domestic currency investment.

Underlying earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *core earnings*, or *persistent earnings*.

Unexpected earnings The difference between reported EPS and expected EPS. Also referred to as an *earnings surprise*.

Unit root A time series that is not covariance stationary is said to have a unit root.

Unsupervised learning Machine learning that does not make use of labeled data.

Upfront payment The difference between the credit spread and the standard rate paid by the protection buyer if the standard rate is insufficient to compensate the protection seller. Also called *upfront premium*.

Upfront premium See *upfront payment*.

Upstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary company) such that the associate company records a profit on its income statement. An example is a sale of inventory by the associate to the investor company or by a subsidiary to a parent company.

Validation sample A data sample that is used to validate and tune a model.

Valuation The process of determining the value of an asset or service either on the basis of variables perceived to be related to future investment returns or on the basis of comparisons with closely similar assets.

Value additivity An arbitrage opportunity when the value of the whole equals the sum of the values of the parts.

Value at risk (VaR) The minimum loss that would be expected a certain percentage of the time over a certain period of time given the assumed market conditions.

Value of growth The difference between the actual value per share and the no-growth value per share.

Variance error Describes how much a model's results change in response to new data from validation and test samples. Unstable models pick up noise and produce high variance error, causing overfitting and high out-of-sample error.

Vasicek model A partial equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is constant.

Vega The change in a given derivative instrument for a given small change in volatility, holding everything else constant. A sensitivity measure for options that reflects the effect of volatility.

Venture capital investors Private equity investors in development-stage companies.

Vertical merger A merger involving companies at different positions of the same production chain; for example, a supplier or a distributor.

Vertical ownership Ownership structure in which a company or group that has a controlling interest in two or more holding companies, which in turn have controlling interests in various operating companies.

Vested benefit obligation The actuarial present value of vested benefits.

Vesting date The date that employees can first exercise stock options.

Visibility The extent to which a company's operations are predictable with substantial confidence.

Voting caps Legal restrictions on the voting rights of large share positions.

Web spidering (scraping or crawling) programs Programs that extract raw content from a source, typically web pages.

Weighted average cost of capital (WACC) A weighted average of the after-tax required rates of return on a company's common stock, preferred stock, and long-term debt, where the weights are the fraction of each source of financing in the company's target capital structure.

Weighted harmonic mean See *harmonic mean*.

White-corrected standard errors A synonym for robust standard errors.

White knight A third party that is sought out by the target company's board to purchase the target in lieu of a hostile bidder.

White squire A third party that is sought out by the target company's board to purchase a substantial minority stake in the target—enough to block a hostile takeover without selling the entire company.

Winner's curse The tendency for the winner in certain competitive bidding situations to overpay, whether because of overestimation of intrinsic value, emotion, or information asymmetries.

Winsorization The process of replacing extreme values and outliers in a dataset with the maximum (for large value outliers) and minimum (for small value outliers) values of data points that are not outliers.

Write-down A reduction in the value of an asset as stated in the balance sheet.

Yield curve factor model A model or a description of yield curve movements that can be considered realistic when compared with historical data.

Zero A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

Zero-coupon bond A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

CFA Institute®
CFA Program

CORPORATE ISSUERS AND EQUITY

CFA® Program Curriculum
2022 • LEVEL II • VOLUME 3

WILEY

CONTENTS

How to Use the CFA Program Curriculum

Congratulations on your decision to enter the Chartered Financial Analyst (CFA®) Program. This exciting and rewarding program of study reflects your desire to become a serious investment professional. You are embarking on a program noted for its high ethical standards and the breadth of knowledge, skills, and abilities (competencies) it develops. Your commitment should be educationally and professionally rewarding.

The credential you seek is respected around the world as a mark of accomplishment and dedication. Each level of the program represents a distinct achievement in professional development. Successful completion of the program is rewarded with membership in a prestigious global community of investment professionals. CFA charterholders are dedicated to life-long learning and maintaining currency with the ever-changing dynamics of a challenging profession. CFA Program enrollment represents the first step toward a career-long commitment to professional education.

The CFA exam measures your mastery of the core knowledge, skills, and abilities required to succeed as an investment professional. These core competencies are the basis for the Candidate Body of Knowledge (CBOK™). The CBOK consists of four components:

- A broad outline that lists the major CFA Program topic areas (www.cfainstitute.org/programs/cfa/curriculum/cbok);
- Topic area weights that indicate the relative exam weightings of the top-level topic areas (www.cfainstitute.org/programs/cfa/curriculum);
- Learning outcome statements (LOS) that advise candidates about the specific knowledge, skills, and abilities they should acquire from readings covering a topic area (LOS are provided in candidate study sessions and at the beginning of each reading); and
- CFA Program curriculum that candidates receive upon exam registration.

Therefore, the key to your success on the CFA exams is studying and understanding the CBOK. The following sections provide background on the CBOK, the organization of the curriculum, features of the curriculum, and tips for designing an effective personal study program.

BACKGROUND ON THE CBOK

CFA Program is grounded in the practice of the investment profession. CFA Institute performs a continuous practice analysis with investment professionals around the world to determine the competencies that are relevant to the profession, beginning with the Global Body of Investment Knowledge (GBIK®). Regional expert panels and targeted surveys are conducted annually to verify and reinforce the continuous feedback about the GBIK. The practice analysis process ultimately defines the CBOK. The CBOK reflects the competencies that are generally accepted and applied by investment professionals. These competencies are used in practice in a generalist context and are expected to be demonstrated by a recently qualified CFA charterholder.

© 2021 CFA Institute. All rights reserved.

The CFA Institute staff—in conjunction with the Education Advisory Committee and Curriculum Level Advisors, who consist of practicing CFA charterholders—designs the CFA Program curriculum in order to deliver the CBOK to candidates. The exams, also written by CFA charterholders, are designed to allow you to demonstrate your mastery of the CBOK as set forth in the CFA Program curriculum. As you structure your personal study program, you should emphasize mastery of the CBOK and the practical application of that knowledge. For more information on the practice analysis, CBOK, and development of the CFA Program curriculum, please visit www.cfainstitute.org.

ORGANIZATION OF THE CURRICULUM

The Level II CFA Program curriculum is organized into 10 topic areas. Each topic area begins with a brief statement of the material and the depth of knowledge expected. It is then divided into one or more study sessions. These study sessions should form the basic structure of your reading and preparation. Each study session includes a statement of its structure and objective and is further divided into assigned readings. An outline illustrating the organization of these study sessions can be found at the front of each volume of the curriculum.

The readings are commissioned by CFA Institute and written by content experts, including investment professionals and university professors. Each reading includes LOS and the core material to be studied, often a combination of text, exhibits, and in-text examples and questions. End of Reading Questions (EORQs) followed by solutions help you understand and master the material. The LOS indicate what you should be able to accomplish after studying the material. The LOS, the core material, and the EORQs are dependent on each other, with the core material and EORQs providing context for understanding the scope of the LOS and enabling you to apply a principle or concept in a variety of scenarios.

The entire readings, including the EORQs, are the basis for all exam questions and are selected or developed specifically to teach the knowledge, skills, and abilities reflected in the CBOK.

You should use the LOS to guide and focus your study because each exam question is based on one or more LOS and the core material and practice problems associated with the LOS. As a candidate, you are responsible for the entirety of the required material in a study session.

We encourage you to review the information about the LOS on our website (www.cfainstitute.org/programs/cfa/curriculum/study-sessions), including the descriptions of LOS "command words" on the candidate resources page at www.cfainstitute.org.

FEATURES OF THE CURRICULUM

End of Reading Questions/Solutions *All End of Reading Questions (EORQs) as well as their solutions are part of the curriculum and are required material for the exam.* In addition to the in-text examples and questions, these EORQs help demonstrate practical applications and reinforce your understanding of the concepts presented. Some of these EORQs are adapted from past CFA exams and/or may serve as a basis for exam questions.

Glossary For your convenience, each volume includes a comprehensive Glossary. Throughout the curriculum, a **bolded** word in a reading denotes a term defined in the Glossary.

Note that the digital curriculum that is included in your exam registration fee is searchable for key words, including Glossary terms.

LOS Self-Check We have inserted checkboxes next to each LOS that you can use to track your progress in mastering the concepts in each reading.

Source Material The CFA Institute curriculum cites textbooks, journal articles, and other publications that provide additional context or information about topics covered in the readings. As a candidate, you are not responsible for familiarity with the original source materials cited in the curriculum.

Note that some readings may contain a web address or URL. The referenced sites were live at the time the reading was written or updated but may have been deactivated since then.

Some readings in the curriculum cite articles published in the *Financial Analysts Journal®*, which is the flagship publication of CFA Institute. Since its launch in 1945, the *Financial Analysts Journal* has established itself as the leading practitioner-oriented journal in the investment management community. Over the years, it has advanced the knowledge and understanding of the practice of investment management through the publication of peer-reviewed practitioner-relevant research from leading academics and practitioners. It has also featured thought-provoking opinion pieces that advance the common level of discourse within the investment management profession. Some of the most influential research in the area of investment management has appeared in the pages of the *Financial Analysts Journal*, and several Nobel laureates have contributed articles.

Candidates are not responsible for familiarity with *Financial Analysts Journal* articles that are cited in the curriculum. But, as your time and studies allow, we strongly encourage you to begin supplementing your understanding of key investment management issues by reading this, and other, CFA Institute practice-oriented publications through the Research & Analysis webpage (www.cfainstitute.org/en/research).

Errata The curriculum development process is rigorous and includes multiple rounds of reviews by content experts. Despite our efforts to produce a curriculum that is free of errors, there are times when we must make corrections. Curriculum errata are periodically updated and posted by exam level and test date online (www.cfainstitute.org/en/programs/submit-errata). If you believe you have found an error in the curriculum, you can submit your concerns through our curriculum errata reporting process found at the bottom of the Curriculum Errata webpage.

DESIGNING YOUR PERSONAL STUDY PROGRAM

Create a Schedule An orderly, systematic approach to exam preparation is critical. You should dedicate a consistent block of time every week to reading and studying. Complete all assigned readings and the associated problems and solutions in each study session. Review the LOS both before and after you study each reading to ensure that

you have mastered the applicable content and can demonstrate the knowledge, skills, and abilities described by the LOS and the assigned reading. Use the LOS self-check to track your progress and highlight areas of weakness for later review.

Successful candidates report an average of more than 300 hours preparing for each exam. Your preparation time will vary based on your prior education and experience, and you will probably spend more time on some study sessions than on others.

You should allow ample time for both in-depth study of all topic areas and additional concentration on those topic areas for which you feel the least prepared.

CFA INSTITUTE LEARNING ECOSYSTEM (LES)

As you prepare for your exam, we will email you important exam updates, testing policies, and study tips. Be sure to read these carefully.

Your exam registration fee includes access to the CFA Program Learning Ecosystem (LES). This digital learning platform provides access, even offline, to all of the readings and End of Reading Questions found in the print curriculum organized as a series of shorter online lessons with associated EORQs. This tool is your one-stop location for all study materials, including practice questions and mock exams.

The LES provides the following supplemental study tools:

Structured and Adaptive Study Plans The LES offers two ways to plan your study through the curriculum. The first is a structured plan that allows you to move through the material in the way that you feel best suits your learning. The second is an adaptive study plan based on the results of an assessment test that uses actual practice questions.

Regardless of your chosen study path, the LES tracks your level of proficiency in each topic area and presents you with a dashboard of where you stand in terms of proficiency so that you can allocate your study time efficiently.

Flashcards and Game Center The LES offers all the Glossary terms as Flashcards and tracks correct and incorrect answers. Flashcards can be filtered both by curriculum topic area and by action taken—for example, answered correctly, unanswered, and so on. These Flashcards provide a flexible way to study Glossary item definitions.

The Game Center provides several engaging ways to interact with the Flashcards in a game context. Each game tests your knowledge of the Glossary terms a in different way. Your results are scored and presented, along with a summary of candidates with high scores on the game, on your Dashboard.

Discussion Board The Discussion Board within the LES provides a way for you to interact with other candidates as you pursue your study plan. Discussions can happen at the level of individual lessons to raise questions about material in those lessons that you or other candidates can clarify or comment on. Discussions can also be posted at the level of topics or in the initial Welcome section to connect with other candidates in your area.

Practice Question Bank The LES offers access to a question bank of hundreds of practice questions that are in addition to the End of Reading Questions. These practice questions, only available on the LES, are intended to help you assess your mastery of individual topic areas as you progress through your studies. After each practice question, you will receive immediate feedback noting the correct response and indicating the relevant assigned reading so you can identify areas of weakness for further study.

Mock Exams The LES also includes access to three-hour Mock Exams that simulate the morning and afternoon sessions of the actual CFA exam. These Mock Exams are intended to be taken after you complete your study of the full curriculum and take practice questions so you can test your understanding of the curriculum and your readiness for the exam. If you take these Mock Exams within the LES, you will receive feedback afterward that notes the correct responses and indicates the relevant assigned readings so you can assess areas of weakness for further study. We recommend that you take Mock Exams during the final stages of your preparation for the actual CFA exam. For more information on the Mock Exams, please visit www.cfainstitute.org.

PREP PROVIDERS

You may choose to seek study support outside CFA Institute in the form of exam prep providers. After your CFA Program enrollment, you may receive numerous solicitations for exam prep courses and review materials. When considering a prep course, make sure the provider is committed to following the CFA Institute guidelines and high standards in its offerings.

Remember, however, that there are no shortcuts to success on the CFA exams; reading and studying the CFA Program curriculum *is* the key to success on the exam. The CFA Program exams reference only the CFA Institute assigned curriculum; no prep course or review course materials are consulted or referenced.

SUMMARY

Every question on the CFA exam is based on the content contained in the required readings and on one or more LOS. Frequently, an exam question is based on a specific example highlighted within a reading or on a specific practice problem and its solution. To make effective use of the CFA Program curriculum, please remember these key points:

1. All pages of the curriculum are required reading for the exam.
2. All questions, problems, and their solutions are part of the curriculum and are required study material for the exam. These questions are found at the end of the readings in the print versions of the curriculum. In the LES, these questions appear directly after the lesson with which they are associated. The LES provides immediate feedback on your answers and tracks your performance on these questions throughout your study.
3. We strongly encourage you to use the CFA Program Learning Ecosystem. In addition to providing access to all the curriculum material, including EORQs, in the form of shorter, focused lessons, the LES offers structured and adaptive study planning, a Discussion Board to communicate with other candidates, Flashcards, a Game Center for study activities, a test bank of practice questions, and online Mock Exams. Other supplemental study tools, such as eBook and PDF versions of the print curriculum, and additional candidate resources are available at www.cfainstitute.org.
4. Using the study planner, create a schedule and commit sufficient study time to cover the study sessions. You should also plan to review the materials, answer practice questions, and take Mock Exams.
5. Some of the concepts in the study sessions may be superseded by updated rulings and/or pronouncements issued after a reading was published. Candidates are expected to be familiar with the overall analytical framework contained in the assigned readings. Candidates are not responsible for changes that occur after the material was written.

FEEDBACK

At CFA Institute, we are committed to delivering a comprehensive and rigorous curriculum for the development of competent, ethically grounded investment professionals. We rely on candidate and investment professional comments and feedback as we work to improve the curriculum, supplemental study tools, and candidate resources.

Please send any comments or feedback to info@cfainstitute.org. You can be assured that we will review your suggestions carefully. Ongoing improvements in the curriculum will help you prepare for success on the upcoming exams and for a lifetime of learning as a serious investment professional.

Corporate Issuers

STUDY SESSIONS

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to evaluate capital budget projects, capital structure policy, dividend policy, corporate governance, and mergers and acquisitions.

Capital investments, corporate structure, payout policies, corporate governance and ESG considerations, as well as mergers and acquisitions can significantly affect a company's operations, financials, and performance. Companies having strong leadership, well managed operations, sound corporate governance policies, and profitable investment activities are more likely to add value for their shareholders and other stakeholders.

© 2021 CFA Institute. All rights reserved.

CORPORATE ISSUERS
STUDY SESSION

6

Corporate Issuers (1)

This study session covers the capital budgeting process with emphasis on its principles and investment decision criteria. Project evaluation through the use of spreadsheet modeling is presented. Other income and valuation model approaches are compared. The subject of capital structure is introduced with the classic Modigliani–Miller irrelevance theory, which proposes that capital structure decisions should have no effect on company value. Additional considerations of taxes, agency costs, and financial distress are introduced. The session concludes with discussion on dividend policies, factors affecting distribution or reinvestment, and dividend payout or share repurchase decisions.

READING ASSIGNMENTS

Reading 15	Capital Structure by Raj Aggarwal, PhD, CFA, Pamela Peterson Drake, PhD, CFA, Adam Kobor, PhD, CFA, and Gregory Noronha, PhD, CFA
Reading 16	Analysis of Dividends and Share Repurchases by Gregory Noronha, PhD, CFA, and George H. Troughton, PhD, CFA

© 2021 CFA Institute. All rights reserved.

READING

15

Capital Structure

by Raj Aggarwal, PhD, CFA, Pamela Peterson Drake, PhD, CFA, Adam Kobor, PhD, CFA, and Gregory Noronha, PhD, CFA

Raj Aggarwal, PhD, CFA, is at the Kent State University Foundation Board (USA). Pamela Peterson Drake, PhD, CFA, is at James Madison University (USA). Adam Kobor, PhD, CFA, is at New York University (USA). Gregory Noronha, PhD, CFA, is at the University of Washington, Tacoma (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** explain the Modigliani–Miller propositions regarding capital structure;
☐	**b.** explain the effects on costs of capital and capital structure decisions of taxes, financial distress, agency costs, and asymmetric information;
☐	**c.** explain factors an analyst should consider in evaluating the effect of capital structure policy on valuation;
☐	**d.** describe international differences in the use of financial leverage, factors that explain these differences, and implications of these differences for investment analysis.

INTRODUCTION

1

The most important decision a company makes in pursuit of maximizing its value is typically the decision concerning what products to manufacture and/or what services to offer. The decision on how to finance investments (e.g., in factories and equipment), the so-called capital structure decision, is often seen as less important, even secondary. As we will see in this reading, the importance of the capital structure decision depends on the assumptions one makes about capital markets and the agents operating in it.

Under the most restrictive set of assumptions, the capital structure decision—the choice between how much debt and how much equity a company uses in financing its investments—is irrelevant. That is, any level of debt is as good as any other. The capital structure decision is not only secondary but also irrelevant. However, as some of the underlying assumptions are relaxed, the choice of how much debt to have in

© 2020 CFA Institute. All rights reserved.

the capital structure becomes meaningful. Under a particular set of assumptions, it is even possible to have an optimal level of debt in the capital structure—that is, a level of debt at which company value is maximized.

In this reading, we first discuss the capital structure decision and the assumptions and theories that lead to alternative capital structures. We then present important practical issues for the analyst, such as differences in capital structure policies arising from country-specific factors. We conclude with a summary of key points from the reading.

2 MODIGLIANI-MILLER PROPOSITION I WITHOUT TAXES: CAPITAL STRUCTURE IRRELEVANCE

a explain the Modigliani–Miller propositions regarding capital structure;

A company's **capital structure** is the mix of debt and equity the company uses to finance its business. The goal of a company's capital structure decision is to determine the financial leverage or capital structure that maximizes the value of the company by minimizing the weighted average cost of capital. The **weighted average cost of capital (WACC)** is given by the weighted average of the marginal costs of financing for each type of financing used. For a company with both debt and equity in its capital structure for which interest expense is tax deductible at a rate t, the WACC, which we will denote r_{wacc}, is

$$r_{wacc} = \left(\frac{D}{V}\right) r_d (1 - t) + \left(\frac{E}{V}\right) r_e \quad (1)$$

where

r_{wacc} = the weighted average cost of capital of the company
r_d = the before-tax marginal cost of debt capital
r_e = the marginal cost of equity capital
t = the company's marginal tax rate
D = the market value of shareholders' outstanding debt
E = the market value of shareholders' outstanding equity
V = the value of the company, which is equal to $D + E$

The ratios of D/V and E/V are the proportions of debt and equity in the capital structure, respectively. These are weights applied to the respective costs and are often indicated as w_d, representing the proportion of debt in the capital structure, and w_e, representing the proportion of equity. For simplicity, this discussion ignores preferred stock. You will notice that the debt and equity costs of capital and the tax rate are all understood to be "marginal" rates. The overall cost of capital is therefore a marginal cost also: what it costs the company to raise additional capital using the specified mixture of debt and equity. Further, this is the current cost: what it would cost the company today. What it cost in the past is not relevant. Therefore, the cost of equity, the cost of debt, and the tax rate that we use throughout the remainder of this reading are marginal: the cost or tax rate for additional capital.

In the following section, we first consider the theoretical relationship between leverage and a company's value. We then examine the practical relationship between leverage and company value in equal depth.

2.1 Proposition I without Taxes: Capital Structure Irrelevance

In a now-classic paper, Nobel Prize–winning economists Franco Modigliani and Merton Miller (1958) argued the important theory that given certain assumptions, a company's choice of capital structure does not affect its value. The assumptions relate to expectations and markets:

1 Investors agree on the expected cash flows from a given investment. This means that all investors have the same expectations with respect to the cash flows from an investment in bonds or stocks. In other words, expectations are homogeneous.

2 Bonds and shares of stock are traded in **perfect capital markets**. This means that there are no transaction costs, no taxes, no bankruptcy costs, and everyone has the same information. In a perfect capital market, any two investments with identical cash flow streams and risk must trade for the same price.

3 Investors can borrow and lend at the risk-free rate.

4 There are no agency costs. This means that managers always act to maximize shareholder wealth.

5 The financing decision and the investment decision are independent of each other. This means that operating income is unaffected by changes in the capital structure.

Many of these assumptions are unrealistic, and we will examine the consequences of relaxing some of them later in this section. The important point is that Modigliani and Miller provided a basis for thinking about capital structure and the starting point for analysis. Consider the capital of a company to be a pie: Each slice represents how much of total capital is provided by a specific type of capital (e.g., common equity). One can split it in any number of ways, but the size of the pie remains the same. Saying that the pie remains the same size is equivalent to saying that the present value of cash flows to the company remains the same. This can only happen if the future cash flow stream is expected to remain the same and if the risk of that cash flow stream, as reflected by the cost of capital, remains the same. Modigliani and Miller prove that under these conditions, and given their assumptions, changing the capital structure (i.e., how the pie is sliced) does not affect value. In other words, in a perfect capital market with risk-free borrowing and lending and with investment and financing decisions independent of each other, investors can create the capital structure which they individually prefer for the company by borrowing and lending on their own accounts. The capital structure chosen by management does not matter because it can be adjusted to the desired capital structure by investors at no cost.

Suppose that a company has a capital structure consisting of 50% debt and 50% equity and that an investor would prefer that the company's capital structure be 70% debt and 30% equity. The investor uses borrowed money to finance his or her share purchases so that effectively ownership of company assets reflects the preferred 70% debt financing. To the extent this changing capital structure has no effect on the company's expected operating cash flows, the capital structure decision has no impact on company value. Modigliani and Miller use the concept of arbitrage to demonstrate their point: If the value of an unlevered company—that is, a company without any debt—is not equal to that of a levered company, investors could make an arbitrage profit. The arbitrage operation (selling the overvalued asset and using the proceeds to buy the undervalued asset) would quickly force the values to be equivalent.

The importance of the Modigliani and Miller theory is that it demonstrates that managers cannot create value simply by changing the company's capital structure. Consider why this might be true. The operating earnings of a business are available to the providers of its capital. In an all-equity company (that is, a company with no debt), all of the operating earnings are available to the equityholders and the value of the company is the present value of these operating earnings. If a company is partially financed by debt, however, these operating earnings are split between the providers of capital: the equityholders and the debtholders. Under market equilibrium, the sum of the values of debt and equity in such a case should equal the value of the all-equity company. In other words, the value of a company is determined solely by its cash flows, not by the relative reliance on debt and equity capital.

This principle does not change the fact of the relative risks of leverage to debtholders versus equityholders. Adding leverage does increase the risk faced by the equityholders. In such a case, equityholders seek compensation for this extra risk by requiring a higher return. Indeed, in equilibrium the increase in equity returns is exactly offset by increases in the risk and the associated increase in the required rate of return on equity; there is no change in the value of the company. Finally, in the absence of taxes the capital structure irrelevance result holds whether debt is assumed to be risk-free (as Modigliani and Miller assumed in their 1958 article) or risky—as long as there are no bankruptcy costs.

Modigliani and Miller (MM) first illustrated the capital structure irrelevance proposition under the condition of no taxes:

> *MM Proposition I:*
>
> The market value of a company is not affected by the capital structure of the company.

In other words, the value of the company levered (V_L) is equal to the value unlevered (V_U), or $V_L = V_U$. A crucial implication of MM Proposition I is that the weighted average cost of capital for a company in the no-tax case is unaffected by its capital structure.

To understand this proposition, we can think about two companies with the same expected, perpetual cash flows and uncertainty and, hence, the same discount rate applied to value these cash flows. Even if the companies have different capital structures, these two companies must have the same present value using discounted cash flow models. If capital structure changes were to have any effect on a company's value, there would be an arbitrage opportunity to make riskless profits.

In a perfect market, investors can substitute their own leverage for a company's leverage by borrowing or lending appropriate amounts in addition to holding shares of the company. Because this process is costless for investors (assuming perfect markets), a company's financial leverage should have no impact on its value. Therefore, a company's capital structure is irrelevant in perfect markets (which assume no taxes).

3 MODIGLIANI-MILLER PROPOSITION II WITHOUT TAXES: HIGHER FINANCIAL LEVERAGE RAISES THE COST OF EQUITY

a explain the Modigliani–Miller propositions regarding capital structure;

Modigliani and Miller's second proposition focuses on the cost of capital of the company:

> *MM Proposition II:*
>
> The cost of equity is a linear function of the company's debt-to-equity ratio.

Assuming that financial distress has no costs and that debtholders have prior claim to assets and income relative to equityholders, the cost of debt is less than the cost of equity. According to this proposition, as the company increases its use of debt financing, the cost of equity rises. We know from MM Proposition I that the value of the company is unchanged and the weighted average cost of capital remains constant if the company changes its capital structure. What Proposition II means is that the cost of equity increases in such a manner as to exactly offset the increased use of cheaper debt in order to maintain a constant WACC.

The risk of the equity depends on two factors: the risk of the company's operations (business risk) and the degree of financial leverage (financial risk). Business risk determines the cost of capital, whereas the capital structure determines financial risk.

The weighted average cost of capital, or r_{wacc}, *ignoring taxes*, is simply Equation 1 with t (the company's marginal tax rate) = 0:

$$r_{wacc} = \left(\frac{D}{V}\right)r_d + \left(\frac{E}{V}\right)r_e \tag{2}$$

Let us define r_0 as the cost of capital for a company financed only by equity (an "all-equity company"). Then, by MM Proposition I, $r_{wacc} = r_0$, so

$$r_{wacc} = \left(\frac{D}{V}\right)r_d + \left(\frac{E}{V}\right)r_e = r_0 \tag{3}$$

Recalling that $D + E = V$ and using this to substitute for V, we can rearrange Equation 3 to solve for the cost of equity:

$$r_e = r_0 + \left(r_0 - r_d\right)\frac{D}{E} \tag{4}$$

Equation 4 is the precise expression for the cost of equity in MM Proposition II. As shown in Equation 4, the cost of equity is a linear function of the debt-to-equity ratio (D/E), with the intercept equal to r_0 and the slope coefficient equal to the positive quantity ($r_0 - r_d$). We know that ($r_0 - r_d$) is positive because the cost of equity must be an increasing function of the debt-to-equity ratio for WACC to be unchanged as the use of debt in financing is increased, as required by Proposition I. Thus, as the debt-to-equity ratio increases, the cost of equity capital also increases. See Exhibit 1 later in this section.

Consider the example of the Leverkin Company, which currently has an all-equity capital structure. Leverkin has an expected operating income of \$5,000 and a cost of equity, which is also its WACC, of 10%. Adopting a common practice, we represent operating income by earnings before interest and taxes, EBIT. For simplicity, we will assume that the EBIT and other cash flows are perpetual. Let us suppose that Leverkin is planning to issue \$15,000 in debt at a cost of 5% in order to buy back \$15,000 worth of its equity.

Because there are no taxes and the EBIT is a perpetuity, we can compute the value of the all-equity Leverkin as the present value of its expected cash flows:

$$V = \frac{\text{EBIT}}{r_{wacc}} = \frac{\$5{,}000}{0.10} = \$50{,}000$$

Under MM Proposition I, because $V_L = V_U$, the value of Leverkin remains the same whether it is all-equity financed or has \$15,000 of debt. When it issues the debt, Leverkin pays an interest charge of 5% on this debt. That is, Leverkin's interest payment is 0.05(\$15,000) = \$750.

By MM Proposition II, the cost of Leverkin's equity when it has \$15,000 debt and \$50,000 – \$15,000 = \$35,000 equity is

$$r_e = 0.10 + \left(0.10 - 0.05\right)\frac{\$15{,}000}{\$35{,}000} \approx 0.12143 = 12.143\%$$

The value of Leverkin with \$15,000 debt in its capital structure must equal the sum of the present value of cash flows to debtholders and equityholders. With \$15,000 debt, Leverkin makes an interest payment of \$750 to debtholders, leaving \$5,000 – \$750 = \$4,250 for equityholders.

$$V = D + E = \frac{\$750}{0.05} + \frac{\$4{,}250}{0.12143} = \$15{,}000 + \$34{,}999.59 \approx \$50{,}000$$

It is straightforward to demonstrate that Leverkin's value remains at \$50,000 at any level of debt. Note that this statement is true even with 100% debt financing because in that case the debtholders are effectively the company's owners (i.e., equityholders). We can also confirm, using Equation 3, that Leverkin's WACC with the new capital structure remains at 10% as required by Proposition I:

$$r_{wacc} = \left(\frac{\$15{,}000}{\$50{,}000}\right)0.05 + \left(\frac{\$35{,}000}{\$50{,}000}\right)0.12143 = 0.10 = 10\%$$

Just as we can express the beta of any investment portfolio as a market-value-weighted average of the betas of the investments in that portfolio, we can express the systematic risk of each of the sources of a company's capital in a similar manner using the Hamada equation (see Hamada 1972). In other words, we can represent the systematic risk of the assets of the entire company as a weighted average of the systematic risk of the company's debt and equity:

$$\beta_a = \left(\frac{D}{V}\right)\beta_d + \left(\frac{E}{V}\right)\beta_e \tag{5}$$

where β_a is the asset's systematic risk, or **asset beta**, β_d is the beta of debt, and β_e is the equity beta. The asset beta represents the amount of the assets' risk that is non-diversifiable (i.e., cannot be eliminated by holding those assets as part of a large, well-diversified portfolio).

According to Modigliani and Miller, the company's cost of capital does not depend on its capital structure but rather is determined by the business risk of the company. As the level of debt rises, however, the risk of the company defaulting on its debt increases. These costs are borne by the equityholders. So, as the proportionate use of debt rises, the equity's beta, β_e, also rises. By reordering the formula of β_a to solve for β_e, we get

$$\beta_e = \beta_a + (\beta_a - \beta_d)\left(\frac{D}{E}\right) \tag{6}$$

In the next section, we look at the decision to use debt financing given the taxes and market imperfections found in the real world.

4 MODIGLIANI-MILLER PROPOSITIONS WITH TAXES: TAXES, COST OF CAPITAL AND VALUE OF THE COMPANY

a explain the Modigliani–Miller propositions regarding capital structure;

b explain the effects on costs of capital and capital structure decisions of taxes, financial distress, agency costs, and asymmetric information;

Taxes are the first practical consideration in modifying the results of the MM propositions. In the following discussion, we present MM Propositions I and II with taxes.

Because interest paid is deductible from income for tax purposes in most countries, the use of debt provides a tax shield that translates into savings that enhance the value of a company. Indeed, ignoring other practical realities of costs of financial distress and bankruptcy, the value of the company increases with increasing levels of debt. In effect, by making the interest costs deductible for income taxes, the government subsidizes companies' use of debt. The actual cost of debt is reduced by the level of the company's tax benefit:

After-tax cost of debt = Before-tax cost of debt × (1 − Marginal tax rate).

Modigliani and Miller show that in the presence of corporate taxes (but not personal taxes), the value of the company with debt is greater than that of the all-equity company by an amount equal to the tax rate multiplied by the value of the debt. That is, MM Proposition I with corporate taxes is:

$$V_L = V_U + tD, \quad (7)$$

where t is the marginal tax rate. The term tD is often referred to as the debt tax shield.

By introducing corporate tax, we adjust the weighted average cost of capital formula to reflect the impact of the tax benefit:

$$r_{wacc} = \left(\frac{D}{V}\right) r_d (1 - t) + \left(\frac{E}{V}\right) r_e \quad (8)$$

Because by Proposition I with taxes the value of a company with debt is greater than that of the same company without debt *for the same level of operating income*, it must follow that the WACC for the company with debt *must be lower* than that for the all-equity company. If we continue to define r_0 as the cost of capital for an all-equity company, MM show that the cost of equity for the same company with debt is:

$$r_e = r_0 + (r_0 - r_d)(1 - t)\frac{D}{E} \quad (9)$$

This is MM Proposition II when there are corporate taxes. Notice that the difference between Equation 9 and MM Proposition II in the no-tax case (Equation 4) is the presence of the term $(1 - t)$. When t is not zero, the term $(1 - t)$ is less than 1 and serves to lower the cost of leveraged (or levered) equity when compared to the no-tax case. That is, the cost of equity becomes greater as the company increases the amount of debt in its capital structure, but the cost of equity does not rise as fast as it does in the no-tax case. Equivalently, the slope coefficient is $(r_0 - r_d)(1 - t)$, which is smaller than the slope coefficient $(r_0 - r_d)$ in the case of no taxes. As a consequence, the WACC for the leveraged company falls as debt increases and overall company value increases. Therefore, if taxes are considered but financial distress and bankruptcy costs are not, debt financing is highly advantageous. In the extreme, a company's optimal capital structure is all debt.

Let us return to the example of the Leverkin Company, which is currently all-equity, has an EBIT of \$5,000, and a WACC (which is also its cost of equity) of 10%. As before, Leverkin is planning to issue \$15,000 of debt in order to buy back an equivalent amount of equity. Now, however, Leverkin pays corporate taxes at a rate of 25%.

Because Leverkin must pay taxes, the after-tax cash flow available to its shareholders is earnings before tax, EBT, times $(1 - t)$, or EBT$(1 - t)$. EBT$(1 - t)$ is the same here as EBIT$(1 - t)$ because I = 0. If we continue to assume perpetual cash flows, the value of the all-equity, or unlevered, Leverkin is:

$$V_U = \frac{\text{EBT}(1 - t)}{\text{WACC}} = \frac{\$5{,}000(1 - 0.25)}{0.10} = \$37{,}500$$

Note that the value of Leverkin when there are corporate taxes is less than its value in the no-tax case. This is simply because a new claimant on Leverkin's cash flows, the government through its ability to impose taxes, has entered the picture.

Let us now see what happens to Leverkin's value when it issues \$15,000 in debt and buys back stock. According to MM Proposition I, when there are corporate taxes (i.e., Equation 7), then

$$V_L = V_U + tD = \$37{,}500 + 0.25(\$15{,}000) = \$41{,}250.$$

Because the value of the debt is \$15,000, the value of the equity (after the buyback) must be (\$41,250 – \$15,000) = \$26,250. According to MM Proposition II with corporate taxes (Equation 9), the cost of the levered equity is:

$$r_e = 0.10 + (0.10 - 0.05)(1 - 0.25)\frac{\$15{,}000}{\$26{,}250} = 0.12143 = 12.143\%$$

Because the value of the company must equal the present value of cash flows to debt and to equity,

$$\begin{aligned} V_L &= D + E = \frac{r_d D}{r_d} + \frac{(\text{EBIT} - r_d D)(1 - t)}{r_e} \\ &= \frac{\$750}{0.05} + \frac{(\$5{,}000 - \$750)(1 - 0.25)}{0.12143} \approx \$41{,}250 \end{aligned}$$

This is the value of the company as given by MM Proposition I. As a further check, using Equation 8 the WACC for the levered Leverkin is:

$$\begin{aligned} r_{wacc} &= \frac{\$15{,}000}{\$41{,}250}(0.05)(1 - 0.25) + \frac{\$26{,}250}{\$41{,}250}(0.12143) \\ &= 0.09091 = 9.091\% \end{aligned}$$

As expected, this is lower than the unlevered WACC of 10%. Because after taxes are paid whatever is left of the cash flows can be claimed by debtholders and equityholders, we must also have:

$$V_L = \frac{\text{EBIT}(1 - t)}{\text{WACC}} = \frac{\$5{,}000(1 - 0.25)}{0.09091} \approx \$41{,}250$$

We can see the effect of taxes on the cost of capital in Exhibit 1. Here, we see that if there are no taxes, as shown in Panel B, the cost of capital is constant at $r_{wacc} = r_0$. If interest is tax deductible, however, the cost of capital declines for ever-increasing use of debt financing, as shown in Panel C.

Exhibit 1

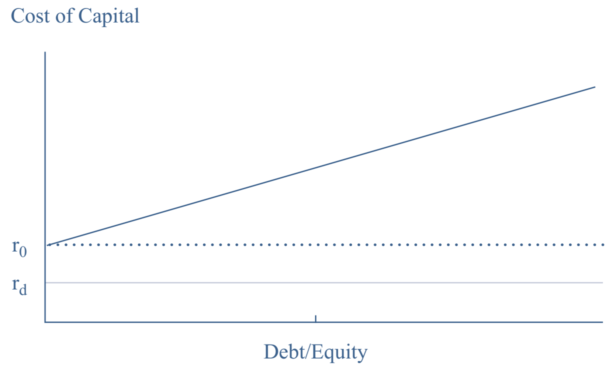
Modigliani and Miller Propositions

A. Value of the Company and Cost of Capital for Propositions without and with Taxes

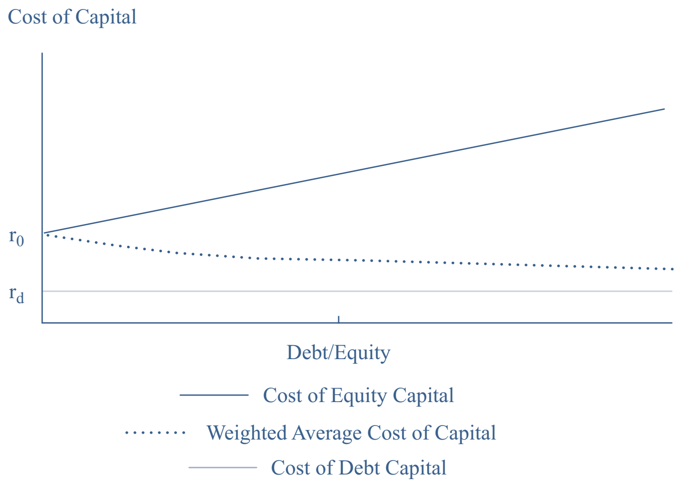
	Without Taxes	With Taxes
Proposition I	$V_L = V_U$	$V_L = V_U + tD$
Proposition II	$r_e = r_0 + (r_0 - r_d)\frac{D}{E}$	$r_e = r_0 + (r_0 - r_d)(1 - t)\frac{D}{E}$

Exhibit 1 (Continued)

B. Costs of Capital with No Taxes



C. Costs of Capital with Taxes



Miller (1977) introduced another aspect into the analysis of the tax benefits from the use of debt financing. He argued that if investors face different tax rates on dividend and interest income for their personal taxes, this situation may reduce the advantage of debt financing somewhat. If investors face a higher personal rate of tax on income from debt investments relative to stock investments, they will demand a higher return on debt—driving up the cost of debt to the company. Thus in the Miller model, whether or not financing with debt adds value to the company depends on the corporate tax rate, the personal tax rate on interest income, and the personal tax rate on dividend income. It is therefore possible in the Miller model, depending on the levels of the various tax rates, for debt to add value, lower value, or to have no effect on value.

In practice, however, the value of a levered company is affected by more than the tax issues surrounding the use of debt. The analysis gets more complicated once we introduce such factors as the cost of financial distress, agency costs, and asymmetric information. We address these additional factors next.

5 OTHER CAPITAL STRUCTURE COST CONSIDERATIONS: COSTS OF FINANCIAL DISTRESS

b explain the effects on costs of capital and capital structure decisions of taxes, financial distress, agency costs, and asymmetric information;

The downside of operating and financial leverage is that earnings are magnified downward during economic slowdowns. Lower or negative earnings put companies under stress, and this **financial distress** adds costs—both explicit and implicit—to a company. Even before taking the drastic step of filing for bankruptcy, companies under stress may lose customers, creditors, suppliers, and valuable employees to more secure competitors.

EXAMPLE 1

Costs of Financial Distress

The Carillion Plc liquidation provides an example of the costs of financial distress and the loss of shareholder value. Carillion was a UK-based construction and services entity that employed 43,000, had revenues of almost £4.4 billion, and was involved in many projects. These projects included ongoing projects with the UK government for the maintenance of roads and for catering and cleaning contracts at hundreds of schools. Carillion's share price was £308 per share at the end of 2015 and £236 at the end of 2016.

The first sign of trouble came in mid-2017 when Carillion warned on its full-year profit, citing issues with payment problems in several of its construction contracts. This announcement was followed soon after with a 30% loss of shareholder value in the market. Even as it faced these collection problems on its contracts, however, it was awarded more contracts, including those for railway and military projects.

Subsequent profit warnings sent share value further downward. The second profit warning of that year was followed by new credit facilities and deferrals of debt repayments. The third profit warning, which included concern over violating debt covenants, resulted in a further decline in the value of its shares, with shares trading around £21 per share.

Discussions with creditors in December 2017 and January 2018 did not result in an agreement. The company was put in compulsory liquidation on 15 January 2018, and share trading was suspended.

The primary causes of Carillion's problems were its many overreaching, unprofitable projects and the inability to collect quickly on its projects. Moreover, goodwill—an intangible asset associated with the purchase of one company by another—comprised a significant portion of assets. In FY2016, alone, goodwill was over 35% of the company's assets.

The costs of this financial distress included:

- the loss of all shareholder value from a market capitalization of over £2 billion in 2016;[1]

1 BBC News, "Carillion Collapse to Cost Taxpayers £148m" (7 June 2018): https://www.bbc.com/news/business-44383224.

- £44.2 million paid to an accounting firm as special managers of the insolvency;[2]
- costs to the UK government as it grappled with nationalizing contracts, at least on a temporary basis;[3] and
- creditors paid potentially a small fraction of what was owed to them.

The expected cost of financial distress is composed of two key ingredients: 1) the costs of financial distress and bankruptcy, in the event they happen, and 2) the probability that financial distress and bankruptcy happen. We can classify the costs of financial distress into direct and indirect costs. Direct costs of financial distress include the actual cash expenses associated with the bankruptcy process, such as legal and administrative fees. Indirect costs of financial distress include forgone investment opportunities, impaired ability to conduct business, and agency costs associated with the debt during periods in which the company is near or in bankruptcy.

Companies whose assets have a ready secondary market have lower costs associated with financial distress. Companies with relatively marketable tangible assets—such as airlines, shipping companies, and steel manufacturers—incur lower costs from financial distress because such assets are usually more readily marketable. In contrast, companies with few tangible assets—such as high-tech growth companies, pharmaceutical companies, information technology companies, and others in the service industry—have less to liquidate and therefore have a higher cost associated with financial distress.

The probability of bankruptcy increases as the degree of leverage increases. The probability of bankruptcy for a given company depends on how the fixed costs of debt service interact with the instability of the business environment and the reserves available to the company to delay bankruptcy. In other words, the probability of bankruptcy depends, in part, on the company's business risk. Other factors that affect the likelihood of bankruptcy include the company's corporate governance structure and the management of the company.

OTHER CAPITAL STRUCTURE COSTS: AGENCY COSTS AND COSTS OF ASYMMETRIC INFORMATION 6

b explain the effects on costs of capital and capital structure decisions of taxes, financial distress, agency costs, and asymmetric information;

Agency costs are the costs associated with the fact that all public companies and the larger private companies are managed by non-owners. Agency costs are the incremental costs arising from conflicts of interest when an agent makes decisions for a principal. In the context of a corporation, agency costs arise from conflicts of interest between managers, shareholders, and bondholders. In the following, "perquisite consumption" refers to items that executives may legally authorize for themselves that have a cost to shareholders, such as subsidized dining, a corporate jet fleet, and chauffeured limousines.

---

2 Mark Cobley,, "PwC to Reap £44m from Carillion Bankruptcy," *Financial News* (6 February 2019): https://www.fnlondon.com/articles/pwc-to-reap-44m-from-carillion-bankruptcy-20190206.

3 Hallie Detrick, "What You Need to Know about the Collapse of Carillion, a U.K. Construction Giant," *Fortune* (15 January 2018): http://fortune.com/2018/01/15/what-you-need-to-know-about-the-collapse-of-carillion-a-u-k-construction-giant/.

The smaller the stake managers have in the company, the less their share in bearing the cost of excessive perquisite consumption—consequently, the less their desire to give their best efforts in running the company. The costs arising from this conflict of interest have been called the **agency costs of equity**. Given that outside shareholders are aware of this conflict, they will take actions to minimize the loss, such as requiring audited financial statements. The net agency costs of equity, therefore, have three components (Jensen and Meckling 1976):

1 **Monitoring costs**. These are the costs borne by owners to monitor the management of the company and include the expenses of the annual report, board of director expenses, and the cost of the annual meeting.

2 **Bonding costs**. These are the costs borne by management to assure owners that they are working in the owners' best interest. These include the implicit cost of noncompete employment contracts and the explicit cost of insurance to guarantee performance.

3 **Residual loss**. This consists of the costs that are incurred even when there is sufficient monitoring and bonding, because monitoring and bonding mechanisms are not perfect.

The better a company is governed, the lower the agency costs. Good governance practices translate into higher shareholder value because managers' interests are better aligned with those of shareholders. Additionally, agency theory predicts that a reduction in net agency costs of equity results from an increase in the use of debt versus equity. That is, there are savings in the agency costs of equity associated with the use of debt. Similarly, the more financially leveraged a company is, the less freedom managers have to either take on more debt or unwisely spend cash. This is the foundation of Michael Jensen's **free cash flow hypothesis**. According to Jensen (1986), higher debt levels discipline managers by forcing them to manage the company efficiently so the company can make its interest and principal payments and by reducing the company's free cash flow and thus management's opportunities to misuse cash. Harvey, Lins, and Roper (2004) observe that this discipline is especially important in emerging markets, in which there is a tendency to overinvest.

6.1 Costs of Asymmetric Information

Asymmetric information (an unequal distribution of information) arises from the fact that managers have more information about a company's performance and prospects (including future investment opportunities) than do outsiders, such as owners and creditors. Whereas all companies have a certain level of asymmetric information, companies with comparatively high asymmetry in information are those with complex products. These include high-tech companies, companies with little transparency in financial accounting information, or companies with lower levels of institutional ownership. Providers of both debt and equity capital demand higher returns from companies with higher asymmetry in information because they have a greater likelihood of agency costs.

Some degree of asymmetric information always exists because investors never know as much as managers and other insiders. Consequently, investors often closely watch manager behavior for insight into insider opinions on the company's future prospects. Aware of this scrutiny, managers take into account how their actions might be interpreted by outsiders. The signaling model of capital structure suggests a hierarchy ("pecking order") to the selection of methods for financing new investments.

The **pecking order theory**, developed by Myers and Majluf (1984), suggests that managers choose methods of financing according to a hierarchy that gives first preference to methods with the least potential information content (internally generated funds) and lowest preference to the form with the greatest potential information content (public equity offerings). Public equity offerings are, in general, closely scrutinized because investors are typically skeptical that existing owners would share ownership of a company with a great future with other investors. In brief, managers prefer internal financing. If internal financing is insufficient, managers next prefer debt, then equity. Another implication of the work of Myers and Majluf is that financial managers tend to issue equity when they believe the stock is overvalued but are reluctant to issue equity if they believe the stock is undervalued. Thus, additional issuance of stock is often interpreted by investors as a negative signal.

We can read the signals that managers provide in their choice of financing method. For example, commitments to fixed payments, such as dividends and debt service payments, may be interpreted as the company's management having confidence in the company's future prospects of making payments. Such signals are considered too costly for poorly performing companies to afford. Alternatively, the signal of raising money at the top of the pecking order and issuing equity at the bottom of the pecking order holds other clues. If, for instance, the company's cost of capital increases after an equity issuance, we may interpret this effect as an indication that management needed capital beyond what comes cheaply; in other words, this is a negative signal regarding the company's future prospects.

OPTIMAL CAPITAL STRUCTURE: STATIC TRADE-OFF THEORY

7

c explain factors an analyst should consider in evaluating the effect of capital structure policy on valuation;

When companies make decisions about financial leverage, they must weigh the value-enhancing effects of leverage from the tax deductibility of interest against the value-reducing impact of the costs of financial distress or bankruptcy, agency costs of debt, and asymmetric information. Putting together all the pieces of the Modigliani and Miller theory—along with the taxes, costs of financial distress, debt agency costs, and asymmetric information—we see that as financial leverage is increased, there comes a point beyond which further increases in value from value-enhancing effects are offset completely by value-reducing effects. This point is known as the **optimal capital structure**. In other words, the optimal capital structure is that capital structure at which the value of the company is maximized.

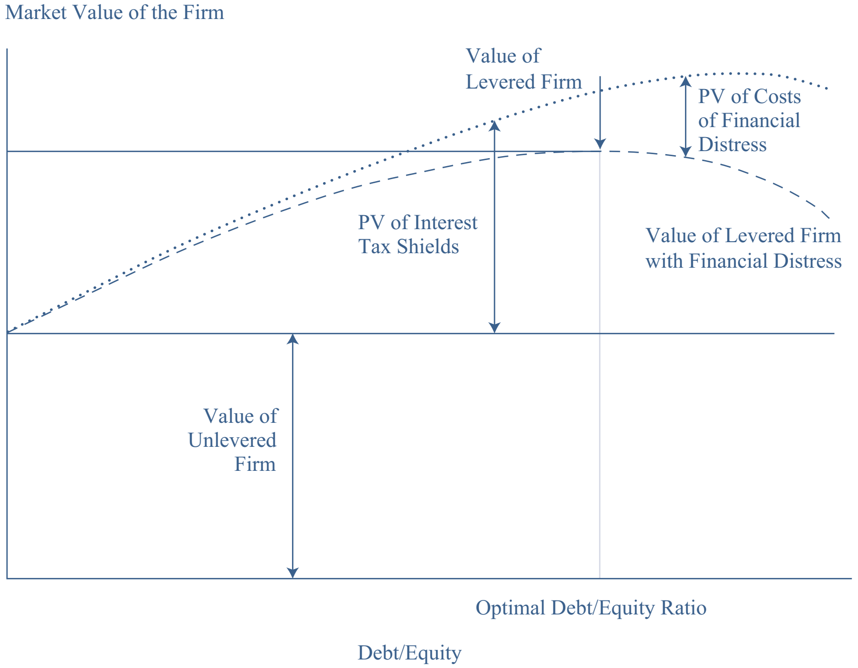
Considering only the tax shield provided by debt and the costs of financial distress, the expression for the value of a leveraged company becomes

$$V_L = V_U + tD - PV(\text{Costs of financial distress}). \quad (10)$$

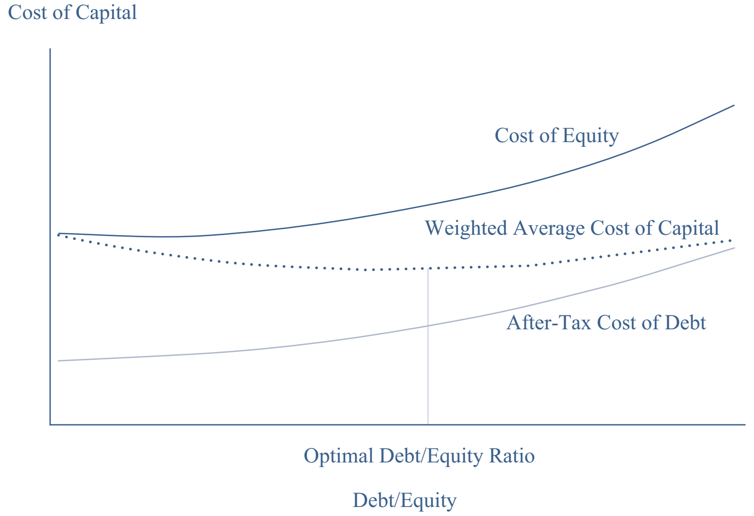
Equation 10 represents the **static trade-off theory of capital structure**. It results in an optimal capital structure such that debt constitutes less than 100% of a company's capital structure. We diagram this optimum in Exhibit 2.

Exhibit 2 Trade-off Theory with Taxes and Cost of Financial Distress

A. Value of the Company and the Debt/Equity Ratio



B. Cost of Capital and the Debt/Equity Ratio



The static trade-off theory of capital structure is based on balancing the expected costs from financial distress against the tax benefits of debt service payments, as shown in Panel A of Exhibit 2. Unlike the Modigliani and Miller proposition of no optimal capital structure, or a structure with almost all debt when the tax shield is considered, static trade-off theory puts forth an optimal capital structure with an optimal proportion of debt. Optimal debt usage is found at the point where any additional debt would cause the costs of financial distress to increase by a greater amount than the benefit of the additional tax shield.

We cannot say precisely at which level of debt financing a company reaches its optimal capital structure. The optimal capital structure depends on the company's business risk combined with its tax situation, corporate governance, and financial accounting information transparency, among other factors. However, what we can say, based on this theory, is that a company should consider a number of factors, including its business risk and the possible costs of financial distress, in determining its capital structure.

A company's management uses these tools to decide the level of debt appropriate for the company. The tax benefit from the deductibility of the interest expense on debt must be balanced against the risk associated with the use of debt. The extent of financial leverage used should thus depend on owners' and management's appetites for risk and the stability of the company's business environment. Indeed, as we show in Panel B of Exhibit 2, as the proportion of debt in a business rises, the costs of both debt and equity are likely to rise to offset the higher risks associated with higher levels of debt. These cost increases reduce or even negate the cost savings due to the greater use of debt, the cheaper source of financing. The result is a U-shaped weighted average cost of capital curve.

When the company recognizes its most appropriate or best capital structure, it may adopt this as its **target capital structure**. Because management may exploit short-term opportunities in one or another financing source and because market-value fluctuations continuously affect the company's capital structure, a company's capital structure at any point in time may differ from the target. In addition, it may be impractical (due to market conditions making it inadvisable to raise capital) and expensive (because of flotation costs) for a company to continuously maintain its target structure. Nevertheless, so long as the assumptions of the analysis and the target are unchanged, analysts and management should focus on the target capital structure.

CAPITAL STRUCTURE POLICY: PRACTICAL ISSUES AND EVALUATION 8

c explain factors an analyst should consider in evaluating the effect of capital structure policy on valuation;

Although capital structure theories should serve to inform an analyst's decision-making process in valuing a company, there are several practical aspects of capital structure to consider. These include an assessment of company creditworthiness, an understanding of the industry to which the company belongs, and an analysis of the legal, institutional, and macroeconomic environment in which the company operates. We consider these factors next.

8.1 Debt Ratings

Debt ratings are an important consideration in the practical management of leverage. As leverage rises, rating agencies tend to lower the ratings of the company's debt to reflect the higher credit risk resulting from the increasing leverage. Lower ratings signify higher risk to both equity and debt capital providers, who therefore demand higher returns.

Most large companies pay one or more rating services to provide debt ratings for their bonds. Debt issues are rated for creditworthiness by these credit rating agencies after they perform a financial analysis of the company's ability to pay the promised cash flows as well as an analysis of the bond's indenture (i.e., the set of complex legal documents associated with the issuance of debt instruments).

These agencies evaluate a wealth of information about the issuer and the bond, including the bond's characteristics and indenture agreement, and provide investors with an assessment of the company's ability to pay the interest and principal on the bond as promised.

Thus, in practice, most managers consider the company's debt rating in their policies regarding capital structure because the cost of capital is tied so closely to bond ratings.

8.2 Evaluating Capital Structure Policy

In evaluating a company's capital structure, the financial analyst must look at the capital structure of the company over time, the company's business risk, the capital structure of competitors that have similar business risk, and company-specific factors, such as the quality of corporate governance, that may affect agency costs. Good corporate governance should lower the net agency costs of equity. The financial analyst is not privy to the company's target capital structure but rather can evaluate the company's ability to handle its financial obligations and the potential role of costs of financial distress in determining how much financial leverage a company can handle.

Several practical considerations are important in this regard, such as the industry in which a company operates, the volatility of the company's cash flows, and its need for financial flexibility. Regulatory aspects can also play a role. For example, companies in the utility industry have relatively stable cash flows because they have a natural monopoly. Such companies usually also have a low degree of information asymmetry. As a result, utility companies tend to have much more debt than companies in other industries. Similarly, the guarantee afforded by deposit insurance in the United States, for example, allows banks in the United States to have debt levels in excess of 80% of assets.[4] In contrast, companies in the technology or pharmaceutical industries tend to have little or no debt for the following three reasons. 1) They have few tangible assets (their assets are chiefly human capital, patents, ideas, etc.). 2) There is a high degree of information asymmetry (such companies spend a lot on research and development and are very secretive about their products). 3) They have a great need for financial flexibility (they need to respond quickly to competitive and other changes in their operating environment).

A common goal of capital structure decisions is to finance at the lowest cost of capital. Analysts can use a scenario approach to assess this point for a particular company, starting with the current cost of capital for a company and considering various changes, to answer the following questions:

1 How do possible changes in capital structure affect changes in bond ratings, and how may this likely affect the cost of capital?

2 Are there anticipated events, such as an acquisition or merger, that may alter the company's business risk?

3 Is the company's capital structure in line with its peers and competitors? Why or why not?

4 Board of Governors of the Federal Reserve System, "Assets and Liabilities of Commercial Banks in the United States – H.8" (Release Date 22 March 2019): https://www.federalreserve.gov/releases/h8/20190322/.

4 Does the company tend to maintain a consistent capital structure, or does this appear to be changing over time to be more or less levered?

5 Do any recent or anticipated changes in tax rates affect the costs of capital?

INTERNATIONAL CAPITAL STRUCTURE DIFFERENCES: COUNTRY-SPECIFIC FACTORS

9

d describe international differences in the use of financial leverage, factors that explain these differences, and implications of these differences for investment analysis.

Modigliani and Miller told us that under several conditions the market value of a company is independent of its capital structure. However, we know that a company's capital structure is indeed relevant in the real world because of the effects of taxation, the costs of financial distress, and agency costs. The static trade-off theory suggests that the optimal level of leverage should be the level at which the value of the company is maximized; this is the level of debt financing at which any additional debt increases the costs of financial distress by an amount greater than the benefit from interest deductibility.

A company's capital structure largely depends on company-specific factors, such as the probability of bankruptcy, profitability, quality and structure of assets, and growth opportunities. Beyond these factors, the company's industry affiliation and the characteristics of the country where the company operates can also account for differences in capital structure.

The general business environment differs from one country to another. While researchers show that country-specific factors are becoming less of an influence on capital structure decisions over time, a meaningful comparison of financial leverage indicators of a US-based energy company and a Japanese energy company still requires us to take country-specific differences into account. Tradition, tax policy, and regulation may largely explain the different degrees of leverage in the two countries.

In examining the capital structure and debt maturity structure of corporations in an international context, researchers generally find that differences in the capital structures exist between developed and emerging markets as well as across the developed countries. Moreover, the debt maturity structure—another important capital structure decision—also tends to vary across the international setting. Therefore, when analysts focus on the capital structure of companies in an international setting, they must consider both the relative use of debt and the maturity structure of debt.

In fact, short-term and long-term debt ratios follow very different patterns in international comparisons. Whereas such characteristics of corporations as profitability and the degree to which total assets are tangible affect capital structure decisions globally, country-specific characteristics—and, to a lesser degree, global factors—affect capital structure decisions. For example, greater growth in GDP and greater property rights are associated with longer bond maturities. Further, capital structure decisions are affected by such global factors as risk aversion, where greater risk aversion is associated with shorter maturities.

Beyond the pure comparison of the capital structures, it is equally or even more important to identify and understand the country-specific factors that explain the cross-country differences. Three major types of factors may be used to explain most capital structure differences in an international comparison:

1 *Institutional and legal environment*: These factors represent the legal and regulatory environment in which companies operate and the requirements related to financial reporting. These institutional factors—including accounting standards, information transparency, taxation, and even the presence or lack of corruption—may affect a company's optimal capital structure.

2 *Financial markets and banking sector*: These factors include characteristics of the banking sector and the size and activity of the financial markets. Financial institutions are crucial for companies' access to financing.

3 *Macroeconomic environment*: These factors capture the general economic and business environment, addressing the influence of economic growth and inflation on the capital structure.

9.1 Institutional and Legal Environment

Taxation, financial legislation, the content of laws (e.g., bankruptcy law), and the quality of enforcement all differ from one country to another. These differences may influence the capital structures of companies and explain many of the differences that we observe across countries.

The apparent conflict of interest between a company's management and outside investors has already been addressed as the agency problem. This problem is, in fact, one of the key determinants of a company's ability to obtain capital; hence, agency costs are one of the major factors determining the capital structure. This conflict may be mitigated by carefully prepared contracts. The quality of investors' legal protections depends on both the content and the enforcement of the contracts and laws. As a result, we expect to see higher financial leverage in those countries that have weaker, less-efficient legal systems. Further, in countries with weaker legal systems, we expect a greater use of short-term debt financing versus long-term debt financing. Researchers find that companies operating in countries with an efficient legal system tend to use more long-term debt than short-term debt and exhibit lower leverage than comparable companies in countries with weaker legal systems.

Some researchers assume that legal systems based on common law offer external capital providers (both equity and debt providers) better protection compared to the legal systems of civil-law countries. Common law originated in England and is also followed in other countries, such as the United States, Canada, Australia, New Zealand, and Singapore. Civil law has origins going back to ancient Rome; the countries of continental Europe and South America and most of the rest of the world have legal systems based on this tradition. Researchers find mixed and limited evidence that companies operating in common-law countries tend to have longer debt maturity structures compared to their peers in civil-law countries. Further, in countries with a legal system based on common law, companies tend to use less debt and more equity in their capital structure. Further, some researchers observe that companies in countries with greater property rights tend to have longer debt maturities. In countries with stronger creditor rights, researchers find that companies tend to use less debt.

Similar to the rationale described in the case of legal system efficiency, a high level of information asymmetry favoring insiders over outsiders encourages a greater use of debt relative to equity as well as a greater reliance on short-term debt than on long-term debt in the capital structure. This is likely due to the fact that enforcing the debt contract is easier than enforcing the less clearly contracted shareholders' rights. Auditors and financial analysts can help reduce information asymmetries and increase the level of transparency. Researchers confirm that the presence of auditors and analysts is associated with lower financial leverage. The importance of auditors is usually strongest in emerging markets, whereas the presence of analysts is more important in developed markets. Researchers also observe that companies tend to use more equity to finance investments in countries with stronger information transparency.

As discussed earlier, taxes affect the capital structure decision by lowering the cost of debt financing to the issuer in those jurisdictions in which interest expense is tax deductible. In the absence of debt agency costs and bankruptcy costs, the benefit from the tax deductibility of interest encourages companies to use debt financing instead of equity financing. However, if dividend income is taxed at lower rates than interest income, some of the advantage of debt versus equity financing may be reduced from the corporate perspective because the price at which equity can be sold should reflect that advantage. In general, researchers observe that both corporate and personal tax rates influence a company's capital structure decisions.

9.2 Financial Markets and the Banking Sector

The size, activity, and liquidity of capital markets are crucial for corporations' access to capital. Several researchers have analyzed the impact of capital markets' characteristics on companies' capital structures. Some find that liquid and active capital markets affect companies' debt maturity structure. Specifically, they find that companies in countries that have liquid and active capital markets tend to use more long-term (as opposed to short-term) debt with longer maturity (30-year maturity is preferred to 15-year maturity). Researchers attribute this finding to the heightened external monitoring of companies by financial analysts and other market participants in active markets.

The banking sector is one of the primary sources of funds for the corporate sector in many countries, and its role is especially significant in countries that do not have a corporate bond market. The importance of the banking sector relative to the capital markets can vary from one country to another, however. Countries with a common-law tradition in which the shareholders' rights are stronger tend to be more market-based, whereas civil-law countries tend to be more bank-based. Because the relationship between a bank and a company is stronger and closer than between a company and a bondholder, banks can handle information asymmetries more efficiently. This effect may partly explain why civil-law countries are more bank-oriented. Researchers find a difference in the role of capital structure determinants based on whether the country's economy is bank-oriented (such as in France, Germany, and Japan) or capital market-oriented (such as in the United States and the United Kingdom). For example, the tangibility of assets plays a bigger role in capital structure decisions in bank-oriented economies; that is, bank-oriented economies focus on such tangible assets as plants and equipment rather than on such intangible assets as patents or goodwill.

An element of the role of the type of market—bank-oriented or market-oriented—is the resultant information available to stakeholders. De Fiore and Uhlig (2011) find that the increased availability of information about credit risk in the euro area may result in high debt, while the better availability of public information on market participants in the United States may reduce agency costs.

The presence of institutional investors may also affect the companies' capital structure choice. Some institutional investors may have preferred debt maturities ("preferred habitats"), and these preferences may affect companies' debt maturity structure. Insurance companies and pension plans, for example, may prefer investing in long-term debt securities to align with the interest rate risk of their long-term liabilities.

9.3 Macroeconomic Environment

Macroeconomic factors may affect both the relative use of debt and debt maturity. For example, researchers examining the effects of the 2007–2008 global financial crisis observed reduced reliance on debt financing and reduced debt maturity in both developed and developing countries as a result of the crisis, even for countries not directly affected by the crisis. Additionally, researchers observe that countries with companies that have higher debt and a greater ability to substitute bonds for loans recover faster from the trough of a business cycle.

Inflation is a widely recognized macroeconomic indicator. High inflation has a negative impact on both the level of debt financing and the use of long-maturity debt. Companies in higher-inflation countries usually exhibit lower levels of financial leverage, rely more on equity financing, and have a shorter debt maturity structure compared to their peers in lower-inflation countries.

Researchers have also found that the growth in gross domestic product is associated with longer debt maturity in developed markets. In addition, researchers focusing on developing countries find that companies in countries with high growth rely more on equity financing.

9.4 Conclusions

Financial analysts must consider country-specific factors when analyzing and comparing companies that operate in different countries. We have summarized these factors in Exhibit 3.

Exhibit 3 Country-Specific Factors and Their Assumed Impacts on the Companies' Capital Structure

Country-Specific Factor	If a Country	. . . Then D/E Ratio Is Potentially	. . . And Debt Maturity Is Potentially
Institutional framework			
Legal system efficiency	Is more efficient	Lower	Longer
Legal system origin	Has common law as opposed to civil law	Lower	
Information asymmetries	Has greater shareholder rights	Lower	Longer
Information intermediaries	Has auditors and analysts	Lower	Longer
Taxation	Has taxes that favor equity	Lower	
Banking system, financial markets			
Equity and bond markets	Has active bond and stock markets		Longer
Bank-based or market-based country	Has a bank-based financial system	Higher	
Investors	Has large institutional investors	Lower	Longer
Macroeconomic environment			
Inflation	Has high inflation	Lower	Shorter
Growth	Has high GDP growth	Lower	Longer

These factors include the differences in the business and legal environments in other countries, taxes, and macroeconomic factors, among others. Companies' optimal capital structures may differ simply as a consequence of these many country-specific differences. In addition to presenting challenges for international financial and credit analysis, these international differences in debt ratios present some challenges in developing debt policies for the foreign subsidiaries of multinational companies. Theory provides little guidance, and corporate practices in this area seem to vary widely.

SUMMARY

In this reading, we have reviewed theories of capital structure and considered practical aspects that an analyst should examine when making investment decisions.

- The goal of the capital structure decision is to determine the financial leverage that maximizes the value of the company (or minimizes the weighted average cost of capital).
- In the Modigliani and Miller theory developed without taxes, capital structure is irrelevant and has no effect on company value.
- The deductibility of interest lowers the cost of debt and the cost of capital for the company as a whole. Adding the tax shield provided by debt to the Modigliani and Miller framework suggests that the optimal capital structure is all debt.
- In the Modigliani and Miller propositions with and without taxes, increasing a company's relative use of debt in the capital structure increases the risk for equity providers and, hence, the cost of equity capital.
- When there are bankruptcy costs, a high debt ratio increases the risk of bankruptcy.
- Using more debt in a company's capital structure reduces the net agency costs of equity.
- The costs of asymmetric information increase as more equity is used versus debt, suggesting the pecking order theory of leverage in which new equity issuance is the least preferred method of raising capital.
- According to the static trade-off theory of capital structure, in choosing a capital structure, a company balances the value of the tax benefit from deductibility of interest with the present value of the costs of financial distress. At the optimal target capital structure, the incremental tax shield benefit is exactly offset by the incremental costs of financial distress.
- A company may identify its target capital structure, but its capital structure at any point in time may not be equal to its target for many reasons.
- Many companies have goals for maintaining a certain credit rating, and these goals are influenced by the relative costs of debt financing among the different rating classes.
- In evaluating a company's capital structure, the financial analyst must look at such factors as the capital structure of the company over time, the business risk of the company, the capital structure of competitors that have similar business risk, and company-specific factors (e.g., the quality of corporate governance, which may affect agency costs).

- Good corporate governance and accounting transparency should lower the net agency costs of equity.
- When comparing capital structures of companies in different countries, an analyst must consider a variety of characteristics that might differ and affect both the typical capital structure and the debt maturity structure. The major characteristics fall into three categories: institutional and legal environment, financial markets and banking sector, and macroeconomic environment.

REFERENCES

Antoniou, Antonios, Yilmaz Guney, and Krishna Paudyal. 2008. "The Determinants of Capital Structure: Capital Market-Oriented versus Bank-Oriented Institutions." *Journal of Financial and Quantitative Analysis* 43 (1): 59–92.

Berk, Jonathan, and Peter DeMarzo. 2017. *Corporate Finance*. 4th ed., Pearson.

Chen, Sophia, Paola Ganum, Lucy Liu, Leonardo Martinez, and Soledad Martinez Peria Maria. 2019. "*Debt Maturity and the Use of Short-Term Debt: Evidence from Sovereigns and Firms.*" International Monetary Fund Policy Paper 19/30.

Dang, Tung Lam, Hanh Huynh Thi Hong, Manh Toan Nguyen, and Hue Nguyen Thi Minh. 2017. "The Firm Information Environment and Capital Structure: International Evidence." *Applied Economics* 49 (44): 4482–500.

De Fiore, Fiorella, and Harald Uhlig. 2011. "Bank Finance versus Bond Finance." *Journal of Money, Credit and Banking* 43 (7): 1399–421.

Demirgüç-Kunt, Asli, Maria Soledad Martinez-Peria, and Thierry Tressel. 2015. "*The Impact of the Global Financial Crisis on Firms' Capital Structure.*" World Bank Group Policy Research Working Paper 7522. https://openknowledge.worldbank.org/handle/10986/23623.

Domowitz, Ian, Jack Glen, and Ananth Madhavan. 2001. "International Evidence on Aggregate Corporate Financing Decisions." In *Financial Structure and Economic Growth: A Cross-Country Comparison of Banks, Markets, and Development*, ed. Demirgüç-Kunt, Asli, and Ross Levine. Cambridge, MA: MIT Press.

Faccio, Mara, and Jin Xu. 2015. "Taxes and Capital Structure." *Journal of Financial and Quantitative Analysis* 50 (3): 277–300.

Fan, J.P.H.Sheridan Titman, and Garry J. Twite. 2012. "An International Comparison of Capital Structure and Debt Maturity Choices." *Journal of Financial and Quantitative Analysis* 47 (1): 23–56.

Gao, Wenlian, and Feifei Zhu. 2015. "Information Asymmetry and Capital Structure around the World." *Pacific-Basin Finance Journal* 32 (April): 131–59.

Grjebine, Thomas, Urszula Szczerbowicz, and Fabien Tripier. 2017. "*Corporate Debt Structure and Economic Recoveries.*" Working paper, Banque de France (September). https://www.banque-france.fr/sites/default/files/medias/documents/wp-646.pdf.

Hamada, Robert. 1972. "The Effect of the Firm's Capital Structure on the Systematic Risk of Common Stocks." *Journal of Finance* 27 (2): 435–52.

Harvey, Campbell R., Karl V. Lins, and Andrew H. Roper. 2004. "The Effect of Capital Structure When Expected Agency Costs Are Extreme." *Journal of Financial Economics* 74 (1): 3–30.

Jensen, Michael C. 1986. "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review* 76 (2): 323–29.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Company: Managerial Behavior, Agency Costs, and Ownership Structure." *Journal of Financial Economics* 3 (4): 305–60.

Miller, Merton H. 1977. "Debt and Taxes." *Journal of Finance* 32 (2): 261–75. doi:

Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance, and the Theory of Investment." *American Economic Review* 48 (3): 261–97.

Mokhova, Natalia, and Marek Zinecker. 2014. "Macroeconomic Factors and Corporate Capital Structure." *Procedia: Social and Behavioral Sciences* 110:530–40.

Myers, Stewart, and Nicholas S. Majluf. 1984. "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have." *Journal of Financial Economics* 13:187–221.

Ariss, TurkRima. . 2016. "Legal Systems, Capital Structure, and Debt Maturity in Developing Countries." *Corporate Governance* 24 (2): 130–44.

Vincent, Christopher, and Roni Michaely. 2011. "*Do Institutional Investors Influence Capital Structure Decisions?*" Harvard Law School Forum on Corporate Governance and Financial Regulation (2 December). https://corpgov.law.harvard.edu/2011/12/02/do-institutional-investors-influence-capital-structure-decisions/.

Yu, Bing. 2012. "*Agency Costs of Stakeholders and Capital Structure: International Evidence.*" Managerial Finance, vol. 38, no. 3 (special issue): 303–324.

PRACTICE PROBLEMS

1 If investors have homogeneous expectations, the market is efficient, and there are no taxes, no transaction costs, and no bankruptcy costs, the Modigliani and Miller Proposition I states that:

A bankruptcy risk rises with more leverage.

B managers cannot change the value of the company by using more or less debt.

C managers cannot increase the value of the company by employing tax saving strategies.

2 According to Modigliani and Miller's Proposition II without taxes:

A the capital structure decision has no effect on the cost of equity.

B investment and the capital structure decisions are interdependent.

C the cost of equity increases as the use of debt in the capital structure increases.

3 The current weighted average cost of capital (WACC) for Van der Welde is 10%. The company announced a debt offering that raises the WACC to 13%. The *most likely* conclusion is that for Van der Welde:

A the company's prospects are improving.

B equity financing is cheaper than debt financing.

C the company's debt/equity has moved beyond the optimal range.

4 All else equal, the use of long-maturity debt is expected to be *greater* in those markets in which:

A inflation is low.

B capital markets are passive and illiquid.

C the legal system's protection of bondholders' interests is weak.

5 According to the pecking order theory:

A new debt is preferable to new equity.

B new debt is preferable to internally generated funds.

C new equity is always preferable to other sources of capital.

6 According to the static trade-off theory:

A debt should be used only as a last resort.

B companies have an optimal level of debt.

C the capital structure decision is irrelevant.

The following information relates to Questions 7–9

Barbara Andrade is an equity analyst who covers the entertainment industry for Greengable Capital Partners, a major global asset manager. Greengable owns a significant position with a large unrealized capital gain in Mosely Broadcast Group (MBG). On a recent conference call, MBG's management states that they plan to increase the

© 2020 CFA Institute. All rights reserved.

proportion of debt in the company's capital structure. Andrade is concerned that any changes in MBG's capital structure will negatively affect the value of Greengable's investment.

To evaluate the potential impact of such a capital structure change on Greengable's investment, she gathers the information about MBG given in Exhibit 1.

Exhibit 1 Current Selected Financial Information for MBG

Yield to maturity on debt	8.00%
Market value of debt	$100 million
Number of shares of common stock	10 million
Market price per share of common stock	$30
Cost of capital if all equity-financed	10.3%
Marginal tax rate	35%

Andrade expects that an increase in MBG's financial leverage will increase its costs of debt and equity. Based on an examination of similar companies in MBG's industry, Andrade estimates MBG's cost of debt and cost of equity at various debt-to-total capital ratios, as shown in Exhibit 2.

Exhibit 2 Estimates of MBG's Before-Tax Costs of Debt and Equity

Debt-to-Total Capital Ratio (%)	Cost of Debt (%)	Cost of Equity (%)
20	7.7	12.5
30	8.4	13.0
40	9.3	14.0
50	10.4	16.0

7 MBG is *best* described as currently:

A 25% debt-financed and 75% equity-financed.

B 33% debt-financed and 66% equity-financed.

C 75% debt-financed and 25% equity-financed.

8 Holding operating earnings constant, an increase in the marginal tax rate to 40% would:

A result in a lower cost of debt capital.

B result in a higher cost of debt capital.

C not affect the company's cost of capital.

9 According to the pecking order theory, MBG's announced capital structure change may be optimal:

A because debt is cheaper than equity on an after-tax basis.

B if new equity is issued before new debt as a source of capital.

C if new debt is issued after all internally generated funds are first used as sources of capital.

The following information relates to Questions 10–11

Lindsay White, CFA, is an analyst with a firm in London. She is responsible for covering five companies in the Consumer Staples industry. White believes that over the next two years the domestic and global economies will grow slightly below average, but she is also concerned about the possibility of a mild recession during the same period. She has been asked to review the companies that she covers, and she has collected information about them, presented in Exhibit 1. White has estimated that earnings before interest and taxes (EBIT) will remain constant for all five companies for the foreseeable future. Currency is in terms of the British pound (£). The marginal corporate tax rate is 30% for all five companies.

Exhibit 1 Selected Company Financial Data

	Aquarius	Bema	Garth	Holte	Vega
EBIT (£)	600,000	600,000	400,000	400,000	400,000
Debt-to-equity ratio (market value)	0.60	0.00	0.00	0.71	0.62
Debt (market value) (£)	2,000,000	0	0	2,000,000	2,000,000
S&P debt rating	A+	n.a.	n.a.	A–	A
Weighted average cost of capital	—	10%	10%	—	—

Based on conversations with management of the five companies and her own independent research and analysis, White notes the following:

Aquarius:

- has lower bonding costs than does Bema.
- has a higher percentage of tangible assets to total assets than does Bema.
- has a higher degree of operating leverage than does Bema.

Garth:

- invests significantly less in research and development than does Holte.
- has a more highly developed corporate governance system than does Holte.
- has more business risk than does Holte.

In addition, White has reached various conclusions regarding announcements by Bema, Garth, and Vega:

Announcement	Bema has announced that it will issue debt and use the proceeds to repurchase shares. As a result of this debt-financed share repurchase program, Bema indicates that its debt/equity will increase to 0.6 and its before-tax cost of debt will be 6%.
Conclusion	As a result of the announced program, Bema's total market value should decrease relative to Aquarius's.

Announcement	Garth has announced that it plans to abandon the prior policy of all-equity financing by the issuance of £1 million in debt in order to buy back an equivalent amount of equity. Garth's before-tax cost of debt is 6%.
Conclusion	This change in capital structure is reasonable, but Garth should take care subsequently to maintain a lower D/E than Holte.
Announcement	Vega has announced that it intends to raise capital next year, but it is unsure of the appropriate method of raising capital.
Conclusion	White has concluded that Vega should apply the pecking order theory to determine the appropriate method of raising capital.

10 Based on Exhibit 1 and White's notes, which of the following is *most* consistent with White's conclusion regarding Garth's announcement?

A Garth has more business risk than does Holte.

B Garth invests significantly less in research and development than does Holte.

C Garth has a more highly developed corporate governance system than does Holte.

11 Based on White's third conclusion, regarding determining the appropriate method of raising capital, Vega should raise capital in the following order:

A debt, internal financing, equity.

B equity, debt, internal financing.

C internal financing, debt, equity.

SOLUTIONS

1 B is correct. Proposition I, or the capital structure irrelevance theorem, states that the level of debt versus equity in the capital structure has no effect on company value in perfect markets.

2 C is correct. The cost of equity rises with the use of debt in the capital structure (e.g., with increasing financial leverage).

3 C is correct. If the company's WACC increases as a result of taking on additional debt, the company has moved beyond the optimal capital range. The costs of financial distress may outweigh any tax benefits to the use of debt.

4 A is correct. The use of long-maturity debt is expected to be inversely related to the level of inflation.

5 A is correct. According to the pecking order theory, internally generated funds are preferable to both new equity and new debt. If internal financing is insufficient, managers next prefer new debt, then new equity.

6 B is correct. The static trade-off theory indicates that there is a trade-off between the tax shield from interest on debt and the costs of financial distress, leading to an optimal amount of debt in a company's capital structure.

7 A is correct. The market value of equity is ($30)(10,000,000) = $300,000,000. With the market value of debt equal to $100,000,000, the market value of the company is $100,000,000 + $300,000,000 = $400,000,000. Therefore, the company is $100,000,000/$400,000,000 = 0.25 or 25% debt-financed.

8 A is correct. The after-tax cost of debt decreases as the marginal tax rate increases.

9 C is correct. If internally generated funds have already been fully used, the use of new debt may be optimal according to the pecking order theory of capital structure.

10 A is correct. The statement implies that Garth possesses a lower ability to assume debt than does Holte, all else being equal.

11 C is correct. According to the pecking order theory, managers prefer internal financing. If internal financing is not sufficient, managers next prefer debt, then equity.

READING

16

Analysis of Dividends and Share Repurchases

by Gregory Noronha, PhD, CFA, and George H. Troughton, PhD, CFA

Gregory Noronha, PhD, CFA, is at the University of Washington, Tacoma (USA). George H. Troughton, PhD, CFA (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe the expected effect of regular cash dividends, extra dividends, liquidating dividends, stock dividends, stock splits, and reverse stock splits on shareholders' wealth and a company's financial ratios;
☐	**b.** compare theories of dividend policy and explain implications of each for share value given a description of a corporate dividend action;
☐	**c.** describe types of information (signals) that dividend initiations, increases, decreases, and omissions may convey;
☐	**d.** explain how agency costs may affect a company's payout policy;
☐	**e.** explain factors that affect dividend policy in practice;
☐	**f.** calculate and interpret the effective tax rate on a given currency unit of corporate earnings under double taxation, dividend imputation, and split-rate tax systems;
☐	**g.** compare stable dividend with constant dividend payout ratio, and calculate the dividend under each policy;
☐	**h.** compare share repurchase methods;
☐	**i.** calculate and compare the effect of a share repurchase on earnings per share when 1) the repurchase is financed with the company's surplus cash and 2) the company uses debt to finance the repurchase;
☐	**j.** calculate the effect of a share repurchase on book value per share;
☐	**k.** explain the choice between paying cash dividends and repurchasing shares;

(continued)

© 2020 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**l.** describe broad trends in corporate payout policies;
☐	**m.** calculate and interpret dividend coverage ratios based on 1) net income and 2) free cash flow;
☐	**n.** identify characteristics of companies that may not be able to sustain their cash dividend.

1 DIVIDENDS: FORMS AND EFFECTS ON SHAREHOLDER WEALTH AND FINANCIAL RATIOS

a describe the expected effect of regular cash dividends, extra dividends, liquidating dividends, stock dividends, stock splits, and reverse stock splits on shareholders' wealth and a company's financial ratios;

This reading covers the features and characteristics of dividends and share repurchases as well as the theory and practice of corporate payout policy. A **dividend** is a distribution paid to shareholders. Dividends are declared (i.e., authorized) by a corporation's board of directors, whose actions may require approval by shareholders (e.g., in most of Europe) or may not require such approval (e.g., in the United States). Shares trading **ex-dividend** refers to shares that no longer carry the right to the next dividend payment. The **ex-dividend date** is the first date that a share trades without (i.e., "ex") this right to receive the declared dividend for the period. All else holding constant, on the ex-dividend date the share price can be expected to drop by the amount of the dividend. In contrast to the payment of interest and principal on a bond by its issuer, the payment of dividends is discretionary rather than a legal obligation and may be limited in amount by legal statutes and debt contract provisions. Dividend payments and interest payments in many jurisdictions are subject to different tax treatment at both the corporate and personal levels.

In this reading, we focus on dividends on common shares (as opposed to preferred shares) paid by publicly traded companies. A company's **payout policy** is the set of principles guiding cash dividends and the value of shares repurchased in any given year. Payout policy (also called distribution policy) is more general than dividend policy because it reflects the fact that companies can return cash to shareholders by means of share repurchases and cash dividends. One of the longest running debates in corporate finance concerns the impact of a company's payout policy on common shareholders' wealth. Payout decisions, along with financing (capital structure) decisions, generally involve the board of directors and senior management and are closely watched by investors and analysts.

Dividends and share repurchases concern analysts because, as distributions to shareholders, they affect investment returns and financial ratios. The contribution of dividends to total return for stocks is formidable. For example, the total compound annual return for the S&P 500 Index with dividends reinvested from the beginning of 1926 to the end of 2018 was 10.0%, as compared with 5.9% on the basis of price alone. Similarly, from 1950 to 2018 the total compound annual return for the Nikkei 225 Index with dividends reinvested was 11.1%, as compared with 8.0% on the basis of price alone. Dividends also may provide important information about future company performance and investment returns. Analysts should strive to become familiar with all investment-relevant aspects of dividends and share repurchases.

This reading is organized as follows. Section 1 reviews the features and characteristics of cash dividends, liquidating dividends, stock dividends, stock splits, and reverse stock splits and describes their expected effect on shareholders' wealth and a company's financial ratios. Sections 2–4 present theories of the effects of dividend policy on company value. In Section 5, we discuss factors that affect dividend policy in practice and presents global trends in payout policy. In Section 6, we cover three major types of dividend policies. Sections 7–9 present share repurchases, including their income statement and balance sheet effects and equivalence to cash dividends (under certain assumptions). Section 10 covers analysis of dividend safety. The reading concludes with a summary.

1.2 Dividends: Forms and Effects on Shareholder Wealth and Issuing Company's Financial Ratios

Companies can pay dividends in a number of ways. Cash dividends can be distributed to shareholders through regular, extra (also called special or irregular), or liquidating dividends. Other forms of dividends include stock dividends and stock splits. In this section, we review the different forms that dividends can take and explain their impact on both the shareholder and the issuing company.

1.2.1 *Regular Cash Dividends*

Many companies choose to distribute cash to their shareholders on a regular schedule. The customary frequency of payment, however, may vary among markets. In the United States and Canada, most companies that pay dividends choose a quarterly schedule of payments, whereas in Europe and Japan, the most common choice is to pay dividends twice a year (i.e., semiannually). Elsewhere in Asia, companies often favor paying dividends once a year (i.e., annually). Exhibit 1 summarizes typical dividend payment schedules for selected markets.

Exhibit 1 Geographic Differences in Frequency of Payment of Cash Dividends

Market	Most Common Frequency
Canada, United States	Quarterly
Australia, Japan, Saudi Arabia	Semiannually
Egypt, Germany, Thailand	Annually

Most companies that pay cash dividends strive to maintain or increase their dividends. A record of consistent dividends over a long period of time is important to many companies and shareholders because it is widely interpreted as evidence of consistent profitability. At a minimum, most dividend-paying companies strive not to reduce dividends when they are experiencing temporary problems.

Regular dividends, and especially increasing regular dividends, also signal to investors that their company is growing and will share profits with its shareholders. Perhaps more importantly, management can use dividend announcements to communicate confidence in the company's future. Accordingly, an increase in the regular dividend (especially if it is unexpected) often has a positive effect on share price.

1.2.2 *Extra or Special (Irregular) Dividends*

An **extra dividend** or **special dividend** (also known as an irregular dividend) is either a dividend paid by a company that does not pay dividends on a regular schedule or a dividend that supplements regular cash dividends with an extra payment. These extra dividend payments may be brought about by special circumstances. For example, in December 2018 Hong Kong Stock Exchange (HKEX)-listed Tencent Holdings, a leading provider of internet value-added services, declared a special dividend of HKD250 million to its shareholders after its spin-off Tencent Music went public in New York. This special dividend was approximately 3.5% of Tencent's annual dividend. Like many high-growth technology companies, Tencent had a history of paying very low dividends—with a yield of just 0.26% for 2018 (compared to an average of 4.6% for all stocks listed on the Hong Kong Stock Exchange).

Companies, particularly in cyclical industries, have sometimes chosen to use special dividends as a means of distributing more earnings only during strong earnings years. During economic downturns, when earnings are low or negative, cash that might otherwise be used for dividends is conserved. For example, a company may choose to declare a small regular dividend, and then when operating results are good, it may declare an extra dividend at the end of the year. In May 2018, Mumbai-listed Ingersoll-Rand (India) Ltd, a diversified industrial manufacturer, declared a special "second interim" dividend of Rs202 in addition to the regular annual Rs6 dividend, whereas for the prior 2 decades, the company had paid only the regular Rs6 dividend (excepting a special 2011 Rs24 dividend). The 2018 second interim dividend was paid out of current year profits and accumulated surpluses from earlier years. At the time, the company's reported year-on-year net profit growth was 25%.

Example 1 concerns a hypothetical company with a stated **dividend policy**—the strategy a company follows to determine the amount and timing of dividend payments—regarding the payment of extra dividends. In the example, the **dividend payout ratio** refers to common share cash dividends divided by net income available to common shares over the same time period.

EXAMPLE 1

AfriSage Technologies' Dividend Policy

AfriSage Technologies (AST), a hypothetical company, is a leading provider of commercial and enterprise software solutions in Southern African Development Community (SADC) countries. AST's financial data are reported in South African Rand (ZAR). In November 2017, AfriSage's board of directors modified its dividend policy, stating:

> The company will target an investment-grade, long-term credit rating to secure strategic financial flexibility for investments in future growth. The ordinary dividend shall be at least 35% of net income. Excess capital will be returned to shareholders after the board has taken into consideration the company's cash at hand, projected cash flow, and planned investment from a medium-term perspective as well as capital market conditions.

Selected AfriSage Financial per Share Data

	2018	2017
Shares outstanding	632.5 million	632.5 million
Earnings per share	ZAR14.23	ZAR12.65
Cash dividends per share	ZAR7.61	ZAR10.68

1. Calculate the cash dividend payout ratio for 2018 and 2017.
2. Assuming the board's new dividend policy became effective in 2018, calculate the amount of the annual ordinary dividend on the basis of AfriSage's minimum payout policy in 2018 and the amount that could be considered an extra dividend.

Solution to 1:

With the same number of shares outstanding, the dividend payout ratio on a per share basis is dividends per share divided by earnings per share.

For 2018: ZAR7.61/ZAR14.23 = 53.5%.

For 2017: ZAR10.68/ZAR12.65 = 84.4%.

Solution to 2:

Under a policy of 35% of earnings, the minimum amount of dividends would be ZAR14.23 × 0.35 = ZAR4.98. The amount of the extra dividend would then be ZAR7.61 − ZAR4.98 = ZAR2.63.

1.2.3 *Liquidating Dividends*

A dividend may be referred to as a **liquidating dividend** when a company:

- goes out of business and the net assets of the company (after all liabilities have been paid) are distributed to shareholders;
- sells a portion of its business for cash and the proceeds are distributed to shareholders; or
- pays a dividend that exceeds its accumulated retained earnings (impairs stated capital).

These points illustrate that a liquidating dividend is a return of capital rather than a distribution from earnings or retained earnings.

1.2.4 *Stock Dividends*

Stock dividends are a non-cash form of dividends. With a **stock dividend** (also known as a **bonus issue of shares** or a scrip dividend), the company distributes additional shares (typically 2–10% of the shares then outstanding) of its common stock to shareholders instead of cash. Although the shareholder's total cost basis remains the same, the cost per share held is reduced. For example, if a shareholder owns 100 shares with a purchase price of US$10 per share, the total cost basis would be US$1,000. After a 5% stock dividend, the shareholder would own 105 shares of stock at a total cost of US$1,000. However, the cost per share would decline to US$9.52 (US$1,000/105).

Superficially, the stock dividend might seem an improvement on the cash dividend from both the shareholders' and the company's point of view. Each shareholder ends up with more shares, which did not have to be paid for, and the company did not have to spend any actual money issuing a dividend. Furthermore, stock dividends are generally not taxable to shareholders because a stock dividend merely divides the "pie" (the market value of shareholders' equity) into smaller pieces. The stock dividend, however, does not affect the shareholder's proportionate ownership in the company because other shareholders receive the same proportionate increase in shares. Additionally, the stock dividend does not change the value of each shareholder's ownership position because the increase in the number of shares held is accompanied by an offsetting decrease in earnings per share, and other measures of value per share, resulting from the greater number of shares outstanding.

The second point is illustrated in Exhibit 2, which shows the impact of a 3% stock dividend to a shareholder who owns 10% of a company with a market value of US$20 million. As one can see, the market value of the shareholder's wealth does not change, assuming an unchanged **price-to-earnings ratio** (the ratio of share price, P, to earnings per share, E, or P/E). That assumption is reasonable because a stock dividend does not alter a company's asset base or earning power. (As the reader will see shortly, the same is true of a stock split.) The total market value of the company is unaffected by the stock dividend because the decrease in the share price is exactly offset by the increase in the number of shares outstanding.

Exhibit 2 Illustration of the Effect of a Stock Dividend

	Before Dividend	After Dividend
Shares outstanding	1,000,000	1,030,000
Earnings per share	US$1.00	US$0.97 (1,000,000/1,030,000)
Stock price	US$20.00	US$19.4175 (20 × 0.9709)
P/E	20	20
Total market value	US$20 million	US$20 million (1,030,000 × US$19.4175)
Shares owned	100,000 (10% × 1,000,000)	103,000 (10% × 1,030,000)
Ownership value	US$2,000,000 (100,000 × US$20)	US$2,000,000 (103,000 × US$19.4175)

Note: The exhibit shows intermediate results rounded to four decimal places, but final results are based on carrying intermediate results at full precision.

Companies that regularly pay stock dividends see some advantages to this form of dividend payment. It favors long-term investors, which, in turn, may lower the company's cost of equity financing. The payment of a stock dividend also helps increase the stock's float, which improves the liquidity of the shares and dampens share price volatility.

A traditional belief is that a lower stock price will attract more investors, all else equal. US companies often view the optimal share price range as US$20 to US$80. For a growing company, payment of a regular stock dividend is more likely to help keep the stock in the "optimal" range. In February 2019, for example, Massmart—the second-largest distributor of consumer goods in Africa—changed its established policy of paying interim and final dividends in cash and instead declared a scrip dividend for the 2018 final dividend. When the company pays the same dividend rate on the new shares as it did on the old shares, a shareholder's dividend income increases; however, the company could have accomplished the same result by increasing the cash dividend.

From a company's perspective, the key difference between a stock dividend and a cash dividend is that a cash dividend affects a company's capital structure, whereas a stock dividend has no economic impact on a company. Cash dividends reduce assets (because cash is being paid out) and shareholders' equity (by reducing retained earnings). All else equal, liquidity ratios, such as the cash ratio (cash and short-term marketable securities divided by current liabilities) and current ratio (current assets divided by current liabilities), should decrease, reflecting the reduction in cash. Financial leverage ratios, such as the debt-to-equity ratio (total debt divided by total shareholders' equity) and debt-to-assets ratio (total debt divided by total assets), should also increase. Stock dividends, on the other hand, do not affect assets or shareholders' equity. Although retained earnings are reduced by the value of the stock dividends paid (i.e., by the number of shares issued × price per share), contributed capital increases by the same amount (i.e., the value of the shares issued). As a result, total shareholders' equity does not change. Neither stock dividends nor stock splits (which are discussed in the next section) affect liquidity ratios or financial leverage ratios.

1.2.5 *Stock Splits*

Stock splits are similar to stock dividends in that they have no economic effect on the company, and the shareholders' total cost basis does not change. For example, if a company announces a two-for-one stock split, each shareholder will be issued an additional share for each share currently owned. Thus, a shareholder will have twice as many shares after the split as before the split. Therefore, earnings per share (and all other per share data) will decline by half, leaving the P/E and equity market value unchanged. Assuming the corporation maintains the same dividend payout ratio as before the split, **dividend yield** (annual dividends per share divided by share price) will also be unchanged. Apart from the effect of any information or benefit that investors perceive a stock split to convey, stock splits (like stock dividends) should be neutral in their effect on shareholders' wealth.

Although two-for-one and three-for-one stock splits are the most common, such unusual splits as five-for-four or seven-for-three sometimes occur. It is important for shareholders to recognize that their wealth is not changed by the stock split (just as it was not changed for a stock dividend, all else equal). Exhibit 3 shows an example of a two-for-one split and its impact on stock price, earnings per share, dividends per share, dividend payout ratio, dividend yield, P/E, and market value.

Exhibit 3 Before and After a Two-for-One Stock Split

	Before Split	After Split
Number of shares outstanding	4 million	8 million
Stock price	€40.00	€20.00 (€40/2)
Earnings per share	€1.50	€0.75 (€1.50/2)
Dividends per share	€0.50	€0.25 (€0.50/2)
Dividend payout ratio	1/3	1/3
Dividend yield	1.25%	1.25% (€0.25/€20.00)
P/E	26.7	26.7 (€20.00/€0.75)
Market value of equity	€160 million	€160 million (€20.00 × 8 million)

As can be seen, a two-for-one stock split is basically the same as a 100% stock dividend because all per share data have been reduced by 50%. The only difference is in the accounting treatment: Although both stock dividends and stock splits have no effect on total shareholders' equity, a stock dividend is accounted for as a transfer of retained earnings to contributed capital. A stock split, however, does not affect any of the balances in shareholder equity accounts.

A company may announce a stock split at any time. Typically, a split is announced after a period in which the stock price has risen. Many investors view the announcement of a stock split as a positive sign pointing to future stock price increases. More often, however, announced stock splits merely recognize that the stock has risen enough to justify a stock split to return the stock price to a lower, more marketable price range.

Several of the largest companies in the world (as measured by market value) had stock splits in the last decade. For example, Schneider Electric SA (France) had a two-for-one split in 2011; Whole Foods Market (United States) had a two-for-one split in 2013. In each case, the stock split came after a significant rise in stock price but was not, in and of itself, a meaningful predictor of future price action. However, data show that stock splits have been on the decline in the United States. Although S&P 500 constituent stock splits averaged 45 per year between 1980 and 2017, they reached the maximum of 114 splits in 1986 and have steadily declined since 2015 (e.g., only 5 splits in 2017). This decline in stock splits has been attributed to greater use of funds and exchange-traded funds (ETFs) by individual investors and to changes in market microstructure that have de-linked such transaction costs as commissions paid to number of shares traded. Thus, the concept of a "marketable price range" of a company's stock has become less important.

Much less common than stock splits are reverse stock splits. A **reverse stock split** increases the share price and reduces the number of shares outstanding—again, with no effect on the market value of a company's equity or on shareholders' total cost basis. Just as a high stock price might lead a company to consider a stock split, so too a low stock price may lead a company to consider a reverse stock split. The objective of a reverse stock split is to increase the price of the stock to a higher, more marketable range. As reported in *Barron's*, companies execute reverse splits "to attract institutional investors and mutual funds that often shy from buying stocks trading below US$5." Reverse stock splits are perhaps most common for companies in, or coming out of, financial distress. Kitov Pharma, an Israeli drug developer, announced a 1-for-20 reverse split in December 2018, reducing its issued shares to 16 million, in order to meet minimum share price listing criteria to begin trading on the Tel Aviv Stock Exchange and to begin the trading of its ADRs on the NASDAQ in January 2019.

Reverse splits, historically less common in Asia, are becoming more popular. For example, reverse stock splits were not permitted in Japan under Corporation Law until 2001, but since 2007, they have been actively encouraged by the Tokyo Stock Exchange to meet the Exchange's objective of standardizing trading lot size to 100 shares for listed companies by 1 October 2018. While most companies were compliant by the deadline, on that date 23 companies reduced their trading lot size to 100 shares by carrying out reverse stock splits. As an example, in May 2018 Fuji Electric Co. Ltd announced that it would conduct a 5-for-1 reverse stock split on 1 October 2018 to adjust the unit of investment in the company to a level deemed desirable by the TSE (between ¥50,000 and ¥500,000).

EXAMPLE 2

Globus Maritime Announces a Reverse Split

In May 2018, Globus Maritime Ltd, a Greek dry bulk shipping company providing worldwide maritime transportation services, was warned by NASDAQ that it no longer met the continuing listing requirements once its share price had traded below the US$1 a share minimum price requirement for 30 consecutive business days. Globus was given until the end of October 2018 to regain compliance. Globus announced a 1 for 10 reverse split to occur on 15 October. On 12 October, shares were trading at US$4.25 before the reverse split had taken place.

1 If the reverse split were to take place when the share price was US$4.25, find the expected stock price after a 1-for-10 reverse split, assuming no other factors affect the split.
2 Comment on the following statement: “Shareholder wealth is negatively affected by a reverse stock split.”

Solution to 1:

If the price was US$4.25 before the reverse split, for every 10 shares, a shareholder would have 1 share priced at 10 × US$4.25 = US$42.50.

Solution to 2:

The statement is not generally correct. Considering the reverse split on its own, the market capitalization of the common equity would be unchanged. If the reverse split was interpreted as a good decision (e.g., because the company will be able to retain the advantages of being listed on the NASDAQ), its price and thus market capitalization might increase. But other factors—such as continued limited growth of its operations or continued small share float and turnover—could drive down the stock’s value.

DIVIDEND POLICY AND COMPANY VALUE: THEORY 2

b compare theories of dividend policy and explain implications of each for share value given a description of a corporate dividend action;

Since the early 1960s, financial theorists have debated the extent to which dividend policy (decisions about whether, when, and in what amount to pay dividends) should and does matter to a company’s shareholders. One group of theorists believes that dividend policy is irrelevant to shareholders. This group typically holds that only the decisions of the company that are directly related to investment in working and fixed capital affect shareholders’ wealth. A second group holds that dividend policy does matter to investors, for one or more reasons, and that a company can affect shareholders’ wealth through its dividend policy. Typically, dividend relevance is attributed to either the belief that investors value a unit of dividends more highly than an equal amount of uncertain capital gains or to one or more market imperfections. Such imperfections include taxes (because dividends may be taxed differently than capital gains), asymmetric information (corporate insiders are better informed about their company’s prospects than outside investors), and agency costs (management has a tendency to squander extra cash). We examine these positions and the assumptions that underlie them in the following subsections.

2.1 Dividend Policy Does Not Matter

In a 1961 paper, Miller and Modigliani (“MM”) argued that in a world without taxes, transaction costs, and equal (“symmetric”) information among all investors—that is, under **perfect capital market** assumptions—a company’s dividend policy should have no impact on its cost of capital or on shareholder wealth. Their argument begins by assuming a company has a given capital budget (e.g., it accepts all projects with a positive net present value, or NPV) and that its current capital structure and debt ratio are optimal. Another way of stating this argument is that the dividend decision is independent of a company's investment and financing decisions. For example, suppose that an all-equity financed company decided to pay as a dividend the investment amount it required for its capital budget. To finance capital projects, the company could issue additional common shares in the amount of its capital budget (such financing would leave its capital structure unchanged). The value of the newly issued shares would exactly offset the value of the dividend. Thus, if a company paid out a dividend that represented 5% of equity, its share price would be expected to drop by 5%. If a common stock in Australia is priced at A$20 before an A$1 per share dividend, the implied new price would be A$19. The shareholder has assets worth A$20 if the dividend is not paid or assets worth A$20 if the stock drops to A$19 and an A$1 dividend is paid.

Note that under the MM assumptions, there is no meaningful distinction between dividends and share repurchases (repurchases of outstanding common shares by the issuing company): They are both ways for a company to return cash to shareholders. If a company had few investment opportunities such that its current cash flow was more than that needed for positive NPV projects, it could distribute the excess cash flow via a dividend or a share repurchase. Shareholders selling shares would receive A$20 a share, and shareholders not selling would hold shares whose value continued to be A$20. To see this, suppose the company being discussed has 10,000 shares outstanding, a current free cash flow of A$10,000, and a present value of future cash flows of A$190,000. Thus, the share price is (A$10,000 + A$190,000)/10,000 = A$20. Now if the company uses the free cash flow to repurchase shares, in lieu of paying a dividend of A$1, it will repurchase 500 shares (A$10,000/A$20 = 500). The 9,500 shares left outstanding have a claim on the A$190,000 future cash flow, which results in a share price of A$20 (A$190,000/9,500 = A$20).

An intuitive understanding of MM dividend irrelevance also follows from the concept of a "homemade dividend." In a world with no taxes or transaction costs, if shareholders wanted or needed income, they could construct their own dividend policy by selling sufficient shares to create their desired cash flow stream. Using the example above, assume the company did not pay the A$1 dividend and the stock remained at A$20. A holder of 1,000 shares who desired A$1,000 in cash could sell 50 shares at A$20, thus reducing his or her holdings to 950 shares. Note that by reducing share holdings, second-period dividend income is reduced; higher dividend income in one period is at the expense of exactly offsetting lower dividend income in subsequent periods. The irrelevance argument does not state that dividends per se are irrelevant to share value but that dividend *policy* is irrelevant. By taking the earning power of assets as a given and assuming perfect capital markets, policy alternatives merely involve tradeoffs of different dividend streams of equal present value.

In the real world, market imperfections create some problems for MM's dividend policy irrelevance propositions. First, both companies and individuals incur transaction costs. A company issuing new shares incurs **flotation costs** (i.e., costs in selling shares to the public that include underwriters' fees, legal costs, registration expenses, and possible negative price effects) often estimated to be as much as 4% to 10% of the capital raised, depending on the size of the company and the size of the issue. Shareholders selling shares to create a "homemade" dividend would incur transaction costs and, in some countries, capital gains taxes (of course, cash dividends incur taxes in most countries). Furthermore, selling shares on a periodic basis to create an income stream of dividends can be problematic over time if share prices are volatile. If share prices decline, shareholders have to sell more shares to create the same dividend stream.

2.2 Dividend Policy Matters: The Bird in the Hand Argument

Financial theorists have argued that, even under perfect capital markets assumptions, investors prefer a dollar of dividends to a dollar of potential capital gains from reinvesting earnings because they view dividends as less risky. A related viewpoint is that "the typical dollar of reinvestment has less economic value to the shareholder than a dollar paid in dividends" (Graham, Dodd, Cottle, and Tatham 1962). These arguments are similar and have sometimes been called the "bird in the hand" argument, a reference to the proverb "a bird in the hand is worth two in the bush." By assuming that a given amount of dividends is less risky than the same amount of capital gains, the argument is that a company that pays dividends will have a lower cost of equity capital than an otherwise similar company that does not pay dividends; the lower cost of equity should result in a higher share price. MM contend that this argument is incorrect because, under their assumptions, paying or increasing the dividend today does not affect the risk of future cash flows. Such actions only lower the ex-dividend price of the share.

2.3 Dividend Policy Matters: The Tax Argument

In some countries, dividend income has traditionally been taxed at higher rates than capital gains. In the United States since 2012, for instance, dividends on shares held at least 60 days have been taxed at a maximum rate of 20%, which exceeds the long-term capital gains tax rate of 15%. In mainland China, there is no capital gains tax on shares; however, dividend income is taxed at 20% for shares held less than a month, 10% for shares held between one month and a year, and since 2015 at 0% for shares held longer than a year.

An argument could be made that in a country that taxes dividends at higher rates than capital gains, taxable investors should prefer companies that pay low dividends and reinvest earnings in profitable growth opportunities. Presumably, any growth in earnings in excess of the opportunity cost of funds would translate into a higher share price. If, for any reason, a company lacked growth opportunities sufficient to consume its annual retained earnings, it could distribute such funds through share repurchases (again, the assumption is that capital gains are taxed more lightly than dividends). Taken to its extreme, this argument would advocate a *zero* dividend payout ratio. Real world market considerations may complicate the picture. For example, in some jurisdictions governmental regulation may require companies to distribute excess earnings as dividends or to classify share repurchases as dividends if the repurchases appear to be ongoing in lieu of dividend payments.

OTHER THEORETICAL ISSUES: CLIENTELE EFFECT AND THE INFORMATION CONTENT OF DIVIDEND ACTIONS: SIGNALING

3

c describe types of information (signals) that dividend initiations, increases, decreases, and omissions may convey;

d explain how agency costs may affect a company's payout policy;

In the following section, we present additional perspectives related to the theory of dividend policy.

3.1 The Information Content of Dividend Actions: Signaling

MM assumed that all investors—including outside investors—have the same information about the company: a situation of symmetric information. In reality, corporate managers typically have access to more detailed and extensive information about the company than do outside investors.

A situation of asymmetric information raises the possibility that dividend increases or decreases may affect share price because they may convey new information about the company. A company's board of directors and management, having more information than outside investors, may use dividends to signal to investors about (i.e., convey information on) the company's prospects. A company's decision to initiate, maintain, increase, or cut a dividend may convey more credible information than positive words from management because cash is involved. For a signal to be effective, it must be difficult or costly to mimic by another entity without the same attributes. Dividend increases are costly to mimic because a company that does not expect its cash flows to increase will not be able to maintain the dividend at increasingly high levels in the long run. (In the short run, a company may be able to borrow to fund dividends.)

Empirical studies broadly support the thesis that dividend initiations or increases convey positive information and are associated with future earnings growth, whereas dividend omissions or reductions convey negative information and are associated with future earnings problems. A dividend declaration can help resolve some of the information asymmetry between insiders and outsiders and help close any gap between the market price of shares and their intrinsic value. Evidence in both developed and emerging market equities suggests the presence of an earnings and return effect following dividend initiation announcements. In general, company earnings increase in the year of dividend initiation and in the following several years, and then the announcement of the initiation of a regular cash dividend is accompanied by an excess return. By looking at two historical examples of signaling, Example 3 provides further support for the idea that dividend initiations contain value-relevant information.

EXAMPLE 3

Historical Examples: Information on Dividend Initiations

Following are two examples of the information content of dividend initiations following the 2008 global financial crisis.

A Oracle Corporation, a leading business software maker, initiated a US$0.05 quarterly dividend in May 2009. Oracle's annual US$0.20 dividend amounts to about US$1 billion, a relatively small amount compared with operating cash flow of US$8 billion and another US$9 billion in cash and cash-equivalent assets on its balance sheet at the end of fiscal year 2009. An analyst who follows Oracle for institutional investors saw the Oracle announcement as a signal that the company was well positioned to ride out the downturn and also gain market share.

B In mid-2009, Paris-based Groupe Eurotunnel announced its first ever dividend after it completed a debt restructuring and received insurance proceeds resulting from a fire that had closed the Channel Tunnel. In a 2 June 2009 press release, Eurotunnel's CEO said that this "marked a turning point for the company as its business has returned to the realm of normality"; the company anticipated a return to profitability.

Some researchers have argued that a company's dividend initiation or increase tends to be associated with share price increases because it attracts more attention to the company. Managers have an incentive to increase the company's dividend if they believe the company to be undervalued because the increased scrutiny will lead to a positive price adjustment. In contrast, according to this line of reasoning, managers of overvalued companies have little reason to mimic such a signal because increased scrutiny would presumably result in a downward price adjustment to their shares.

EXAMPLE 4

Signaling with Dividends and the Costs of Mimicking

Suppose that the management of a company with poor future prospects recommends to the board of directors an increase in its dividend. Management explains to the board that investors may then believe that the company has positive future prospects, leading to an increase in share value and shareholder wealth.

1 State whether such imitation is likely to achieve the stated objective over the long term.
2 Justify your answer to Question 1.

Solution to 1:

No, such dividend increases are not likely to achieve the stated objective over the long term for the company described.

Solution to 2:

Dividend increases are costly to mimic because a company that does not expect its cash flows to increase will not be able to maintain the increased dividend. The company will have to either cut the dividend in the future or go to the market to obtain new equity or debt funding to pay the dividend. Both these alternatives are costly for the company because they result in downward revisions, on average, to the stock price.

Many companies take pride in their record of consistently increasing dividends over a long period of time. Standard & Poor's, for example, identifies companies in its US-based S&P 500 Index, Europe 350 Index, Pan Asia Index, and S&P/TSX Canadian Index that have increased their dividend for a number of consecutive years (at least 25 years in the case of the S&P 500, at least 10 years in the case of the Europe 350, at least 7 years in the case of Pan Asia Index, and at least 5 years in the case of the S&P/TSX). These companies are in various industries. When a company's earnings and cash flow outlook has been and continues to be positive, it often views a policy of increasing dividends as an important tool to convey that information to existing and potential shareholders. Companies that consistently increase their dividends seem to share certain characteristics:

- Dominant or niche positions in their industry
- Global operations
- Relatively less volatile earnings
- Relatively high returns on assets
- Relatively low debt ratios (dividend payouts unlikely to be affected by restrictions in debt covenants)

Dividend cuts or omissions, in contrast, present powerful and often negative signals. For companies under financial or operating stress, the dividend declaration date may be viewed with more than usual interest. Will they cut the dividend? Will they omit the dividend altogether? In these instances, merely maintaining the dividend or not cutting it as much as expected is usually viewed as good news (i.e., that current difficulties are transitory and manageable), unless investors view managers as trying to convey erroneous information to the market.

In principle, although difficult in practice, management can attempt to send a positive signal by cutting the dividend. Telstra, a major Australian telecoms company with an enviable record of paying close to 90% of profits as dividends, announced in 2017 a 30% cut in its dividend—its first cut in more than 20 years. Telstra's management explained it intended to use the funds conserved to reinvest in the business. It was planning for the longer term and retaining financial flexibility as a priority because the company faced significant challenges from rising competition and competing technologies. Although management's message was met with an initial 12% share price decline as disappointed yield-focused investors exited the stock, it was, in retrospect, a positive signal. Telstra was viewed by institutional investors as successfully using its cash flow to reorganize to meet business challenges, and it was regarded as one of the few cases in which a large Australian dividend payer was not cutting payouts as a result of extreme financial pressure.

EXAMPLE 5

Dividend Reductions and Price Increases

In November 2018, BT Group Plc, one of the world's largest providers of communication services and solutions operating in over 170 countries, announced it would cut its interim dividend from 4.85 pence a share to 4.62 pence a share. The company also revealed that net cash flow from operating activities had plunged 71% to £754 million and that revenue had fallen 2% to £11.6 billion, with declines across all divisions.

All this despite the fact that in the first six months of the year, the company reported a pretax profit increase to £1.3 billion from £1.1 billion a year prior and a 2% increase in adjusted earnings (EBITDA) to £3.7 billion from £3.6 billion as the telecoms giant cut costs as part of its restructuring. One analyst commented that while the dividend decrease was an "unwelcome surprise," it was also a "prudent move" given the 71% decline in net cash and thus "should not take too much sheen on a dividend yield, which previously stood at an attractive 6.4%." It was also noted that BT Group was replacing its chief executive in February 2019; thus, future dividends would depend on decisions made by the new leadership. As the market digested this information, the telecoms company's share price rose 6.9% to 257 pence per share.

Source: Renae Dyer, "BT Shares Surge Despite Dividend Cut as It Expects Earnings to Hit Top End of Guidance," Proactive Investors (1 November 2018).

Another example of the signaling content of dividends can be found in the actions of eBay, the e-commerce multinational corporation, and its initial dividend declaration in 2019 (24 years after the online retailer was established in 1995 during the dot-com boom). Technology companies have among the lowest dividend yields and below-average dividend payout ratios. This is because most technology companies have high R&D requirements, and some (e.g., integrated circuit manufacturers) are capital intensive. Those that are profitable often achieve returns on assets and owners' equity that are well above average. In addition, business risk is considerable as discoveries and unforeseen advances change the product landscape. All of these considerations would suggest a policy of low (or no) dividend payments so that internally generated funds are directed toward new product development and capital investment that will maintain high growth and returns. Some companies in the technology sector, however, do mature. Legacy tech companies that initiated dividends as their businesses matured and growth slowed include Apple in 2012, Cisco in 2011, Oracle in 2009, and Microsoft in 2003. At the time of eBay's dividend initiation, such non-dividend-paying tech companies as Alibaba, Weibo, Baidu, and JD.com remained the norm in markets where the technology sector was still growing.

In early 2019, eBay declared its first-ever dividend and announced that it would begin paying quarterly dividends of US$0.14 a share, which represented a yield of 1.6% (for comparison, Microsoft's dividend yield at the time was 1.9% and Cisco's was 2.9%). At the same time, eBay announced an increase in its existing share repurchase program to US$4 billion. Investor reaction was mixed. Some believed that eBay was signaling an interest in broadening its investor focus by attracting a new group of shareholders focused on income over growth while refraining from undertaking unprofitable expansion. Others viewed the dividend declaration as an admission that it was becoming a mature company—that it could no longer deliver high returns from reinvesting its earnings. The future growth prospects for the stock, they would argue, had been diminished. In other words, although the dividend initiation showed confidence in eBay's cash flow generation, investors preferred for management's use of internal investments to regenerate eBay's core business. Regardless, few could argue that eBay's dividend initiation declaration in 2019 was not a corporate event of some importance.

3.2 Agency Costs and Dividends as a Mechanism to Control Them

Large, publicly traded corporations typically have a substantial separation between the professional managers who control the corporation's operations and the outside investors who own it. When agents (the managers) and owners (the shareholders) are two separate parties, managers may have an incentive to maximize their own welfare at the company's expense because they own none or relatively small percentages of the company for which they work and thus do not bear all the costs of such actions. This incentive is ultimately also a problem of unequal (asymmetric) information between managers and outside investors because if outside investors could perfectly observe managers, managers would be dissuaded from such actions. One managerial incentive of particular concern is the potential private benefit managers may obtain from investment in negative net present value (NPV) projects. Such projects will generate negative economic returns; but because they may grow the size of the company (measured in sales or assets) and thus enlarge the manager's span of control, the manager may have the incentive to invest in them. This is a particular problem when management's compensation is tied to assets or sales rather than value enhancement, a flaw in the firm's corporate governance. The potential overinvestment agency problem might be alleviated by the payment of dividends. In particular, by paying out all free cash flow to equity in dividends, managers would be constrained in their ability to overinvest by taking on negative NPV projects. This concern or hypothesis that management may create an overinvestment agency cost is known as Jensen's free cash flow hypothesis.

The potential for managers to squander free cash flow by undertaking unprofitable projects is a consideration to be evaluated on a case-by-case basis. Prior to initiating its dividend in 2003, for example, Microsoft accumulated increasingly large cash positions but was not observed to squander monies on unprofitable projects. In some cases, such cash positions may provide financial flexibility to respond quickly to changes in the environment, to grasp unforeseen opportunities, or to survive periods of restricted credit, as in the case of Ford Motor Company's accumulation of cash during profitable years in the 1990s and similarly by Japanese automotive parts manufacturer Denso Corporation in the late 2000s and 2010s. Clearly, there are industry-specific and life-cycle conditions to consider. In general, it makes sense for growing companies in industries characterized by rapid change to hold cash and pay low or no dividends, but it does not make sense for large, mature companies in relatively non-cyclical industries. In general, there is empirical support for the market reaction to dividend change announcements to be stronger for companies with greater potential for overinvestment than for companies with lesser potential for overinvestment.

Another concern when a company is financed by debt as well as equity is that paying dividends can exacerbate the agency conflict between shareholders and bondholders. When a company has debt outstanding, the payment of dividends reduces the cash cushion available to the company for the disbursement of fixed required payments to bondholders. The payment of large dividends, with the intention of transferring wealth from bondholders to shareholders, could lead to underinvestment in profitable projects. All else equal, both dividends and share repurchases increase the default risk of debt. Reflecting bondholders' concern, the bond **indenture** (contract) often includes a covenant restricting distributions to shareholders that might impair the position of bondholders. A typical form of this restriction is to define the maximum allowable amount of distributions to shareholders during the life of the bond. This amount of funds is usually a positive function of the company's current and past earnings and issues of new equity and a negative function of dividends paid since the bonds were issued. Such covenants often do not really restrict the level of dividends as long as those dividends come from new earnings or from new issues of stock. What the covenant attempts to do is prevent the payment of dividends financed by the sale of the company's existing assets or by the issuance of new debt. Covenants that specify minimum levels of EBITDA and/or EBIT coverage of interest charges are frequently used as well. These covenants provide some assurance that operating earnings include a cushion for the payment of fixed charges. Other covenants focus on balance sheet strength—for example, by specifying a maximum value for the ratio of debt to tangible net worth.

EXAMPLE 6

Agency Issues and Dividends

Two dividend-paying companies A and B directly compete with each other. Both companies are all-equity financed and have recent dividend payout ratios averaging 35%. The corporate governance practices at Company B are weaker than at Company A. For example, at B but not A, the chief executive officer is also chair of the board of directors. Recently, profitable investment opportunities for B have become fewer, although operating cash flow for both A and B is strong.

Based only on the information given, investors who own shares in both A and B are *most likely* to press for a dividend increase at:

A Company A, because it has better growth prospects than Company B.

B Company B, because a dividend increase may mitigate potential overinvestment agency problems.

C Company B, because a dividend increase may mitigate potential underinvestment agency problems.

Solution:

B is correct. Company B's strong operating cash flow in an environment of fewer profitable growth opportunities may tempt Company B's management to overinvest. The concern is increased because of Company B's relatively weak corporate governance.

The final example in this section illustrates the complex agency considerations that may affect dividend policy.

EXAMPLE 7

Electric Utilities, Agency Costs, and Dividends

Electric utilities often have above average dividend yields. A distinctive characteristic of many utility companies is that they pay a high percentage of earnings as dividends, while periodically issuing new equity to invest in the many projects necessitated by the capital-intensive nature of their business. This practice of financing dividends with new equity appears unwise because new equity is expensive. Researchers examining a set of US-based electric utilities, however, have demonstrated that there may be a good reason for paying dividends and then issuing equity: the mitigation of the agency problems between managers and shareholders and between utility regulators and utility shareholders.

Because electric utilities are typically monopolies in the sense that they are usually the only providers of electricity in a given area, they are regulated so they are not able to set electricity rates at monopolistically high levels. The regulators are expected to set rates such that the company's operating expenses are met and investors are provided with a fair return. The regulators, however, are usually elected, or are political appointees, and view ratepayers as potential voters. Thus, utility shareholders, in addition to facing potential manager–shareholder agency issues because managers have incentives to consume perquisites or to overinvest, also face a regulator–shareholder conflict in which regulators set rates low to attract the votes of individuals being served by the utility.

In the utility industry, therefore, dividends and the subsequent equity issue are used as mechanisms to monitor managers and regulators. The company pays high dividends and then goes to the capital markets to issue new equity. If the market does not think that shareholders are getting a fair return because regulators are setting rates too low, or because managers are consuming too many perks, the price at which new equity can be sold will fall until the shareholder expectations for returns are met. As a result, the company may not be able to raise sufficient funds to expand its plant to meet increasing electricity demand—the electric utility industry is very capital intensive—and, in the extreme, customer needs may not be met. Faced with this possibility, and potentially angry voters, regulators have incentives to set rates at a fair level. Thus, the equity market serves to monitor and arbitrate conflicts between shareholders and both managers and regulators.

4 OTHER THEORETICAL ISSUES: AGENCY COSTS AND DIVIDENDS AS A MECHANISM TO CONTROL THEM AND DIVIDEND THEORY: SUMMARY

d explain how agency costs may affect a company's payout policy;

What can we conclude about the link between dividends and valuation? In theory, in the absence of market imperfections Miller and Modigliani (1961) find that dividend policy is irrelevant to the wealth of a company's investors. But in reality, the existence of market imperfections makes matters more complicated. In addition, some investors are led, by logic or custom, to prefer dividends.

Unfortunately, in the search for the link between dividend policy and value, the evidence is inconclusive. It is difficult to show an exact relationship between dividends and value because so many variables affect value. We have presented factors that would seem to explain why some companies put emphasis on dividends and others do not. Financial theory predicts that reinvestment opportunities should be the dominant factor. Indeed, no matter where they are located in the world, small, fast-growing companies pay out little or none of their earnings. Regardless of jurisdiction, more mature companies with fewer reinvestment opportunities tend to pay dividends. For these mature companies, taxes, regulations/laws, tradition, signaling, ownership structure, and attempts to reconcile agency conflicts all seem to play a role in determining the dividend payout ratio. At a minimum, in looking at a company an analyst should evaluate whether a given company's dividend policy matches its reinvestment opportunities and legal/financial environment.

5 FACTORS AFFECTING DIVIDEND POLICY IN PRACTICE

e explain factors that affect dividend policy in practice;

f calculate and interpret the effective tax rate on a given currency unit of corporate earnings under double taxation, dividend imputation, and split-rate tax systems;

In Sections 2–4 we discussed theories of dividend policy and value and concluded that the issue is, at best, unresolved. In this section we explore six factors that affect a company's dividend policy, which we defined earlier as decisions about whether, when, and in what amount to pay dividends:

- Investment opportunities
- The expected volatility of future earnings
- Financial flexibility
- Tax considerations
- Flotation costs
- Contractual and legal restrictions

Boards of directors and managers spend considerable time setting dividend policy despite the lack of clear guidance from theory to inform their deliberations. The factors listed are, however, often mentioned by managers themselves as relevant to dividend policy selection in practice. Some of the factors we explore, such as taxation, are not company-specific, whereas other factors, such as possible contractual restrictions on dividend payments and the expected volatility of future earnings, are more company-specific. The factors may be interrelated, and the presence of one may enhance or diminish the effect of another. Importantly, the independence between the investment, financing, and dividend decisions assumed by MM may no longer hold when such market imperfections as information effects, agency problems, and taxes are recognized.

5.1 Investment Opportunities

All else equal, a company with many profitable investment opportunities will tend to pay out less in dividends than a company with fewer opportunities because the former company will have more uses for internally generated cash flows. Internally generated cash flow is generally a cheaper source of equity funding than new equity issuance. Opportunities for new investments, and the speed with which a company needs to respond to them, are influenced by the industry in which the company operates. A company with the ability to delay the initiation of projects without penalty may be willing to pay out more in dividends than a company that needs to act immediately to exploit profitable investment opportunities. Technology companies tend to have much lower average dividend yields than utilities. The chief explanation may be the size and time horizon of profitable investment opportunities in relation to annual operating cash flow generated. For technology companies, the pace of change is rapid, so having internally generated funds available to react to profitable opportunities affords them valuable flexibility. For utility companies, for which there are typically fewer such opportunities and for which change is much slower, higher dividend payouts are indicated.

5.2 The Expected Volatility of Future Earnings

Several important factors in the dividend payout decision have been identified as important to managers. Most managers

- had a target payout ratio based on long-run sustainable earnings;
- focused more on dividend changes (increases or decreases) than on dividend levels; and
- were reluctant to increase the dividend if the increase might soon need to be reversed.

Findings in the United States, United Kingdom, and other countries suggest that managers are reluctant to cut dividends—preferring to smooth them over time. Smoothing takes the form of relating dividend increases to the long-term earnings growth rate, even if short-term earnings are volatile. All else equal, the more volatile earnings are, the greater the risk that a given dividend increase may not be covered by earnings in a future time period. Thus, when earnings are volatile, we expect companies to be more cautious in the size and frequency of dividend increases. These findings also hold for other countries, although variation between countries has been noted in managers' willingness to decrease dividends based on available investment opportunities.

5.3 Financial Flexibility

Companies may not initiate, or may reduce or omit, dividends to obtain the financial flexibility associated with having substantial cash on hand. A company with substantial cash holdings is in a relatively strong position to meet unforeseen operating needs and to exploit investment opportunities with minimum delay. Having a strong cash position can be particularly valuable during economic contractions when the availability of credit may be reduced. Financial flexibility may be viewed as a tactical consideration that is of greater importance when access to liquidity is critical and when the company's dividend payout is relatively large.

A classic example of explaining a dividend decision in terms of the need to preserve financial flexibility occurred with Skanska AB, based in Sweden. On 8 February 2019, Skanska AB, one of the world's biggest construction and development companies, announced its board's suggestion to cut Skanska's dividend going forward by 30% to SKr6.00. This would allow for continued expansion of its project development business while maintaining its financial ability to deliver sustainable shareholder returns. Skanska's Chief Executive Anders Danielsson stated:

> As we enter 2019, there are political and macroeconomic uncertainties which are likely to increase further. In many of our home geographies and sectors, the markets are levelling out and it is difficult to predict how long this relatively favourable environment will last.
>
> *Source*: "Skanska Warns of 'Increasing Uncertainties' and Proposes Dividend Cut," *Financial Times* (8 February 2019): https://www.ft.com/content/9201486e-2b81-11e9-a5ab-ff8ef2b976c7.

The cut was expected to conserve SKr920 million on an annual basis. With approximately SKr19 billion of cash on hand at the time of the statement and with operating cash flows at least covering the previous dividend, the dividend reduction appeared to be accurately characterized as "precautionary." Although the dividend cut announcement was accompanied by a 9% decline in Skanska's share price, the share price quickly recovered. Within two months, it had risen 7% above its pre-dividend cut announcement value, indicating the market's favorable response to Skanska's decision to cut the dividend arising from uncertainty in its operating environment and the desire to maintain financial flexibility.

When increasing financial flexibility is an important consideration, a company may decide to distribute money to shareholders primarily by means of share repurchases (covered in Sections 7–9) rather than regular dividends. A program to repurchase shares in the open market does not involve a formal requirement that any repurchases be executed, and share repurchases in general do not establish the same expectations for continuation in the future as regular dividends.

5.4 Tax Considerations

Taxation is an important factor that affects investment decisions for taxable investors, in particular, because it is the after-tax return that is most relevant to investors. Different jurisdictions tax corporate dividends in a wide variety of ways. Some tax both capital gains and dividend income. Others tax dividends but not capital gains. Even within a given country, taxation can be quite complex. In addition, because taxation is a major fiscal policy tool that is subject to politics, governments have a tendency to "re-address" tax issues, sometimes with great frequency. As with other aspects of taxation, governments use the taxation of dividends to address a variety of goals: to encourage or discourage the retention or distribution of corporate earnings; to redistribute income; or to address other political, social, and/or investment goals.

For the global investor, foreign taxes can be as important as domestic taxes. Foreign tax credits in the investor's home country also may figure importantly into the overall taxation issue. For example, France requires companies domiciled in France to withhold dividends paid to foreign investors at the corporate tax rate (reduced to 25% by 2022), but investors in other countries can usually claim a tax credit on their home country tax return for the amount of that tax, especially where a double tax agreement exists.

5.4.1 *Taxation Methods*

We look at three main systems of taxation that determine dividends: double taxation, imputation, and split-rate. Other tax systems can be a combination of these.

In a **double taxation system**, corporate pretax earnings are taxed at the corporate level and then taxed again at the shareholder level if they are distributed to taxable shareholders as dividends. Exhibit 4 illustrates double taxation, where the individual tax rate on dividends is an assumed maximum of 15%.

Exhibit 4 Double Taxation of Dividends at 15% Personal Tax Rate (per US$100)

	15.0%
Net income before taxes	US$100
Corporate tax rate	35%
Net income after tax	US$65
Dividend assuming 100% payout	US$65
Shareholder tax on dividend	US$9.75
Net dividend to shareholder	US$55.25
Double tax rate on dividend distributions*	44.8%

* Based on pretax income.

Investors will clearly prefer a lower tax rate on dividends, but it is not clear whether they prefer a higher or lower payout. Payout preferences will depend on whether there is a tax on long-term capital gains for shareholders in their country and whether the tax rate on capital gains is higher or lower than the tax rate on dividends. Later, we will discuss a company's decision with respect to the dividend payout ratio.

A second major taxation system is the **dividend imputation tax system**, which effectively ensures that corporate profits distributed as dividends are taxed just once, at the shareholder's tax rate. Australia and New Zealand use a dividend imputation tax system. Under this system, a corporation's earnings are first taxed at the corporate level. When those earnings are distributed to shareholders in the form of dividends, however, shareholders receive a tax credit, known as a **franking credit**, for the taxes that the corporation paid on those distributed earnings (i.e., corporate taxes paid are imputed to the individual shareholder). If the shareholder's marginal tax rate is higher than the company's, the shareholder pays the difference between the two rates. Exhibit 5 shows one variation of a tax imputation system in which a shareholder with a lower marginal tax bracket than the company's actually receives a tax credit for the difference between the corporate rate and his own rate.

Exhibit 5 Taxation of Dividends Based on Tax Imputation System (A$)

	Marginal Shareholder Tax Rate	
	15%	47%
Pretax income	A$100	A$100
Taxes at 30% corporate tax rate	30	30
Net income after tax	70	70
Dividend assuming 100% payout	70	70
Shareholder tax on pretax income	15	47

(continued)

Exhibit 5 (Continued)

	Marginal Shareholder Tax Rate	
	15%	**47%**
Less tax credit for corporate payment	30	30
Tax due from shareholder	(15)	17
Effective tax rate on dividend	15/100 = 15%	47/100 = 47%

A **split-rate tax system** is a third taxation system of greater historical than current importance. Under this system, corporate earnings that are distributed as dividends are taxed at a lower rate at the corporate level than earnings that are retained. At the level of the individual investor, dividends are taxed as ordinary income. Earnings distributed as dividends are still taxed twice, but the relatively low corporate tax rate on earnings mitigates that penalty. Exhibit 6 depicts this split-rate tax system for dividends.

Exhibit 6 Taxation of Dividends Based on Split-Rate System (per €100)

Pretax earnings	€200
Pretax earnings retained	100
35% tax on retained earnings	35
Pretax earnings allocated to dividends	100
20% tax on earnings allocated to dividends	20
Dividends distributed	80
Shareholder tax rate	35%
After tax dividend to shareholder	[(1 − 0.35) × 80] = 52
Effective tax rate on dividend	[20% + (80 × 0.35)%] = 48%

5.4.2 *Shareholder Preference for Current Income versus Capital Gains*

All other things being equal, one could expect that the lower an investor's tax rate on dividends relative to his or her tax rate on capital gains, the stronger the investor's preference for dividends. But other issues also impinge on this preference. The investor may buy high-payout shares for a tax-exempt retirement account. Even if dividends are taxed at a lower rate than capital gains, it is not clear that shareholders will necessarily prefer higher dividends. After all, capital gains taxes do not have to be paid until the shares are sold, whereas taxes on dividends must be paid in the year received even if reinvested. In addition, in some countries, such as the United States and Australia, shares held at the time of death benefit from a step-up valuation or tax exemption as of the death date. Finally, tax-exempt institutions, such as pension funds and endowment funds, are major shareholders in most industrial countries. Such institutions are typically exempt from both taxes on dividends and taxes on capital gains. Hence, all other things being equal, they are indifferent as to whether their return comes in the form of current dividends or capital gains.

5.5 Flotation Costs

Another factor that affects a company's dividend policy is flotation cost. Flotation costs include 1) the fees that the company pays (to investment bankers, attorneys, securities regulators, auditors, and others) to issue shares and 2) the possible adverse market price impact from a rise in the supply of shares outstanding. Aggregate flotation costs are proportionally higher (in terms of percentage of gross proceeds) for smaller companies (which issue fewer shares) than for larger companies. Flotation costs make it more expensive for companies to raise new equity capital than to use their own internally generated funds. As a result, many companies try to avoid establishing a level of dividends that would create the need to raise new equity to finance positive NPV projects.

EXAMPLE 8

A Company That Needs to Reinvest All Internally Generated Funds

Boar's Head Spirits Ltd., based in the United Kingdom, currently does not pay a dividend on its common shares. Boar's Head has an estimated operating cash flow of £500 million. The company's financial analyst has calculated its cost of capital as 12%. The same analyst has evaluated modernization and expansion projects with a positive NPV that would require £800 million. The cost of positive NPV projects exceeds estimated operating cash flow by £300 million (£800 million – £500 million). Having an above average debt ratio for its industry, Boar's Head is reluctant to increase its long-term debt in the next year. Discuss whether you would expect Boar's Head to initiate a dividend based on the above facts.

Solution:

One would expect Boar's Head would not initiate a dividend. As things stand, internally generated funds, as represented by operating cash flow, are not sufficient to fund positive NPV projects. So, payment of a dividend would be at the expense of rejecting positive NPV projects unless the balance of such projects and the dividend were both financed by debt. Given its concern about debt levels, the company would not be expected to pay a dividend that needs to be financed by debt. Because the company has unfunded positive NPV projects, it could consider issuing new shares to fund those projects. The company, however, would not be expected to issue shares solely for the purpose of paying dividends.

5.6 Contractual and Legal Restrictions

The payment of dividends is often affected by legal or contractual restrictions or rules. In some countries, such as Brazil, the distribution of dividends is legally mandated (with certain exceptions). In other countries (e.g., Canada and the United States) the payment of a dividend not specifically indicated to be a liquidating dividend may be restricted by an **impairment of capital rule**. Such a rule requires that the net value of the remaining assets as shown on the balance sheet be at least equal to some specified amount (related to the company's capital).

Contractual restrictions on the amount of dividends that can be paid are often imposed by bondholders in bond indentures. These restrictions require that the company maintain certain ratios (interest coverage ratios, current ratio, etc.) or fulfill certain conditions before dividend payments can be made. Debt covenants in a bond indenture are a response to the agency problems that exist between shareholders and bondholders and are put in place to limit the ability of the shareholders to expropriate wealth from bondholders. As an extreme example, in the absence of covenants or legal restrictions management could liquidate the company's assets and pay the proceeds to the shareholders as a liquidating dividend, leaving the bondholders with nothing to settle their claims.

If a company has issued preference shares, dividends on common shares may not be paid until preference share dividends are paid. In addition, if the preference dividends are cumulative, then preference dividends that are in arrears must be paid before any common dividend can be paid.

5.7 Factors Affecting Dividend Policy: Summary

Several factors of varying degrees of importance can affect a company's dividend policy. In the following example, we explore how these factors affect the dividend policy of a hypothetical company named Makinasi Appliances Company.

EXAMPLE 9

Makinasi Appliances Company Cuts Its Dividend

In September 2018, Makinasi Appliances Company, a hypothetical global home appliances manufacturer, announced it would cut its dividend for the first time in its history. The company, which pays quarterly dividends, said the dividend would be reduced to US$0.70 a share from the US$1.60 paid a year earlier. The 2017 total dividend was US$6.50 a share. The dividend cut ends a 400% cumulative increase in the dividend over 10 years. Faced with plunging global demand for appliances (Makinasi's sales were forecasted to fall 19%) and ongoing competition in the white goods industry, Makinasi was expecting a loss as high as US$32.5 million (operating loss of US$46 million) for fiscal year ending March 2019, compared with the analyst forecasted loss of US$18.3 million for the same period. The company already had a loss of US$28.6 million in fiscal year 2018 (the operating loss was US$30.4 million). Makinasi's plans are to aggressively cut costs: It plans to cut production-related costs by US$18 million and fixed costs by US$21 million. The company has said that the lower dividend is because of the difficulty in maintaining the dividend at its previous level. Board member bonuses have been eliminated, and manager bonuses have been reduced by 40%. Capital spending will be cut by 30% to US$27 million, and R&D spending will be cut by 13.5% to US$24million.

The company announced plans to raise capital via a bond issue for up to US$50 million. The national credit rating agency has cut Makinasi's bond rating from A to A−.

Discuss Makinasi's decision to cut its dividend in light of the factors affecting dividend policy covered in this section.

Solution:

Of the six factors discussed in this section, the *volatility of future earnings* and preservation of *financial flexibility* are the major factors influencing Makinasi's decision to cut its dividend. Paying the full dividend would have lowered Makinasi's liquidity ratios and forced it to raise even more external capital. In addition, paying the full dividend probably would likely have resulted in a more severe downgrade in its bond rating and an increase in the cost of debt financing. Paying the full dividend when faced with huge, larger than expected operating losses also might have sent a signal to investors that Makinasi was not serious about cutting costs and curtailing losses. *Flotation costs* could also play a role in

Makinasi's case. Flotation costs on new equity are typically higher than those on new debt; it is possible that if it paid a dividend of more than US$0.70 a share, it would have to issue new equity in addition to the US$50 million in debt.

PAYOUT POLICIES

6

g compare stable dividend with constant dividend payout ratio, and calculate the dividend under each policy;

l describe broad trends in corporate payout policies;

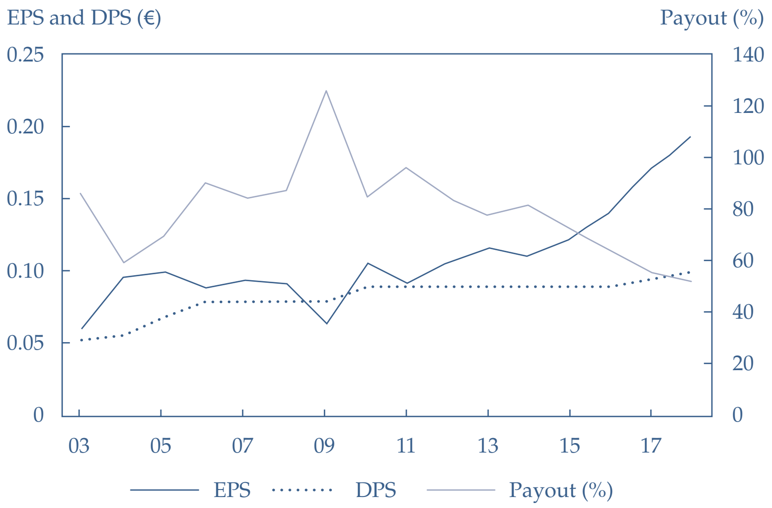
In the following sections we discuss two types of dividend policies: stable dividend and constant dividend payout ratio policies. A **stable dividend policy** is one in which regular dividends are paid that generally do not reflect short-term volatility in earnings. This type of dividend policy is the most common because managers are very reluctant to cut dividends, as discussed earlier. A **constant dividend payout ratio policy** is the policy of paying out a constant percentage of net income in dividends. In Sections 7–9, we discuss share repurchases as an alternative to the payment of cash dividends.

6.1 Stable Dividend Policy

This dividend policy is the most common. Companies that use a stable dividend policy base dividends on a long-term forecast of sustainable earnings and increase dividends when earnings have increased to a sustainably higher level. Thus, if the long-term forecast for sustainable earnings is slow growth, the dividends would be expected to grow slowly over time, more or less independent of cyclical upward or downward spikes in earnings. If sustainable earnings were not expected to grow over time, however, the corresponding dividends would be level (i.e., not growing). Compared with the constant payout ratio policy, a stable dividend policy typically involves less uncertainty for shareholders about the level of future dividends. This is so because the constant payout ratio policy reflects to a higher degree short-term volatility in earnings and/or in investment opportunities.

Many companies pride themselves on a long record of gradually and consistently increasing dividends. Exhibit 7 shows the record of Gruppo Hera (Hera), an Italian multi-utility company that operates in waste management, water, gas, electricity and central heating distribution, and energy trading and electricity generation. Between 2003 and 2018, dividends per share (DPS) show an upward trajectory. Earning declines during this period were accompanied by stable or increasing dividends, underscoring the company's longer-term stated policy of a stable and growing dividend, irrespective of yearly earnings. Consequently, Hera's payout ratio varies widely, between 52% to 125%, over the period shown. For the long term, Hera's management appeared notably optimistic about earnings prospects. In 2019, they committed to a continuing increase in annual dividends per share from €0.10 up to €0.11 by 2022.

Exhibit 7 Gruppo Hera Earnings and Dividends



Source: https://eng.gruppohera.it/group/investor_relations/investor_proposition/hera_share/dividends/

As the example shows, dividends over the period were either stable or rising—even while earnings experienced considerable variability.

A stable dividend policy can be modeled as a process of gradual adjustment toward a target payout ratio based on long-term sustainable earnings. A **target payout ratio** is a goal that represents the proportion of earnings that the company intends to distribute (pay out) to shareholders as dividends over the long term.

A model of gradual adjustment (which may be called a "target payout adjustment model") was developed by John Lintner (1956). The model reflects three basic conclusions from his study of dividend policy: 1) Companies have a target payout ratio based on long-term, sustainable earnings; 2) managers are more concerned with dividend changes than with the level of the dividend; and 3) companies will cut or eliminate a dividend only in extreme circumstances or as a last resort.

A simplified version of Lintner's model can be used to show how a company can adjust its dividend. For example, suppose that the payout ratio is below the target payout ratio and earnings are expected to increase. The expected increase in the dividend can be estimated as a function of four variables: expected earnings next year, the target payout ratio, the previous dividend, and the adjustment factor (one divided by the number of years over which the adjustment in dividends should take place). Suppose that the current dividend is US$0.40, the target payout ratio is 50%, the adjustment factor is 0.2 (i.e., the adjustment is to occur over five years), and expected earnings are US$1.50 for the year ahead (an increase from the US$1 value of last year). The expected increase in dividends is US$0.07, as shown here:

Expected increase in dividends

= (Expected earnings × Target payout ratio − Previous dividend) × Adjustment factor

= (US$1.50 × 0.5 − US$0.40) × 0.2

= US$0.07

Therefore, even though earnings increased 50% from US$1.00 to US$1.50, the dividend would only incrementally increase by about 17.5% from US$0.40 to US$0.47.

By using this model, note that if in the following year earnings temporarily fell from US$1.50 to US$1.34, the dividend might well be increased by up to US$0.04 [(US$1.34 × 0.5 − US$0.47) × 0.2 = US$0.04] a share, because the implied new dividend of US$0.51 would still be moving the company toward its target payout ratio of 50%. Even if earnings were to fall further or even experience a loss, the company would be reluctant to cut or eliminate the dividend (unless its estimate of sustainable earnings or target payout ratio were lowered); instead, it would rather opt to maintain the current dividend until future earnings increases justified an increase in the dividend.

EXAMPLE 10

Determining Dividends by Using a Target Payout Adjustment Model

Last year Luna Inc. had earnings of US$2.00 a share and paid a regular dividend of US$0.40. For the current year, the company anticipates earnings of US$2.80. It has a 30% target payout ratio and uses a 4-year period to adjust the dividend. Compute the expected dividend for the current year.

Solution:

Expected dividend = Previous dividend + [(Expected earnings × Target payout ratio − Previous dividend) × Adjustment factor)]

= US$0.40 + [(US$2.80 × 0.3 − US$0.40) × (1/4)]
= US$0.40 + [(US$0.84 − US$0.40) × (1/4)]
= US$0.51 dividend, an US$0.11 increase

Thus, although earnings are expected to increase by 40%, the increase in the dividend would be 27.5%. Despite the adjustment process, the payout ratio would fall from 20% (US$0.40/US$2.00) to 18.2% (US$0.51/US$2.80). The firm would move toward its target payout ratio if earnings growth were slower and the adjustment time period were shorter (i.e., the adjustment factor higher).

6.2 Constant Dividend Payout Ratio Policy

In this type of policy, a dividend payout ratio decided on by the company is applied to current earnings to calculate the dividend. With this type of dividend policy, dividends fluctuate with earnings in the short term. Constant dividend payout ratio policies are infrequently adopted in practice. Example 11 illustrates this type of policy with Pampas Fertilizer, a hypothetical company.

EXAMPLE 11

Pampas Fertilizer Changes from a Stable to a Constant Dividend Payout Ratio Policy

Pampas Fertilizer, a hypothetical company, is the leading fertilizer producer in Argentina. Its earnings tend to be highly volatile. Demand for fertilizer is seasonal, typically being higher in summer than in winter. On the supply side, costs are primarily driven by ammonia prices that are subject to business cycle influences and are thus very volatile. In consideration of earnings volatility, Pampas might have difficulty sustaining a steadily rising dividend level. In view of such considerations, Pampas changed its dividend policy from a stable dividend policy to a constant dividend payout ratio policy (called a "variable dividend policy" by management) in its fiscal year 2018. The following is the explanation by the company:

> Pampas has paid cash dividends on our common stock since 2003. The annual dividend rate of ARS0.50 per share of common stock, or ARS1.50 per quarter, was paid each fiscal quarter, as shown in the following table, through the second quarter of fiscal year 2018.
>
> Effective 30 November 2017, the company's board of directors approved the use of a variable dividend policy to replace the company's fixed dividend policy. Beginning with the third quarter of fiscal year 2018, Pampas began to pay a dividend to shareholders of its common stock on a quarterly basis for each quarter for which the company reports net income in an amount equal to 25% of such quarterly income.
>
> The board of Pampas implemented the variable dividend policy to more accurately reflect the results of the company's operations while recognizing and allowing for the cyclicality of the fertilizer industry.

Exhibit 8 shows quarterly data for fiscal years 2019 and 2018 in Argentine pesos (ARS).

Exhibit 8 Earnings per Share (EPS) and Dividends per Share (DPS) for Pampas Fertilizer (Fiscal Years Ending 31 March)

Fiscal Period	EPS(ARS)	DPS(ARS)
2019:Q4	9.32	2.350
2019:Q3	4.60	1.152
2019:Q2	15.41	3.852
2019:Q1	10.53	2.636
2018:Q4	7.84	1.961
2018:Q3	18.65	4.660
2018:Q2	26.30	1.500
2018:Q1	21.22	1.500

1. From the table, identify the fiscal quarter when Pampas first applied a constant dividend payout ratio policy.
2. Demonstrate that the dividend for 2019:Q4 reflects the stated current dividend policy.

Solution to 1:

Pampas first used that policy in the third quarter of fiscal year 2018. Until then, a quarterly dividend of ARS1.500 was paid irrespective of quarterly earnings. The payout ratios in all subsequent quarters round to approximately 25%.

Solution to 2:

(EPS ARS9.32)/4 = ARS2.330, which differs only slightly from the reported dividend of ARS2.350 (EPS are rounded to two decimal places, so rounding error is expected).

6.3 Global Trends in Payout Policy

An interesting question is whether corporations are changing their dividend policies in response to changes in the economic environment and in investor preferences. Dividend policy practices have international differences and change through time, even within one market, consistent with the idea that companies adapt their dividend policy over time to changing investor tastes. Typically, fewer companies in a given US stock market index have paid dividends than have companies in a comparable European stock market index. In some Asian markets, companies have significantly increased their dividend payouts, albeit from a lower base, as these companies and markets mature. In addition, the following broad trends in dividend policy have been observed:

- The fraction of companies paying cash dividends has been in long-term decline in most developed markets (e.g., the United States, Canada, the European Union overall, the United Kingdom, and Japan). In Asia-Pacific, however, the value paid out in annual dividends tripled from 2009 to 2019. In the rest of the world, the value of annual dividend payouts only doubled over the same period.
- Since the early 1980s in the United States and the early 1990s in the United Kingdom and continental Europe, the fraction of companies engaging in share repurchases has trended upward. Since the late 2010s, share repurchases by major companies in Asia, particularly in mainland China and Japan, have been substantial (following a history of little to no prior share repurchase activity).

Research on dividend behavior globally shows that aggregate dividend amounts as well as payout ratios have generally increased over time, although the fraction of dividend payers has decreased. For example, studies using data from around the world substantiate the proportion of cash dividend paying firms declining over the long term, with aggregate dividend payments concentrated in a relatively small number of firms. Post-global financial crisis, there has been some reversal in the long-term downward trend in the fraction of dividend payers and payout ratios. The dividend payers are, on average, larger, more profitable, have fewer growth opportunities, and spend less on R&D compared to the non-dividend payers.

Moreover, researchers have documented internationally a negative relationship between dividend initiations/increases and enhanced corporate governance and transparency (such as mandatory adoption of IFRS rules and enforcement of new insider trading laws). This is consistent with the notion of the decreasing information content of dividends and their reduced signaling role as governance and transparency of markets improves. Similarly, findings show less generous dividend payout policies in countries requiring detailed corporate disclosures and having strong investor protection. The reduction in both information asymmetry and agency issues resulting from improved corporate governance, along with the flexibility offered by share repurchases, appear to explain the long-term decline in dividend payers.

7 SHARE REPURCHASES, METHODS AND FINANCIAL STATEMENT EFFECTS

h compare share repurchase methods;

i calculate and compare the effect of a share repurchase on earnings per share when 1) the repurchase is financed with the company's surplus cash and 2) the company uses debt to finance the repurchase;

j calculate the effect of a share repurchase on book value per share;

A **share repurchase** (or **buyback**) is a transaction in which a company buys back its own shares. Unlike stock dividends and stock splits, share repurchases use corporate cash. Hence, share repurchases can be viewed as an alternative to cash dividends. Shares that have been issued and subsequently repurchased are classified as **treasury shares/stock** if they may be reissued or **canceled shares** if they will be retired; in either case, they are not then considered for dividends, voting, or computing earnings per share.

In contrast to the case of cash dividends, usage or growth in usage of share repurchases has historically required enabling regulation. In the United Kingdom, share repurchases became legal in 1981. They were never explicitly illegal in the United States, but usage became substantial only subsequent to US Securities and Exchange Commission rule 10b–18 in 1982. (That rule protected repurchasing companies from charges of share manipulation if repurchases were conducted consistent with the terms of the rule.) Other markets in continental Europe and Asia have also followed with enabling regulation (e.g., 1995 for Japan, 1998 for Germany and Singapore, 1999 for India and Norway, 2000 for Denmark and Sweden). Share repurchases in many markets remain subject to more restrictions than in the United States. Restrictions include requiring shareholder approval of share repurchase programs, limiting the percent of share repurchases to a certain fraction (often 10%) of outstanding shares, allowable repurchase mechanisms, and other restrictions to protect creditors. In many markets, use of share repurchases is becoming increasingly common.

In general, when an amount of share repurchases is authorized, the company is not strictly committed to following through with repurchasing shares. This situation contrasts with the declaration of dividends, where that action does commit the company to pay the dividends. Another contrast with cash dividends is that whereas cash dividends are distributed to shareholders proportionally to their ownership percentage, share repurchases in general do not distribute cash in such a proportionate manner. For example, if repurchases are executed by a company via buy orders in the open market, cash is effectively being received by only those shareholders with concurrent sell orders.

The next section presents the means by which a company may execute a share repurchase program.

7.1 Share Repurchase Methods

Following are the four main ways that companies repurchase shares, listed in order of importance.

1 **Buy in the open market.** This method of share repurchase is the most common, with the company buying its own shares as conditions warrant in the open market. The open market share repurchase method gives the company maximum flexibility. Open market repurchases are the most flexible option for a company because there is no legal obligation to undertake or complete the repurchase program; a company may not follow through with an announced program for various reasons, such as unexpected cash needs for liquidity, acquisitions, or capital expenditures. In the United States, open market transactions do not require shareholder approval, whereas in Europe, shareholder approval is required for buybacks. After studying buybacks in 32 countries, findings by Manconi, Peyer, and Vermaelen (2015) suggest that all companies have shareholder authorization in place to allow management the opportunity to buy back undervalued shares in the future. They conclude that the need for shareholder approval does not compensate for poor corporate governance and instead limits management's flexibility to time buybacks to create long-term shareholder value. Authorizations to repurchase stock can last for years. In many shareholders' minds, the announcement of a repurchase policy provides support for the

share price. If the share repurchases are competently timed to minimize price impact and to exploit perceived undervaluation in the marketplace, this method is also relatively cost effective.

2 **Buy back a fixed number of shares at a fixed price.** Sometimes a company will make a **fixed price tender offer** to repurchase a specific number of shares at a fixed price that is typically at a premium to the current market price. For example, in Australia, if a stock is selling at A$37 a share, a company might offer to buy back 5 million shares from current shareholders at A$40. If shareholders are willing to sell more than 5 million shares, the company will typically buy back a pro rata amount from each shareholder. By setting a fixed date, such as 30 days in the future, a fixed price tender offer can be accomplished quickly.

3 **Dutch auction.** A Dutch auction is also a tender offer to existing shareholders, but instead of specifying a fixed price for a specific number of shares, the company stipulates a range of acceptable prices. A Dutch auction uncovers the minimum price at which the company can buy back the desired number of shares with the company paying that price to all qualifying bids. For example, if the stock price is A$37 a share, the company would offer to buy back 5 million shares in a range of A$38 to A$40 a share. Each shareholder would then indicate the number of shares and the lowest price at which he or she would be willing to sell. The company would then begin to qualify bids beginning with those shareholders who submitted bids at A$38 and continue to qualify bids at higher prices until 5 million shares had been qualified. In our example, that price might be A$39. Shareholders who bid between A$38 and A$39, inclusive, would then be paid A$39 per share for their shares. Like Method 2, Dutch auctions can be accomplished in a short time period.

4 **Repurchase by direct negotiation.** In some markets, a company may negotiate with a major shareholder to buy back its shares, often at a premium to the market price. The company may do this to keep a large block of shares from overhanging the market (and thus acting to dampen the share price). A company may try to prevent an "activist" shareholder from gaining representation on the board of directors. In some of the more infamous cases, unsuccessful takeover attempts have ended with the company buying back the would-be suitor's shares at a premium to the market price, referred to as a **greenmail** transaction, often to the detriment of remaining shareholders. Private repurchases can also be made at discounts to the market price, reflecting the relatively weaker negotiating position of large investors with liquidity needs.

Outside the United States and Canada, almost all share repurchases occur in the open market (Method 1). Note that not all the methods listed may be permissible according to local regulations.

EXAMPLE 12

BCII Considers Alternative Methods of Share Repurchase

The board of directors of British Columbia Industries, Inc. (BCII) is considering a 5 million common share repurchase program. BCII has a sizable cash and marketable securities portfolio. BCII's current stock price is C$37. The company's chief financial officer wants to accomplish the share repurchases in a cost-effective manner. Some board members want repurchases accomplished as quickly as possible, whereas other board members mention the importance of flexibility. Discuss the relative advantages of each of the following methods with respect to cost, flexibility, and speed:

1 Open market share repurchases.
2 A fixed price tender offer.
3 Dutch auction tender offer.

Solution to 1:

Open market share repurchases give the company the most flexibility. BCII can time repurchases, making repurchases when the market prices its stock below its perceived intrinsic value. BCII can also change amounts repurchased or even not execute the repurchase program. Open market repurchases are typically made opportunistically, with cost a more important consideration than speed. Because open market repurchases can be conducted so as to minimize any effects on price and can be timed to exploit prices that are perceived to be below intrinsic value, this method is also relatively cost effective.

Solution to 2:

A fixed price tender offer can be accomplished quickly, but the company usually has to offer a premium. Obviously, this raises the cost of the buyback; however, the premium may provide a positive signal to investors regarding management's view of the value of the stock.

Solution to 3:

Dutch auctions generally enable a company to do the buyback at a lower price than with a fixed price tender offer. For example, a fixed price tender offer for 5 million shares at C$40 would cost BCII C$200 million. If the Dutch auction were successful at C$38, the cost would be C$190 million, a savings of C$10 million. Dutch auctions can be accomplished quickly, though usually not as quickly as fixed price tender offers

7.2 Financial Statement Effects of Repurchases

Share repurchases affect both the balance sheet and income statement. Both assets and shareholders' equity decline if the repurchase is made with surplus cash. As a result, leverage increases. Debt ratios (leverage) will increase even more if the repurchase is financed with debt.

On the income statement, fewer shares outstanding could increase earnings per share (i.e., by reducing the denominator) depending on how and at what cost the repurchase is financed. We discuss the effects on the income statement and balance sheet in the following sections.

7.2.1 *Changes in Earnings per Share*

One rationale for share repurchases often cited by corporate financial officers and some investment analysts is that reducing the number of shares outstanding can increase earnings per share (EPS). Assuming a company's net income does not change, a smaller number of shares after the buyback will produce a higher EPS. If a company's share repurchase is financed by high-cost borrowing, the resulting lower net income can offset the effect of the reduced shares outstanding, producing a lower EPS.

Examples 13 and 14 show changes in EPS resulting from alternative methods of financing a share repurchase.

EXAMPLE 13

Share Repurchase Using Surplus Cash

Takemiya Industries, a Japanese company, has been accumulating cash in recent years with a plan of expanding in emerging Asian markets. Takemiya's management and directors believe that such expansion is no longer practical, and they are considering a share repurchase using surplus cash. Takemiya has 10 million shares outstanding, and its net income is ¥100 million. Takemiya's share price is ¥120. Cash not needed for operations totals ¥240 million and is invested in Japanese government short-term securities that earn close to zero interest. For a share repurchase program of the contemplated size, Takemiya's investment bankers think the stock could be bought in the open market at a ¥20 premium to the current market price, or ¥140 a share. Calculate the impact on EPS if Takemiya uses the surplus cash to repurchase shares at ¥140 per share.

Solution:

First, note that current EPS = (¥100 million net income)/(10 million shares) = ¥10.00. If Takemiya repurchases shares, net income is unchanged at ¥100 million. A share repurchase at ¥140 a share reduces share count by approximately 1.7 million shares (¥240,000,000/¥140) so that 8.3 million shares remain outstanding. Thus, after the share repurchase, EPS should be (¥100 million)/(8.3 million shares) = ¥12.00, approximately. EPS would increase by 20% as a result of the share repurchase. Note that EPS would increase even more if the open market purchases were accomplished at the prevailing market price without the premium.

In the absence of surplus cash and equivalents, companies may fund share repurchases by using long-term debt. Example 14 shows that any increase in EPS is dependent on the company's after-tax borrowing rate on the funds used to repurchase stock.

EXAMPLE 14

Share Repurchases Using Borrowed Funds

Selamat Plantations, Inc., plans to borrow Malaysian ringgit (MYR)12 million, which it will use to repurchase shares. The following information is given:

- Share price at time of share repurchase = MYR60
- Earnings after tax = MYR6.6 million
- EPS before share repurchase = MYR3
- Price/Earnings (P/E) = MYR60/MYR3 = 20
- Earnings yield (E/P) = MYR3/MYR60 = 5%
- Shares outstanding = 2.2 million
- Planned share repurchase = 200,000 shares

1 Calculate the EPS after the share repurchase, assuming the after-tax cost of borrowing is 5%.

2 Calculate the EPS after the share repurchase, assuming the company's borrowing rate increases to 6% because of the increased financial risk of borrowing the MYR12 million.

Solution to 1:

EPS after buyback = (Earnings − After-tax cost of funds)/Shares outstanding after buyback
= [MYR6.6 million − (MYR12 million × 0.05)]/2 million shares
= [MYR6.6 million − (MYR0.6 million)]/2 million shares
= MYR6.0 million/2 million shares
= MYR3.00

With the after-tax cost of borrowing at 5%, the share repurchase has no effect on the company's EPS. Note that the stock's earnings yield, the ratio of earnings per share to share price or E/P, was MYR3/MYR60 = 0.05 or 5%, equal to the after-tax cost of debt.

Solution to 2:

EPS after buyback = (Earnings − After-tax cost of funds)/Shares outstanding after buyback
= [MYR6.6 million − (MYR12 million × 0.06)]/2 million shares
= [MYR6.6 million − (MYR0.72 million)]/2 million shares
= MYR5.88 million/2 million shares
= MYR2.94

Note that in this case, the after-tax cost of debt, 6%, is greater than the 5% earnings yield; thus, a reduction in EPS resulted.

In summary, a share repurchase may increase, decrease, or have no effect on EPS. The effect depends on whether the repurchase is financed internally or externally. In the case of internal financing, a repurchase increases EPS only if the funds used for the repurchase would *not* earn their cost of capital if retained by the company. In the case of external financing, the effect on EPS is positive if the earnings yield exceeds the after-tax cost of financing the repurchase. In Example 14, when the after-tax borrowing rate equaled the earnings yield of 5%, EPS was unchanged as a result of the buyback. Any after-tax borrowing rate above the earnings yield would result in a decline in EPS, whereas an after-tax borrowing rate less than the earnings yield would result in an increase in EPS.

These relationships should be viewed with caution so far as any valuation implications are concerned. Notably, to infer that an increase in EPS indicates an increase in shareholders' wealth would be incorrect. For example, the same surplus cash could also be distributed as a cash dividend. Informally, if one views the total return on a stock as the sum of the dividend yield and a capital gains return, any capital gains as a result of the boost to EPS from the share repurchase may be at the expense of an offsetting loss in dividend yield.

7.2.2 *Changes in Book Value per Share*

Price-to-book value per share is a popular ratio used in equity valuation. The following example shows the impact of a share repurchase on book value per share (BVPS).

EXAMPLE 15

The Effect of a Share Repurchase on Book Value per Share

The market price of both Company A's and Company B's common stock is US$20 a share, and each company has 10 million shares outstanding. Both companies have announced a US$5 million buyback. The only difference is that Company A has a market price per share greater than its book value per share, whereas Company B has a market price per share less than its book value per share:

- Company A has a book value of equity of US$100 million and BVPS of US$100 million/10 million shares = US$10. *The market price per share of US$20 is greater than BVPS of US$10.*
- Company B has a book value of equity of US$300 million and BVPS of US$300 million/10 million shares = US$30. *The market price per share of US$20 is less than BVPS of US$30.*

Both companies:

- buy back 250,000 shares at the market price per share (US$5 million buyback/US$20 per share = 250,000 shares) and
- are left with 9.75 million shares outstanding (10 million pre-buyback shares − 0.25 million repurchased shares = 9.75 million shares).

After the share repurchase:

- Company A's shareholders' equity at book value falls to US$95 million (US$100 million − US$5 million), and its *book value per share decreases* from US$10 to US$9.74 (shareholders' equity/shares outstanding = US$95 million/9.75 million shares = US$9.74).
- Company B's shareholders' equity at book value falls to US$295 million (US$300 million − US$5 million), and its *book value per share increases* from US$30 to US$30.26 (shareholders' equity/shares outstanding = US$295 million/9.75 million = US$30.26).

This example shows that when the market price per share is greater than its book value per share, BVPS will decrease after the share repurchase. When the market price per share is less than BVPS, however, BVPS will increase after a share repurchase.

8 VALUATION EQUIVALENCE OF CASH DIVIDENDS AND SHARE REPURCHASE: THE BASELINE

k explain the choice between paying cash dividends and repurchasing shares;

A share repurchase should be viewed as equivalent to the payment of cash dividends of equal amount in terms of the effect on shareholders' wealth, all other things being equal. "All other things being equal" in this context is shorthand for assumptions that the taxation and information content of cash dividends and share repurchases do not differ. Understanding this baseline equivalence result permits more advanced analysis for when taxation and/or information content do differ between cash dividends and share repurchases. Example 16 demonstrates the claim of equivalence in the "all other things being equal" case.

EXAMPLE 16

The Equivalence of Share Repurchases and Cash Dividends

Rohit Chemical Industries, Inc. (RCII) has 10 million shares outstanding with a current market value of Rs200 per share. WCII's board of directors is considering two ways of distributing RCII's current Rs500 million free cash flow to equity. The first method involves paying an irregular or special cash dividend of Rs500 million/10 million = Rs50 per share. The second method involves repurchasing Rs500 million worth of shares. For simplicity, we make the assumptions that dividends are received when the shares go ex-dividend and that any quantity of shares can be bought at the market price of Rs200 per share. We also assume that the taxation and information content of cash dividends and share repurchases, if any, do not differ. How would the wealth of a shareholder be affected by RCII's choice of method in distributing the Rs500 million?

Solution:

Cash Dividend

After the shares go ex-dividend, a shareholder of a single share would have Rs50 in cash (the dividend) and a share worth Rs200 – Rs50 = Rs150. The ex-dividend value of Rs150 can be demonstrated as the market value of equity after the distribution of Rs500 million divided by the (unchanged) number of shares outstanding after the dividend payment, or [(10 million)(Rs200) – Rs500 million]/10 million = Rs1,500 million/10 million = Rs150. Total wealth from ownership of one share is, therefore, Rs50 + Rs150 = Rs200.

Share Repurchase

With Rs500 million, RCII could repurchase Rs500 million/Rs200 = 2.5 million shares. The post-repurchase share price would be unchanged at Rs200, which can be calculated as the market value of equity after the Rs500 million share repurchase divided by the shares outstanding after the share repurchase, or [(10 million) (Rs200) – Rs500 million]/(10 million – 2.5 million) = Rs1,500 million/7.5 million = Rs200. Total wealth from ownership of one share is, therefore, Rs200—exactly the same as in the case of a cash dividend. Whether the shareholder actually sold the share back to RCII in the share repurchase is irrelevant for a shareholder's wealth: If the share was sold, Rs200 in cash would be realized; if the share was not sold, its market value of Rs200 would count equally toward the shareholder's wealth.

The theme of Example 16 is that a company should not expect to create or destroy shareholder wealth merely by its method of distributing money to shareholders (i.e., by share repurchases as opposed to cash dividends). Example 17 illustrates that if a company repurchases shares from an individual shareholder at a negotiated price representing a premium over the market price, the remaining shareholders' wealth is reduced.

EXAMPLE 17

Direct Negotiation: A Share Repurchase That Transfers Wealth

AfriCitrus (AC) common shares sell at South African rand (ZAR)200, and there are 10 million shares outstanding. Management becomes aware that Kirk Mzazi recently purchased a major position in its outstanding shares with the intention of influencing the business operations of AC in ways the current board does not approve. An adviser to the board has suggested approaching Mzazi privately with an offer to buy back ZAR500 million worth of shares from him at ZAR250 per share, which is a ZAR50 premium over the current market price. The board of AC declines to do so because of the effect of such a repurchase on AC's other shareholders. Determine the effect of the proposed share repurchase on the wealth of shareholders other than Mzazi.

Solution:

With ZAR500 million, AC could repurchase ZAR500 million/ZAR250 = 2 million shares from Mzazi. The post-repurchase share price would be ZAR187.50, which can be calculated as the market value of equity after the ZAR500 million share repurchase divided by the shares outstanding after the share repurchase, or [(10 million)(ZAR200) − ZAR500 million]/(10 million − 2 million) = ZAR1,500 million/8 million = ZAR1875.50. Shareholders other than Mzazi would lose ZAR200 − ZAR187. 50 = ZAR12.50 for each share owned. Although this share repurchase would conserve total wealth (including Mzazi's), it effectively transfers wealth to Mzazi from the other shareholders.

THE DIVIDEND VERSUS SHARE REPURCHASE DECISION

9

k explain the choice between paying cash dividends and repurchasing shares;

The question of the valuation implications of share repurchases and dividends is of great interest to investors. Many investors and corporate managers believe that share repurchases have, on average, a net positive effect on shareholder value Studies have found that share repurchase announcements are accompanied by significant positive excess returns both around the announcement date and for the next two years—and in some studies, five years. An explanation consistent with that finding is that managements tend to buy back their stock when it is undervalued in the marketplace and issue stock when it is overvalued.

Theory concerning the dividend–share repurchase decision generally concludes that share repurchases are equivalent to cash dividends of equal amount in their effect on shareholders' wealth, all other things being equal. Further discussion about the choice revolves around what might not "be equal" and what might cause one distribution mechanism to be preferred over the other. The use of share repurchases also may be legally restricted.

In general, share repurchases can be considered part of a company's broad policy on distributing earnings to shareholders. Also, a company may engage in share repurchases for reasons similar to those mentioned in connection with cash dividends—for example, to distribute free cash flow to equity to common shareholders. A number of additional reasons for share repurchases include the following:

- Potential tax advantages
- Share price support/signaling that the company considers its shares a good investment
- Added managerial flexibility
- Offsetting dilution from employee stock options
- Increasing financial leverage

In jurisdictions that tax shareholder dividends at higher rates than capital gains, share repurchases have a tax advantage over cash dividends. Even if the two tax rates are equal, the option to defer capital gains taxes—by deciding not to participate in the share repurchase—will be valuable to many investors.

Management of a company may view its own shares as undervalued in the marketplace and hence a good investment. Although management's stock market judgment can be just as good or bad as that of any other market participant, corporate management typically does have more information about the company's operation and future prospects than does any outside investor or analyst. Furthermore, share repurchases via open market purchase, the dominant repurchase mechanism, allow management to time share repurchases with respect to market price. The announcement of a share repurchase program is often understood as a positive signal about the company's prospects and attractiveness as an investment. An unexpected announcement of a meaningful share repurchase program can often have the same positive impact on share price as would a better-than-expected earnings report or similar positive event. In the days following the global stock market crash of October 1987, a number of prominent companies announced huge buybacks in an effort to halt the slide in the price of their shares and show confidence in the future. It may have been an important aspect in the stock market recovery that followed. Some investment analysts, however, take issue with the notion that initiation of share repurchases is a positive signal, because a repurchase program could mean that the company has no new profitable investment opportunities and is thus returning cash to shareholders.

Unlike regular cash dividends, share repurchase programs appear not to create the expectation among investors of continuance in the future. Furthermore, in contrast to an announced dividend, the announcement of a share repurchase by open market purchase does not typically create an obligation to follow through with repurchases. Additionally, the timing of share repurchases via open market activity is at managers' discretion. Share repurchases also afford shareholders flexibility because participation is optional, which is not the case with the receipt of cash dividends.

For some companies, share repurchases are used to offset the possible dilution of earnings per share that may result from the exercise of employee stock options. Whether stated or not, many companies try to repurchase at least as many shares as were issued in the exercise of stock options—even though the options are typically exercised at lower prices than the repurchase price.

Another reason for repurchasing shares is to modify the company's capital structure because share repurchases can be used to increase leverage. Share buybacks funded by newly issued debt increase leverage more than those funded by surplus cash.

Among other reasons mentioned for share repurchases by corporate managers is the objective of increasing EPS. This objective, however, is problematic for two reasons. First, even when share repurchases result in an EPS increase, the required rate of return will likely increase, reflecting higher leverage. Second, according to finance theory, changing EPS by changing the number of shares outstanding does not affect shareholder wealth given that total free cash flow is unchanged.

EXAMPLE 18

Share Repurchase to Increase Financial Leverage

Deira Oasis Holdings Inc. (DOHI), with debt and a debt ratio of United Arab Emirates durham (AED)30 million and 30%, respectively, plans a share repurchase program involving AED7 million or 10% of the market value of its common shares.

1 Assuming nothing else changes, what debt ratio would result from financing the repurchases using cash on hand?
2 Assuming nothing else changes, what debt ratio would result from financing the repurchases using new debt?
3 Discuss the effect on value of equity from financing the repurchases using cash on hand, assuming DOHI's net income and P/E remain the same.
4 Discuss the effect on value of equity from financing the repurchases using new debt, assuming DOHI's after-tax cost of debt is greater than its E/P, which remains the same.
5 Discuss the effect on value of debt from financing the repurchases using new debt, assuming the conditions in question 4 and knowing that DOHI is in imminent danger of a credit rating downgrade.

Solution to 1:

Assuming nothing else changes, if DOHI uses cash on hand to make the share repurchase, the debt ratio would increase to 32% (AED30 million/AED93million = 0.3226 or 32.3%).

Solution to 2:

Assuming nothing else changes, if DOHI uses debt to finance the share repurchase, the debt ratio would increase to 37% (AED37 million/AED100 million = 0.3700 or 37.0%).

Solution to 3:

After repurchase, DOHI's equity stands at AED63 million. However, with the same net income and fewer shares outstanding, its EPS would increase. Then, with the same P/E, DOHI's market value of equity would be expected to increase above AED63 million.

Solution to 4:

After repurchase, DOHI's equity stands at AED63 million. However, with the after-tax cost of debt exceeding the E/P, its EPS would decrease. Then, with the same P/E, DOHI's market value of equity would be expected to decrease below AED63 million.

Solution to 5:

After repurchase, DOHI's debt stands at AED37 million. However, with the real threat of a credit rating downgrade, spreads for DOHI's debt versus government treasuries would widen. Then, DOHI's market value of debt would be expected to decrease below AED37 million.

Note that with the assumptions in questions 4 and 5, the post-repurchase market values of both equity and debt would be expected to decrease. Therefore, the proportion of each in DOHI's post-repurchase capital structure is indeterminate based on the information given.

Exhibit 9 shows the results. By either means of financing the share repurchase, financial leverage increases.

Exhibit 9 Estimated Impact on Capital Structure (AED millions)

	Before Buyback		After Buyback			
			All Cash		All Debt	
	AED	%	AED	%	AED	%
Debt	30	30	30	32	37	37
Equity (at market)	70	70	63	68	63	63
Total Cap	100	100	93	100	100	100

Deira Oasis Holdings' beginning debt ratio was 30%. If Deira Oasis Holdings uses borrowed funds to repurchase equity, the debt ratio at market value will increase to 37%, which is significantly more than if it used excess cash (32%).

EXAMPLE 19

ITOCHU Corporation Announces Share Buyback to Improve ROE

In October 2018, ITOCHU Corporation, a leading Japanese *sogo shosha* (general trading company), reported that in order to improve its return on equity (ROE) it would repurchase shares by fiscal year-end March 2019 to achieve a target medium-to-long-term ROE of 13% or higher. Accordingly, ITOCHU said it could repurchase shares in the amount up to ¥30 billion. In February 2019, ITOCHU announced it was increasing its share repurchase target up to ¥100 billion. ITOCHU repurchases in these first two tranches are shown in Exhibit 10.

Exhibit 10 Share Buyback Activities, October 2018 to March 2019

Period	Shares Repurchased	Average Price (¥)	Total Value (¥)
December 2018 – January 2019	15,097,200	1,987	30 billion
February – March 2019	19,024,200	1.997	38 billion
Sum	**34,121,400**	**1,993**	**68 billion**

Source: Annual Report 2019 (online version), ITOCHU Corporation: https://www.itochu.co.jp/en/ir/financial_statements/2020/__icsFiles/afieldfile/2019/08/09/20_1st_03_e.pdf.

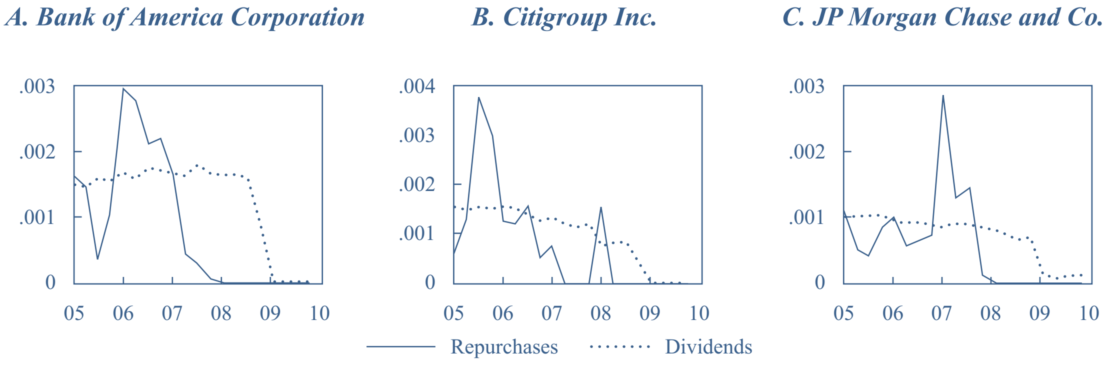
ITOCHU was followed by many other large Japanese companies—including SoftBank, Sony, Haseko, Tokyo Tatemono, and Toppan Printing. Also in February 2019, these companies announced large share buyback programs to improve ROE in response to shareholder activist pressure to improve shareholder returns and governance.

A company can use both special cash dividends and share repurchases as a supplement to regular cash dividends. These means of distributing cash are often used in years when there are large and extraordinary increases in cash flow that are not expected to continue in future years. In making these types of payments, the company essentially communicates that the distribution, like the increase in cash flow, should not be expected to continue in the future. In this context, a share repurchase is effectively an alternative to paying a special cash dividend.

Some companies initiate payouts to shareholders using share repurchases rather than cash dividends. As with the case of a share repurchase substituting for a special cash dividend, the use of share repurchases is again with the expectation that it will not be viewed as creating a fixed commitment.

Although all of the preceding can be the stated or unstated reasons for share repurchases, in general, share repurchases increase when the economy is strong and companies have more cash. During recessions, when cash is often short, share repurchases typically fall. From the fourth quarter of 2004 to the fourth quarter of 2008, the 500 companies in the S&P 500 spent US$1.8 trillion on share repurchases as compared with US$2 trillion on capital expenditures and US$1 trillion on cash dividends. In the market crash of 2008–2009, share repurchases plummeted. Major companies (particularly in the global financial sector) that had made large share repurchases encountered challenges to their financial viability in 2008 and 2009. This caused them to abandon their share repurchases and then to drastically curtail, or even eliminate, their dividends. The predominance of large US banks abandoning their share repurchase programs following the 2008 global financial crisis is shown in Exhibit 11.

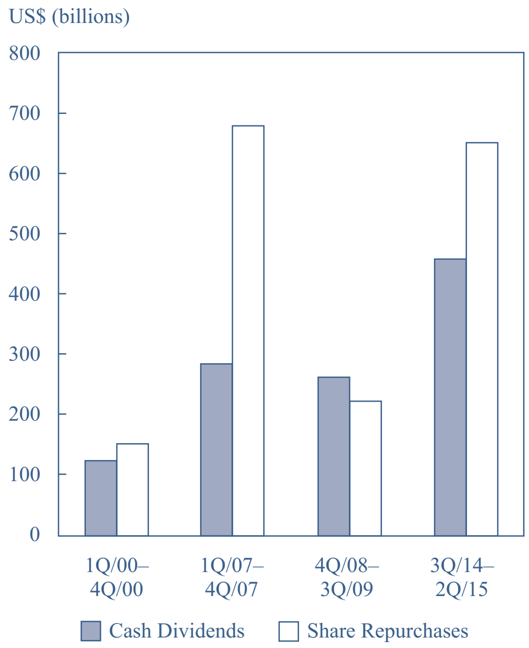
Exhibit 11 Historical Example: Share Repurchases and Dividends for Several Large US Banks



Source: Hirtle (2016).

The curtailing of share repurchases following the 2008 global financial crisis was a general occurrence; it was not restricted to the banking sector. As can be seen in Exhibit 12, data for the companies in the Russell 1000 Index, a broader US stock index than the S&P 500, show that share repurchases grew at almost twice the rate of cash dividends between 2000 and 2007, 25.0% compared to 13.0%. However, during the financial crisis of 2008–2009, companies cut back sharply on their discretionary share repurchases, from US$680 billion to US$223 billion, because many faced shrinking operating cash flows or even financial distress. Although cash dividends were also cut, the decline was much less considerable (US$286 billion to US$262 billion). By 2015, corporate operating cash flows had recovered to the point where total distributions (cash dividends plus share repurchases) reached US$1,102 billion, surpassing their previous peak of US$966 billion in 2007. Share repurchases increased nearly three times from their 2009 levels to reach US$650 billion. However, cash dividends reached US$452 billion, or over 40% of total distributions; this compares to slightly less than 30% of total distributions (US$286 billion/US$966 billion) in 2007. The higher proportion of dividends in total distributions may reflect investors' increased appetite for dividend yield during the extended period of low (or even negative) interest rates on many fixed-income securities that has prevailed in many developed countries since the end of the financial crisis.

Exhibit 12 Historical Example: Share Repurchases and Cash Dividends: Russell 1000 Companies (2000 to 2015)



Time Period	Cash Dividends*	Share Repurchases	CAGR Cash Dividends	CAGR Repurchases
	(US$ billions)		(Base Year is 2000)	
1Q2000–4Q2000	126	152	—	—
1Q2007–4Q2007	286	680	13.0%	25.0%
4Q2008–3Q2009	262	223	9.0%	4.0%
3Q2014–2Q2015	452	650	10.0%	11.0%

* Includes special dividends.
Source: JP Morgan, "2015 Distribution Policy" (September 2015).

Example 20, in which a hypothetical company's board of directors initiates a cash dividend, integrates a number of themes related to cash dividends, stock dividends (in which additional shares are distributed to shareholders instead of cash), and share repurchases.

EXAMPLE 20

Shenzhen Medical Devices' Dividend Policy Decision

Shenzhen Medical Devices Ltd. (SMDL) is a hypothetical company based in Shenzhen, China. SMDL is emerging as a leader in providing medical testing equipment to the pharmaceutical and biotechnology industries. SMDL's primary markets are growing, and the company is spending ¥100 million a year on research and development to enhance its competitive position. SMDL is highly profitable and has substantial positive free cash flow after funding positive NPV projects. During the past three years, SMDL has made significant share repurchases. Subsequent to the removal of tax on cash dividends from shares held more than a year in mainland China, SMDL management is proposing the initiation of a cash dividend. The first dividend is proposed to be an annual dividend of ¥0.40 a share to be paid during the next fiscal year. Based on estimated earnings per share of ¥3.20, this dividend would represent a payout ratio (DPS/EPS) of 0.125 or 12.5%. The proposal that will be brought before the board of directors is the following:

> "Proposed: Shenzhen Medical Devices Ltd. will institute a program of cash dividends. The first dividend will be an annual dividend of ¥0.40 a share, to be paid at a time to be determined during the next fiscal year. Thereafter, an annual dividend will be paid, equal to or above this amount, consistent with the intention of reaching a target payout ratio of 25% in line with management's expectation for long-term sustainable earnings—thereby retaining funds sufficient to finance profitable capital projects."

The company's board of directors will formally consider the dividend proposal at its next meeting in one month's time. Although some directors favor the dividend initiation proposal, other directors, led by Director Z, are skeptical of it. Director Z has stated:

> "The initiation of a cash dividend will suggest to investors that SMDL is no longer a growth company."

As a counterproposal, Director Z has offered his support for the initiation of an annual 2% stock dividend. Director W, a director who is neutral to both the cash dividend and stock dividend ideas, has told Director Z the following:

> "A 2% stock dividend will not affect the wealth of our shareholders."

Exhibit 13 presents selected *pro forma* financials of SMDL, if the directors approve the initiation of a cash dividend.

Exhibit 13 Shenzhen Medical Devices Ltd. Pro Forma Financial Data Assuming Cash Dividend (¥ millions)

Income Statement		Statement of Cash Flows	
Sales	1,200	Cash flow from operations	135
Earnings before taxes	155	Cash flow from investing activities	(84)
Taxes	35	Cash flow from financing activities	
Net income	120	Debt repayment	(4)
		Share repurchase	(32)
		Proposed dividend	(15)
		Estimated change in cash	0

Ratios		Five-Year Forecasts	
Current ratio	2.1	Sales growth	8% annually
Debt/Equity (at market)	0.27	Earnings growth	11% annually
Interest coverage	10.8x	Projected cost of capital	10%
ROA	10.0%		
ROE	19.3%		
P/E	20x		
E/P	5.0%		

Using the information provided, address the following:

1 Critique Director Z's statement.

2 Justify Director W's statement.

3 Identify and explain the dividend policy that the proposed ¥0.40 a share cash dividend reflects.

Solution to 1:

The following points argue against the thesis of Director Z's statement:

- As discussed in the text, dividend initiations and increases are on average associated with higher future earnings growth.
- Forecasted sales and earnings growth rates are relatively high.
- SMDL still has considerable positive NPV projects available to it, as shown by the cash flow from investing activities of negative ¥84 million. This fact is consistent with SMDL being a company with substantial current growth opportunities.
- For the past three years, SMDL has been making share repurchases, so investors are already cognizant that management is distributing cash to shareholders. The initiation of a dividend as a continuation of that policy is less likely to be interpreted as an information signaling event.

Solution to 2:

A stock dividend has no effect on shareholder wealth. A shareholder owns the same percentage of the company and its earnings as it did before the stock dividend. All other things being equal, the price of a stock will decline to reflect the stock dividend, but the decline will be exactly offset by the greater number of shares owned.

Solution to 3:

As shown in the statement of cash flows, the ¥0.40 a share annual dividend reflects a total amount of ¥15 million, fully using SMDL's free cash flow after acceptance of positive NPV projects. However, the proposal brought before the board also states a commitment to maintain the annual dividend at ¥0.40 a share (or greater), as a stable dividend policy would typically imply. Further, the proposal refers to a target payout ratio based on long-term sustainable earnings. These facts taken together are most consistent with a stable dividend policy based on a target payout adjustment model. (The relatively low target payout ratio of 25% of long-term sustainable earnings allows for sufficient funding of profitable capital projects, suitable for maintaining growth as a pharmaceutical company.)

ANALYSIS OF DIVIDEND SAFETY 10

m calculate and interpret dividend coverage ratios based on 1) net income and 2) free cash flow;

n identify characteristics of companies that may not be able to sustain their cash dividend.

The global recession that began in late 2007 gave rise to the largest number of dividend cuts and suspensions since the Great Depression of the 1930s. By mid-2009, S&P 500 dividends for US companies were down by 25% from the prior year, and, as indicated earlier in Exhibit 13, by 3Q 2009 dividends for companies in the broader Russell 1000 index declined by over 8% from 2007 levels. Other markets experienced similar dividend cuts following the global financial crisis; for example, UK companies reduced dividends by 15% and Australian companies by 9% in 2009. In this section, we discuss how an analyst can form a judgment on the likelihood that a company's cash dividend may be cut.

The traditional way of looking at dividend safety is the dividend payout ratio (dividends/net income) and its inverse, the **dividend coverage ratio** (net income/dividends). A higher dividend payout ratio or a lower dividend coverage ratio tends to indicate, all else equal, higher risk of a dividend cut. The logic is that with a relatively high dividend payout ratio, a relatively small percentage decline in earnings could cause the dividend not to be payable out of earnings.

EXAMPLE 21

Traditional Measures of Dividend Safety

Given the following data, calculate the dividend payout and coverage ratios:

Mature European SA	FY2019
Net income available for common stock	€100 million
Dividends paid	€40 million

Solution:

Dividend payout ratio	40/100 = 40%
Dividend coverage ratio	100/40 = 2.5x

In judging these ratios, various generalizations may be stated based on observed practice. In stating these generalizations, we emphasize that they should be confirmed for the particular market and time period being addressed.

Small, young companies generally do not pay dividends, preferring to reinvest internally for growth. However, as such companies grow, they typically initiate dividends and their payout ratios tend to increase over time. Large mature companies often target dividend payout ratios of 40% to 60% so that dividend coverage ratios range from about 1.7x to 2.5x, excluding "extra" payments. Mature companies are expected to be in this range over the course of a 5- to 10-year business cycle. Higher dividend payout ratios (or lower dividend coverage ratios) often constitute a risk factor that a dividend may be cut if earnings decline. High dividend payout ratios in relation to those of peer group companies may also point to dividend safety concerns. When a dividend coverage ratio drops to 1.0, the dividend is considered to be in jeopardy unless non-recurring events, such as an employee strike or a typhoon, are responsible for a temporary decline in earnings. In judging safety, qualitative pluses are awarded for companies that have had stable or increasing dividends, while minuses accrue to companies that have reduced their dividend in the past. Indeed, concerning this issue, Graham et al. (1962) stated that "[t]he absence of rate reduction in the past record is perhaps as important as the presence of numerous rate advances."

Free cash flow to equity represents the cash flow available for distribution as dividends after taking account of working and fixed capital expenditure needs. If those needs are ignored, distribution of dividends may be at cross-purposes with shareholder wealth maximization. Cash flow—specifically, free cash flow to equity (FCFE)—not reported net income, should be viewed as the source of cash dividend payments from that perspective. Thus, analysis of dividend safety can properly include payout and coverage ratios based on FCFE rather than net income. Other cash flow definitions besides FCFE have also been used in such ratios. Examining the correlation of dividends with cash flow measures may also provide insights.

Payouts should be considered in terms of share repurchases as well as dividends because they both represent cash distributions to shareholders. Arguably, a comprehensive measure of dividend safety would relate FCFE to both cash dividends and share repurchases:

FCFE coverage ratio = FCFE/[Dividends + Share repurchases].

If that ratio is 1, the company is returning all available cash to shareholders. If it is significantly greater than 1, the company is improving liquidity by using funds to increase cash and/or marketable securities. A ratio significantly less than 1 is not sustainable because the company is paying out more than it can afford by drawing down existing cash/marketable securities, thereby decreasing liquidity. At some point the company will have to raise new equity or cut back on capital spending.

Fundamental risk factors with regard to dividend safety include above-average financial leverage. Additional issuance of debt, whether to fund projects or to finance the dividend, may be restricted during business downturns.

Example 22 shows an analysis of dividend sustainability for Lygon Resources Ltd. (Lygon), a hypothetical company that is one of the world's largest producers of fertilizer products. The analysis includes the traditional earnings/dividend coverage approach and an alternative FCFE approach that considers total cash payouts to shareholders—dividends and share repurchases.

EXAMPLE 22

Lygon's Coverage Ratios

Lygon Resources Ltd. is a lithium miner and producer with operations in Australia, South America, and South Africa, and export markets worldwide. The company has paid dividends since 1995. Exhibit 14 shows financial information for the company.

Exhibit 14 Lygon Resources

Years Ending 31 December (A$ millions)	2015	2016	2017	2018
Net income (earnings)	540	458	399	341
Cash flow from operations	837	824	679	628
FCInv (capital expenditures)	554	417	296	327
Net borrowing	(120)	(39)	79	(7)
Dividends paid	121	256	277	323
Stock repurchases	0	105	277	0

1 Using the above information, calculate the following for 2015, 2016, 2017, and 2018:
   - A Dividend/earnings payout ratio.
   - B Earnings/dividend coverage ratio.
   - C Free cash flow to equity (FCFE).
   - D FCFE/[Dividend + Stock repurchase] coverage ratio.

2 Discuss the trends in earnings/dividend coverage and in FCFE/[Dividend + Stock repurchase] coverage.

3 Comment on the sustainability of Lygon's dividend and stock repurchase policy after 2017/2018.

Solution to 1:

A Dividend/earnings payout = A\$121/A\$540 = 0.224 or 22.4% in 2015; A\$256/A\$458 = 0.559 or 55.9% in 2016; 0.694 or 69.4% in 2017; and 0.947 or 94.7% in 2018.

B Earnings/dividend coverage = A\$540/A\$121 = 4.46x in 2015; A\$458/A\$256 = 1.79x in 2016; 1.44x in 2017; and 1.06x in 2018.

C FCFE = Cash flow from operations (CFO) − FCInv + Net borrowing = A$837 − A$554 + (A$120) = A$163 in 2015; A$824 − A$417 + (A$39) = A$368 in 2016; A$462 in 2017; and A$294 in 2018.

D FCFE coverage of dividends + Share repurchases = FCFE/[Dividends + Stock repurchases] = A$163/(A$121 + 0) = 1.35x in 2015, and A$368/(A$256 + A$105) = 1.02x in 2016. Similar calculations result in 0.83x in 2017 and 0.91x in 2018.

These results are summarized in Exhibit 15.

Exhibit 15 Lygon Resources Coverage Ratios

Years Ending 31 December	2015	2016	2017	2018
A. Dividend-to-earnings payout ratio	22.4%	55.9%	69.4%	94.7%
B. Earnings-to-dividend coverage ratio (x)	4.46	1.79	1.44	1.06
C. Free cash flow to equity (FCFE) (mil.)	163	368	462	294
D. FCFE/[div. + stock repurch.] cover. (x)	1.35	1.02	0.83	0.91

Solution to 2:

Although earnings/dividend coverage was nearly 4.5x in 2015, it declined steadily over the four years. By 2018, accounting earnings were just sufficient to pay the dividend (1.06x earnings-to-dividend coverage ratio). An analyst who looked at this metric should have suspected problems.

The FCFE coverage ratio was 1.35x in 2015, a year before the stock repurchase program began. In 2016, the FCFE coverage of dividends and stock repurchases declined to 1.02x. Lower capital expenditures were offset by increased dividends and the new stock repurchase program. Despite declining capital expenditures and positive net borrowings, the FCFE coverage ratio continued to fall substantially to 0.83x in 2017 as the company elected to increase distributions to shareholders. Despite completing the stock repurchase program the previous year, by 2018 FCFE had deteriorated so much that FCFE coverage of dividends was still less than 1.0x (0.91x).

Solution to 3:

With the FCFE coverage ratio falling to 0.83x in 2017, management likely realized that it was not prudent to undertake any new discretionary stock repurchases. By 2018, net income was still declining and FCFE coverage of the dividend at less than 1.0x meant that management should probably consider cutting the dividend.

The deterioration over time of Lygon's earnings/dividend coverage and FCFE coverage (of dividends and stock repurchases) was clear. There may be other instances when the earnings-to-dividend coverage ratio declines but still appears healthy. This is why it is important for analysts to closely examine the level and trend of the FCFE coverage ratio and the components of FCFE. Analysts should be particularly alert to companies that support their dividends and stock repurchases by reducing productive capital spending or by adding net debt or by some combination of the two because these neither are sustainable policies.

Whether based on a company's net income or free cash flow, past financial data do not always predict dividend safety. Surprise factors and other unexpected events can confound the most rigorous analysis of past data. Equity and debt markets were shaken in 2008–2009 by the losses taken by almost all US and European banks. These losses led to the cutting and, in some cases, virtual elimination of cash dividends. Not all 21st century investors would agree with Graham et al.'s 1962 assertion that "for the vast majority of common stocks, the dividend record and prospects have always been the most important factor controlling investment quality and value." But most investors would agree that when the market even begins to suspect a decrease or suspension of a company's cash dividend, that expectation is likely to weigh unfavorably on that company's common stock valuation. Therefore, many analysts look for external stock market indicators of market expectations of dividend cuts.

Extremely high dividend yields in comparison with a company's past record and forward-looking earnings is often another warning signal that investors are predicting a dividend cut. For example, the dividend yield on Singapore-listed telecoms company StarHub shares was 9.4% just prior to its fixed-to-variable dividend cut in 2019. After the announced dividend cut to a variable 80% of net profit for 2019 onwards, StarHub shares were still projected to yield about 5.6%, relatively high compared to its yield in recent years prior to the fixed dividend (which were generally about 5%). At the time, shareholder equity value was anticipated to go to zero by 2020 if the fixed dividend continued. In such cases, investors bid down the price of shares such that after the expected cut the expected total return on the shares remains adequate.

The observations of Madden (2008) support an attitude of caution with respect to very high dividend yields. Madden examined yields for the 1,963 stocks in the MSCI World Index. His company classified 865 companies out of the 1,963 companies as a "High Dividend Universe" (HDU). In the early months of the economic decline, Madden found that 78.6% of the companies in the HDU had questionable ability to maintain their dividend payments as compared with 30.7% of all the companies in the MCSI World Index. This point is supported by more recent evidence. Research using data for the S&P 500 Index stocks from 2005 to 2015 shows that the top 5% of dividend-yielding stocks accounted for over 8% of the bottom decile of performance. This over-representation of very high dividend-yielding stocks in the bottom decile of performance is likely attributable to deteriorating corporate fundamentals resulting in non-sustainable dividends. Similarly, in 2016 analysts became concerned that many European companies' dividends were unsustainable because they were paying out the highest proportion of their earnings as dividends in decades (a 60% payout ratio) at a time when their earnings were declining. This caused some companies to change their policies and cut dividends for future reinvestment and balance sheet improvement.

SUMMARY

A company's cash dividend payment and share repurchase policies constitute its payout policy. Both entail the distribution of the company's cash to its shareholders affect the form in which shareholders receive the return on their investment. Among the points this reading has made are the following:

- Dividends can take the form of regular or irregular cash payments, stock dividends, or stock splits. Only cash dividends are payments to shareholders. Stock dividends and splits merely carve equity into smaller pieces and do not create wealth for shareholders. Reverse stock splits usually occur after a stock has dropped to a very low price and do not affect shareholder wealth.

- Regular cash dividends—unlike irregular cash dividends, stock splits, and stock dividends—represent a commitment to pay cash to stockholders on a quarterly, semiannual, or annual basis.
- There are three general theories on investor preference for dividends. The first, MM, argues that given perfect markets dividend policy is irrelevant. The second, "bird in hand" theory, contends that investors value a dollar of dividends today more than uncertain capital gains in the future. The third theory argues that in countries in which dividends are taxed at higher rates than capital gains, taxable investors prefer that companies reinvest earnings in profitable growth opportunities or repurchase shares so they receive more of the return in the form of capital gains.
- An argument for dividend irrelevance given perfect markets is that corporate dividend policy is irrelevant because shareholders can create their preferred cash flow stream by selling the company's shares ("homemade dividends").
- Dividend declarations may provide information to current and prospective shareholders regarding management's confidence in the prospects of the company. Initiating a dividend or increasing a dividend sends a positive signal, whereas cutting a dividend or omitting a dividend typically sends a negative signal. In addition, some institutional and individual shareholders see regular cash dividend payments as a measure of investment quality.
- Payment of dividends can help reduce the agency conflicts between managers and shareholders, but it also can worsen conflicts of interest between shareholders and debtholders.
- Empirically, several factors appear to influence dividend policy, including investment opportunities for the company, the volatility expected in its future earnings, financial flexibility, tax considerations, flotation costs, and contractual and legal restrictions.
- Under double taxation systems, dividends are taxed at both the corporate and shareholder level. Under tax imputation systems, a shareholder receives a tax credit on dividends for the tax paid on corporate profits. Under split-rate taxation systems, corporate profits are taxed at different rates depending on whether the profits are retained or paid out in dividends.
- Companies with outstanding debt often are restricted in the amount of dividends they can pay because of debt covenants and legal restrictions. Some institutions require that a company pay a dividend to be on their "approved" investment list. If a company funds capital expenditures by borrowing while paying earnings out in dividends, it will incur flotation costs on new debt issues.
- Using a stable dividend policy, a company tries to align its dividend growth rate to the company's long-term earnings growth rate. Dividends may increase even in years when earnings decline, and dividends will increase at a lower rate than earnings in boom years.
- A stable dividend policy can be represented by a gradual adjustment process in which the expected dividend is equal to last year's dividend per share plus [(Expected earnings × Target payout ratio − Previous dividend) × Adjustment factor].
- Using a constant dividend payout ratio policy, a company applies a target dividend payout ratio to current earnings; therefore, dividends are more volatile than with a stable dividend policy.
- Share repurchases, or buybacks, most often occur in the open market. Alternatively, tender offers occur at a fixed price or at a price range through a Dutch auction. Shareholders who do not tender increase their relative position

in the company. Direct negotiations with major shareholders to get them to sell their positions are less common because they could destroy value for remaining stockholders.

- Share repurchases made with excess cash have the potential to increase earnings per share, whereas share repurchases made with borrowed funds can increase, decrease, or not affect earnings per share depending on the company's after-tax borrowing rate and earnings yield.
- A share repurchase is equivalent to the payment of a cash dividend of equal amount in its effect on total shareholders' wealth, all other things being equal.
- If the buyback market price per share is greater (less) than the book value per share, then the book value per share will decrease (increase).
- Companies can repurchase shares in lieu of increasing cash dividends. Share repurchases usually offer company management more flexibility than cash dividends by not establishing the expectation that a particular level of cash distribution will be maintained.
- Companies can pay regular cash dividends supplemented by share repurchases. In years of extraordinary increases in earnings, share repurchases can substitute for special cash dividends.
- On the one hand, share repurchases can signal that company officials think their shares are undervalued. On the other hand, share repurchases could send a negative signal that the company has few positive NPV opportunities.
- Analysts are interested in how safe a company's dividend is, specifically whether the company's earnings and, more importantly, its cash flow are sufficient to sustain the payment of the dividend.
- Early warning signs of whether a company can sustain its dividend include the dividend coverage ratio, the level of dividend yield, whether the company borrows to pay the dividend, and the company's past dividend record.

REFERENCES

Abor, Joshua, and Godfred A. Bokpin. 2010. "Investment Opportunities, Corporate Finance, and Dividend Payout Policy: Evidence from Emerging Markets." *Studies in Economics and Finance* 27 (3): 180–94.

Archbold, Stuart, and Dra Elisabete Fátima Simões Vieira. 2008. "*Corporate Dividend Policies: Survey Evidence from Finance Directors in the UK and Portugal*": 19–20. https://pdfs.semanticscholar.org/61f6/18a673a03cc9a58515df016cbdf339b0f6e5.pdf.

Bildik, Recep, Ali Fatemi, and Iraj Fooladi. 2015. "Global Dividend Payout Patterns: The US and the Rest of the World and the Effect of Financial Crisis." *Global Finance Journal*, vol. 28, no. 3: 38–67.

Brav, Alon, John Graham, Campbell Harvey, and Roni Michaely. 2005. "Payout Policy in the 21st Century." *Journal of Financial Economics*, vol. 77, no. 3: 483–527.

Brivanlou, Iman H. 2016. "High Dividend Equities—Fertile Grounds, Just Beware the Landmines." Los Angeles: TCW.

Fatemi, Ali, and Recep Bildik. 2012. "Yes, Dividends Are Disappearing: Worldwide Evidence." *Journal of Banking & Finance*, vol. 36, no. 3: 662–677.

Ferris, Stephen, Nilanjan Sen, and Emre Unlu. 2009. "An International Analysis of Dividend Payment Behavior." *Journal of Business Finance & Accounting*, vol. 36, no. 3–4: 496–522.

Gordon, Myron J. 1963. "Optimal Investment and Financing Policy." *Journal of Finance*, vol. 18, no. 2: 264–272.

Graham, Benjamin, and David L. Dodd. 1934. *Security Analysis*. New York: McGraw-Hill.

Graham, Benjamin, David Dodd, Sidney Cottle, and Charles Tatham. 1962. *Security Analysis*, 4th edition. New York: McGraw-Hill.

Hail, Luzi, Ahmed Tahoun, and Clare Wang. 2014. "Dividend Payouts and Information Shocks." *Journal of Accounting Research*, vol. 52, no. 2: 403–456.

Hansen, Robert, Raman Kumar, and Dilip Shome. 1994. "Dividend Policy and Corporate Monitoring: Evidence from the Regulated Electric Utility Industry." *Financial Management*, vol. 23, no. 1: 16–22.

Healy, P., and K. Palepu. 1988. "Earnings Information Conveyed by Dividend Initiations and Omissions." *Journal of Financial Economics*, vol. 21, no. 2: 149–175.

Hirtle, Beverly. 2016. "Bank Holding Company Dividends and Repurchases during the Financial Crisis." *Federal Reserve Bank of New York Staff Reports*, no. 666.

Jensen, Michael C. 1986. "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review*, vol. 76, no. 2: 323–329.

Lease, Ronald, Kose John, Avner Kalay, Uri Loewenstein, and Oded Sarig. 2000. *Dividend Policy: Its Impact on Firm Value*. Boston: Harvard Business School Press.

Lintner, John. 1956. "Distribution of Incomes of Corporations among Dividends, Retained Earnings and Taxes." *American Economic Review*, vol. 46, no. 2: 97–113.

Lintner, John. 1962. "Dividends, Earnings, Leverage, Stock Prices and the Supply of Capital to Corporations." *Review of Economics and Statistics*, vol. 44, no. 3: 243–269.

Madden, Ian. 2008. "High Dividend Stocks: Proceed with Caution."

Manconi, Alberto, Urs Peyer, and Theo Vermaelen. 2015. "*Buybacks around the World: Market Timing, Governance and Regulation.*" INSEAD (September): 24. https://knowledge.insead.edu/sites/www.insead.edu/files/images/1bb_around_the_world_revised_-_september_8_2015-2.pdf

Miller, Merton H., and Franco Modigliani. 1961. "Dividend Policy, Growth, and the Valuation of Shares." *Journal of Business*, vol. 34, no. 4: 411–433.

PRACTICE PROBLEMS

1 The payment of a 10% stock dividend by a company will result in an increase in that company's:

A current ratio.

B financial leverage.

C contributed capital.

2 If a company's common shares trade at very low prices, that company would be *most likely* to consider the use of a:

A stock split.

B stock dividend.

C reverse stock split.

3 In a recent presentation, Doug Pearce made two statements about dividends:

Statement 1 "A stock dividend will increase share price on the ex-dividend date, all other things being equal."

Statement 2 "One practical concern with a stock split is that it will reduce the company's price-to-earnings ratio."

Are Pearce's two statements about the effects of the stock dividend and stock split correct?

A No for both statements.

B Yes for Statement 1, and no for Statement 2.

C No for Statement 1, and yes for Statement 2.

4 All other things being equal, the payment of an internally financed cash dividend is *most likely* to result in:

A a lower current ratio.

B a higher current ratio.

C the same current ratio.

5 Match the phrases in Column A with the corresponding dividend theory in Column B. Note that you may use the answers in Column B more than once.

Column A	Column B
1. Bird in the hand	a) Dividend policy matters
2. Homemade dividends	b) Dividend policy is irrelevant
3. High tax rates on dividends	

6 Which of the following assumptions is *not* required for Miller and Modigliani's (MM) dividend theory?

A Shareholders have no transaction costs when buying and selling shares.

B There are no taxes.

C Investors prefer dividends over uncertain capital gains.

7 Sophie Chan owns 100,000 shares of PAT Company. PAT is selling for €40 per share, so Chan's investment is worth €4,000,000. Chan reinvests the gross amount of all dividends received to purchase additional shares. Assume that the

© 2018 CFA Institute. All rights reserved.

clientele for PAT shares consists of tax-exempt investors. If PAT pays a €1.50 dividend, Chan's new share ownership after reinvesting dividends at the ex-dividend price is *most* likely to be closest to:

A 103,600.

B 103,750.

C 103,900.

8 Which of the following is *most* likely to signal negative information concerning a company?

A Share repurchase.

B Decrease in the quarterly dividend rate.

C A two-for-one stock split.

9 WL Corporation is located in a jurisdiction that has a 40% corporate tax rate on pretax income and a 30% personal tax rate on dividends. WL distributes all its after-tax income to shareholders. What is the effective tax rate on WL pretax income distributed in dividends?

A 42%.

B 58%.

C 70%.

10 Which of the following factors is *least likely* to be associated with a company having a low dividend payout ratio?

A High flotation costs on new equity issues.

B High tax rates on dividends.

C Low growth prospects.

11 The dividend policy of Berkshire Gardens Inc. can be represented by a gradual adjustment to a target dividend payout ratio. Last year Berkshire had earnings per share of US$3.00 and paid a dividend of US$0.60 a share. This year it estimates earnings per share will be US$4.00. Find its dividend per share for this year if it has a 25% target payout ratio and uses a five-year period to adjust its dividend.

A US$0.68.

B US$0.80.

C US$0.85.

12 Beta Corporation is a manufacturer of inflatable furniture. Which of the following scenarios best reflects a stable dividend policy for Beta?

A Maintaining a constant dividend payout ratio of 40–50%.

B Maintaining the dividend at US$1.00 a share for several years given no change in Beta's long-term prospects.

C Increasing the dividend 5% a year over several years to reflect the two years in which Beta recognized mark-to-market gains on derivatives positions.

13 A company has 1 million shares outstanding and earnings are £2 million. The company decides to use £10 million in surplus cash to repurchase shares in the open market. The company's shares are trading at £50 per share. If the company uses the entire £10 million of surplus cash to repurchase shares at the market price, the company's earnings per share will be *closest* to:

A £2.00.

B £2.30.

C £2.50.

14 Devon Ltd. common shares sell at US$40 a share, and their estimated price-to-earnings ratio (P/E) is 32. If Devon borrows funds to repurchase shares at its after-tax cost of debt of 5%, its EPS is *most likely* to:

A increase.

B decrease.

C remain the same.

15 A company can borrow funds at an after-tax cost of 4.5%. The company's stock price is US$40 per share, earnings per share is US$2.00, and the company has 15 million shares outstanding. If the company borrows just enough to repurchase 2 million shares of stock at the prevailing market price, that company's earnings per share is *most likely* to:

A increase.

B decrease.

C remain the same.

16 Crozet Corporation plans to borrow just enough money to repurchase 100,000 shares. The following information relates to the share repurchase:

Shares outstanding before buyback	3.1 million
Earnings per share before buyback	US$4.00
Share price at time of buyback	US$50
After-tax cost of borrowing	6%

Crozet's earnings per share after the buyback will be *closest* to:

A US$4.03.

B US$4.10.

C US$4.23.

17 A company with 20 million shares outstanding decides to repurchase 2 million shares at the prevailing market price of €30 per share. At the time of the buyback, the company reports total assets of €850 million and total liabilities of €250 million. As a result of the buyback, that company's book value per share will *most likely*:

A increase.

B decrease.

C remain the same.

18 An analyst gathered the following information about a company:

Number of shares outstanding	10 million
Earnings per share	US$2.00
P/E	20
Book value per share	US$30

If the company repurchases 1 million shares at the prevailing market price, the resulting book value per share will be *closest* to:

A US$26.

B US$27.

C US$29.

19 If a company's objective is to support its stock price in the event of a market downturn, it would be advised to authorize:

A an open market share repurchase plan to be executed over the next five years.

B a tender offer share repurchase at a fixed price effective in 30 days.

C a Dutch auction tender offer effective in 30 days.

20 A company has positive free cash flow and is considering whether to use the entire amount of that free cash flow to pay a special cash dividend or to repurchase shares at the prevailing market price. Shareholders' wealth under the two options will be equivalent unless the:

A company's book value per share is less than the prevailing market price.

B company's book value per share is greater than the prevailing market price.

C tax consequences and/or information content for each alternative is different.

21 Assume that a company is based in a country that has no taxes on dividends or capital gains. The company is considering either paying a special dividend or repurchasing its own shares. Shareholders of the company would have:

A greater wealth if the company paid a special cash dividend.

B greater wealth if the company repurchased its shares.

C the same wealth under either a cash dividend or share repurchase program.

22 Investors may prefer companies that repurchase their shares instead of paying a cash dividend when:

A capital gains are taxed at lower rates than dividends.

B capital gains are taxed at the same rate as dividends.

C the company needs more equity to finance capital expenditures.

The following information relates to Questions 23–24

Janet Wu is treasurer of Wilson Chemical Company, a manufacturer of specialty chemicals used in industrial manufacturing and increasingly in technology applications. Wilson Chemical is selling one of its older divisions for US$70 million cash. Wu is considering whether to recommend a special dividend of US$70 million or a repurchase of 2 million shares of Wilson common stock in the open market. She is reviewing some possible effects of the buyback with the company's financial analyst. Wilson has a long-term record of gradually increasing earnings and dividends.

23 Wilson's share buyback could be a signal that the company:

A is decreasing its financial leverage.

B views its shares as undervalued in the marketplace.

C has more investment opportunities than it could fund internally.

24 The most likely tax environment in which Wilson Chemical's shareholders would prefer that Wilson repurchase its shares (share buybacks) instead of paying dividends is one in which:

A the tax rate on capital gains and dividends is the same.

B capital gains tax rates are higher than dividend income tax rates.

C capital gains tax rates are lower than dividend income tax rates.

The following information relates to questions 25–29

John Ladan is an analyst in the research department of an international securities firm. Ladan is currently analyzing Yeta Products, a publicly traded global consumer goods company located in the United States. Selected data for Yeta are presented in Exhibit 1.

Exhibit 1 Selected Financial Data for Yeta Products

Most Recent Fiscal Year		Current	
Pretax income	US$280 million	Shares outstanding	100 million
Net income after tax	US$182 million	Book value per share	US$25.60
Cash flow from operations	US$235 million	Share price	US$20.00
Capital expenditures	US$175 million		
Earnings per share	US$1.82		

Yeta currently does not pay a dividend, and the company operates with a target capital structure of 40% debt and 60% equity. However, on a recent conference call, Yeta's management indicated that they are considering four payout proposals:

Proposal #1: Issue a 10% stock dividend.

Proposal #2: Repurchase US$40 million in shares using surplus cash.

Proposal #3: Repurchase US$40 million in shares by borrowing US$40 million at an after-tax cost of borrowing of 8.50%.

Proposal #4: Initiate a regular cash dividend.

25 The implementation of Proposal #1 would generally lead to shareholders:

A having to pay tax on the dividend received.

B experiencing a decrease in the total cost basis of their shares.

C having the same proportionate ownership as before implementation.

26 If Yeta's management implemented Proposal #2 at the current share price shown in Exhibit 1, Yeta's book value per share after implementation would be *closest* to:

A US$25.20.

B US$25.71.

C US$26.12.

27 Based on Exhibit 1, if Yeta's management implemented Proposal #3 at the current share price, earnings per share would:

A decrease.

B remain unchanged.

C increase.

28 Based on Yeta's target capital structure, Proposal #4 will *most likely:*

A increase the default risk of Yeta's debt.

B increase the agency conflict between Yeta's shareholders and managers.

C decrease the agency conflict between Yeta's shareholders and bondholders.

29 The implementation of Proposal #4 would *most likely* signal to Ladan and other investors that future earnings growth can be expected to:

A decrease.

B remain unchanged.

C increase.

SOLUTIONS

1 C is correct. A stock dividend is accounted for as a transfer of retained earnings to contributed capital.

2 C is correct. A reverse stock split would increase the price per share of the stock to a higher, more marketable range that could possibly increase the number of investors who would consider buying the stock.

3 A is correct. Both statements are incorrect. A stock dividend will decrease the price per share, all other things being equal. A stock split will reduce the price and earnings per share proportionately, leaving the price-to-earnings ratio the same.

4 A is correct. By reducing corporate cash, a cash dividend reduces the current ratio, whereas a stock dividend (whatever the size) has no effect on the current ratio.

5 The appropriate matches are as follows:

Column A	Column B
1. Bird in the hand	a) Dividend policy matters
2. Homemade dividends	b) Dividend policy is irrelevant
3. High tax rates on dividends	a) Dividend policy matters

6 C is correct. The MM dividend theory assumes no taxes or transaction costs, but it does not assume investors have a preference for dividends over capital gains.

7 C is correct. Because the clientele for PAT investors has the same tax rate (zero) for dividends and capital gains, the ex-dividend stock price of PAT should decline by the amount of the dividend to €40 – €1.50 = €38.50. Chan will purchase €150,000/€38.50 = 3,896 additional shares. This increases her total shares owned to 103,896. Chan's new share ownership is closest to 103,900.

8 B is correct. A decrease in the quarterly dividend rate is likely to signal negative information. A decrease is typically understood as signaling poor future business prospects.

9 B is correct. The effective tax rate can be computed as 1 minus the fraction of 1 unit of earnings that investors retain after all taxes, or 1 – (1 – 0.40)(1 – 0.30) = 0.58 or 58% effective tax rate. Another way to obtain the solution: Corporate taxes = 1.00 × 0.40 = 0.40 and Personal taxes = 0.60 in dividends × 0.30 = 0.18, so Total tax = 0.40 + 0.18 = 0.58, or 58% effective rate.

10 C is correct. With low growth prospects, a company would typically have a high payout ratio, returning funds to its shareholders rather than retaining funds.

11 A is correct. The estimated dividend per share is US$0.68.

Previous DPS = US$0.60

Expected EPS = US$4

Target payout ratio = 0.25

Five-year adjustment factor = 1/5 = 0.2

Expected dividend = Previous dividend + (Expected earnings × Target payout ratio − Previous dividend) × Adjustment factor
= US$0.60 + [(US$4.00 × 0.25 − US$0.60) × 0.2]
= US$0.60 + US$0.08
= US$0.68

12 B is correct. Choice A is consistent with a constant dividend target payout ratio policy. Choice C is not correct because the earnings increases described are not sustainable long term.

13 C is correct. At the current market price, the company can repurchase 200,000 shares (£10 million/£50 = 200,000 shares). The company would have 800,000 shares outstanding after the repurchase (1 million shares − 200,000 shares = 800,000 shares).

EPS before the buyback is £2.00 (£2 million/1 million shares = £2.00). Total earnings after the buyback are the same because the company uses surplus (nonearning) cash to purchase the shares, but the number of shares outstanding is reduced to 800,000. EPS increases to £2.50 (£2 million/ 800,000 shares = £2.50).

14 B is correct. If the P/E is 32, the earnings-to-price ratio (earnings yield or E/P) is 1/32 = 3.125%. When the cost of capital is greater than the earnings yield, earnings dilution will result from the buyback.

15 A is correct. The company's earnings yield (E/P) is US$2/US$40 = 0.05. When the earnings yield is greater than the after-tax cost of borrowed funds, EPS will increase if shares are repurchased using borrowed funds.

16 A is correct.

Total earnings before buyback: US$4.00 × 3,100,000 shares = US$12,400,000

Total amount of borrowing: US$50 × 100,000 shares = US$5,000,000

After-tax cost of borrowing the amount of funds needed: US$5,000,000 × 0.06 = US$300,000

Number of shares outstanding after buyback: 3,100,000 − 100,000 = 3,000,000

EPS after buyback: (US$12,400,000 − US$300,000)/3,000,000 shares = US$4.03

The P/E before the buyback is US$50/US$4 = 12.5; thus, the E/P is 8%. The after-tax cost of debt is 6%; therefore, EPS will increase.

17 C is correct. The company's book value before the buyback is €850 million in assets − €250 million in liabilities = €600 million. Book value per share is €600 million/20 million = €30 per share. The buyback will reduce equity by 2 million shares at the prevailing market price of €30 per share. The book value of equity will be reduced to €600 million − €60 million = €540 million, and the number of shares will be reduced to 18 million; €540 million/18 million = €30 book value per share. If the prevailing market price is equal to the book value per share at the time of the buyback, book value per share is unchanged.

18 C is correct. The prevailing market price is US$2.00(20) = US$40.00 per share; thus, the buyback would reduce equity by US$40 million. Book value of equity before the buyback is US$300 million. Book value of equity after the buyback would be US$300 million − US$40 million = US$260 million. The number of

shares outstanding after the buyback would be 9 million. Thus, book value per share after the buyback would be US$260 million/9 million = US$28.89 ≈ US$29.

19 A is correct. Of the three methods, only an authorized open market share repurchase plan allows the company the flexibility to time share repurchases to coincide with share price declines.

20 C is correct. For the two options to be equivalent with respect to shareholders' wealth, the amount of cash distributed, the taxation, and the information content must be the same for both options.

21 C is correct. When there are no taxes, there are no tax differences between dividends and capital gains. All other things being equal, the effect on shareholder wealth of a dividend and a share repurchase should be the same.

22 A is correct. When capital gains are taxed at lower rates than dividends, investors may prefer companies that return cash to shareholders through share repurchases rather than dividends.

23 B is correct. Management sometimes undertakes share repurchases when it views shares as being undervalued in the marketplace.

24 C is correct. Shareholders would prefer that the company repurchase its shares instead of paying dividends when the tax rate on capital gains is lower than the tax rate on dividends.

25 C is correct. The implementation of Proposal #1, a stock dividend, would not affect a shareholder's proportionate ownership because all shareholders would receive the same proportionate increase in shares. Stock dividends, which are generally not taxable to shareholders, do not impact an investor's total cost basis (they merely reduce the cost basis per share).

A is incorrect because stock dividends are generally not taxable to shareholders. A stock dividend merely divides the "pie" (the market value of shareholders' equity) into smaller pieces.

B is incorrect because an investor's total cost basis will not be affected by a stock dividend; a stock dividend merely reduces the cost basis per share.

26 B is correct. If Yeta implemented Proposal #2, a repurchase of US$40 million in shares, the resulting book value per share (BVPS) would be US$25.71, calculated as follows:

1 Yeta has a current BVPS of US$25.60; therefore, total book value of equity is US$2,560 million (= US$25.60 × 100,000,000 shares).

2 The number of shares Yeta would repurchase is US$40 million/US$20.00 per share = 2 million shares.

3 Yeta shareholders' book value of equity after the buyback would be US$2,520 million (= US$2,560 million – US$40 million).

4 The number of shares after the buyback would be 98 million (= 100 million – 2 million).

5 The BVPS after the buyback would be US$2,520 million/98 million = US$25.71.

A is incorrect because US$25.20 incorrectly uses 100 million shares instead of 98 million shares in calculating BVPS after the buyback: US$2,520 million/100 million = US$25.20.

C is incorrect because US$26.12 incorrectly uses US$2,560 million (current book value) instead of US$2,520 million as the book value of equity in calculating BVPS after the buyback. The BVPS after the buyback is incorrectly calculated as US$2,560 million/98 million = US$26.12.

27 C is correct. In the case of external funding, a company's earnings per share will increase if the stock's earnings yield, which is the ratio of earnings per share to share price, exceeds the after-tax cost of borrowing. Yeta's earnings yield is 9.10% (= US$1.82/US$20.00), which exceeds the after-tax cost of borrowing of 8.50%.

A is incorrect because EPS will increase (not decrease) if the stock's earnings yield (= US$1.82/US$20.00) exceeds the after-tax cost of borrowing. Yeta's earnings yield of 9.10% exceeds the after-tax cost of borrowing of 8.50%.

B is incorrect because EPS will increase (not remain unchanged) if the stock's earnings yield (= US$1.82/US$20.00) exceeds the after-tax cost of borrowing. Yeta's earnings yield of 9.10% exceeds the after-tax cost of borrowing of 8.50%.

28 A is correct. Yeta is financed by both debt and equity; therefore, paying dividends can increase the agency conflict between shareholders and bondholders. The payment of dividends reduces the cash cushion available for the disbursement of fixed required payments to bondholders. All else equal, dividends increase the default risk of debt.

B is incorrect because the agency conflict between shareholders and managers would decrease (not increase) with the payment of dividends. Paying out free cash flow to equity in dividends would constrain managers in their ability to overinvest by taking on negative net present value (NPV) projects.

C is incorrect because paying dividends can increase (not decrease) the agency conflict between shareholders and bondholders. The payment of dividends would reduce the cash cushion available to Yeta for the disbursement of fixed required payments to bondholders. The payment of dividends transfers wealth from bondholders to shareholders and increases the default risk of debt.

29 C is correct. Dividend initiations convey positive information and are associated with future earnings growth, whereas dividend omissions or reductions convey negative information and are associated with future earnings problems.

A is incorrect because dividend initiations convey positive information and are associated with an expected increase (not a decrease) in future earnings growth. Dividend omissions or reductions convey negative information and are associated with future earnings problems.

B is incorrect because dividend initiations convey positive information and are associated with an expectation that future earnings growth will increase (not remain unchanged). In contrast, dividend omissions or reductions convey negative information and are associated with future earnings problems.

CORPORATE ISSUERS

STUDY SESSION 7

Corporate Issuers (2)

This study session presents two major organizational topics of corporate finance. The first topic presented is environmental, social, and governance (ESG) considerations in investment analysis. The process for identifying ESG-related risks and opportunities relevant to security analysis are described. ESG considerations provide analysts with a broader perspective of the risks and investment opportunities of a company's securities. Next, mergers, acquisitions, and corporate restructurings—which create changes in ownership and control—are examined to determine whether 1) value is created from the transaction and 2) acquisition price is justified by the transaction's benefits.

READING ASSIGNMENTS

Reading 17	Environmental, Social, and Governance (ESG) Considerations in Investment Analysis by Deborah S. Kidd, CFA, Young Lee, CFA, and Johan Vanderlugt
Reading 18	Mergers and Acquisitions by Rosita P. Chang, PhD, CFA, and Keith M. Moore, PhD, CFA
Reading 19	Capital Budgeting by John D. Stowe, PhD, CFA, and Jacqués R. Gagne, FSA, CFA, CIPM

© 2021 CFA Institute. All rights reserved.

READING

17

Environmental, Social, and Governance (ESG) Considerations in Investment Analysis

by Deborah S. Kidd, CFA, Young Lee, CFA, JD, and Johan Vanderlugt

Deborah S. Kidd, CFA, is at CFA Institute (USA). Young Lee, CFA, JD, is at MacKay Shields (USA and Europe), MacKay Shields Europe Investment Management Ltd. (Ireland), and MacKay Shields UK LLP (United Kingdom). Johan Vanderlugt is at NN Investment Partners (Netherlands).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe global variations in ownership structures and the possible effects of these variations on corporate governance policies and practices;
☐	**b.** evaluate the effectiveness of a company's corporate governance policies and practices;
☐	**c.** describe how ESG-related risk exposures and investment opportunities may be identified and evaluated;
☐	**d.** evaluate ESG risk exposures and investment opportunities related to a company.

INTRODUCTION

1

Environmental, social, and governance (ESG) considerations are increasingly being integrated into investment analysis. Evaluating how ESG factors potentially affect a company may provide analysts with a broader perspective on the risks and investment opportunities of a company's securities. Although corporate governance has long been recognized as having a significant impact on a company's long-term performance, investors have become increasingly concerned with environmental and social factors and how companies manage their resources and risk exposures that relate to such factors. Mismanagement of these resources has led to a number of high-profile corporate events that have negatively affected security prices. Increasingly stringent

CFA Institute would like to thank Hardik Sanjay Shah, CFA, for his contributions to the 2022 update of this reading.

© 2021 CFA Institute. All rights reserved.

regulatory environments, potentially finite supplies of natural resources, and global trends toward energy conservation and waste reduction have led many investors to place greater emphasis on the management of environmental risks. Similarly, such issues as worker health and safety policies, community impact, and marketing practices have increased the visibility of how a company manages its social capital.

This reading provides an overview of ESG considerations in investment analysis. Section 2 provides an overview of the global variations in corporate ownership structures, as well as how these ownership structures may affect corporate governance outcomes. In Section 3, we discuss company-specific factors that should be considered when evaluating corporate governance in the investment process. Section 4 discusses the identification of ESG-related risks and opportunities that are relevant to security analysis. Section 5 demonstrates the evaluation of ESG-related risks and opportunities through several examples. The reading concludes with a summary of the key points discussed.

2 OWNERSHIP STRUCTURES AND THEIR EFFECTS ON CORPORATE GOVERNANCE

a describe global variations in ownership structures and the possible effects of these variations on corporate governance policies and practices.

The global corporate governance landscape comprises a vast range of ownership structures that reflect unique economic, political, social, legal, and other forces in each country and/or region. Within any of these distinct ownership structures, one may find a variety of complex relationships involving shareholders and other stakeholders who have an interest in the company. Those other stakeholders include creditors, managers (executives), employees, directors, customers, suppliers, governments, and regulators. An understanding of the variation of ownership structures, the conflicts that arise within these structures, types of influential shareholders, and the effects of ownership structure on corporate governance are important considerations for analyzing corporate governance in the investment process.

2.1 Dispersed vs. Concentrated Ownership

Corporate ownership structures are generally classified as *dispersed*, *concentrated*, or a hybrid of the two. **Dispersed ownership** reflects the existence of many shareholders, none of which has the ability to individually exercise control over the corporation. In contrast, **concentrated ownership** reflects an individual shareholder or a group (called *controlling shareholders*) with the ability to exercise control over the corporation. In this context, a group is typically a family, another company (or companies), or a sovereign entity.

On a global basis, concentrated ownership structures are considerably more common than dispersed ownership structures. A global corporate governance report by the Organisation for Economic Co-operation and Development (OECD)[1] noted that 38 out of 47 jurisdictions analyzed have predominantly concentrated ownership structures. Among the other nine jurisdictions, four were characterized as having dispersed ownership structures (Australia, Ireland, the United Kingdom, and the

1 OECD (2017).

United States) and five were characterized as having "hybrid" corporate ownership structures (Canada, Germany, Japan, the Netherlands, and Switzerland). The OECD's classification of corporate ownership structure by jurisdiction is shown in Exhibit 1.

Exhibit 1 Corporate Ownership Classifications

Jurisdictions with Concentrated Ownership	
Austria, Belgium, Brazil, Chile, China, Colombia, Czech Republic, Denmark, Estonia, Finland, France, Greece, Hungary, Iceland, India, Indonesia, Israel, Italy, Latvia, Mexico, New Zealand, Norway, Poland, Portugal, Russia, Singapore, Slovenia, South Africa, South Korea, Spain, Sweden, Turkey, United Arab Emirates	State ownership is characteristic of certain countries, such as China, Norway, and Sweden. In other countries, including Brazil, Mexico, Portugal, and South Korea, families are the predominant shareholders. Company groups are prevalent in a number of additional countries, such as India and Russia.

Jurisdictions with Dispersed Ownership	
Australia, Ireland, United Kingdom, United States	Among the largest companies in Australia, the majority of shares are held (albeit dispersed) by financial institutions. In Ireland, ownership shares tend to be widely dispersed, although there are a few family-controlled companies. Among UK companies, few have major shareholders owning 25% or more of shares. In the United States, ownership of public companies is generally characterized by dispersed shareholdings; listed companies are rarely under the control of a major shareholder.

Hybrid Jurisdictions	
Canada, Germany, Japan, Netherlands, Switzerland	In Canada, among the largest listed firms, a meaningful minority have controlling shareholders. In Germany, a significant number of companies are under "tight control," but in many cases shares are broadly distributed (especially for listed companies). In Japan, a small minority of listed companies have a shareholder that owns a majority of shares. The Netherlands has a more dispersed ownership structure than most continental European countries; however, when accounting for "trust offices," ownership is somewhat more concentrated. In Switzerland, the largest listed companies have more dispersed ownership than medium-sized and smaller companies.

Source: OECD (2017).

The degree of share ownership alone may not necessarily reflect whether the control of a company is dispersed or concentrated. This is true because controlling shareholders may be either **majority shareholders** (i.e., own more than 50% of a corporation's shares) or **minority shareholders** (i.e., own less than 50% of shares). In certain ownership structures, shareholders may have disproportionately high control of a corporation relative to their ownership stakes as a result of horizontal and/or vertical ownership arrangements. **Horizontal ownership** involves companies with mutual business interests (e.g., key customers or suppliers) that have cross-holding share arrangements with each other. This structure can help facilitate strategic alliances and foster long-term relationships among such companies. **Vertical ownership** (or pyramid ownership) involves a company or group that has a controlling interest in two or more holding companies, which in turn have controlling interests in various operating companies.

The existence of *dual-class* (or multiple-class) shares can also serve to disconnect the degree of share ownership from actual control. **Dual-class shares** grant one share class superior or sole voting rights, whereas the other share class has inferior or no voting rights. When used in connection with vertical ownership arrangements, the company or group at the top of the pyramid can issue to itself all or a disproportionately high number of shares with superior voting rights and thus maintain control of the operating companies with relatively fewer total shares of a company owned.

2.2 Conflicts within Different Ownership Structures

The type of corporate ownership structure affects corporate governance policies and practices because of the potentially different set of conflicts that may exist between shareholders and managers, as well as among shareholders themselves.

The combination of *dispersed* ownership and *dispersed* voting power is generally associated with shareholders who lack the power to exercise control over managers. These shareholders are referred to as *weak shareholders*, and such managers are referred to as *strong managers*. Under this combination, conflict between the shareholders and managers of a corporation may be significant. Shareholders are interested in maximizing shareholder value. There is a risk, however, that managers will seek to use a company's resources to pursue their own interests. In corporate governance, this conflict is known as a *principal–agent* problem. This problem can be mitigated if controlling shareholders are present because they may be able to control the board of directors (and, in turn, the appointment of managers) and have the incentive to monitor management.

The combination of *concentrated* ownership and *concentrated* voting power often results in controlling shareholders maintaining a position of power over both managers and minority shareholders; these controlling shareholders are referred to as *strong shareholders*, and such managers are referred to as *weak managers*. In this scenario, controlling shareholders can effectively monitor management because they are able to control the board of directors and, in turn, the appointment of managers. With concentrated ownership and concentrated voting power, however, controlling owners may also be able to allocate company resources to their own benefit at the expense of minority owners. This conflict is known as a *principal–principal* problem.

The combination of *dispersed* ownership and *concentrated* voting power generally leads to the principal–principal problem as well. The one difference, however, is that the strong controlling shareholders do not own a majority of the shares of a company. In this scenario, controlling shareholders with less than majority ownership can exert control over other minority owners through certain mechanisms, such as dual-class share structures and pyramid structures, and can also monitor management owing to their outsized voting power.

Finally, the combination of *concentrated* ownership and *dispersed* voting power arises when there are legal restrictions on the voting rights of large share positions, known as **voting caps**. A number of sovereign governments have imposed voting caps to deter foreign investors from obtaining controlling ownership positions in strategically important local companies.

EXAMPLE 1

Conflicts between Shareholders and Managers

The managers of Company A, a widely held conglomerate, collectively own approximately 30% of the outstanding shares. No other shareholder owns more than a 1% share. Each ownership share has equivalent voting rights. Describe the potential conflict between the shareholders and managers of Company A given its ownership structure and voting rights.

Solution:

Company A has dispersed ownership and dispersed voting power. In this ownership structure, shareholders do not appear to have the ability to control or monitor managers; that is, there are weak shareholders and strong managers. In this case, a risk exists that managers may seek to use company resources to prioritize their own interests rather than to maximize shareholder value. This type of conflict is known as the *principal–agent* problem.

2.3 Types of Influential Shareholders

In different parts of the world, the types of corporate shareholders that have a significant influence on corporate governance vary. Each of these shareholder types possesses its own unique set of motivations, interests, and agendas. By identifying these shareholders, an investment analyst is in a position to further assess corporate governance risks.

2.3.1 *Banks*

In several regions, notably in Europe and Asia, banks often have considerable control over corporations with which they have a lending relationship as well as an equity interest. A conflict of interest could arise if banks have loan exposures to a corporation in addition to their equity investment. For example, if a bank has both a lending relationship with and an equity interest in a corporation, it could seek to influence the corporation to take out large loans, and perhaps on less favorable terms, to the potential detriment of other shareholders. In this situation, appropriate corporate governance controls could ensure that banks that are both creditors and investors appropriately balance their interests as lenders against their interests as shareholders.

2.3.2 *Families*

Family ownership is the predominant form of corporate structure in some parts of the world, notably Latin America and, to a slightly lesser extent, Asia and Europe. In some cases, also commonly in Latin America, individuals serve on the board of directors of multiple corporations. This situation, known as **interlocking directorates**, typically results in the same family or the same member of a corporate group controlling several corporations. A benefit of family control is lower risks associated with principal–agent problems as a result of families having concentrated ownership and management responsibility. Conversely, drawbacks of family ownership may include poor transparency, lack of management accountability, modest consideration for minority shareholder rights, and difficulty in attracting quality talent for management positions.

2.3.3 *State-Owned Enterprises*

State-owned enterprises (SOEs) often exist in corporate sectors that are strategically important to a sovereign government, have minimum initial or ongoing capital requirements that are beyond the private sector's funding ability, or provide certain products or services (e.g., power generation or health services) that the state believes should be provided at a certain price or minimum standard. Listed SOEs are partially owned by sovereign governments but also have shares traded on public stock markets. This structure is called a *mixed-ownership model*. This model tends to have lower market scrutiny of management than that of corporate ownership models, which have implicit or explicit state guarantees to prevent corporate bankruptcy. In some cases, SOEs may pursue policies that enhance social or public policy considerations at the expense of maximizing shareholder value.

2.3.4 *Institutional Investors*

In many countries, institutional investors—typically mutual funds, pension funds, insurance companies, and hedge funds—collectively represent a significant proportion of equity market ownership. Because these investors tend to have considerable resources and market expertise, they can use informed judgment in exercising their shareholder rights. In markets with widely dispersed ownership, institutional investors do not typically control a large enough ownership position to qualify as a controlling shareholder. Institutional investors can promote good corporate governance, however, by holding a company's board and management accountable when the board or management does not appear to be acting in the best interests of shareholders.

2.3.5 *Group Companies*

Some ownership structures, such as the previously mentioned horizontal and vertical ownership structures, may result in shareholders having disproportionately high control relative to their ownership stakes. Cross-holding share arrangements and long-term relationships between these group companies may restrict the potential for a transfer of share ownership—as well as create a potential obstacle for outsiders to purchase a significant portion of shares in companies. Without appropriate corporate governance policies/procedures or regulatory protections, there is a greater risk that corporations controlled by groups engage in related-party transactions at the expense of minority shareholders. Examples of group companies are Samsung (South Korea), Sanwa (Japan), and Grupo Carso (Mexico).

2.3.6 *Private Equity Firms*

Private equity firms, notably those involved in venture capital and leveraged buyouts, are strategic owners that invest in privately owned companies or in public companies with the intent to take them private. Venture capital firms invest in the early stages of a company and provide oversight of portfolio companies. Similarly, leveraged buyout (LBO) firms typically have majority control in mature companies. The involvement of venture capital and LBO firms in the management of corporations may bring important changes to companies' corporate governance, such as the development of corporate codes and implementation of performance-based manager compensation.

2.3.7 *Foreign Investors*

Foreign investors, particularly when investing in emerging market countries, can have a significant influence on local companies when they own more shares than domestic investors own. Foreign investors from countries that have more stringent standards may demand higher levels of transparency and accountability. If a local company chooses to cross-list its shares in another country with greater transparency requirements and investor protections, local minority shareholders may benefit from the arrangement.

2.3.8 *Managers and Board Directors*

When managers and board directors are also shareholders of a company, they are known as **insiders**. As their ownership positions increase, insiders are more likely to dedicate company resources toward long-term profitability because their economic interests in the company have become more aligned with the interests of external shareholders. Large ownership positions, however, may also provide insiders with increased power and an accompanying desire to protect their own interests at the expense of other shareholders.

2.4 Effects of Ownership Structure on Corporate Governance

This subsection highlights the effects of ownership structures on corporate governance policies and practices. Key considerations include board independence; board structure; special voting arrangements; corporate governance codes, laws, and listing requirements; and stewardship codes.

2.4.1 *Director Independence*

Independent board directors (or independent board members) are defined as those with no material relationship with the company with regard to employment, ownership, or remuneration. The percentage of independent board directors tends to be higher in jurisdictions with generally dispersed ownership structures relative to those countries with generally concentrated ownership structures. Independent directors originated in dispersed ownership jurisdictions as a means to strengthen the board's monitoring role over managers. The proportion of independent directors on boards has increased over time amid regulatory responses to corporate scandals (e.g., the Enron Corporation scandal in the early 2000s).

Independent directors generally serve a narrower role in concentrated ownership structures than in dispersed ownership structures. For example, the United States requires that some committees (such as the audit, nomination, and compensation committees) be composed entirely of independent directors. Conversely, in most jurisdictions with concentrated ownership structures, nomination and remuneration committees are not mandatory; when these committees do exist, jurisdictions typically recommend that the committees be wholly or largely composed of independent directors. In short, the principal–agent problem is generally less of a concern in a concentrated ownership structure than in a dispersed ownership structure.

Almost all OECD countries have introduced a requirement or recommendation for the level of independent directors serving on boards. These requirements and recommendations vary by jurisdiction, however. Some countries impose or recommend a minimum number of independent directors (typically ranging from one to three), whereas others impose or recommend a minimum ratio of independent directors (typically ranging from 20% to 50% or greater).

2.4.2 *Board Structures*

A corporation's board of directors is typically structured as either one tier or two tier. A **one-tier board** structure consists of a single board of directors, composed of executive (internal) and non-executive (external) directors. A **two-tier board** structure consists of a supervisory board that oversees a management board. A one-tier board is the most common structure, but a number of jurisdictions mandate a two-tier board structure (e.g., Argentina, Germany, and Russia), whereas other jurisdictions offer the choice of a one-tier or two-tier board (e.g., Brazil and France). The supervisory board of a two-tier board can serve as a control function through activities such as inspecting the corporation's books and records, reviewing the annual report, overseeing the work of external auditors, analyzing information provided by the management board, and setting or influencing management compensation. In certain countries, such as Germany, the supervisory boards comprise representatives from key stakeholders, such as banks and labor or other groups.

2.4.3 *Special Voting Arrangements*

Several jurisdictions have special voting arrangements to improve the position of minority shareholders. For example, Brazil, India, Portugal, Turkey, Italy, Israel, and the United Kingdom have special arrangements that facilitate engagement of minority shareholders in board nomination and election processes. When a UK company has a controlling shareholder, a condition for obtaining a "premium listing" (i.e., meeting the United Kingdom's highest standards of regulation and corporate governance) on the London Stock Exchange is that independent directors must be separately approved by both the entire shareholder base and non-controlling shareholders.

2.4.4 *Corporate Governance Codes, Laws, and Listing Requirements*

Many countries have adopted national corporate governance codes in which companies disclose their adoption of recommended corporate governance practices or explain why they have not done so. In some jurisdictions, companies are required to go beyond this "comply or explain" approach. In Japan, for example, companies with no outside directors must justify why appointing outside directors is not appropriate. Some jurisdictions do not have national corporate governance codes but make use of company law or regulation (e.g., Chile) or stock exchange listing requirements (e.g., India) to achieve similar objectives.

2.4.5 *Stewardship Codes*

Many countries have introduced voluntary codes, known as *stewardship codes*, that encourage investors to exercise their legal rights and increase their level of engagement in corporate governance. In some cases, stewardship codes are not entirely voluntary. As an example, the UK Stewardship Code includes a duty for institutional investors to monitor the companies in which they invest and requires that UK asset managers investing in the shares of UK companies publish a "comply or explain" statement of commitment to the UK Stewardship Code.

3 EVALUATING CORPORATE GOVERNANCE POLICIES AND PROCEDURES

b evaluate the effectiveness of a company's corporate governance policies and practices.

Effective corporate governance is critical for a company's reputation and competitiveness. Benefits of effective corporate governance may include higher profitability, growth in return on equity (or other return metrics), better access to credit, higher and sustainable dividends, favorable long-term share performance, and a lower cost of capital. In contrast, companies with ineffective corporate governance may experience reputational damage, reduced competitiveness, potential share price weakness/volatility, reduced profitability, and a higher cost of capital.

Corporate governance factors are often difficult to quantify. However, an understanding of these factors and their impact on governance policies and procedures can be important for investors to consider. Understanding the disclosed corporate governance policies and procedures is a key starting point for investors. Regular dialogue and engagement efforts with companies can help investors better understand corporate governance policies and procedures. In some situations, shareholder activism can be used to attempt to compel a company to act in a desired manner. **Shareholder activism** refers to strategies used by shareholders to attempt to compel a company to act in a desired manner.

The quality of corporate governance is typically reflected in a company's behavior in the market and toward its stakeholders. To that end, an evaluation of a corporation's board of directors is a starting point for investors. We discuss several of the considerations relating to boards of directors in this section. In addition, a company's policies regarding business ethics, bribery and corruption, whistleblower protection, and related-party transactions can help analysts evaluate a company's corporate governance. In practice, analysts typically adjust the risk premium (cost of capital) or credit spread of a company to reflect their assessment of corporate governance considerations.

3.1 Board Policies and Practices

A starting point for evaluating a board's effectiveness is its policies and practices. An oversight role is one aspect of a board's effectiveness—for example, whether the board is high-performing or dysfunctional. Each capital market is subject to different corporate governance issues, depending on its predominant ownership structure, history, legal environment, culture, and industry diversity. For example, boards of companies with concentrated family ownership structures and concentrated voting power may engage in related-party transactions that benefit family members or affiliates at the expense of outside shareholders.

3.1.1 *Board of Directors Structure*

Generally, when evaluating board structure, investors consider whether the organization and structure of the board—whether it is a one-tier or two-tier structure—provide sufficient oversight, representation, and accountability to shareholders. A related topic is "CEO duality," whereby the chief executive officer (CEO) also serves as chairperson of the board. CEO duality may raise concerns that the monitoring and oversight role of the board may be compromised relative to independent chairperson and CEO roles. When the chairperson is not independent or the role is combined, a company may appoint a lead independent director to help protect investor interests.

3.1.2 *Board Independence*

The independence of the directors, which we discussed previously, is a relevant consideration for investors. The absence or presence of a minority of independent directors is a negative aspect of corporate governance. Without independent directors, the potential exists for management to act in a self-serving manner. Consequently, a lack of independent directors on a board may increase investors' perception of the corporation's risk.

3.1.3 *Board Committees*

The number of board committees and how the committees operate are relevant considerations in an investor's analysis of governance. Committees vary by corporation and industry but generally include audit, governance, remuneration (or compensation), nomination, and risk and compliance committees. When evaluating a company's board committees, investors assess whether there are sufficiently independent committees that focus on key governance concerns, such as audit, compensation, and the selection of directors. The presence of non-independent committee members or executive directors may prompt the consideration of potential conflicts of interest or biases, such as those relating to compensation decisions (remuneration committee), management selection (nomination committee), and the integrity of financial reporting (audit committee).

3.1.4 *Board Skills and Experience*

The underlying skill set and experience of board directors are important investor considerations. A board with concentrated skills and experience may lack sufficient expertise to govern, as may a board with diverse skills and expertise that are not directly related to the company's core operations. In certain sectors/industries that rely on natural resources or face potentially large ESG risks, board members typically have expertise in environmental, climate, or social issues.

An issue related to skills and experience is board tenure. According to many corporate governance codes, a board director's tenure is considered long if it exceeds 10 years. Long tenure of a board member could be viewed positively or negatively. On the positive side, a board member with a long tenure may have a comprehensive understanding of how the corporation's business operates, as well as how effective company management has been during the director's tenure. On the negative side, long tenure may affect the independence of board members (i.e., they could be too closely aligned with management) or may result in directors being less willing to embrace changes in the corporation's business.

3.1.5 *Board Composition*

Board composition primarily reflects the number and diversity of directors, including their professional, cultural, and geographical background, as well as gender, age, and tenure. Boards with too many members or that lack diversity may govern less effectively than boards that are smaller or more diverse. For example, a board with long-tenured board members could become controlling, self-serving, or resistant to the introduction of new practices or policies that may benefit stakeholders.

3.1.6 *Other Considerations in Board Evaluation*

Board evaluation is necessary to maintain a company's competitive position and to meet the expectations of investors, as indicated by the widely recognized Cadbury Report, issued in the United Kingdom in 1992. Dimensions of the board evaluation process may include who evaluates the board, what should be evaluated, to whom the evaluation is targeted, and how the evaluation will be accomplished.

A board evaluation can be performed by the board itself (self-evaluation) or by an outsider on behalf of the board (external review). Some boards may decide to evaluate their performance on an "as needed" basis, whereas others will prefer to conduct a periodic external review. A board evaluation typically covers how the board performs its duties, its leadership, its structure (including the committees), and the interaction between board members and management (including culture). Apart from internal stakeholders, the evaluation may be targeted to the company's shareholders, regulators, or other external stakeholders.

EXAMPLE 2

Evaluating the Board of Directors

A junior analyst is analyzing the board of directors of Style, a fictional global clothing retailer based in Italy. Style was founded by the Donato family and is publicly traded. Style's 11-member board of directors has a chairperson—who is not the CEO—and two independent directors. Among the six non-independent directors, the Donato family accounts for four of them. All these family members have served on the board for at least 20 years. The gender and age characteristics of the board are both diverse, with women representing five of the board's directors—including its chair, Leila Donato—and the directors ranging in age from 35 to 75 years old.

Describe considerations that the junior analyst would use in evaluating the effectiveness of Style's board of directors.

Solution:

The CEO and chairperson roles are separate for Style (no CEO duality), which can be considered a sign of effective corporate governance. In addition, the board appears to be diverse in terms of age and gender, which is typically considered a positive attribute. Conversely, board independence appears to be substandard: Only two board directors are independent, whereas four Donato family members, including the chairperson (Leila Donato), are board members. The tenure of the family board members is also likely to be considered a negative attribute (it far exceeds the typical 10 years).

3.2 Executive Remuneration

Executive remuneration involves such issues as transparency of compensation, performance criteria for incentive plans (both short term and long term), the linkage of remuneration with the company strategy, and the pay differential between the CEO and the average worker. When a corporation has a "say-on-pay" provision, shareholders can vote and/or provide feedback on remuneration issues. A clawback policy allows a company to recover previously paid remuneration if certain events, such as financial restatements, misconduct, breach of the law, or risk management deficiencies, are uncovered.

There is increasing concern among investors regarding "excessive" remuneration, often represented by the ratio of CEO pay to average-worker pay. In evaluating a company's executive remuneration, investors typically consider whether the company's remuneration policies and practices provide appropriate incentives for management to drive the value of a corporation. Company disclosures such as those metrics (also known as key performance indicators, or KPIs) used in executive incentive plans may be useful tools for analysis.

3.3 Shareholder Voting Rights

Shareholder voting rights are important investor considerations. Under **straight voting** share structures, shareholders are granted the right of one vote for each share owned. Dual-class share structures differ from straight voting in that company founders and/or management typically have shares with more voting power than the class of shares available to the general public. That is, dual-class share structures—in contrast to the one share, one vote principle of straight voting—can benefit one group of shareholders over another. Because a potential conflict of interest may exist between minority shareholders and the company's founders and management (some of whom may also serve on the board of directors), it is important for investors to be aware of dual-class share structures when investing.

4 IDENTIFYING ESG-RELATED RISKS AND OPPORTUNITIES

c describe how ESG-related risk exposures and investment opportunities may be identified and evaluated.

A primary challenge when integrating ESG factors into investment analysis is identifying and obtaining information that is relevant and decision-useful. In practice, ESG-related data are generally obtained from publicly available corporate filings, documents, and communications such as corporate sustainability reports that may or may not be assured by a third party. Some of the challenges analysts face are related to inconsistent reporting of ESG information and metrics as well as the fact that the level of disclosure varies because most ESG-related disclosures are voluntary. ESG-related disclosure has generally increased over time, however, because of increased stakeholder and shareholder interest in understanding whether a company effectively manages its ESG risks and opportunities.

4.1 Materiality and Investment Horizon

When considering ESG factors in investment analysis, analysts need to evaluate the *materiality* of the underlying data. In an ESG context, materiality typically refers to ESG-related issues that are expected to affect a company's operations, its financial performance, and the valuation of its securities. In overall financial reporting, information is considered to be material if omission or misstatement of the information could influence users' decisions. Companies' as well as stakeholders' definitions of materiality in an ESG context may differ. Some companies may use the term "material" in emphasizing positive ESG information, although such information may have little impact on the company's operations or financial performance. In contrast, a company may minimize or not report negative ESG information that investors might consider material.

Analysts also consider their investment horizon and holding period when deciding which ESG factors to consider in their analysis, especially credit analysts, because of the different maturities of bonds. Some ESG issues may affect a company's performance in the short term, whereas other issues may be more long term in nature. It is important to note that the time horizon of ESG factors' impact can move from the long term to the short term and vice versa depending on a wide variety of external factors, such as a sudden change in regulation or an ESG-related controversy such as an oil spill. An investor with a short-term investment horizon may find that longer-term ESG issues can have little effect on a security's market value in the near term. Consider a manufacturing company operating in an industry that is expected to face stricter environmental regulations in the future. An investor with a short-term horizon may expect that the company's profitability will not be affected in the short term. An investor with a long-term horizon, however, may anticipate costly upgrades to plants and equipment or significant regulatory fines that are likely to reduce profitability over the longer term.

4.2 Relevant ESG-Related Factors

Corporate governance considerations, such as the structure of the board of directors, are often reasonably consistent across most companies, although best practices vary greatly regionally. In contrast, there is no globally accepted best practice with regard to environmental and social considerations. When identifying a company's specific

ESG risks and opportunities, analysts must determine the relevant factors that affect its industry. For example, energy companies are clearly more affected by environmental factors, whereas banking institutions are typically more affected by social factors (e.g., data security and privacy issues or customer satisfaction) than by environmental factors. Meanwhile, both industries are subject to governance factors. Once an analyst has determined which ESG-related factors are relevant to a company's industry, the analyst can identify applicable qualitative and quantitative data.

Approaches used to identify a company's (or industry's) ESG factors include (1) proprietary methods, (2) ratings and analysis from ESG data providers, and (3) not-for-profit industry initiatives and sustainability reporting frameworks. For example, Access to Nutrition Index evaluates the world's largest food and beverage manufacturers' policies and performance related to the most pressing nutrition challenges: obesity and undernutrition. Each of the above approaches can be used independently, or a combination of approaches can be used.

The first way of identifying company and industry ESG factors is the proprietary method approach. In this approach, analysts use their own judgment or their firm's proprietary tools to identify ESG information by researching companies, news reports, industry associations, environmental groups, financial markets, labor organizations, industry experts, and government organizations. Company-specific ESG data are generally publicly available from such sources as annual reports, corporate citizenship or sustainability reports, proxy reports, and regulatory filings (e.g., the annual 10-K report required by the US Securities and Exchange Commission). Company disclosures can generally be found on company websites.

Exhibit 2 illustrates an example of how management of one key ESG-related issue—climate change—is disclosed by City Developments Limited (CDL) in its sustainability report. Note that other real estate companies may report this information differently. In fact, ESG disclosures in general can range from minimal reporting to comprehensive data and information that span several pages, thus potentially creating comparability issues for analysts. As we discuss later in this section, a number of organizations and initiatives are working toward voluntary or mandatory standardization of various ESG-related metrics.

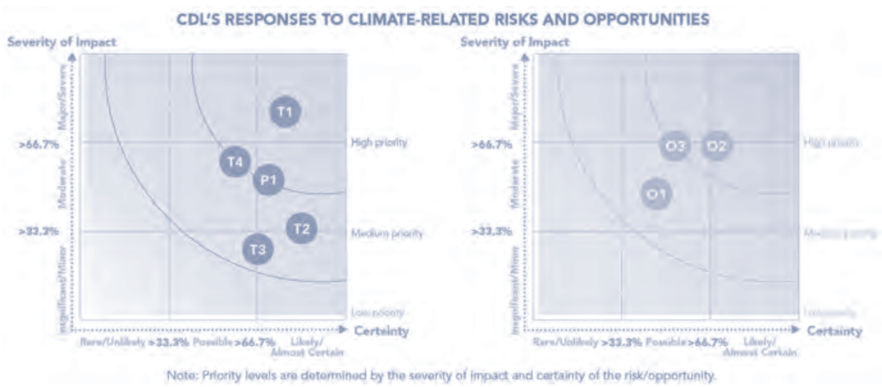
Exhibit 2 Climate Change Scenario Planning for City Developments Limited

Aligned with the recommendations of Task Force on Climate-related Financial Disclosures (TCFD) and Intergovernmental Panel on Climate Change (IPCC), CDL aims to better prepare its business for the potential financial impacts of both physical and transition risks of climate change.

CDL approached the study with two scenarios by 2030: one in which it assumed the world would decarbonize fast enough to meet the Paris Agreement's goal of limiting climate change to a global average surface temperature rise of 2°C; and another scenario that used a more ambitious 1.5°C above pre-industrial level rise. A systematic and cohesive approach was used to holistically assess and quantify all potential impacts on CDL's selected portfolio from material climate-related risks and opportunities.

(continued)

Exhibit 2 (Continued)



Source:

	Transition Risks
T1	Climate-related policy risks (e.g., increased carbon taxes and more-stringent building standards) increase operating and construction costs
T2	Water security risks increase operating costs and disrupt business continuity
T3	Call for companies to take greater responsibility of their waste production, leading to increased operating costs
T4	Climate risks lead to higher insurance premiums, lower coverage, and expose uninsurable assets

	Physical Risks
P1	Increased frequency and severity of climate events such as floods and heatwaves increase the risk of stranded assets

	Opportunities
O1	Consumer activism is on the rise globally
O2	Global shift to low-carbon growth is gaining steam
O3	Pioneering adoption of green finance in Singapore

Source: CDL, "Integrated Sustainability Report 2020."

The second approach in identifying company/industry ESG factors—ESG data providers—involves the use of information supplied by an ESG data provider (vendor), such as MSCI or Sustainalytics. These vendors obtain publicly available corporate ESG disclosures and translate them into individual ESG analyses, scores, and/or rankings for each company in the vendor's universe, often with subjective assessments by ESG analysts. In addition, vendors may score and/or rank companies within their industries and provide detailed industry analyses relating to ESG considerations.

The third approach in identifying ESG factors involves the consideration of not-for-profit initiatives and sustainability reporting frameworks that provide data and insights on ESG issues. These include the International Integrated Reporting Council (IIRC), the Global Reporting Initiative (GRI), the Sustainable Accounting Standards Board (SASB), and the 2° Investing Initiative (2DII), to name a few. The IIRC is a coalition of industry participants that promotes a standardized framework of ESG disclosures in corporate reporting. The GRI has worked with various stakeholder groups to develop sustainability reporting standards. These standards include a list of business activity groups (industries) with relevant sustainability topics that correspond to each group. A GRI report excerpt relating to the consumer durables and household and personal products sector is shown in Exhibit 3. The exhibit indicates the proposed ESG-related topics for this sector as well as additional specifications on these topics, if available. The SASB seeks to promote uniform accounting standards for sustainability reporting. In doing so, it has developed the SASB Materiality Map, which lists relevant ESG-related, sector-specific factors that the organization and industry working groups deem to be material. Exhibit 4 displays a sample SASB Materiality Map that shows the key ESG factors (shaded boxes) for the health care sector.

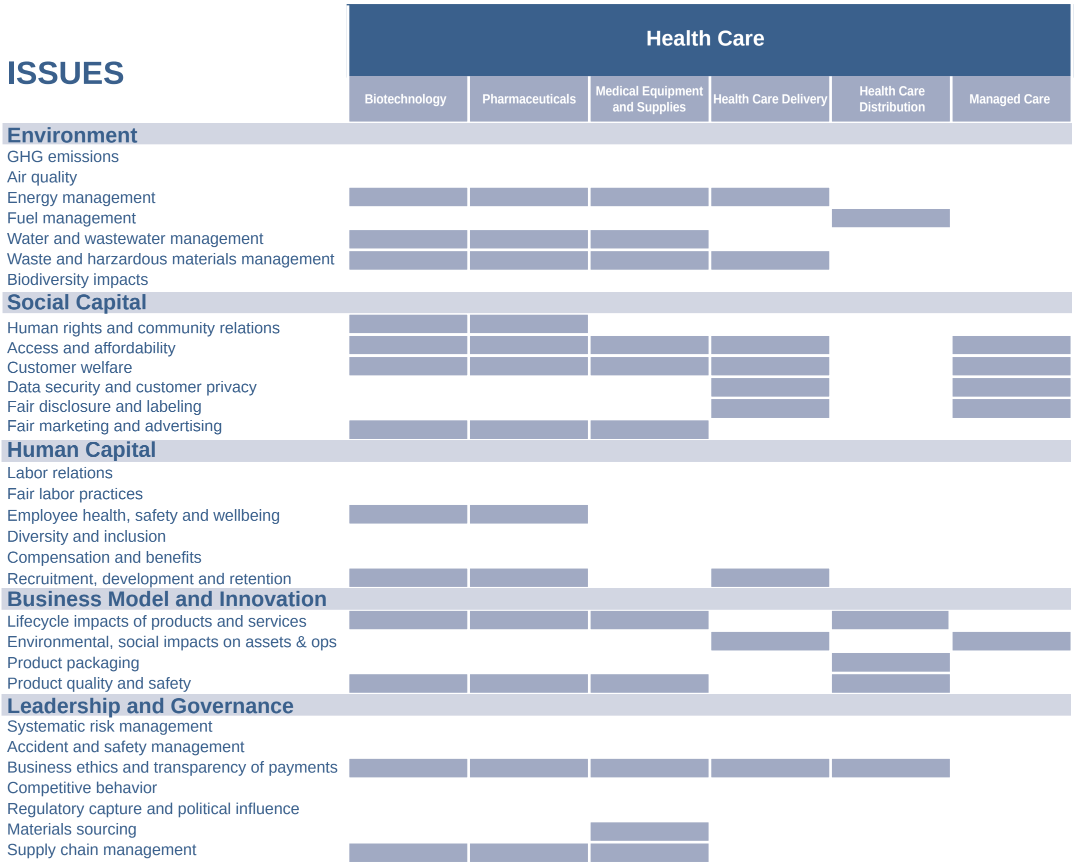
As well as providing data and analysis, ESG service providers and not-for-profit initiatives provide a variety of tools to help integrate relevant ESG factors.

Exhibit 3 GRI Sustainability Topics—Consumer Durables and Household and Personal Products Sector

Category	Proposed Topic	Topic Specification (where applicable)
Environmental	Materials sourcing	Rare metals; Sourcing standards for raw materials; Sourcing standards on animal testing; Wood-based products from responsibly managed forests
	Product packaging	Not applicable
	Plastic use	Product and packaging
	Chemicals use	International and national chemical safe use regulations; Personal care products; Phthalates and parabens
	Energy efficiency of end products	Consumer electronics
	Life cycle assessment of products	Not applicable
	Product transport efficiency	Not applicable
Social	Migrant workers	Recruitment and employment
	Product safety	Personal care products—human health and the environment
	Transparent product information and labeling	Not applicable
	Access to products, technologies, and services	Consumers with disabilities
	Electronic waste (e-waste) management	Consumer awareness
	Product design	Eco-friendly personal care products
	Product innovation	Energy consumption, GHG emissions and packaging
Other	Corporate governance	Executive board compensation; Gender participation on governance bodies
	Supplier screening	Environmental and social standards in the supply chain

Source: GRI, "Sustainability Topics for Sectors: What Do Stakeholders Want to Know?" (2013).

Exhibit 4 SASB Materiality Map—Health Care Sector



Source: Sustainability Accounting Standards Board.

From a risk/reward perspective, the use of **ESG integration**—the implementation of qualitative and quantitative ESG factors in traditional security and industry analysis as well as portfolio construction—typically differs for equity and fixed-income (debt) analysis. In equity analysis, ESG integration is used to both identify potential opportunities and mitigate downside risk, whereas in fixed-income analysis, ESG integration is generally focused on mitigating downside risk as the bond redeems at par on maturity.

The process of identifying and evaluating relevant ESG-related factors is reasonably similar for both equity and corporate credit analysis, because they share the same above-mentioned proprietary methods although material factors may differ based on relevance to credit. ESG integration techniques are also reasonably similar, such as adjustments to forecasted financial metrics and ratios, although the implication differs in practice.

In equity security analysis, ESG-related factors are often analyzed in the context of forecasting financial metrics and ratios, adjusting valuation model variables (e.g., discount rate), or using sensitivity and/or scenario analysis. For example, an analyst might increase her forecast of a hotel company's operating costs because of the impacts of excessive employee turnover—lost productivity, reduced customer satisfaction, and increased expenses for employee searches, temporary workers, and training programs. As another example, an analyst might choose to lower the discount rate for a snack food company that is expected to gain a competitive advantage by transitioning to a sustainable source of a key ingredient in its products.

In credit analysis, ESG factors may be integrated using internal credit assessments, forecasting financial ratios, and relative credit ranking of companies (or governments). In terms of valuation, relative value, spread, duration, and sensitivity/scenario analysis are often used. For example, an analyst may include the effect of lawsuits on the credit ratios, cash flow, or liquidity of a toy company. The same analyst may also estimate the potential for the credit spreads of the toy company's bonds to widen from these lawsuits. Generally speaking, the effect on the credit spreads of an issuer's debt obligations or its credit default swaps (CDSs) may differ depending on maturity. As a different example, consider an analyst who believes that a coal company faces long-term risk from potential **stranded assets**—that is, assets that are obsolete or not economically viable, often owing to changes in regulatory or government policy and/or shifts in demand. In this case, the analyst may believe that valuation of the coal company's 10-year-maturity notes would be considerably more negatively affected than its 1-year-maturity notes.

One particular type of bond an analyst might encounter is a **green bond**. The sidebar "Green Bonds" provides more detail about these securities and how investors typically analyze them. Increasingly, investors use scenario analysis and stress tests to assess the potential impact of key factors, such as physical risks of climate change.

Green Bonds

Green bonds are bonds in which the proceeds are designated by issuers to fund a specific project or portfolio of projects that have environmental or climate benefits. The first green bond, the Climate Awareness Bond, was issued by the European Investment Bank in 2007. Issuers have the primary decision for labeling their bonds "green." This decision is made in close cooperation with the lead underwriter. At a minimum level, issuers provide detail to the investors about the green eligibility criteria for the use of proceeds, in line with the Green Bond Principles (discussed in the next paragraph). Issuers are responsible for providing investors with details on the criteria used to classify the bonds as green and how the bond's proceeds are used. In some cases, issuers may commission independent reviews of the green criteria to provide investors with greater transparency. Issuers of green bonds typically incur additional costs related to the monitoring and reporting of the use of the bond's proceeds. However, these issuers may benefit from a more diversified investor base and potentially a new-issue premium if demand is strong.

The Green Bond Principles, a set of voluntary standards to guide issuers in the determination of labeling a bond as green, were developed in 2014 by a consortium of investment banks. Ongoing monitoring and further development of the Green Bond Principles is the responsibility of the International Capital Market Association, a global securities self-regulatory organization. As the green bond market has evolved, index providers, credit rating agencies, and the not-for-profit Climate Bonds Initiative have developed their own methodologies or standards to assess labeled green bonds. In addition, the European Commission is exploring the feasibility of imposing specific criteria that must be met for a bond to be labeled green.

Green bonds typically resemble an issuer's conventional bonds, with the exception that the bond proceeds are earmarked for green projects. Green bonds normally have the same credit ratings and bondholder recourse as conventional bonds of the same issuer (all else being equal). In addition to conventional or "plain vanilla" corporate bonds, other types of green bonds include project bonds, mortgage-backed and asset-backed securities, and municipal bonds. For example, the state of California's $300 million general obligation 2014 green bond issue is backed by the state's General Fund, just as California's other general obligation bonds are.

Because only the use of proceeds differs, the analysis and valuation of green bonds are essentially the same as those of conventional bonds. Some green bonds, however, may command a premium, or tighter credit spread, versus comparable conventional bonds because of market demand. One unique risk of green bonds is **greenwashing**, which is the risk that the bond's proceeds are not actually used for a beneficial environmental or climate-related project. Greenwashing can result in an investor overpaying for a bond (if the investor paid a premium for the bond's green feature) or holding a bond that does not satisfy a prescribed environmental or climate investment mandate. Liquidity risk may also be a consideration for green bonds, given that they are often purchased by buy-and-hold investors.

5 EVALUATING ESG-RELATED RISKS AND OPPORTUNITIES

d evaluate ESG risk exposures and investment opportunities related to a company.

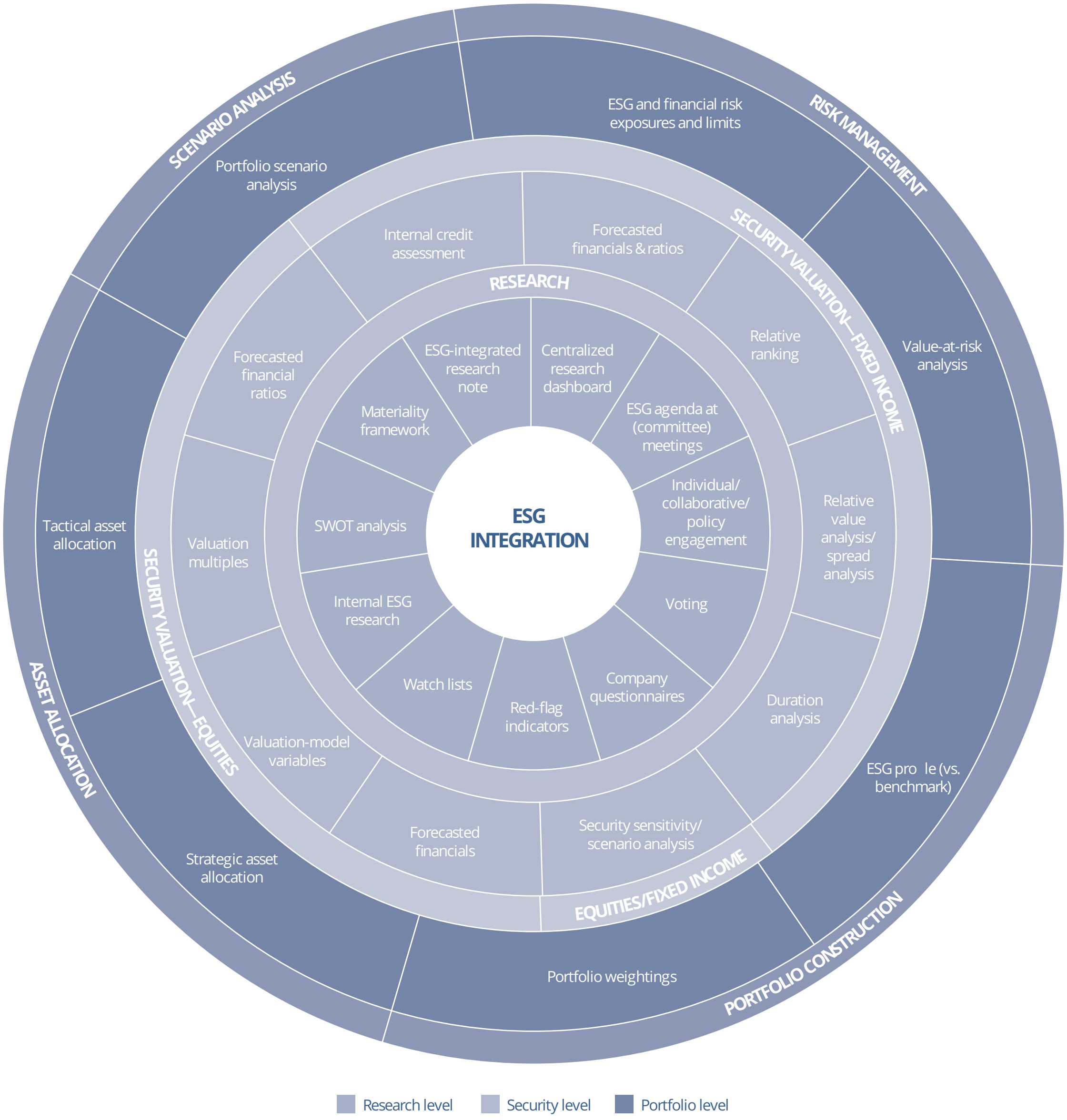
By integrating ESG considerations into the investment process, investors can take a broader perspective of company and industry analysis. In this way, the potential effects of ESG factors on a company's financial statements and valuation can be assessed and, in turn, can help drive investment decisions. In this section, we discuss examples of how ESG considerations can be integrated into financial analysis and valuation, from both an equity and a corporate bond perspective.

5.1 ESG Integration

A typical starting point for ESG integration is the identification of material qualitative and quantitative ESG factors that pertain to a company or its industry. An analyst may evaluate these factors on both a historical and a forecast basis, as well as relative to a company's peers, and then make relevant adjustments to a company's financial statements or valuation. ESG-related adjustments to a company's income statement and cash flow statement typically relate to projected revenues, operating/non-operating costs, operating margins, earnings, capital expenditures, or other items. ESG-related adjustments to a company's balance sheet often reflect an analyst's estimate of impaired assets. For equities, valuation adjustments often include adjusting a company's cost of capital using the discount rate or a multiple of price or terminal value. For bonds, an analyst may adjust an issuer's credit spread or CDS to reflect anticipated effects from ESG considerations.

The use of qualitative and quantitative research, as well as securities valuation of equities and fixed income, are key elements of the "ESG Integration Framework" (see Exhibit 5). Portfolio construction, asset allocation, scenario analysis, and risk management form the remainder of this framework.

Exhibit 5 The ESG Integration Framework



Source: Guidance and Case Studies for ESG Integration: Equities and Fixed Income, 2018

5.2 Examples of ESG Integration

This section provides examples of ESG integration for three fictitious companies in different industries: beverages, pharmaceuticals, and banks. For simplicity, each integration example focuses on either environmental, social, or governance factors—largely depending on which is most relevant for that company or its industry. Note that although specific industries are used in the examples, the underlying concepts can be applied to other industries as well. Finally, given the scope of this reading, we focus on the *effects* of ESG integration on financial analysis and valuation rather than the computations involved.

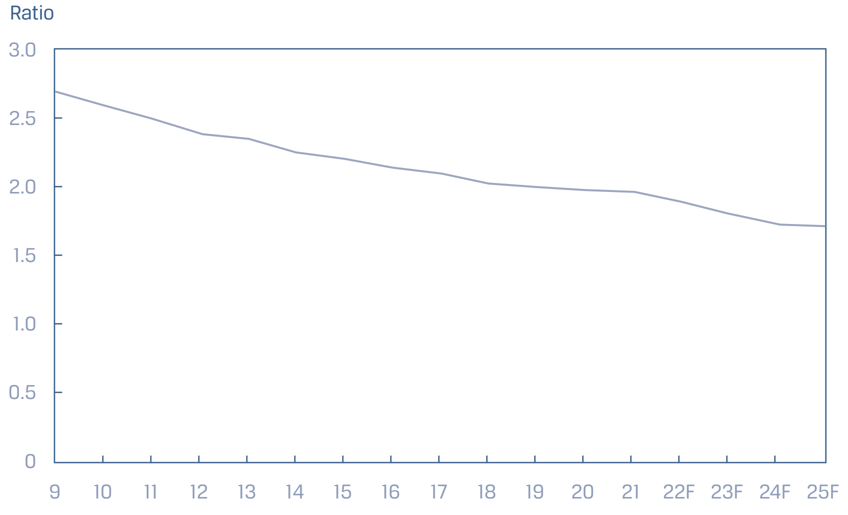
EXAMPLE 3

ESG Integration—Environmental Factors (Beverage Company)

Based in the United States, Frizzle Drinks (Frizzle) is a fictitious non-alcoholic beverage company that ranks among the largest in the world. Frizzle operates in both developed and emerging markets, including countries where water is scarce. Frizzle is a significant user of water in its operations. Given that water is a key ingredient in Frizzle's beverages, the continued availability of water is critical to the company's manufacturing process. Because of its extensive use of water, Frizzle faces ongoing regulatory scrutiny for pollution and effects on climate change. Ultimately, how Frizzle conserves and manages its water usage has implications for product pricing and company/brand reputation.

Sam Smith, CFA, is analyzing the effects of environmental factors on Frizzle's financial statements. Based on his research, Smith considers "water intensity" to be a key ESG metric for the beverage industry. Water intensity is defined as the ratio of total liters of water used per one liter of a beverage product. Exhibit 6 illustrates the trend of Frizzle's water intensity ratio from 2009 to 2021, as well as the consensus forecast ratio for the subsequent four years. Frizzle has steadily decreased its water usage over the past several years. From 2009 to 2021, its water intensity ratio declined by 27%. By the end of 2025(F), the company aims to reduce its water intensity by another 13%.

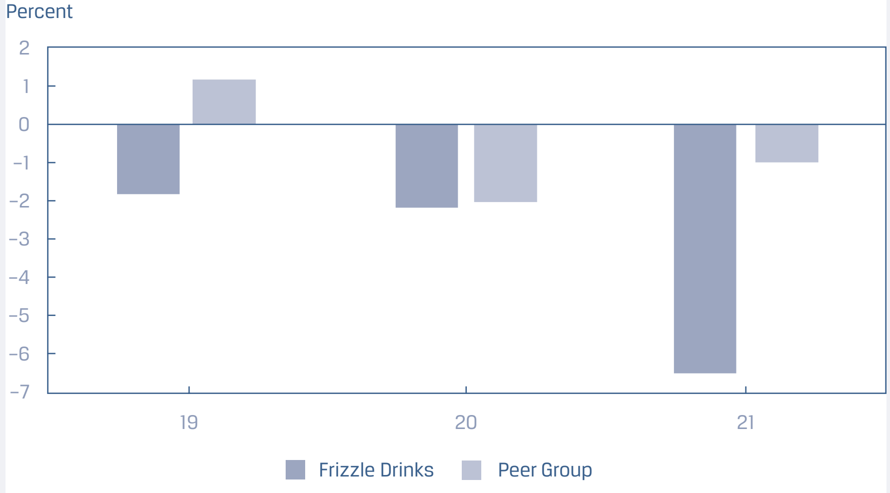
Exhibit 6 Water Intensity Ratio (in liters)



Note: (F) indicates forecast year.

Exhibit 7 compares the year-over-year change in Frizzle's water intensity ratio with that of its peer group over the past three years. To facilitate comparison among companies of varying sizes, Smith normalized the reported water intensity ratios by calculating the water intensity ratio per $1 million of revenue. Exhibit 7 illustrates that Frizzle's water intensity has decreased considerably relative to its peers over the past few years, particularly in the last reported year, 2021.

Exhibit 7 Water Intensity Ratio Change per $1 Million of Revenue



Next, Smith analyzes the effects of Frizzle's water intensity on its overall financial performance and compares it with the adjusted financial performance of its peers. As one example, Smith adjusts Frizzle's operating costs to account for the improved effects of water intensity (i.e., reduced usage). For the first projected year, 2022, Smith expects that Frizzle's cost of goods sold as a percentage of revenues (before any ESG adjustment) will be 40% and its peer group average will be 42%. For the same forecast period, Smith assumes that Frizzle's reduction in water intensity will result in a 1% reduction in its cost of goods sold/revenues, whereas the peer group average will remain the same. Exhibit 8 demonstrates this improvement in cost of goods sold/revenues on both an absolute and a relative basis. By extension, Exhibit 9 shows the absolute and relative improvement in Frizzle's gross margin (sales minus cost of goods sold) percentage.

Exhibit 8 Cost of Goods Sold as a Percentage of Revenue

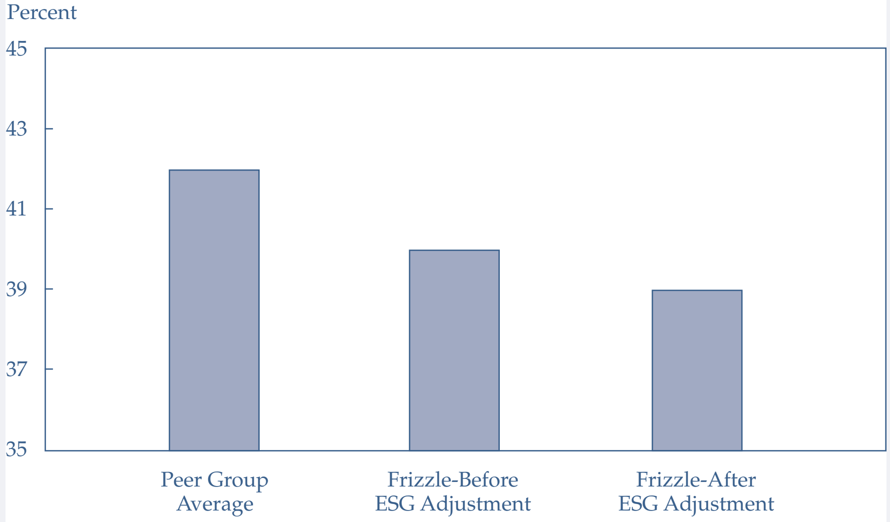
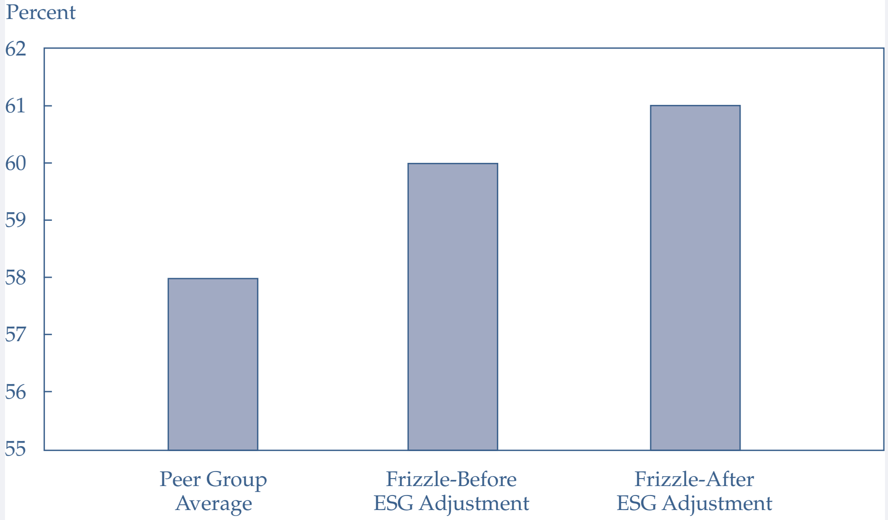


Exhibit 9 Gross Margin



In the last step of the integration analysis, Smith incorporates Frizzle's adjusted financial performance in valuing Frizzle's stock, bonds, and, if applicable, CDSs. In this example, Smith judges that Frizzle's lower cost of goods sold from the adjustment would result in higher forecast earnings and, all else being equal, a theoretically higher fair value for Frizzle's stock. With respect to Frizzle's bonds and CDSs, Frizzle's operating cash flow would improve through a lower cost of goods sold. When assessing the credit spreads of Frizzle's bonds and/or CDSs, Smith will analyze whether the lower relative ESG risk is already reflected in current spread levels and adjust accordingly.

EXAMPLE 4

ESG Integration—Social Factors (Pharmaceutical Company)

Well Pharma (Well) is a fictitious European pharmaceutical company that manufactures drug products for autoimmune diseases and immune disorders. Over the last five years, Well has had the weakest track record among its peers in terms of product recalls and regulatory warning letters for manufacturing and marketing-related violations. Specifically, the company has been subject to four major drug quality and safety scandals arising from adverse side effects. These scandals have resulted in lost sales, multiple lawsuits, and significant fines. Business disruptions, lawsuits, and fines have reduced revenues and increased costs for the company.

As Well's experience shows, product quality is a material social factor for pharmaceutical companies in general. Smith assumes that a drug company's product quality is a combination of the factors shown in Exhibit 10.

Exhibit 10 Social Factors—Pharmaceuticals

Factor	Description
Product Quality Controversies	Have there been any controversies linked to the company's product or service quality and responsibility?
Regulatory Warning Letters	Number of regulatory warning letters received by the company
Product Recalls	Number and severity of product recalls (voluntary and involuntary)
Regulatory Fines	Level of fines imposed by regulator linked to poor product quality and/or irresponsible behavior
Product Quality Certifications Percentage	Percentage of plants certified according to a widely accepted product safety/quality standard (e.g., ISO 9001 or equivalent)

Exhibit 11 shows the number of regulatory warning letters received, as well as product and marketing controversies faced, by Well and several peers. As the graph shows, Well has received significantly more of these letters than its peers have.

Exhibit 11 Regulatory Warning Letters and Product Quality Controversies

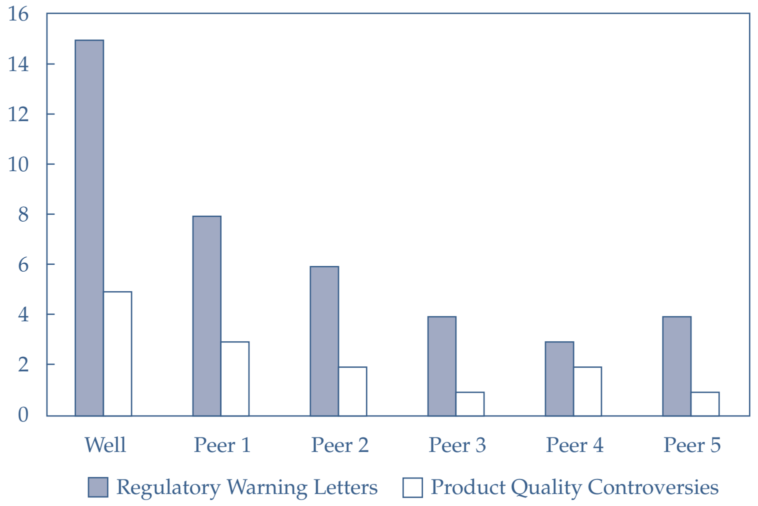


Exhibit 12 demonstrates how the factors listed in Exhibit 10 may affect the financial statements of Well and other pharmaceutical companies.

Exhibit 12 Social Factor Effects on Financial Performance

Factor	Financial Impact
Product Quality Controversies	Damage to brand value resulting in potential decrease in sales
Regulatory Warning Letters	Increased costs to comply with regulatory requirements
Product Recalls	Losses in sales revenue; increased costs of implementing product recalls
Regulatory Fines	Provisions for pharmaceutical sales returns and product-related litigation
Product Quality Certifications Percentage	Lower percentage increases risks of product quality issues, leading to product recalls and related costs

Based on these financial effects, Smith adjusts Well's projected revenues, operating expenses, and non-operating expenses. The nature of these financial statement adjustments will likely differ depending on whether Smith expects these product quality issues to be recurring or non-recurring in nature. Smith assumes that revenues will decrease by 2% over the next year because of existing product quality controversies. For operating expenses, Smith assumes that Well's cost of goods sold relative to revenues will increase by 1.3% to reflect product quality and additional investments in its manufacturing process. Exhibit 13 shows that Well's cost of goods sold as a percentage of revenues is in line with that of its peers, but the additional costs will increase this ratio well above that of the peer group. In addition to operating expenses, Smith forecasts that Well's non-operating expenses, such as restructuring charges, and other non-recurring costs will be an additional 4.5% of operating income. Exhibit 14 shows the current non-operating expense ratio for Well versus its peer group average, as well as the forecast amount.

Exhibit 13 Cost of Goods Sold as a Percentage of Revenue

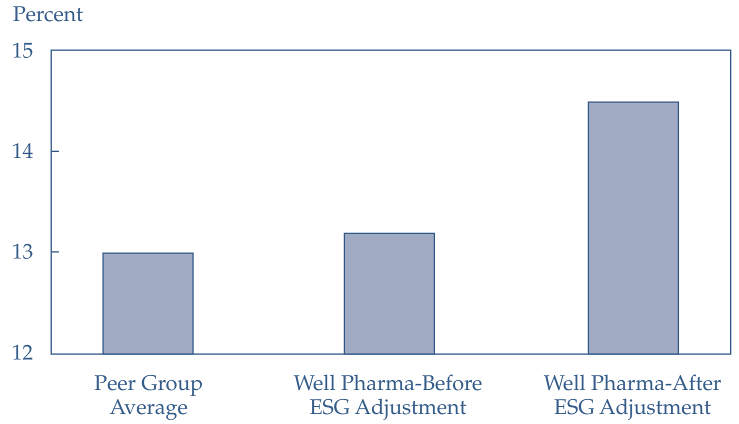
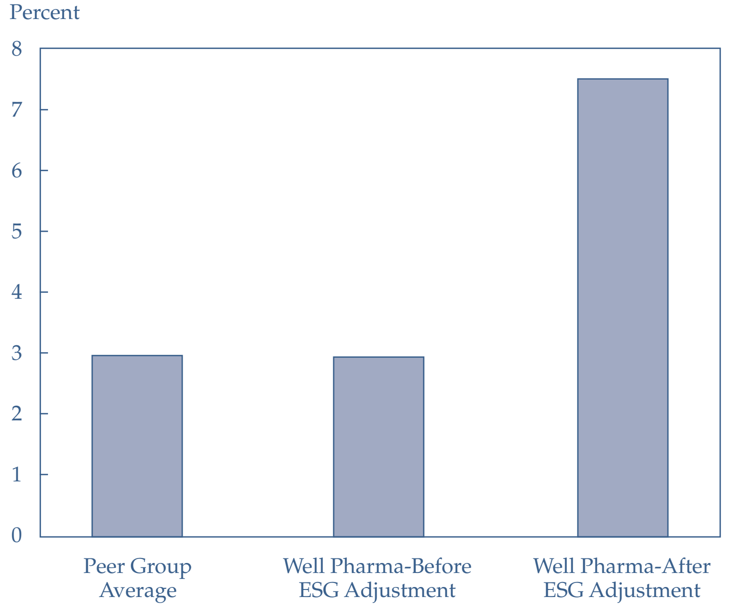


Exhibit 14 Non-Operating Expenses as a Percentage of Operating Income



Smith believes that the valuation implications for Well's stock and bonds could be significant based on its poor product quality and safety track record. Expectations of future poor performance could have a direct impact on earnings and cash flow to the detriment of both shareholders and bondholders. In addition, Smith believes there could be adverse valuation implications if investors view Well's brand value and reputation as impaired.

EXAMPLE 5

ESG Integration—Governance Factors (Bank Holding Company)

Sumiyoshi Banking Group (Sumiyoshi) is a fictitious Japanese bank holding company, with operations in Japan (80% of revenues), the United States, and Southeast Asia. Sumiyoshi's core businesses are commercial banking, leasing, securities, and consumer finance. As with most global banks, corporate governance reforms have become increasingly prominent for Sumiyoshi.

Smith has prepared Exhibit 15 to show how Sumiyoshi's board of directors compares with the majority of its domestic peer group, on the basis of governance factors discussed in Section 2 of this reading.

Exhibit 15 Corporate Governance Factors—Banks

	Domestic peer group	Sumiyoshi Bank
Board type	Two tier	Two tier
Board size, no. of directors	13	14
Total assets/director	JPY14.9 million	JPY13.3 million

(continued)

Exhibit 15 (Continued)

	Domestic peer group	Sumiyoshi Bank
CEO duality	Yes	Yes
Independent chairperson	Yes	No
Board independence %	47%	36%
Board gender diversity	17% female; 83% male	7% female; 93% male
Directors with long tenure (>10 years)	0%	14%
Number of board committees	5	4
Audit, nomination, remuneration, and risk committees in place?	Yes	Yes
Additional board committees?	Yes, governance committee	No
Non-executive directors with industry executive experience/total independent directors	67%	20%
Short-term and long-term incentive plan metrics disclosed?	No	No
Concentrated ownership	No single large shareholder	No single large shareholder
Say-on-pay provision	Yes	No
Straight voting	Yes	Yes
Dual-class shares	No	No

Smith notes that Sumiyoshi lags its peers in in several elements of board composition, such as the lack of an independent chairperson, a lower level of board independence and diversity, fewer board members with industry executive experience, and a number of board directors with long tenures. In addition to board composition, Smith uses credit risk as a proxy for a bank's corporate governance risk. In particular, Smith reviews one key banking credit measure—non-performing loans (NPLs). NPLs are loans that are not current in paying the contractual amounts that are due (i.e., interest or principal payments).

Smith analyzes Sumiyoshi's credit risk by dividing its NPLs by the amount of its total loans outstanding. Smith estimates that Sumiyoshi's ratio of NPLs to total loans is 50 bps higher than its peer group average, reflecting Sumiyoshi's comparatively weaker credit/governance risk. To account for the effect of higher credit risk than that of its peers, Smith may increase the risk premium embedded in his valuation of Sumiyoshi's stock. When valuing Sumiyoshi's corporate bonds, Smith might increase the credit spread relative to peers embedded in the company's outstanding issues.

SUMMARY

- Shareholder ownership structures are commonly classified as dispersed, concentrated, or a hybrid of the two.

- Dispersed ownership reflects the existence of many shareholders, none of which, either individually or collectively, has the ability to exercise control over the corporation. Concentrated corporate ownership reflects an individual shareholder or a group (controlling shareholders) with the ability to exercise control over the corporation.
- Controlling shareholders may be either majority shareholders or minority shareholders.
- Horizontal ownership involves companies with mutual business interests that have cross-holding share arrangements with each other. Vertical (or pyramid) ownership involves a company or group that has a controlling interest in two or more holding companies, which in turn have controlling interests in various operating companies.
- Dual-class (or multiple-class) shares grant one or more share classes superior or even sole voting rights while other share classes have inferior or no voting rights.
- Types of influential owners include banks, families, sovereign governments, institutional investors, group companies, private equity firms, foreign investors, managers, and board directors.
- A corporation's board of directors is typically structured as either one tier or two tier. A one-tier board consists of a single board of directors, composed of executive (internal) and non-executive (external) directors. A two-tier board consists of a supervisory board that oversees a management board.
- CEO duality exists when the chief executive officer also serves as chairperson of the board.
- A primary challenge of integrating ESG factors into investment analysis is identifying and obtaining information that is relevant, comparable, and decision-useful.
- ESG information and metrics are inconsistently reported by companies, and such disclosure is voluntary, which provides additional challenges for analysts.
- In an ESG context, materiality typically refers to ESG-related issues that are expected to affect a company's operations or financial performance and the valuation of its securities.
- Corporate governance considerations, such as the structure of the board of directors, tend to be reasonably consistent across most companies. In contrast, environmental and social considerations often differ greatly.
- Analysts typically use three main sources of information to identify a company's (or industry's) ESG factors: (1) proprietary research, (2) ratings and analysis from ESG data providers, or (3) research from not-for-profit industry organizations and initiatives.
- In equity analysis, ESG integration is used to both identify potential opportunities and mitigate downside risk, whereas in fixed-income analysis, ESG integration is generally focused on mitigating downside risk.
- A typical starting point for ESG integration is the identification of material qualitative and quantitative ESG factors that pertain to a company or its industry.

REFERENCES

CFA Institute/Principles for Responsible Investment. 2018. Guidance and Case Studies for ESG Integration: Equities and Fixed Income.

OECD. 2017. OECD Corporate Governance Factbook 2017. Paris: OECD Publishing.

PRACTICE PROBLEMS

The following information relates to Questions 1 and 2

Liz Kite is a research analyst for a global equity investment firm. She is conducting research on two publicly traded companies, Company A and Company B.

Company A has a large number of shareholders, with no single investor owning more than 5% of the outstanding shares. Company B is managed by a family who owns 60% of the outstanding shares. Both companies offer a single share class with equivalent voting rights.

1 **Determine** the relative level of risk (high or low) of principal–agent problems being present at each company. **Justify** your response.

2 **Discuss** drawbacks *most likely* associated with the ownership structure of Company B.

---

3 Clayton Streett is a consultant specializing in corporate governance issues. His current assignment is to evaluate the effectiveness of the board of Jess-Kait Worldwide. Streett seeks to determine if aspects of the company's corporate governance warrant a higher than average risk premium for the company's shares.

Jess-Kait Worldwide has a one-tier board structure, with the CEO serving as the chair of the board. Of the 20 board members, 8 are executive and 12 are non-executive (i.e., independent). The executive members primarily serve on the compliance and investment committees, while the non-executive members primarily serve on the audit, compensation, and board selection committees. The executive board members have all served on the board for more than 20 years, while the non-executive members' average tenure is only 5 years.

Determine which board considerations would *most likely* warrant a higher than average risk premium for the shares of Jess-Kait Worldwide. **Justify** each response.

i. Structure

ii. Independence

iii. Committees

iv. Composition

Determine which board considerations would *most likely* warrant a higher than average risk premium for the shares of Jess-Kait Worldwide. Justify each response.

Board Consideration	Higher Risk Premium?	Justification
i) Structure	Yes No	

(continued)

© 2020 CFA Institute. All rights reserved.

Determine which board considerations would *most likely* warrant a higher than average risk premium for the shares of Jess-Kait Worldwide. Justify each response.

Board Consideration	Higher Risk Premium?	Justification
ii) Independence	Yes No	
iii) Committees	Yes No	
iv) Composition	Yes No	

The following information relates to Questions 4 and 5

Chambers Carlisle was recently hired as an analyst for a fixed-income fund with a short-term investment horizon. Carlisle focuses on bonds in the materials sector and incorporates environmental, social, and governance (ESG) factors into his analysis. His previous employer was a buy-and-hold equity fund, where Carlisle included ESG factors in his analysis of equities in the materials sector.

4 **Discuss** how Carlisle's ESG analysis in his new position *most likely* differs from that in his previous position.

Carlisle identifies the ESG risk factors that are relevant to companies in the materials sector. He conducts his research by reviewing public documents, such as company annual reports, regulatory filings, and proxy reports.

5 **Discuss** the challenges *most likely* associated with Carlisle's research approach.

The following information relates to Questions 6 and 7

Dalanta Transportation operates a rail transportation business in the Southwest United States. Competition in the region for rail transport has intensified in the last few years, resulting in decreasing profit margins. Dalanta has received multiple warnings from state and federal regulators and faces potential fines for violating clean air regulations due to excessive greenhouse gas emissions (GHG) by the company's aging fleet of trains. As a result, the company has been the target of highly publicized criticism from environmental activist groups.

6 **Discuss** the potential effects of environmental, social, and governance (ESG) factors on Dalanta's financial performance.

Dalanta's CEO hires a consultant to help the company address the ESG issues. The CEO had planned to issue a conventional bond to finance the replacement of the oldest half of the fleet of trains, but the consultant recommends that Dalanta issue a green bond to finance the purchase of new trains.

7 **Discuss** *one* advantage and *one* disadvantage to Dalanta if the company follows the consultant's financing recommendation.

The following scenario applies to Questions 8–15

Karen Maghami is a portfolio manager for a mutual fund. Maghami, working with analyst Marcel Lynbrock, is researching two companies: Syvie Electric and EnileGEN. Located in the same country, both companies produce electricity from conventional and renewable sources. Maghami wants to integrate environmental, social, and governance (ESG) factors into the analysis of each company's financial statements.

Syvie Electric is a stock exchange-listed, state-owned enterprise (SOE). The sovereign government holds 58% of shares; remaining shares are publicly traded. EnileGEN is a public company owned primarily by institutional investors.

Lynbrock starts with an analysis of corporate governance factors. Relevant information about the two companies is summarized in Exhibit 1.

Exhibit 1 Select Corporate Governance Information: Syvie Electric and EnileGEN

	Syvie Electric	EnileGEN
Ownership Information and Voting Policies		
Percentage of shares owned by the largest shareholder	58%	9%
Straight voting	Yes	Yes
Board Information		
Board structure	One-tier	Two-tier
CEO duality	No	Yes
Number of directors	8	11
Percentage of directors with experience in industry	27%	88%
Remuneration Policies		
Clawback policy	No	Yes
Say-on-pay provision	Yes	No

After reviewing the corporate governance information, Maghami and Lynbrock focus on factors related to environmental and social considerations. Lynbrock asks Maghami, "What resource would you recommend using to find a list of ESG factors material to the electricity generation sector?"

Next, Maghami and Lynbrock discuss the effect of expected regulatory changes on stock and bond values for both companies. The government recently announced that it intends to authorize a new environmental regulation requiring that a minimum of 35% of electricity be produced from renewable sources. The exact timing of the new regulation is unknown, but EnileGEN already exceeds the minimum level by a significant margin. Maghami notes that the new regulation should give EnileGEN a competitive edge in the industry, and for scenario analysis purposes, she asks Lynbrock to assume that the regulation will take effect in two years. Lynbrock uses a discounted cash flow model to value EnileGEN stock under the assumption requested by Maghami.

Lynbrock believes that the new regulation will make three of Syvie Electric's coal-fired power stations no longer financially viable within 10 years following implementation. Maghami asks Lynbrock to estimate the potential effect on Syvie Electric's balance sheet and the potential impact on the value of Syvie Electric bonds.

Maghami reads in Syvie Electric's most recent financial statements that the company plans to issue green bonds. Maghami asks Lynbrock to evaluate any possible valuation risks or opportunities associated with green bonds. Lynbrock tells Maghami the following:

Statement 1	Green bonds offer higher protection for an investor because they typically are backed by the income derived from the environmental project they are used to fund.
Statement 2	Issuing green bonds ensures a lower cost of capital because the green feature typically results in a tighter credit spread compared with conventional bonds.
Statement 3	A unique risk related to green bonds is greenwashing, which is the risk that the bond's proceeds are not actually used for a beneficial environmental or climate-related project.

8 Based on Exhibit 1 and their ownership structures, market scrutiny of Syvie Electric's management is *most likely:*

A lower than EnileGEN.

B the same as EnileGEN.

C higher than EnileGEN.

9 Based on the ownership structures of Syvie Electric and EnileGEN, a principal–principal problem is *most likely* to occur in the case of:

A only EnileGEN.

B only Syvie Electric.

C both Syvie Electric and EnileGEN.

10 Based on Exhibit 1, which of the following statements is correct?

A Due to also being on the board, the EnileGEN CEO has two votes for each share owned.

B In the case of managerial misconduct, EnileGEN may be able to recover previously paid remuneration.

C Syvie Electric's corporate structure can be described as having concentrated ownership and dispersed voting power.

11 In response to Maghami's ESG question, Lynbrock should *most likely* recommend using:

A the SASB Materiality Map.

B the Green Bond Principles.

C each company's sustainability report.

12 To reflect the effect of the regulatory scenario requested by Maghami on EnileGEN's stock value, Lynbrock should:

A increase the credit spread.

B decrease the discount rate.

C increase projected operating costs.

13 To reflect the effect of the new regulation on Syvie Electric's balance sheet, Lynbrock should:

A increase terminal value.

B increase impaired assets.

C decrease capital expenditures.

14 Under the regulatory scenario requested by Maghami, Lynbrock should increase the credit spread of a Syvie Electric 10-year bond:

A less than the spread of a 2-year bond.

B by the same amount as the spread of a 2-year bond.

C more than the spread of a 2-year bond.

15 Which of Lynbrock's statements about green bonds is correct?

A Statement 1

B Statement 2

C Statement 3

SOLUTIONS

1 Risk of principal–agent problems:

- Company A has dispersed ownership and dispersed voting power. This results in weak shareholders and strong managers, which suggests a high risk of principal–agent problems.
- Company B has concentrated ownership and concentrated voting power. This results in strong shareholders and weak managers, which suggests a low risk of principal–agent problems.

Company A has a large number of shareholders, with no single shareholder owning more than 5% of the outstanding shares. The combination of dispersed ownership and dispersed voting power is generally associated with shareholders who lack the power to exercise control over managers. These shareholders are referred to as weak shareholders, and such managers are referred to as strong managers. Under this combination, there is a high risk that managers will seek to utilize a company's resources to pursue their own interests rather than those of the shareholders. This conflict is known as a principal–agent problem.

Company B is managed by a family who owns a majority of the outstanding voting shares. One of the benefits of family control through concentrated ownership and management is a low risk of principal–agent problems. Family control can, however, lead to principal–principal problems as the rights of minority shareholders may receive only modest consideration.

2 Drawbacks of Company B's ownership structure include the following:

- Poor transparency
- Lack of management accountability
- Modest consideration for minority shareholder rights
- Difficulty in attracting quality talent for management positions

Company B is managed by a family who owns a majority of the outstanding voting shares. Drawbacks to family ownership may include poor transparency, lack of management accountability, modest consideration for minority shareholder rights, and difficulty in attracting quality talent for management positions.

3

Determine which board considerations would *most likely* warrant a higher than average risk premium for the shares of Jess-Kait Worldwide. Justify each response.

Board Consideration	Higher Risk Premium?	Justification
i) Structure	Yes No	The structure of Jess-Kait's board is one-tier, with the CEO also serving as chair of the board. This CEO duality may raise concerns that the monitoring and oversight role of the board may be compromised. This negative attribute would likely warrant a higher than average risk premium.

Determine which board considerations would *most likely* warrant a higher than average risk premium for the shares of Jess-Kait Worldwide. Justify each response.

Board Consideration	Higher Risk Premium?	Justification
ii) Independence	Yes / **No**	Jess-Kait's board is comprised of 20 members, 12 of whom are independent. Having a majority of the board members be independent is a positive attribute and likely would not warrant a higher than average risk premium.
iii) Committees	Yes / **No**	When evaluating a board's committees, Streett should assess whether the key governance committees, such as the audit, compensation, and board selection committees, are sufficiently independent. With the non-executive board members primarily serving on these three important committees, there appears to be sufficient independence. This positive attribute likely would not warrant a higher than average risk premium.
iv) Composition	**Yes** / No	The executive board members have all served on the board for at least 20 years, which is a long tenure for board members. Long-tenured board members could become controlling, self-serving, or resistant to the introduction of new practices or policies that may be beneficial to stakeholders. This situation is especially troubling due to the fact that the non-executive board members' average tenure is only 5 years. This negative attribute would likely warrant a higher than average risk premium.

4 The investment horizon for the fixed-income securities in his new position is short term in nature, while his prior position had a long-term investment horizon.

ESG integration in fixed-income analysis generally focuses on mitigating downside risks, while ESG integration in equity analysis also includes identifying potential opportunities.

When deciding what ESG factors to consider in their analysis, analysts must consider the investment horizon. Some ESG issues may affect a company's performance in the short term, while other issues may affect it more in the long term. An investor with a short-term investment horizon may find that longer-term issues have little impact on a security's valuation in the near term. Since Carlisle's new job focuses on short-term fixed income, he is likely more concerned with short-term ESG issues than long-term ESG issues. In his previous position, the buy-and-hold nature of the equity fund implies a long-term investment horizon, so Carlisle would likely have considered both short-term and long-term ESG issues.

Carlisle's current position is in fixed-income analysis, while his previous position was in equity analysis. From a risk/reward perspective, the use of ESG integration typically differs for equity and fixed-income analysis. In equity analysis, ESG integration is used to both identify potential opportunities and mitigate downside risk, whereas in fixed-income analysis, ESG integration is generally focused only on mitigating downside risk.

5 Challenges:

- ESG information and metrics may be reported inconsistently by companies.
- ESG-related disclosures are voluntary for many companies, and the level of voluntary disclosure varies.

A primary challenge when integrating ESG factors into investment analysis is identifying and obtaining information that is relevant and useful. Carlisle's research approach involves reviewing public documents, such as company annual reports, regulatory filings, and proxy reports. A challenge he will face is that ESG information and metrics may be reported inconsistently by companies. Another challenge is that a number of ESG-related disclosures are voluntary for many companies, and the level of voluntary disclosure varies.

6 Potential effects:

- Increased costs to comply with regulatory requirements
- Potential fines for violating clean air regulations
- Damage to corporate reputation that could potentially decrease sales

Dalanta must address the clean air regulations. The costs to comply with these regulations could be significant, but the potential fines for failing to comply with these regulations could also be significant. These expenses would have a negative effect on Dalanta's financial performance. Finally, the company received highly publicized criticism from environmental activist groups. Such criticism could damage Dalanta's reputation and have a negative effect on sales, especially considering the increasingly competitive landscape for rail transport in the region.

7 Advantages and disadvantages:

Advantages:

- Green bonds can command a premium over comparable conventional bonds
- Lower cost of capital due to green bond premium

Disadvantages:

- Additional costs related to the monitoring and reporting of the use of the bond's proceeds
- Lack of liquidity of green bonds when purchased and held by buy-and-hold investors

Dalanta's financing of the new trains using green bonds may provide advantages compared to conventional bonds. Some green bonds can command a premium, or tighter credit spread, versus comparable conventional bonds due to market demand. This tighter credit spread could have a positive effect on Dalanta's cost of capital and valuation. However, issuing green bonds could result in Dalanta incurring additional costs related to the monitoring and reporting of the use of the bond's proceeds. In addition, liquidity risk is associated with green bonds that are purchased and held by buy-and-hold investors.

8 A is correct. Syvie Electric is a listed SOE with a mixed-ownership model in which 58% of the shares are owned by a sovereign government. The remaining shares trade on a public stock market. A mixed-ownership model tends to have lower market scrutiny of management than that of corporate ownership models. Additionally, Syvie Electric has a lower percentage of directors with industry experience compared to EnileGEN. A board with experience that is not directly related to the company's core operations may lack sufficient expertise to govern. EnileGEN's shares are owned primarily by institutional investors. Even though EnileGEN has dispersed ownership, institutional investors can promote good corporate governance by holding a company's board and management accountable when the board or management does not appear to be acting in the best interests of shareholders. Consequently, market scrutiny of Syvie Electric's management is most likely lower compared with EnileGEN.

9 B is correct. Syvie Electric's largest shareholder owns 58% of the shares; therefore, the company can be described as having concentrated ownership and concentrated voting power. The combination of concentrated ownership and concentrated voting power is generally associated with controlling shareholders maintaining a position of power over both managers and minority shareholders, known as a principal–principal problem. EnileGEN's largest stockholder owns 9% of shares; hence, the company can be classified as having dispersed ownership. EnileGEN also has a straight voting policy, which results in each of the shareholders having one vote per share; thus, EnileGEN also has dispersed voting power. The combination of dispersed ownership and dispersed voting power is generally associated with shareholders who lack the power to exercise control over managers. This conflict is known as a principal–agent problem. Consequently, the principal–principal problem is most likely to occur in Syvie Electric, not in EnileGEN.

10 B is correct. EnileGEN has a clawback policy, which allows a company to recover previously paid remuneration if certain events—such as financial restatements, misconduct, breach of the law, or risk management deficiencies—are uncovered. Consequently, EnileGEN may be able to recover previously paid remuneration from management in case of misconduct.

A is incorrect because even though EnileGEN's CEO also serves as chair of the board (CEO duality), the CEO does not hold dual-class shares because the company has a straight voting policy (i.e., one share, one vote). Consequently, the CEO of EnileGEN does not have two votes for each share.

C is incorrect because Syvie Electric has concentrated ownership and concentrated voting power (58% of shares are owned by a single shareholder). Consequently, Syvie Electric's ownership structure cannot be described as concentrated ownership with dispersed voting power.

11 A is correct. Maghami wants to find a resource that lists ESG factors material to the electricity generation sector. In an ESG context, materiality typically refers to ESG-related issues that are expected to affect a company's operations, its financial performance, and the valuation of its securities. The Sustainability Accounting Standards Board (the SASB) seeks to promote uniform accounting standards for sustainability reporting. For this purpose, the SASB developed the SASB Materiality Map, which lists relevant ESG-related, sector-specific factors that the organization deems to be material.

B is incorrect because the Green Bond Principles are voluntary standards to guide issuers in the determination of labeling a bond green. The Green Bond Principles do not provide a list of ESG-related factors that are material to a specific sector.

C is incorrect because companies' definitions of materiality in an ESG context may differ in usefulness. For example, some companies may use the term "material" in emphasizing positive ESG information even though such information may have little impact on the company's operations or financial performance. Other companies may minimize or not report negative ESG information that investors might consider material. Consequently, a company's sustainability report is not as likely as a SASB Materiality Map to provide a list of ESG-related factors that are material to a specific sector.

12 B is correct. For scenario analysis purposes, Maghami asks Lynbrock to assume that the new environmental regulation will take effect in two years. In equity security analysis, ESG-related factors are often analyzed in the context of forecasting financial metrics and ratios, by adjusting valuation model variables (e.g., discount rate), or using sensitivity and/or scenario analysis. Lynbrock should decrease the discount rate given that EnileGEN should have a competitive edge over its competitors due to the new environmental regulation.

A is incorrect because Lynbrock is estimating the expected effect of the new environmental regulation on EnileGEN's stock, not bond, value. Additionally, to justify a credit spread increase, the risk for a company would have to increase. Instead, the anticipated environmental regulation should give EnileGEN a competitive edge, thereby reducing risk.

C is incorrect because the new environmental regulation should not impact EnileGEN's operating costs. The company already exceeds the minimum percentage of renewably-sourced electric production expected to be required by the new regulation.

13 B is correct. Syvie Electric is facing risk from stranded assets—that is, assets that are obsolete or not economically viable owing to changes in regulatory or government policy. ESG-related adjustments to a company's balance sheet often reflect an analyst's estimate of impaired assets.

A is incorrect because adjusting terminal value applies to equity valuation; it is not an adjustment to a company's balance sheet.

C is incorrect because decreasing capital expenditures would be reflected in the income and cash flow statements, not the balance sheet.

14 C is correct. Syvie Electric is facing risk from stranded assets—that is, assets that are obsolete or not economically viable owing to changes in regulatory or government policy. The new environmental regulation is expected to be in effect in 2 years, so a 10-year bond will be more negatively impacted than a current 2-year bond. To reflect this, Lynbrock should increase the credit spread of the 10-year bond more than the spread of the 2-year bond.

15 C is correct. Statement 3 is correct because one unique risk of green bonds is greenwashing, which is the risk that the bond's proceeds are not actually used for a beneficial environmental or climate-related project. Greenwashing can result in an investor overpaying for a bond (if the investor paid a premium for the bond's green feature) or holding a bond that does not satisfy a prescribed environmental or climate investment mandate.

A is incorrect because Statement 1 is incorrect. Green bonds are typically similar to an issuer's conventional bonds, with the exception that the bond proceeds are earmarked for green projects. Green bonds normally have the same credit ratings and bondholder recourse as conventional bonds of the same issuer (all else being equal).

B is incorrect because Statement 2 is incorrect. It is true that some green bonds command a premium, or tighter credit spread, versus comparable conventional bonds because of high market demand. This would reduce a company's cost of capital. However, issuing green bonds does not always ensure a lower cost of capital for a company because green bonds are typically similar to an issuer's conventional bonds. Green bonds normally have the same credit ratings and bondholder recourse as conventional bonds of the same issuer. Furthermore, green bonds may be less liquid than conventional bonds, such that they don't command a premium from investors. Green bonds also have higher reporting and monitoring costs than conventional bonds, thus potentially resulting in higher costs for the issuer.

READING

18

Mergers and Acquisitions

by Rosita P. Chang, PhD, CFA, and Keith M. Moore, PhD, CFA

Rosita P. Chang, PhD, CFA, is at Shidler College of Business, University of Hawaii at Manoa (USA). Keith M. Moore, PhD, CFA (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** classify merger and acquisition (M&A) activities based on forms of integration and relatedness of business activities;
☐	**b.** explain common motivations behind M&A activity;
☐	**c.** explain bootstrapping of EPS and calculate a company's post-merger EPS;
☐	**d.** explain, based on industry life cycles, the relation between merger motivations and types of mergers;
☐	**e.** contrast merger transaction characteristics by form of acquisition, method of payment, and attitude of target management;
☐	**f.** distinguish among pre-offer and post-offer takeover defense mechanisms;
☐	**g.** compare the discounted cash flow, comparable company, and comparable transaction analyses for valuing a target company, including the advantages and disadvantages of each;
☐	**h.** evaluate a takeover bid and its effects on the target shareholders versus the acquirer shareholders;
☐	**i.** explain how price and payment method affect the distribution of risks and benefits in M&A transactions;
☐	**j.** describe characteristics of M&A transactions that create value;
☐	**k.** distinguish among equity carve-outs, spin-offs, split-offs, and liquidation;
☐	**l.** explain common reasons for restructuring.

© 2020 CFA Institute. All rights reserved.

1 INTRODUCTION

Companies enter into corporate restructuring activities such as mergers and acquisitions for a variety of reasons. Many companies use mergers as a means to achieve growth or gain operational efficiencies. Others seek to acquire unique capabilities or resources, increase market power, or diversify their businesses. In all cases, it is important for both equity and fixed income analysts to understand the motives for mergers and their financial and operational consequences.

Merger and acquisition (M&A) activities involve a variety of complexities and risks. Analysts should assess these factors, including the expected value arising from a proposed business combination relative to deal price, in addition to the likelihood that the combination will take place and the intended results will be achieved. They should also consider actions taken by other shareholders, regulators, market participants, and competitors, in addition to transaction specifics, given the potential impact of all these factors on deal completion and valuation.

Analysts should be able to answer questions such as: Does the proposed transaction make economic sense and align with management's stated business strategy? How is the transaction being financed, and how will this financing approach affect company financials? How likely is the deal to take place, and what are associated risks? In the case of a hostile bid, does the target company have options to successfully fend off the unwanted bid? What is the anticipated impact for shareholders? Bondholders?

EXAMPLE 1

Takeda–Shire Acquisition

On 5 December 2018, shareholders of Takeda Pharmaceutical, a 237-year-old company and Japan's largest pharmaceutical firm, approved the acquisition of Shire, a UK-listed, Irish-headquartered drug maker of similar size with an operational base in the United States. Announced in May 2018, the US$62 billion deal was aimed at transforming Takeda into a top 10 global pharmaceutical company with a sizable US footprint. It was the largest deal announced worldwide for 2018.

The deal was also the largest international acquisition ever attempted by a Japanese company. Takeda's offer consisted of shares and cash. It also involved a US$30 billion increase in balance sheet debt. Takeda's CEO undertook a nine-month-long campaign to secure the deal. The deal was confirmed in December 2018 with 88% Takeda shareholder approval, despite significant resistance from some Takeda shareholders, including the original Takeda family, who were concerned about the additional risk created by the acquisition.

Takeda–Shire Deal Timeline

Strategy	
September 2016	Takeda's "Vision 2025" business plan envisions sustaining growth year over year.
Rumor	
28 March 2018	Takeda says "considering making an approach to Shire."
6 April	Takeda, Pfizer, and Amgen viewed as likely bidders for Shire.

(Continued)

Courtship	
29 March 2018	Takeda initial offer: £44/share, with £16 in cash.
11 April	Takeda second offer: £45.5/share, £16.75 cash.
13 April	Takeda third offer: £46.5/share, £17.75 cash.
14 April	Shire's board announces it unanimously rejects third offer.
20 April	Takeda fourth offer: £47/share, £21 cash.
24 April	Takeda fifth offer: £49/share, £21.75 cash.
Acquisition	
8 May 2018	Shire accepts fifth offer of £49/share, £21.75 cash, valuing deal at US$62 billion.
14 May	Takeda announces asset disposals to help fund the acquisition.
10 July	US Federal Trade Commission grants regulatory approval for Shire deal.
14 September	Chinese regulator approves Shire deal.
18 October	Japanese regulator approves Shire deal.
20 November	EU regulator approves Shire deal.
20 November	Proxy advisors Glass, Lewis & Co. and Institutional Shareholder Services recommend deal.
May–November 2018	Takeda minority shareholders (including Takeda family) lobby against proposed deal, citing increased risk and loss of Japanese "roots."
Integration	
5 December 2018	Takeda shareholders vote to approve deal.
7 January 2019	Takeda CEO announces integration "well underway."
8 January 2019	Deal closes.
Value creation	
14 May 2019	CEO announces 43% increase in original cost synergies anticipated from acquisition.
December 2019	Takeda's core earnings margin expected to reach 25%, with long-term earnings margin forecast for 35%, up from an earnings margin of 22% in 2018.
March 2020	Takeda sales expected to rise 57% year over year.
Projection for 2023	Divestments of US$10 billion expected to stabilize balance sheet, with the expansion of five key businesses including in neuroscience and in treatments for cancer and rare diseases.

Takeda's motivations for the deal included the potential for sustainable earnings growth believed to be available through Shire's products and markets. Takeda was at that time projecting declining profitability from existing products and limited growth in the Japanese market. Additionally, synergies through new product development, earnings diversification, and cost savings were anticipated from the deal. Despite positive expectations, in the nine months between Takeda's announcement in March 2018 and the time the deal closed, Takeda's share price dropped 26% while Shire's rose 33%.

Following increasing offers in value by Takeda, a series of regulatory approvals post–Shire acceptance, and announcements of a divestments program to help fund the acquisition, the deal finally closed in January 2019. By May 2019, Takeda's management was reporting significant cost savings and projecting major improvements to core earnings margins in the years ahead.

This reading will discuss many of the issues involved in M&A deals, such as form of payment in a merger, legal and contractual issues, and the necessity for regulatory approval. Most importantly, this reading aims to equip you with basic tools and a framework for analyzing such deals. In subsequent sections, we will discuss basic types of M&A, underlying motives, transaction characteristics, governing regulations, and how to evaluate a target company and a proposed merger.

2 MERGERS AND ACQUISITIONS: DEFINITIONS AND CLASSIFICATIONS

a identify merger and acquisition (M&A) activities based on forms of integration and relatedness of business activities;

Business combinations come in different forms. A distinction can be made between acquisitions and mergers. In the context of M&A, an **acquisition** is the purchase of some portion of one company by another. An acquisition might refer to the purchase of assets from another company, the purchase of a definable segment of another entity, such as a subsidiary, or the purchase of an entire company, in which case the acquisition would be known as a merger. A **merger** represents the absorption of one company by another. That is, one of the companies remains and the other ceases to exist as a separate entity. Typically, the smaller of the two entities is merged into the larger, but that is not always the case.

Mergers can be classified by the form of integration. In a **statutory merger**, one of the companies ceases to exist as an identifiable entity and all its assets and liabilities become part of the purchasing company. In a **subsidiary merger**, the company being purchased becomes a subsidiary of the purchaser, which is often done in cases where the company being purchased has a strong brand or good image among consumers that the acquiring company wants to retain. A **consolidation** is similar to a statutory merger except that in a consolidation, *both* companies terminate their previous legal existence and become part of a newly formed company. A consolidation is common in mergers where both companies are approximately the same size.

The parties to a merger are often identified as the target company and the acquiring company. The company being acquired is the **target company**, or simply the **target**. The company acquiring the target is called the **acquiring company**, or the **acquirer**. We will use this terminology throughout the reading.

In practice, many of the terms used to describe various types of transactions are used loosely such that the distinctions between them are blurred. For example, the term "consolidation" is often applied to transactions in which the entities are about the same size, even if the transaction is technically a statutory merger. Similarly, mergers are often described more generally as **takeovers**, although that term is often reserved to describe **hostile transactions**, which are attempts to acquire a company against the wishes of its managers and board of directors. A **friendly transaction**, in contrast, describes a potential business combination that is endorsed by the managers of both companies, although that is certainly no guarantee that the merger will ultimately occur.

An additional way that mergers are classified is based on the relatedness of the merging companies' business activities. Considered this way, there are three basic types of mergers: horizontal, vertical, and conglomerate.

A **horizontal merger** is one in which the merging companies are in the same kind of business, usually as competitors. The 2017 merger between leading chemical companies Dow Chemical (full name: Dow Chemical Company) and DuPont (full name: E.I. du Pont de Nemours and Company) is one example of a horizontal merger. Another example is Disney's acquisition of 21st Century Fox, also in 2017. One of the great motivators behind horizontal mergers is the pursuit of **economies of scale**, which are savings achieved through the consolidation of operations and elimination of duplicate resources. Another common reason for horizontal mergers is to increase market power, because the merger results in a reduction of the number of industry competitors and an increase in the size of the acquiring company.

In a **vertical merger**, the acquirer buys another company in the same production chain, such as a supplier or a distributor. In addition to cost savings, a vertical merger may provide greater control over the production process in terms of quality or procurement of resources or greater control over the distribution of the acquirer's finished goods. If the acquirer purchases a target that is ahead of it in the value chain (a supplier), it is called **backward integration**. An example of backward integration is if a steel manufacturer purchases an iron ore mining company. When an acquirer purchases a company that is further down the value chain (a distributor), it is called **forward integration**. An example of forward integration is the 2018 acquisition of palliative care provider Aspire Health by Anthem, the health insurance provider operating Blue Cross and Blue Shield insurance plans in the United States. By bringing palliative care and support services to the insurer's enrollees and their families, the merger brought the provision of medical insurance and community-based medical care services into one integrated company.

When an acquirer purchases another company that is unrelated to its core business, it may be called a **conglomerate merger**. Amazon's acquisition of Whole Foods in 2017 is an example of a conglomerate merger, in which the leading online retailer acquired a "bricks and mortar" grocery store chain. The concept of company-level diversification has been used as a rationale for inter-industry mergers; by investing in companies from a variety of industries, an organization may reduce its total cash flow volatility. Since the 1990s, however, conglomerates have been viewed as inefficient managers of capital. As a result, many conglomerates have been broken up and fewer conglomerate mergers take place. As we will discuss in the section on merger motivations, company-level diversification is not necessarily in the shareholders' best interests.

MOTIVES FOR MERGERS

3

b explain common motivations behind M&A activity;

c explain bootstrapping of EPS and calculate a company's post-merger EPS;

d explain, based on industry life cycles, the relation between merger motivations and types of mergers;

In the previous section, we mentioned some of the basic motives behind mergers, such as the search for economies of scale (in a horizontal merger) or cost savings through integration (in a vertical merger). In this section, we will expand on this topic and survey some of the reasons companies merge—the motives or rationales for merger.

Most often, mergers are pursued because of perceived economic benefits that will result, although in some cases, management may have other reasons for undertaking a merger.

The topic is important because in assessing a proposed combination, investors and analysts need to carefully evaluate the rationale behind the merger. Does the underlying business strategy and management's stated rationale make sense? Is the merger likely to create value? What is the probability that each of the stated goals for the merger will be attained? Keep in mind that many motives are interrelated and that there are typically several motives, both acknowledged and tacit, behind any given merger.

3.1 Synergy

Among the most common motivations for a merger is the creation of synergy, in which the whole of the combined company will be worth more than the sum of its parts. Generally speaking, synergies created through a merger will either reduce costs or enhance revenues. Cost synergies are typically achieved through economies of scale in research and development, procurement, manufacturing, sales and marketing, distribution, and administration. Revenue synergies are created through the cross-selling of products, expanded market share, or higher prices arising from reduced competition. For example, a bank that acquires its competitors can both increase its market share and realize operating efficiencies by closing duplicate branches and integrating back-office operations.

3.2 Growth

Corporate managers are under constant pressure to grow their companies' revenues, and they often turn to M&A activity to achieve that growth. Companies can grow either by making investments internally (i.e., **organic growth**) or by buying the necessary resources externally (i.e., **external growth**). It is typically faster for companies to grow externally. Growth through M&A activity is common when a company is in a mature industry. For example, the global oil industry is a mature industry, and BP, ExxonMobil, and Chevron Corporation have increased their reserves and output by acquiring smaller competitors.

External growth can also mitigate risk. It is considered less risky to merge with an existing company than to enter an unfamiliar market and establish the resources internally. In the months leading up to 2017, a surge in outbound cross-border M&A transactions (amounting to US$1 trillion in total) occurred in China, many motivated by the desire of Chinese companies to establish footholds in international markets.

3.3 Increasing Market Power

In industries with few competitors or in which market share is sufficiently concentrated, horizontal integration may be a means to increase market power. When a company increases its market power through horizontal mergers, it may develop greater ability to influence market prices. Taken to an extreme, horizontal integration results in a monopoly.

Vertical integration may also result in increased market power. Vertical mergers can lock in a company's sources of critical supplies or create captive markets for its products. Imagine, for example, an industry in which one company supplies raw materials to two separate manufacturing companies. If one of the manufacturers were to acquire the raw materials provider, the acquirer would be in a position to influence industry output and ultimately prices. As we will discuss further in the section on competition law (also referred to as antitrust law in the United States), government regulators routinely block both horizontal and vertical mergers that sufficiently reduce competition in an industry and concentrate market power in the hands of too few companies.

3.4 Acquiring Unique Capabilities and Resources

Many companies undertake a merger or an acquisition either to pursue competitive advantages or to shore up lacking resources. When a company cannot cost-effectively create internally the capabilities needed to sustain its future success, it may seek to acquire them elsewhere. For example, a company may engage in M&A activity in order to acquire specific competencies or resources it lacks, such as a strong research department, nimble sales force, intellectual capital, or creative talent.

3.5 Unlocking Hidden Value

A potential target company may be uncompetitive over a sustained period for a host of reasons, such as poor management, lack of resources, high legacy costs, or poor organizational structure. When a potential target is underperforming, an acquirer may believe it can acquire the company cheaply and then unlock hidden value through reorganization, better management, or synergy. If the target has been underperforming significantly, the acquirer may even believe it can obtain the company for less than its breakup value. A company's **breakup value** is the value that can be achieved if a company's assets are divided and sold separately.

Sometimes mergers are conducted because the acquirer believes that it is purchasing assets for less than their replacement cost. For example, a pharmaceutical company may believe it can acquire another company's research more cheaply than through undergoing a lengthy development process of its own. Or, an oil company may believe it will be less expensive to acquire another oil company's assets than to find and develop its own additional reserves.

3.6 Tax Considerations

It is possible for a profitable acquirer to benefit from merging with a target that has accumulated a large amount of tax losses. Instead of carrying the tax losses forward, the merged company would use the tax losses to immediately lower its tax liability. In many countries, the taxing authority disallows an offset when the primary reason for the merger is tax avoidance. Mergers are typically conducted for a variety of reasons, however, and it is difficult for regulatory authorities to prove that tax considerations are a primary motivator.

3.7 Cross-Border Motivations

The growth of cross-border deals was high during the 1990s, as international M&A became a popular strategy for developed country–based multinational companies seeking to extend their market reach, acquire new manufacturing facilities, develop new sources of raw materials, and tap into the capital markets. Subsequently, significant international M&A activity took place in the other direction, as companies in countries whose economies had grown significantly as a result of this globalization sought to extend their own market reach for the same reasons. In addition to the various factors that drive domestic mergers, cross-border mergers can provide an efficient way of achieving other international business goals.

3.7.1 *Exploiting Market Imperfections*

Cross-border transactions can enable companies to more fully exploit market imperfections. For example, to take advantage of differences in the relative cost of labor, a manufacturer may purchase a company in a country with comparatively lower labor costs.

3.7.2 *Overcoming Adverse Government Policy*

Cross-border mergers can also offer a means by which to overcome disadvantageous government policy—for example, to circumvent protective tariffs, quotas, or other barriers to free trade.

3.7.3 *Technology Transfer*

Companies that possess a new or superior technology may make acquisitions abroad in order to open new markets or otherwise more fully exploit their business advantage. Conversely, it is common for a company to purchase a non-domestic company that possesses a new or superior technology in order to enhance the acquirer's competitive position both at home and abroad.

3.7.4 *Product Differentiation*

Companies often purchase companies in other nations in order to exploit the advantages of having a highly differentiated line of products. Similarly, buying certain intangibles, such as a good reputation, helps to ensure success in the global market.

3.7.5 *Following Clients*

Companies may engage in a cross-border merger to follow and support domestic clients more effectively. For example, since 2005, many Chinese banks have established a presence outside Greater China to provide services abroad to their domestic clients as Chinese companies increased their activity in international markets.

EXAMPLE 2

Takeda–Shire Acquisition Motives

The Takeda–Shire acquisition involved a number of motivations described so far in this section.

Synergy. Takeda expected significant synergies from the deal. On a long-term basis, core earnings margins were anticipated to reach 35%, up from 22% in 2018. Synergies and margin improvements also arose from post-merger disposals of low-revenue-generating products.

New product growth. Traditionally, companies in the pharmaceutical sector have a robust new product pipeline. Takeda had only two new major drugs, in late-stage clinical trials, in its pipeline. The Shire acquisition allowed cash flow from existing Shire products to bolster revenues, allowing time and investment in new product development for the next five years.

Market power. Takeda's strategy was to become a top 10 global pharmaceutical company by acquiring Shire, a company of equivalent size. By doing this, Takeda believed it could gain market power, enabling it to compete longer term in the global pharmaceutical market.

Unique capabilities and resources. With its extremely limited new product pipeline, and sales of existing drugs slowing substantially in response to increasing competition from generic drugs, revenue expectations for Takeda were expected to show significant decline without the deal. The Shire acquisition brought in a portfolio of lucrative rare disease pharmaceuticals, bolstering the combined group's revenue forecasts.

Cross-border motivations. An established company in Japan for more than two centuries, Takeda was also Japan's largest company in the industry. Nevertheless, its CEO believed that to be a successful pharmaceutical company, Takeda needed to be successful outside Japan because Japan accounted for only a small fraction of the global pharmaceutical market. With downward pressure on Japanese drug prices in response to health ministry policies, the Shire acquisition resulted in more than half of Takeda's revenues coming from the US market, where significant growth in drug markets and margins was expected.

3.8 Diversification

Companies sometimes cite diversification as one of the motives behind a merger. The idea behind company-level diversification is to treat the company as a portfolio of investments in other companies. If a conglomerate invests in companies from a variety of industries, then the variability of the conglomerate's total cash flows should decrease, at least to the extent that the industries are uncorrelated.

Although this motive may seem rational, typically it is not in the best interests of the conglomerate's shareholders. In a well-functioning capital market, investors can diversify their own portfolios more easily and at less expense than a company can. Additionally, the desire to diversify has led some companies to lose sight of their major competitive strengths and to expand into businesses in which they lack comparative advantages.

3.9 Managers' Personal Incentives

Various managerial-related theories for mergers have been developed over the years based on evidence of agency problems. **Managerialism theories** posit that because executive compensation is highly correlated with company size, corporate executives are motivated to engage in mergers to maximize the size of their company rather than shareholder value. Additionally, corporate executives may be motivated by self-aggrandizement. For example, being the senior executive of a large company conveys great power and prestige.

3.10 Bootstrapping Earnings

Even when there are no reasons to believe that synergies or growth would result from a merger, it is possible to create the illusion of such synergies or growth. When a company's earnings increase as a consequence of the merger transaction itself rather than because of resulting economic benefits of the combination, it is referred to as the "bootstrap effect" or "bootstrapping earnings." The bootstrap effect occurs when the acquiring company's shares trade at a higher P/E than those of the target and the acquirer's P/E does not decline following the merger.

EXAMPLE 3

Bootstrapping Earnings

Assume two companies are planning a merger. Company A is the acquirer, Company T is the target, and Company A* is the post-merger combination of the two companies. The companies' stock prices and earnings per share are as shown in the following table. Note that the acquirer has a P/E of 25.0 and the target has a P/E of 20.0:

	A	T	A*
Stock price	\$100.00	\$50.00	
EPS	\$4.00	\$2.50	\$4.20
P/E	25.0	20.0	
Total shares outstanding	100,000	50,000	125,000
Total earnings	\$400,000	\$125,000	\$525,000
Market value of equity	\$10,000,000	\$2,500,000	

Given its stock price, the acquirer can issue 25,000 of its own shares and use the proceeds to buy the target company. This amount is determined by dividing the target's market value by the acquirer's stock price (\$2,500,000/\$100 = 25,000). The total shares outstanding of the merged company will be 125,000—the acquirer's initial 100,000 shares plus the 25,000 shares that the acquirer issued to purchase the target. After the merger, the company's combined earnings are divided by the number of shares outstanding to determine the new EPS (\$525,000/125,000 = \$4.20), which is \$0.20 per share higher than the acquirer would have reported without the merger.

If the acquirer's pre-merger stock price had been \$80 instead of \$100, then A's pre-merger P/E would have been 20.0 (\$80/\$4.00). Under that scenario, the acquirer would have issued 31,250 shares to purchase the target. The EPS of the merged company would then have been \$525,000/131,250 = \$4.00, thus illustrating that for bootstrapping to work, the acquirer's P/E must be higher than the target's P/E.

If the market is efficient, the post-merger P/E should adjust to the weighted average of the two companies' contributions to the merged company's earnings. In the previous example, the merged company's P/E would be about 23.8, which implies that the acquirer's stock price would remain at \$100. If, however, the acquiring company's P/E is higher than the target's and management can convince investors to value the merged company using the acquirer's pre-merger P/E, then the stock price of the new company should rise. If the acquirer bootstraps earnings to \$4.20 per share as shown in the preceding example, then the share price should increase to \$105 if investors apply the pre-merger P/E of 25.0 times earnings (\$4.20 × 25.0 = \$105). When there are no expected gains from synergy or other factors, such share price increases are not expected. The market usually recognizes the bootstrapping effect, and post-merger P/Es adjust accordingly.

3.11 Mergers and the Industry Life Cycle

The types of mergers (e.g., horizontal, vertical, or conglomerate) occurring in an industry, as well as the motivations behind those mergers, will vary over time as an industry proceeds through its life cycle. The stages in an industry life cycle are normally categorized by their rates of growth in sales with five growth stages that vary in length. These stages are as follows:

1 Embryonic
2 Growth
3 Shakeout
4 Mature
5 Decline

In the embryonic stage, the industry exhibits substantial development costs and has low, but slowly increasing, sales growth. In this phase, younger, smaller companies may sell themselves to larger companies in mature or declining industries that are looking for ways to enter into a new growth industry. Alternatively, young companies may look to merge with companies that provide them with opportunities to pool management and capital resources. Types of mergers commonly occurring in this stage are conglomerate and horizontal.

When growth is rapidly accelerating, the industry exhibits high profit margins caused by a few market participants. Explosive growth in sales may necessitate large capital requirements to expand existing capacity. Conglomerate and horizontal mergers commonly occur in the rapid growth stage.

In the shakeout stage, the industry experiences a drop in the entry of new competitors, but growth potential remains. Horizontal and vertical mergers may be undertaken to achieve economies of scale, savings, and operational efficiencies.

Upon market maturity, the industry faces increasing competition and capacity constraints. Mergers may be undertaken to achieve economies of scale in research, production, and marketing to match the low costs and price performance of other companies (both domestic and international). Large companies may acquire smaller ones in order to improve management and provide a broader financial base. Horizontal mergers are characteristic of this stage.

In the decline stage, the industry faces overcapacity and eroding profit margins. Horizontal mergers may be undertaken to ensure survival, and vertical mergers may be carried out to increase efficiency and profit margins. Companies in related industries may merge to exploit synergy. Companies in this industry may acquire companies in young industries. Horizontal, vertical, and conglomerate mergers are most common to this stage.

MERGER TRANSACTION CHARACTERISTICS 4

e contrast merger transaction characteristics by form of acquisition, method of payment, and attitude of target management;

The specifics of M&A transactions can vary along many dimensions, including the form of acquisition, financing, timing, control and governance, accounting choices, and numerous details ranging from the post-merger board composition to the location of the new headquarters. In this section, we will focus on the form of acquisition, method of payment, and mindset of target management. These three characteristics play a large role in determining how the transaction will occur, which regulatory rules might apply, how the transaction will be valued, and how it may be taxed.

4.1 Form of Acquisition

There are two basic forms of acquisition: An acquirer can purchase either the target's stock or its assets. The decision will have several consequences, as summarized in Exhibit 1.

Stock purchases are the most common form of acquisition. A **stock purchase** occurs when the acquirer gives the target company's shareholders some combination of cash and securities in exchange for shares of the target company's stock. For a stock purchase to proceed, it must be approved by at least 50% of the target company's shareholders and sometimes more, depending on the legal jurisdiction. Although it can be difficult and time consuming to win shareholder approval, it also stands as an opportunity to circumvent the target company's management in cases where management opposes the merger.

In an **asset purchase**, the acquirer purchases the target company's assets and payment is made directly to the target company. One advantage of this type of transaction is that it can be conducted more quickly and easily than a stock purchase because shareholder approval is not normally required unless a substantial proportion of the assets is being sold, usually more than 50%. Another advantage is that an acquirer can focus on buying the parts of a company of particular interest, such as a specific division, rather than the entire company.

Exhibit 1 Major Differences of Stock versus Asset Purchases

	Stock Purchase	Asset Purchase
Payment	Target shareholders receive compensation in exchange for their shares.	Payment is made to the selling company rather than directly to shareholders.
Shareholder approval	Shareholder approval required.	Shareholder approval might not be required.
Liabilities	Acquirer assumes the target's liabilities.	Acquirer generally avoids the assumption of liabilities.
Complexity	Simpler; less likely to require complex amendment of asset-related legal documents.	More complex; more likely to require reassignment, novation, or renegotiation of contracts, etc.
Regulatory approval	More likely to be subject to regulatory scrutiny and approval.	Less likely to involve regulatory approval.

Another key difference relates to the assumption of liabilities. In stock purchases, the acquiring company assumes the target company's liabilities. Acquiring companies must thus be on guard to avoid assuming unexpected or undisclosed liabilities. With asset purchases, acquiring companies generally avoid assuming the target's liabilities. Purchasing substantially all of a company's assets instead of conducting a stock purchase so as to specifically avoid assuming liabilities is fraught with legal risk, however, because courts have tended to hold acquirers responsible for the liabilities in these cases.

Some of the more dramatic consequences of the decision to pursue one form of acquisition versus another concern taxation. In some countries, in a stock purchase, the target company's shareholders exchange their shares for compensation and must pay tax on their gains, but there are no tax consequences at the corporate level. In some countries, for an asset purchase, there are no direct tax consequences for the target company's shareholders, but the target company itself may be subject to corporate taxes. The complexity of M&A deals, coupled with the complexity and variability of tax laws in different jurisdictions, means tax treatment varies widely.

In addition to shifting the basic tax burden, the form-of-acquisition decision plays a role in determining how tax rules are applied in accounting for the merger. For example, use of a target's accumulated tax losses is allowable in many countries (such as Argentina, Australia, France, Singapore, South Africa, and the United States) for stock purchases but not for asset purchases.

4.2 Method of Payment

The acquirer can pay for the merger with cash, securities, or some combination of the two in what is called a **mixed offering**. In a **cash offering**, the cash might come from the acquiring company's existing assets or from a debt issue. In the most general case of a **securities offering**, the target shareholders receive shares of the acquirer's common stock as compensation. In a consolidation, the target company's shareholders may receive new shares in the surviving entity. Instead of common stock, however, the acquirer might offer other securities, such as preferred shares or even debt securities.

In a stock offering, the **exchange ratio** determines the number of shares that stockholders in the target company receive in exchange for each of their shares in the target company. Because share prices are constantly fluctuating, exchange ratios are typically negotiated in advance for a range of stock prices. The acquirer's cost is the product of the exchange ratio, the number of outstanding shares of the target company, and the value of the stock given to target shareholders. Each shareholder of the target company receives new shares based on the number of target shares he owns multiplied by the exchange ratio.

EXAMPLE 4

Stock Offering

TAJ Education Group, an Indian online education provider, has announced its intended acquisition of Acchaa Prep Company Ltd, a small tutoring company specializing in after-school prep courses. In a press release, TAJ Education outlines the terms of the merger, which specify that Acchaa Prep's shareholders will each receive 0.90 shares of TAJ Education for every share of Acchaa Prep owned. Acchaa Prep has 1 million shares outstanding. On the day of the merger announcement, TAJ Education's stock closed at Rs200.00 and Acchaa Prep's stock closed at Rs150.00. Sunita Jugnauth is an individual investor who owns 500 shares of Acchaa Prep, currently worth Rs75,000 (500 × Rs150.00).

1 Based on the current share prices, what is the cost of the acquisition for TAJ Education?

2 How many shares of TAJ Education will Jugnauth receive, and what is the value of those shares (based on current share prices)?

Solution to 1:

Because there are 1 million shares of Acchaa Prep outstanding and the exchange ratio is 0.90 shares, TAJ Education will need to issue 0.90 × 1 million = 900,000 shares of TAJ Education stock to complete the transaction. Because the cost per share of TAJ Education stock is currently Rs200.00, the cost of the transaction to TAJ Education will be Rs200.00 × 900,000 = Rs180 million.

Solution to 2:

Jugnauth will turn over her 500 shares of Acchaa Prep stock. As compensation, she will receive 0.90 × 500 = 450 shares of stock in TAJ Education. With each share of TAJ Education being worth Rs200.00, the value of those shares to Jugnauth is Rs90,000.

Note that the value of Jugnauth's Acchaa Prep shares was Rs75,000. The Rs15,000 difference in value is a premium paid by TAJ Education for control of Acchaa Prep. The pre-merger value of Acchaa Prep was Rs150 million, but TAJ Education's total cost to purchase the company was Rs180 million. The 20% or Rs30 million difference is the total-control premium paid by TAJ Education.

A variety of factors influence a company's decision to negotiate for one method of payment versus another. As we shall explore in more detail later, the form of payment affects the distribution of risk and reward between acquirer and target shareholders. In a stock offering, target company shareholders assume a portion of the reward as well as a portion of the risk related to the estimated synergies and the target company's value. Consequently, when an acquiring company's management is highly confident both in their ability to complete the merger and in the value to be created by the merger, they are more inclined to negotiate for a cash offering rather than a stock offering.

Another factor in the decision relates to the relative valuations of the companies involved in the transaction. When an acquirer's shares are considered overvalued by the market relative to the target company's shares, stock financing is more appropriate. In effect, the shares are more valuable as a currency. In fact, investors sometimes interpret an acquirer's stock offering as a signal that the company's shares may be overvalued. This effect is similar to the negative market reaction observed in seasoned equity offerings. During the stock market bubble in the late 1990s, stock financing of mergers was popular, accounting for well over half of all deals in the United States. In the early 2000s after merger accounting rules changed, making all-share deals less attractive, stock financing dropped significantly to a low of 10% in 2017. This shift marked a period during which companies had substantial cash on their balance sheets and broad-based investor concerns about equity valuation levels.

Another important consideration when deciding on the payment method is the accompanying change in capital structure. The costs and benefits of different payment structures reflect how the offer will affect the acquirer's capital structure. For instance, on the one hand, borrowing to raise funds for a cash offering increases the acquirer's financial leverage and risk. On the other hand, issuing a significant number of new common shares for a stock offering can dilute the ownership interests of existing shareholders.

4.3 Mindset of Target Management

Mergers are referred to as either friendly or hostile depending on how the target company's senior managers and board of directors view the offer. The distinction is not trivial because an enormous amount of time and resources can be expended by both acquirer and target when the takeover is hostile. Whether a merger is friendly or hostile affects how it is completed, what regulations must be followed, how long the transaction takes, and possibly how much value is created (or destroyed) as a result of the combination.

4.3.1 *Friendly Mergers*

Unless there is cause to think the target will be hostile to a merger, the acquirer will generally start the process by approaching target management directly. The target could approach the acquirer, although this method is much less common. If both management teams are amenable to a potential deal, then the two companies enter into merger discussions. The negotiations revolve around the consideration to be received by the target company's shareholders and the terms of the transaction as well as other aspects, such as the post-merger management structure.

Before negotiations can culminate in a formal deal, each of the parties examines the others' books and records in a process called due diligence. The purpose of due diligence is to protect the companies' respective shareholders by attempting to confirm the accuracy of representations made during negotiations. For example, an acquirer would want to ensure that the target's assets exist and are worth approximately what was claimed by the target. Likewise, a target might want to examine an acquirer's financial records to gauge the likelihood that the acquirer has the capacity to pay for the acquisition as outlined in negotiations. Any deficiencies or problems uncovered during the due diligence process could affect negotiations, resulting in adjustments to the terms or price of the deal. If the issue is large enough, the business combination might be called off entirely.

Once due diligence and negotiations have been completed, the companies enter into a **definitive merger agreement**, a contract written by both companies' attorneys and ultimately signed by each party to the transaction. The agreement contains the details of the transaction, including the terms, warranties, conditions, termination details, and the rights of all parties.

Common industry practice has evolved such that companies typically discuss potential transactions in private and maintain secrecy until the definitive merger agreement is reached. This trend may have been influenced by shifts in securities laws toward more stringent rules related to the disclosure of material developments to the public. Additionally, news of a merger can cause dramatic stock price changes for the parties to the transaction. Premature announcement of a deal can cause volatile swings in the stock prices of the companies as they proceed through negotiations.

After the definitive merger agreement has been signed, the transaction is generally announced to the public through a joint press release by the companies. In a friendly merger, the target company's management endorses the merger and recommends that its stockholders approve the transaction. In cases where a shareholder vote is needed, whether it is the target shareholders approving the stock purchase or the acquirer shareholders approving the issuance of a significant number of new shares, the material facts are provided to the appropriate shareholders in a public document called a **proxy statement**, which is given to shareholders in anticipation of their vote.

After all the necessary approvals have been obtained—from shareholders as well as any other parties, such as regulatory bodies—the attorneys file the required documentation with securities regulators and the merger is officially completed. Target shareholders receive the consideration agreed upon under the terms of the transaction, and the companies are officially and legally combined.

4.3.2 *Hostile Mergers*

In a hostile merger, which is a merger that is opposed by the target company's management, the acquirer may decide to circumvent the target management's objections by submitting a merger proposal directly to the target company's board of directors and bypassing the CEO. This tactic is known as a **bear hug**.

Because bear hugs are not formal offers and have not been mutually agreed upon, there are no standard procedures in these cases. If the offer is high enough to warrant serious consideration, then the board may appoint a special committee to negotiate a sale of the target.

Although unlikely in practice, it is possible that target management will capitulate after a bear hug and enter into negotiations, which may ultimately lead to a friendly merger. If the bear hug is unsuccessful, then the hopeful acquirer may attempt to appeal more directly to the target company's shareholders.

One method for taking a merger appeal directly to shareholders is through a **tender offer**, whereby the acquirer invites target shareholders to submit ("tender") their shares in return for the proposed payment. Although tender offers are often associated with hostile mergers, in which the offer is opposed by the target company's management and board of directors, they can also occur in a friendly context. It is up to the individual shareholders to physically tender shares to the acquiring company's agent in order to receive payment. A tender offer can be made with cash, shares of the acquirer's own stock, other securities, or some combination of securities and cash. Because a cash tender offer can be completed in less time than a cash merger, some acquiring companies use this type of transaction to gain control of a target company quickly.

Another method of taking over a target company involves the use of a proxy fight. In a **proxy fight**, a company or individual seeks to take control of a company through a shareholder vote. Proxy solicitation is approved by regulators and then mailed directly to target company shareholders. The shareholders are asked to vote for the acquirer's proposed slate of directors. If the acquirer's slate is elected to the target's board, then it is able to replace the target company's management. At this point, the transaction may evolve into a friendly merger.

Regardless of how an acquirer seeks to establish control, target managers have a variety of alternatives available for defending the company against unwanted overtures. In these cases, the target usually retains the services of law firms and investment bankers to design a defense against the unwanted takeover attempt. As we will discuss in the next section, target company managers may use a variety of legal and financial defensive maneuvers to ward off a takeover attempt.

5 TAKEOVERS AND THEIR DEFENSE MECHANISMS

f identify and explain pre-offer and post-offer takeover defense mechanisms;

When a target company is faced with a hostile tender offer (takeover) attempt, the target managers and board of directors face a basic choice. They can decide to negotiate and sell the company, either to the hostile bidder or a third party, or they can attempt to remain independent. Aside from the strength of the company's defenses and target management's resolve to stay independent, the premium over the market price offered by the acquirer for the target company's shares is the major driving factor in the decision to support or resist any given takeover.

If the target management decides to resist the unwanted overture, they have a variety of takeover defense mechanisms at their disposal. Once the decision has been reached, the target company generally seeks the counsel of investment bankers and lawyers to explore the fairness of the hostile offer and to advise the board of the alternatives.

A target might use defensive measures to delay, negotiate a better deal for shareholders, or attempt to keep the company independent. Defensive measures can be implemented either before or after a takeover attempt has begun. Most law firms specializing in takeovers recommend that defenses be set up before a company receives or expects any takeover activity.

5.1 Pre-Offer Takeover Defense Mechanisms

In the United States, most hostile takeover attempts result in litigation. The courts generally bless legal pre-offer defense mechanisms but tend to scrutinize post-offer defenses very closely. The target usually assumes the burden of proof in showing that the recently enacted defenses are not simply intended to perpetuate management's tenure at the target company. It is for this reason that most attorneys recommend that target companies put defenses in place prior to any takeover action. Following this policy gives the target more flexibility when defending against a takeover bid.

Takeover regulation in most jurisdictions outside the United States, however, is more likely to emphasize shareholder choice with director-neutrality in hostile takeover situations, thereby involving reduced opportunity for parties to resort to litigation and consequently a lower influence from the courts.

With different twists in takeover strategy come new innovations and variations in takeover defenses. Given the many possible variations, the following is not an exhaustive list but an overview of the more well-known anti-takeover strategies. The two broad varieties of pre-offer defenses are rights-based defenses, such as poison pills and poison puts, and a variety of changes to the corporate charter (e.g., staggered boards of directors and supermajority provisions) that are sometimes collectively referred to as **shark repellents**.

5.1.1 *Poison Pills*

A **poison pill** is a legal device that makes it prohibitively costly for an acquirer to take control of a target without the prior approval of the target's board of directors. Most poison pills make the target company less attractive by creating rights that allow for the issuance of shares of the target company's stock at a substantial discount to market value.

There are two basic types of poison pills: the **flip-in pill** and the **flip-over pill**. When the common shareholder of the target company has the right to buy its shares at a discount, the pill is known as a flip-in. The pill is triggered when a specific level of ownership is exceeded. Because the acquiring company is generally prohibited from participating in the purchase through the pill, the acquirer is subject to a significant level of dilution. Most plans give the target's board of directors the right to redeem the pill prior to any triggering event. If the takeover becomes friendly, the board generally exercises this waiver.

In the case of a flip-over pill, the target company's common shareholders receive the right to purchase shares of the acquiring company at a significant discount from the market price, which causes dilution to all existing acquiring company shareholders. Again, the board of the target generally retains the right to redeem the pill should the transaction become friendly.

Another possible aspect of the poison pill is the **"dead-hand" provision**. This provision allows the board of the target to redeem or cancel the poison pill only by a vote of the continuing directors. Because continuing directors are generally defined as directors who were on the target company's board prior to the takeover attempt, this provision has the effect of making it much more difficult to take over a target without prior board approval.

5.1.2 *Poison Puts*

Whereas poison pills grant common shareholders certain rights in a hostile takeover attempt, **poison puts** give rights to the target company's bondholders. In the event of a takeover, poison puts allow bondholders to put the bonds to the company. In other words, if the provision is triggered by a hostile takeover attempt, then bondholders have the right to sell their bonds back to the target at a redemption price that is pre-specified in the bond indenture, typically at or above par value. The effect of a poison put defense is to require that an acquirer be prepared to refinance the target's debt immediately after the takeover. This defense increases the need for cash and raises the cost of the acquisition.

5.1.3 *Staggered Board of Directors*

Instead of electing the entire board of directors each year at the company's annual meeting, a company may arrange to stagger the terms for board members so that only a portion of the board seats are due for election each year. For example, if the company has a board consisting of nine directors, members could be elected for three-year terms with only three directors coming up for election each year. The effect of this staggered board is that it would take at least two years to elect enough directors to take control of the board.

5.1.4 *Restricted Voting Rights*

Some target companies adopt a mechanism that restricts stockholders who have recently acquired large blocks of stock from voting their shares. Usually, there is a trigger stockholding level, such as 15% or 20%. Shareholders who meet or exceed this trigger point are no longer able to exercise their voting rights without the target company's board releasing the shareholder from the constraint. The possibility of owning a controlling position in the target without being able to vote the shares serves as a deterrent.

5.1.5 *Supermajority Voting Provisions*

Many target companies change their charter and bylaws to provide for a higher percentage approval by shareholders for mergers than normally is required. A typical provision might require a vote of 80% of the outstanding shares of the target company (as opposed to a simple 51% majority). This supermajority requirement is triggered by a hostile takeover attempt and is frequently accompanied by a provision that prevents the hostile acquirer from voting its shares. Thus, even if an acquirer is able to accumulate a substantial portion of the target's shares, it may have great difficulty accumulating enough votes to approve a merger.

5.1.6 *Fair Price Amendments*

Fair price amendments are changes to the corporate charter and bylaws that disallow mergers for which the offer is below some threshold. For example, a fair price amendment might require an acquirer to pay at least as much as the highest stock price at which the target has traded in the public market over a specified period. Fair price amendments protect targets against temporary declines in their share prices by setting a floor value bid. Additionally, fair price amendments protect against two-tiered tender offers, in which the acquirer offers a higher bid in a first step tender offer with the threat of a lower bid in a second step tender offer for those who do not tender right away.

5.1.7 *Golden Parachutes*

Golden parachutes are compensation agreements between the target company and its senior managers. These employment contracts allow the executives to receive lucrative payouts, usually several years' worth of salary, if they leave the target company following a change in corporate control. In practice, golden parachutes do not offer much deterrent, especially for large deals in which the managers' compensation is small relative to the overall takeover price. One reason they persist is that they help alleviate target management's concerns about job loss. Golden parachutes may encourage key executives to stay with the target as the takeover progresses and the target explores all options to generate shareholder value. Some contend that without golden parachutes, target company executives might be quicker to seek employment offers from other companies to secure their financial future. Whether this is actually the case and whether golden parachutes are fair and in shareholders' best interest is the subject of considerable debate among shareholder rights activists and senior managers.

Generally, rights-based defenses (poison pills and puts) are pre-offer defenses that are available in the United States but not in many other jurisdictions, where such "frustrating" actions are often banned. In contrast, charter-based (staggered boards, supermajority, restricted voting rights, and fair price amendments) and contract-based (golden parachute) defenses are pre-offer defenses available in many countries and jurisdictions (in line with the emphasis on shareholder choice), including the United States.

5.2 Post-Offer Takeover Defense Mechanisms

A target also has several defensive mechanisms that can be used once a takeover has already been initiated. Because they may be less successful when used in isolation and because they have historically been subject to greater scrutiny by the courts, post-offer defenses are typically used in conjunction with pre-offer defenses.

5.2.1 *"Just Say No" Defense*

Probably the simplest place for a target company to start when confronted with a hostile takeover bid is to rely on pre-takeover defenses and to decline the offer. If the acquirer attempts a bear hug or tender offer, then target management typically lobbies the board of directors and shareholders to decline and build a case for why the offering price is inadequate or why the offer is otherwise not in the shareholders' best interests. This strategy forces the hopeful acquirer to adjust its bid or further reveal its own strategy in order to advance the takeover attempt.

The Takeda–Shire acquisition provides a good illustration of the "just say no" defense. In the four-week period following Takeda's initial offer, the Shire board was able to extract four more offers from Takeda with increasingly higher prices and cash components. Under the UK takeover code "put up or shut up" rule, Takeda had up to 28 days after public announcement to make its final offer or else walk away from the deal for six months. The Shire board was able to use this rule to best effect by declining early offers on the basis that they substantially undervalued the company and its growth prospects, without having to first refer to shareholders. Just before the 28-day deadline, the Shire board obtained an offer from Takeda that Shire shareholders were satisfied to accept within two weeks. The final outcome was an accepted price almost 10% higher than the initial offer, as well as an increase in the cash component from 36% to 44% of the total offer price.

5.2.2 *Litigation*

A popular technique used by many target companies is to file a lawsuit against the acquiring company based on alleged violations of securities or competition laws. **Competition laws** are laws which promote or maintain market competition by regulating anti-competitive conduct. Unless there is a serious competition law violation, these suits rarely stop a takeover bid. Instead, they often serve as a delaying tactic to create additional time for target management to develop other responses to the unwanted offer. Generally, any securities law violations, even if upheld, can be corrected with additional public disclosures. In many countries, generally most competition law claims that eventually prevent takeover attempts are initiated by either competition, foreign investment, or securities regulators rather than by the target company.

5.2.3 *Greenmail*

This technique involves an agreement allowing the target to repurchase its own shares from the acquiring company, usually at a premium to the market price. Greenmail is usually accompanied by an agreement that the acquirer will not pursue another hostile takeover attempt of the target for a set period. In effect, greenmail is the termination of a hostile takeover through a payoff to the acquirer. The shareholders of the target company do not receive any compensation for their shares.

Greenmail was popular in the United States during the early 1980s, but its use became extremely restricted by the late 1980s after the US Internal Revenue Code was amended to add a 50% tax on profits realized by acquirers through greenmail. Since the mid-2010s, a resurgence in the use of greenmail in the United States has occurred, prompted by activist investors acquiring stakes to drive corporate change. Activist investor strategies in general do not trigger the hostile takeover 50% tax provisions.

In many other countries, such as the United Kingdom and China, greenmail is not permitted as a takeover defense under takeover codes or company law, in line with shareholder-friendly, director-neutral takeover regulatory systems.

5.2.4 *Share Repurchase*

Rather than repurchasing only the shares held by the acquiring company, as in greenmail, a target might use a share repurchase to acquire shares from any shareholder. For example, a target may initiate a cash tender offer for its own outstanding shares. An effective repurchase can increase the potential cost for an acquirer by either increasing the stock's price outright or by causing the acquirer to increase its bid to remain competitive with the target company's tender offer for its own shares. Additionally, a share repurchase often has the effect of increasing the target company's use of leverage because borrowing is typically required to purchase the shares. This additional debt makes the target less attractive as a takeover candidate.

In some cases, a target company buys all of its shares and converts to a privately held company in a transaction called a leveraged buyout. In a **leveraged buyout** (LBO), the management team generally partners with a private equity firm that specializes in buyouts. The new entity borrows a high proportion of the overall purchase price; the financial firm contributes a certain amount of capital; and the management team provides the management expertise to run the business. In exchange for their expertise, management generally receives a payout percentage based on the profitability and success of the company after the LBO is completed. This strategy may allow the target to defend against a hostile bid provided that the LBO provides target shareholders with a level of value that exceeds the would-be acquirer's offer.

5.2.5 *Leveraged Recapitalization*

A technique somewhat related to the leveraged buyout is the leveraged recapitalization. A **leveraged recapitalization** involves the assumption of a large amount of debt that is then used to finance share repurchases (but in contrast to a leveraged buyout, in a recapitalization, some shares remain in public hands). The effect is to dramatically change the company's capital structure while attempting to deliver a value to target shareholders in excess of the hostile bid.

5.2.6 *"Crown Jewel" Defense*

After a hostile takeover is announced, a target may decide to sell off a subsidiary or asset to a third party. If the acquisition of this subsidiary or asset was one of the acquirer's major motivations for the proposed merger, then this strategy could cause the acquirer to abandon its takeover effort. When a target initiates such a sale after a hostile takeover bid is announced, there is a good chance that the courts will declare this strategy illegal.

5.2.7 *Pac-Man® Defense*

The target can defend itself by making a counteroffer to acquire the hostile bidder. This technique is rarely used because, in most cases, it means that a smaller company (the target) is making a bid for a larger entity. Additionally, once a target uses a Pac-Man® defense, it forgoes the ability to use a number of other defensive strategies. For instance, after making a counteroffer, a target cannot very well take the acquirer to court claiming a competition law violation. Further, in many jurisdictions other than the United States, such as the United Kingdom, a counteroffer would require shareholder approval to be obtained first, decreasing the likelihood of this defense being used.

5.2.8 *White Knight Defense*

Often the best outcome for target shareholders is for the target company's board to seek a third party to purchase the company in lieu of the hostile bidder. This third party is called a **white knight** because it is coming to the aid of the target. A target usually initiates this technique by seeking out another company that has a strategic fit with the target. Based on a good strategic fit, the third party can often justify a higher price for the target than what the hostile bidder is offering.

Once a white knight bid is made public, it may elicit an additional higher bid from the hostile bidder. This dynamic can kick off a competitive bidding situation. In some cases, because of the competitive nature of the bidders, the winner's curse can prevail and the target company shareholders may receive a very good deal. **Winner's curse** is the tendency for the winner in certain competitive bidding situations to overpay, whether because of overestimation of intrinsic value, emotion, or information asymmetries. The winner's curse is most likely to occur when the target company has roughly the same value to all bidders but the target's true value is hard to ascertain. The average bid in such cases may represent the best estimate of the target's intrinsic value, whereas the high (winning) bid overestimates its intrinsic value.

5.2.9 *White Squire Defense*

In the **white squire** defense, the target seeks a friendly party to buy a substantial minority stake in the target—enough to block the hostile takeover without selling the entire company. Although the white squire may pay a significant premium for a substantial number of the target's shares, these shares may be purchased directly from the target company and the target shareholders may not receive any of the proceeds. For example, the white squire could purchase shares of convertible preferred stock instead of common stock.

The use of the white squire defense may carry a high litigation risk depending on the details of the transaction and local regulations. Additionally, stock exchange listing requirements sometimes require that target shareholders vote to approve these types of transactions, and shareholders may not endorse any transaction that does not provide an adequate premium to them directly.

REGULATION AND COMPETITION LAW 6

Even when a merger has been accepted by the target company's senior managers, the board of directors, and shareholders, the combination must still be approved by regulatory authorities. Additionally, there are a variety of rules that companies must follow when initiating and completing the merger transaction. This section provides an overview of the key rules and issues that arise from M&A activity.

The two major bodies of jurisprudence relating to mergers are competition law and securities law. Competition laws are intended to ensure that markets remain competitive; securities laws are concerned largely with maintaining both fairness in merger activities and confidence in the financial markets.

Most countries have competition laws, which prohibit mergers and acquisitions that impede competition, generally focusing on restricting monopolies and attempts at restraint of trade. At least 135 jurisdictions have competition laws in place, and these jurisdictions produce roughly 95% of world GDP, which illustrates the significance of competition policy in regulating financial markets. Early competition laws generally focused on horizontal mergers. In subsequent years, however, the laws were updated to address vertical and conglomerate mergers in addition to the acquisition of assets.

To ensure that the laws can be effectively enforced, most jurisdictions have established regulatory agencies with the resources to oversee and enforce competition laws. For example, in the United States, the Federal Trade Commission (FTC) is the regulatory agency that works with the US Department of Justice (DoJ) to enforce antitrust law. China's equivalent is the Anti-Monopoly Enforcement Agency (AEA), which is responsible for enforcing China's Anti-Monopoly Law and where its enforcement powers for merger control are carried out by the Ministry of Commerce (MOFCOM). Most regulators have the opportunity to review and approve mergers in advance, providing an opportunity to halt a merger prior to its completion rather than having to disassemble a company after a merger is later deemed to be anticompetitive.

In Europe, the European Commission (EC) has the authority to review the competition law implications of transactions among companies that generate significant revenues within the European Union. Although the EC member states have jurisdiction on mergers within their respective national borders and their own competition regulators, such as the French Competition Authority (FCA) in France, mergers with significant cross-border effects are subject to EC review. Similar to the requirements in the United States and many other countries, pre-merger notification is required.

In addition to regulatory watchdogs, such as the FTC and the EC, approval may be needed from other regulatory agencies. For example, in most countries, a merger involving banks would require approval from that country's banking regulator, whereas one involving insurance companies would usually require approval from insurance industry regulators. In some cases, if one of the company's businesses is deemed to be of strategic national interest, additional government approvals may be necessary. Each merger must be analyzed by legal experts to determine the specific regulatory approvals required to comply with the relevant rules and laws. This is a very specialized area and can cause significant delays in the closing of some transactions.

The situation can become further complicated when the merging companies have a global presence that falls within multiple jurisdictions of regulatory control. For example, a large trans-Atlantic merger would require approval of both the US regulatory bodies and the EC. Global companies often face dozens of regulatory agencies with different standards and filing requirements. For example, the Dow Chemical–DuPont merger announced in late 2015 involved sales and production in more than 160 countries, requiring competition law approval in more than 40 jurisdictions around the world. The merger was regarded as approved and proceeding once approval had been obtained from the major market regulators of the United States, European Union, China, India, Brazil, and South Africa. The merger was completed in 2017. Regulators typically examine market share data when determining potential competition law and related regulatory guidelines violations. Although there was competition regulation in response to Microsoft's dominant market position, the subsequent emergence of a number of dominant technology companies has largely occurred with little attention from competition regulators. In 2018, however, this sector started to receive greater attention, and in June 2019 the US House of Representatives began looking into possible anti-competitive behavior by these companies, often referred to as "tech giants."

At the same time, the US competition regulators announced they would divide up their investigation, with the DoJ handling potential antitrust investigations of Apple and Google and the FTC taking on Facebook and Amazon.

EXAMPLE 5

Evolving Regulation: Competition Concerns and the Tech Giants

Competition laws continue to evolve in response to global economic and commercial trends, as well as technological developments. Although not subject to a high degree of regulatory intervention to date, the digital sector has become a key area of focus for regulators.

Concerns with respect to competition regulation of the largest technology companies, often referred to as "tech giants" because of their undisputed dominance and market influence in their respective categories, include the following:

Under-regulation. The digital industry represents a significant proportion of economic activity in many countries. This development has occurred without the level of competition regulation and oversight to which other more developed or mature traditional industries have been subject in their activities. There is concern among major regulators as to whether the traditional competition regulation toolkit is "fit for purpose" and appropriate for regulatory oversight of this developing sector.

"Killer" acquisitions. Although the sector has developed based on technological innovation and the related disruption of traditional businesses, it is widely thought that dominant companies in the digital sector may acquire start-ups solely with the intention of closing them down to limit innovation and competition. Competition regulators are considering if and how these "killer" acquisitions should be better policed.

Data aggregation. Regulators are increasingly concerned that the aggregation of data by large private-sector digital companies will provide the basis for inefficient barriers to entry and anti-competitive conduct. In September 2018, the EC said that it will "carefully review transactions which lead to the acquisition of important sets of data."

When reviewing quantitative and qualitative data, one should note that merger guidelines are just that—guidelines. It is possible that under unusual circumstances, the regulator may not challenge one merger that does violate the guidelines and may challenge another merger that does not.

Each transaction must be analyzed carefully to fully explore all potential competition law issues. Recent examples of this level of scrutiny are the different approaches taken by the DoJ regarding the AT&T–Time Warner and Disney–21st Century Fox mergers, both vertical mergers in the same industry. The AT&T–Time Warner merger, valued at US$85 billion, was announced in 2016. In late 2017, the DoJ challenged the merger in the courts, which found in favor of the merger by mid 2018, and the deal moved ahead. The DoJ appealed this decision, but in February 2019, an appeal court decision refused to block the merger. In contrast, the Disney–21st Century Fox merger, announced in early 2018 and valued at US$71 billion, was approved by all entities within six months and closed by March 2019. This transaction was uncharacteristically fast—deals of that size generally take at least a year or more to obtain approvals.

When conflicts between companies and regulators arise, it is often because of disagreements about how the markets are defined. Regulators must consider the market in terms of both geography and product. When considering the industry's geography, regulators must decide whether the relevant competitors are global, national, regional, or local. When considering product offerings, there may be one or multiple relevant product market overlaps. In some cases, the overlap may be clear, and in other transactions it may not be obvious.

Parties to the transaction are usually counseled by attorneys who have relevant experience in competition law. Most companies try to complete their analyses prior to signing a merger agreement in order to avoid entering into a long period of uncertainty while the government decides whether to challenge the transaction. Not only do delays increase costs, they may also cause the companies to lose other important strategic opportunities.

Most regulatory interventions involve horizontal mergers—for example, 88% of competition regulatory interventions in 2017 involved horizontal mergers. Nevertheless, regulatory intervention in vertical and conglomerate mergers, which are often less likely to have adverse competition effects than horizontal mergers, does occur and is increasing. Examples include the EC's prohibition of the Deutsche Börse–London Stock Exchange merger in 2017 (resulting, in part, from the merged entity's ability and incentive to hamper rivals' access to its own securities depository), as well as the DoJ's challenge to the AT&T–Time Warner merger.

7 TARGET COMPANY VALUATION: DISCOUNTED CASH FLOW ANALYSIS

g compare the discounted cash flow, comparable company, and comparable transaction analyses for valuing a target company, including the advantages and disadvantages of each;

In this section, we examine the analysis of merger activity from two perspectives. First, we discuss valuation of the target company, something of key importance for analysts on both sides of the deal as well as for shareholders as they all grapple to determine the fairness and adequacy of an offer. Then, we discuss the analysis of the bid. Analysts can estimate the distribution of benefits in a merger based on expected synergies relative to the premium paid for the target in excess of its intrinsic value.

7.1 Target Company Valuation

The three basic valuation techniques that companies and their advisers use to value companies in an M&A context are discounted cash flow analysis, comparable company analysis, and comparable transaction analysis. An analyst is likely to use some combination of these primary techniques, and possibly others, when gauging a company's fair value.

7.1.1 *Discounted Cash Flow Analysis*

Discounted cash flow (DCF) analysis, as it is generally applied in this context, discounts the company's expected future free cash flows to the present in order to derive an estimate for the company's value. **Free cash flow** (FCF) is the relevant measure in this context because it represents the actual cash that would be available to the company's investors after making all investments necessary to maintain the company as an ongoing enterprise. Free cash flows are the internally generated funds that can be distributed to the company's investors (e.g., shareholders and bondholders) without impairing the value of the company.

There are several variations to the models an analyst might use to estimate and discount free cash flows. Estimating a company's free cash flows begins with the creation of pro forma financial statements. The first step is to select an appropriate time horizon for the first stage. The first stage should include only those years over which the analyst feels capable of generating reasonably accurate estimates of the company's free cash flows. These free cash flow estimates are then discounted to their present value.

To incorporate value deriving from years beyond the first stage, the analyst estimates the value of expected second-stage free cash flows as of the end of the first stage. The result is the so-called terminal value (or continuing value) of the company. The analyst then discounts the terminal value back to the present. The sum of the two pieces (the present value of first-stage expected free cash flows plus the present value of the company's terminal value) is the estimated value of the company.

There is no standard approach for creating pro forma financial statements. The art of financial analysis involves an ability to use the appropriate tools and to exercise good judgment in order to produce the best possible estimates for each financial statement item. In the process, analysts make adjustments to their prior projections based on proposed synergies and the announced plans for the merged company. For example, duplicated resources might result in the sale of one of the target's divisions. Or, the operating costs might be adjusted downward in anticipation of economies of scale. These adjustments are easier to estimate in friendly mergers where the analyst has access to detailed financial data about the target than in hostile mergers. But even in a hostile merger scenario, an analyst with experience in the appropriate industry can still make reasonably good estimates.

Once pro forma financial statements have been generated, the analyst can begin the conversion from pro forma net income to pro forma free cash flow for each year of the first stage. The free cash flows are discounted back to present at the company's weighted average cost of capital (WACC). When evaluating the target from a non-control perspective, we would use the target's WACC, which reflects that company's existing business risk and operating environment. In anticipation of a merger, however, we would adjust that WACC to reflect any anticipated changes in the target's risk from such actions as a redeployment of assets or change in capital structure.

There are two standard methods for calculating a terminal value. The first makes use of the constant growth formula. To apply the constant growth formula, an analyst must select a terminal growth rate, which is the long-term equilibrium growth rate that the company can expect to achieve in perpetuity, accounting for both inflation and real growth. The terminal growth rate is often lower than the growth rate applied during the first stage because any advantages from synergies, new opportunities, or cost reductions are transitory as competitors adjust and the industry evolves over time. The constant growth formula can be applied whenever the terminal growth rate is less than the WACC.

A second method for estimating the terminal value involves applying a multiple at which the analyst expects the average company to sell at the end of the first stage. The analyst might use a free cash flow or other multiple that reflects the expected risk, growth, and economic conditions in the terminal year. Market multiples are rules of thumb applied by analysts, investment bankers, and venture capitalists to produce rough estimates of a company's value. Multiples tend to vary by industry. They can be based on anything applicable to the industry and correlated with market prices. Some service industries tend to be priced as multiples of EBITDA (earnings before interest, taxes, depreciation, and amortization). In contrast, retail stores in some industries might be priced based on multiples applied to floor space. In these cases, the respective multiples can be used directly to produce a terminal value, or they can be incorporated into a pro forma analysis to convert the multiple into a consistent value for free cash flow.

For example, if a company is in an industry where the typical company sells for about 20 times FCF and its free cash flow for the year is $1,799,000, then the company's terminal value estimate would be:

Terminal value = 20 × $1,799,000 = $35,980,000 = $36.0 million

Having established an estimate for the terminal value, the analyst must discount it back from the end of the estimate horizon to the present. The discount rate used is the same WACC estimate applied to discount the free cash flows.

The assumed terminal growth rate and WACC estimate can have a dramatic impact on the terminal value calculation: The final estimate of the company's value will only be as accurate as the estimates used in the model.

Advantages of Using Discounted Cash Flow Analysis

- Expected changes in the target company's cash flows (e.g., from operating synergies and cost structure changes) can be readily modeled.
- An estimate of intrinsic value based on forecast fundamentals is provided by the model.
- Changes in assumptions and estimates can be incorporated by customizing and modifying the model.

Disadvantages of Using Discounted Cash Flow Analysis

- It is difficult to apply when free cash flows do not align with profitability within the first stage. For example, a rapidly expanding company may be profitable but have negative free cash flows because of heavy capital expenditures to the horizon that can be forecast with confidence. The company's free cash flow value will then derive from a later and harder-to-estimate period when free cash flow turns positive.
- Estimating cash flows and earnings far into the future is not an exact science. There is a great deal of uncertainty in estimates for the following year, and even greater uncertainty in perpetuity.
- Estimates of discount rates can change over time because of capital market developments or changes that specifically affect the companies in question. These changes can also significantly affect acquisition estimates.
- Terminal value estimates often subject the acquisition value calculations to a disproportionate degree of estimate error. The estimate of terminal value can differ depending on the specific technique used. Additionally, the range of estimates can be affected dramatically by small changes in the assumed growth and WACC estimates.

8 TARGET COMPANY VALUATION: COMPARABLE COMPANY AND COMPARABLE TRANSACTION ANALYSIS

g compare the discounted cash flow, comparable company, and comparable transaction analyses for valuing a target company, including the advantages and disadvantages of each;

A second approach that investment bankers use to estimate acquisition values is called "comparable company analysis." In this approach, the analyst first defines a set of other companies that are similar to the target company under review. This set may include companies within the target's primary industry as well as companies in similar industries. The sample should be formed to include as many companies as possible that have similar size and capital structure to the target.

Once a set of comparable companies is defined, the next step is to calculate various relative value measures based on the current market prices of the comparable companies in the sample. Such valuation is often based on enterprise multiples. A company's enterprise value is the market value of its debt and equity minus the value of its cash and investments. Examples include enterprise value to free cash flow, enterprise value to EBITDA, enterprise value to EBIT, and enterprise value to sales. Because the denominator in such ratios is pre-interest, they may be preferred when the companies being compared have differences in leverage. The equity can also be valued directly using equity multiples, such as P/E, price to cash flow per share (P/CF), price to sales per share (P/S), and price to book value per share (P/BV).

The specific ratios that the analyst selects are determined by the industry under observation. Often, in addition to common market multiples, analysts will include industry-specific multiples. For instance, in the oil and gas industry, in addition to looking at price paid to earnings and cash flow ratios, many analysts evaluate the price paid per barrel of oil or per thousand cubic feet of natural gas reserves.

Analysts typically review the mean, median, and range for whichever metrics are chosen, and then they apply those values to corresponding estimates for the target to develop an estimated company value. This approach is quite similar to the one we discussed earlier for using multiples to produce a terminal value estimate. In this case, however, we are calculating various relative value metrics rather than using an industry rule of thumb.

Each metric (P/E, P/CF, etc.) is likely to produce a different estimate for the target's value. Analysts hope that these values converge because that increases confidence in the overall estimate. To the extent that they diverge, analysts must apply judgment and experience to decide which estimates are producing the most accurate market values.

It should be noted that the value determined up to this point in the process yields an estimate of where the target company should trade as a stock in the marketplace relative to the companies in the sample. To calculate an acquisition value, the analyst must also estimate a takeover premium. The **takeover premium** is the amount by which the takeover price for each share of stock must exceed the current stock price in order to entice shareholders to relinquish control of the company to an acquirer. This premium is usually expressed as a percentage of the stock price and is calculated as follows:

$$\mathrm{PRM} = \frac{(\mathrm{DP} - \mathrm{SP})}{\mathrm{SP}} \tag{1}$$

where

PRM = takeover premium (as a percentage of stock price)
DP = deal price per share of the target company
SP = stock price of the target company

The analyst must be careful to note any pre-deal jump in the price that may have occurred because of takeover speculation in the market. In these cases, the analyst should apply the takeover premium to a selected representative price from before any speculative influences on the stock price. To calculate the relevant takeover premium for a transaction, analysts usually compile a list of the takeover premiums paid for companies similar to the target. Preferably, the calculations will be from the recent past because acquisition values and premiums tend to vary over time and economic cycles.

Advantages of Using Comparable Company Analysis

- This method provides a reasonable approximation of a target company's value relative to similar companies in the market. It assumes that "like" assets should be valued on a similar basis in the market.
- With this method, most of the required data are readily available.
- The estimates of value are derived directly from the market. This approach is unlike the discounted cash flow method, in which the takeover value is determined based on many assumptions and estimates.

Disadvantages of Using Comparable Company Analysis

- The method is sensitive to market mispricing. To illustrate the issue, suppose that the comparable companies are overvalued. A valuation relative to those companies may suggest a value that is too high in the sense that values would be revised downward when the market corrects.
- Using this approach yields a market-estimated fair *stock* price for the target company. To estimate a fair *takeover* price, analysts must additionally estimate a fair takeover premium and use that information to adjust the estimated stock price.
- The analysis may be inaccurate because it is difficult for the analyst to incorporate any specific plans for the target (e.g., changing capital structure or eliminating duplicate resources) in the analysis.
- The data available for past premiums may not be timely or accurate for the particular target company under consideration.

8.1 Comparable Transaction Analysis

A third common approach to value target companies is known as "comparable transaction analysis." This approach is closely related to comparable company analysis except that the analyst uses details from recent takeover transactions for comparable companies to make direct estimates of the target company's takeover value.

The first step in comparable transaction analysis is to collect a relevant sample of recent takeover transactions. The sample should be as broad as possible but limited to companies in the same industry as the target, or at least closely related. Once the transactions are identified, the analyst can look at the same types of relative value multiples that were used in comparable company analysis (P/E, P/CF, other industry-specific multiples, etc.). In this case, however, we are not comparing the target against market multiples. For this approach we compare the multiples actually paid for similar companies in other M&A deals. As before, analysts typically look at **descriptive statistics**, such as the mean, median, and range for the multiples, and use judgment and experience when applying that information to estimate the target's value.

EXAMPLE 6

Comparable Transaction Analysis

Joel Hofer, an analyst with an investment banking firm, has been asked to estimate a fair price for the General Health Company's proposed acquisition of Medical Services, Inc. He has already taken the initial step and assembled a sample containing companies involved in acquisitions within the same industry in which

Medical Services operates. These companies have all been acquired in the past two years. Details on the acquisition prices and relevant pricing variables are shown in the following table.

Valuation Variables	Acquired Company 1	Acquired Company 2	Acquired Company 3
Acquisition share price ($)	35.00	16.50	87.00
Earnings per share ($)	2.12	0.89	4.37
Cash flow per share ($)	3.06	1.98	7.95
Book value per share ($)	9.62	4.90	21.62
Sales per share ($)	15.26	7.61	32.66

The next step in the process is for Hofer to calculate the multiples at which each company was acquired:

Relative Valuation Ratio	Comparable Company 1	Comparable Company 2	Comparable Company 3	Mean
P/E	16.5	18.5	19.9	18.3
P/CF	11.4	8.3	10.9	10.2
P/BV	3.6	3.4	4.0	3.7
P/S	2.3	2.2	2.7	2.4

After reviewing the distribution of the various values around their respective means, Hofer is confident about using the mean value for each ratio because the range in values above and below the mean is reasonably small. Based on his experience with this particular industry, Hofer believes that cash flows are a particularly important predictor of value for these types of companies. Consequently, instead of finding an equally weighted average, Hofer has decided to apply the following weights for calculating a weighted average estimated price.

Target Company Valuation Variables

	Target Company (a)	Comparable Companies' Valuation Multiples	Mean Multiple Paid for Comparable Companies (b)	Estimated Takeover Value Based on Comparables (c = a × b)	Weight (d)	Weighted Estimates (e = c × d)
Earnings per share	$ 2.62	P/E	18.3	$47.95	20%	$ 9.59
Cash flow per share	$ 4.33	P/CF	10.2	$44.17	40%	$17.67
Book value per share	$12.65	P/BV	3.7	$46.81	20%	$ 9.36
Sales per share	$22.98	P/S	2.4	$55.15	20%	$11.03
Weighted average estimate						**$47.65**

In sum, Hofer multiplied each valuation multiple by the corresponding variable for the target company to produce an estimated takeover value based on each comparable. He then decided to overweight cash flow per share and calculated a weighted average to determine an overall takeover value estimate of

$47.65 per share for Medical Services, Inc. The same procedure could be repeated using the median, high, and low valuations for each of the valuation variables. This process would generate a range of takeover values for Medical Services, Inc.

Advantages of Comparable Transaction Approach

- It is not necessary to separately estimate a takeover premium. The takeover premium is derived directly from the comparable transactions.
- The takeover value estimates come directly from values that were recently established in the market. This approach is unlike the discounted cash flow method, in which the takeover value is determined based on many assumptions and estimates.
- The use of prices established through other recent transactions reduces litigation risk for both companies' board of directors and managers regarding the merger transaction's pricing.

Disadvantages of Comparable Transaction Approach

- Because the value estimates assume that the M&A market has properly determined the intrinsic value of the target companies, there is a risk that the real takeover values in past transactions were not accurate. If true, these inaccurate takeover values are imputed in the estimates based on them.
- There may be no, or an inadequate number of, comparable transactions to use for calculating the takeover value. In these cases, analysts may try to use data from related industries. These derived values may not be accurate for the specific industry under study.
- The analysis may be inaccurate because it is difficult for the analyst to incorporate any specific plans for the target (e.g., changing capital structure or eliminating duplicate resources) into the analysis.

9 MERGER BID ANALYSIS

h evaluate a takeover bid and its effects on the target shareholders versus the acquirer shareholders;

i explain how price and payment method affect the distribution of risks and benefits in M&A transactions;

Assessing the target's value is important, but it is insufficient for an assessment of the deal. Even if both the acquirer and the target separately agree on the target company's underlying value, the acquirer will obviously want to pay the lowest price possible while the target will negotiate for the highest price possible. Both the price and form of payment in a merger will determine the distribution of risks and benefits between the counterparties to the deal.

Acquirers must typically pay a premium to induce the owners of the target company to relinquish control. In an M&A transaction, the premium is the portion of the compensation received by the target company's shareholders that is in excess of the pre-merger market value of their shares. The target company's managers will attempt to negotiate the highest possible premium relative to the value of the target company.

A burst of speculative stock activity typically accompanies merger negotiations. This activity usually results in a higher share price for the target company in anticipation of a takeover premium. When conducting a bid evaluation, the analyst should use some combination of an assessment of the company's intrinsic value and a representative stock price from before any merger speculation.

$$\text{Target shareholders' gain} = \text{Premium} = P_T - V_T \quad (2)$$

where

P_T = price paid for the target company
V_T = pre-merger value of the target company

The acquirer is willing to pay in excess of the target company's value in anticipation of reaping its own gains. The acquirer's gains are derived from the synergies generated by the transaction—usually from some combination of cost reductions and revenue enhancements. All else constant, synergies increase the value of the acquiring company by the value of the synergies minus the premium paid to target shareholders:

$$\text{Acquirer's gain} = \text{Synergies} - \text{Premium} = S - (P_T - V_T) \quad (3)$$

where

S = synergies created by the business combination

The post-merger value of the combined company is a function of the two companies' pre-merger values, the synergies created by the merger, and any cash paid to the target shareholders as part of the transaction:

$$V_{A^*} = V_A + V_T + S - C \quad (4)$$

where

V_{A^*} = post-merger value of the combined companies
V_A = pre-merger value of the acquirer
C = cash paid to target shareholders

When evaluating a bid, the pre-merger value of the target company is the absolute minimum bid that target shareholders should accept. Individual shareholders could sell their shares in the open market for that much instead of tendering their shares for a lower bid. At the other extreme, unless there are mitigating circumstances or other economic justifications, the acquirer's shareholders would not want to pay more than the pre-merger value of the target company plus the value of any expected synergies. If the acquirer were to pay more than that, then the acquirer's post-merger value would be lower than its pre-merger value—therefore, a reduction in shareholder value.

Bidding should thus generally be confined to a range dictated by the synergies expected from the transaction, with each side of the transaction negotiating to capture as much of the synergies as possible. Consequently, analysis of a merger depends not only on an assessment of the target company's value but also on estimates of the value of any synergies that the merged company is expected to attain.

Confidence in synergy estimates will have implications not only for the bid price but also for the method of payment. The reason for this is that different methods of payment for the merger—cash offer, stock offer, or mixed offer—inherently provide varying degrees of risk shifting with respect to misestimating the value of merger synergies. To see why this is the case, we will first walk through the evaluation of an offer for each method of payment.

EXAMPLE 7

Adagio Software Offer

Adagio Software, Inc., and Tantalus Software Solutions, Inc., are negotiating a friendly acquisition of Tantalus by Adagio. The management teams at both companies have informally agreed upon a transaction value of about €12.00 per share of Tantalus Software Solutions stock but are presently negotiating alternative forms of payment. Sunil Agrawal, CFA, works for Tantalus Software Solutions' investment banking team and is evaluating three alternative offers presented by Adagio Software:

1 Cash Offer: Adagio will pay €12.00 per share of Tantalus stock.
2 Stock Offer: Adagio will give Tantalus shareholders 0.80 shares of Adagio stock per share of Tantalus stock.
3 Mixed offer: Adagio will pay €6.00 plus 0.40 shares of Adagio stock per share of Tantalus stock.

Agrawal estimates that the merger of the two companies will result in economies of scale with a net present value of €90 million. To aid in the analysis, Agrawal has also compiled the following data:

	Adagio	Tantalus
Pre-merger stock price	€15.00	€10.00
Number of shares outstanding (millions)	75	30
Pre-merger market value (millions)	€1,125	€300

Based only on the information given, which of the three offers should Agrawal recommend to the Tantalus Software Solutions management team?

Solution:

Alternative 1:

Cash offer of €12.00 per share of Tantalus stock. A cash offer is the most straightforward and easiest to evaluate. The price paid for the target company, P_T, is equal to the cash price per share times the number of target shares: €12.00 × 30 million = €360 million. Because Tantalus' value, V_T, is €300 million, the premium is the difference between the two: €360 million – €300 million = €60 million.

Adagio's gain in this transaction is €30 million, which equals the value of the synergies minus the premium paid to Tantalus shareholders. A longer way to get to the same conclusion is to remember that the value of the post-merger combined company equals the pre-merger values of both companies plus the value of created synergies less the cash paid to target shareholders: $V_{A^*} = V_A + V_T + S - C$ = €1,125 + 300 + 90 – 360 = €1,155 million. Adagio's pre-merger market value was €1,125 million, and Adagio's gain from the transaction is thus €1,155 – 1,125 = €30 million. Agrawal can divide the post-merger market value of €1,155 by the number of shares outstanding to determine Adagio's post-merger stock price. Under a cash offer, Adagio will not issue additional shares of stock, so Agrawal divides €1,155 by 75 million shares to see that, all else constant, Adagio's stock price after the merger should rise to €15.40.

In an all-cash offer, Tantalus shareholders receive €60 million—the premium. Adagio's gain from the transaction equals the expected synergies (€90 million) less the premium paid to Tantalus shareholders (€60 million), which equals €30 million.

Alternative 2:

Stock offer of 0.80 shares of Adagio stock per share of Tantalus stock. A stock offer of 0.80 shares might seem at first glance to be equivalent to a cash offer of €12.00, because Adagio's share price is €15.00 (0.80 × €15 = €12). The results are actually slightly different, however, because Agrawal must account for the dilution that occurs when Adagio issues new shares to Tantalus stockholders. Because there are 30 million shares of the target outstanding, Adagio must issue 30 million × 0.80 = 24 million shares.

To calculate the price paid for Tantalus, Agrawal starts by ascertaining the post-merger value of the combined company. Agrawal uses the same formula as before while using a value of zero for C because this is a stock offer and no cash is changing hands: $V_{A^*} = V_A + V_T + S - C$ = €1,125 + 300 + 90 − 0 = €1,515 million. Next, Agrawal divides Adagio's post-merger value by the post-merger number of shares outstanding. Because Adagio issued 24 million shares to complete the transaction, Agrawal adds 24 million to the original 75 million shares outstanding and arrives at 99 million. Dividing the post-merger market value by the post-merger number of shares outstanding, Agrawal determines that the value of each share given to Tantalus shareholders is actually worth €1,515 million/99 million = €15.30 and that the total value paid to Tantalus shareholders is €15.30 × 24 million = €367 million.

The premium is thus €367 − 300 = €67 million, which is €7 million higher than it was for the cash offer. Because the target shareholders receive €7 million more than in the cash offer, the acquirer's gain is correspondingly less. Because the synergies are valued at €90 million and the premium is €67 million, the acquirer's gain under a stock transaction with these terms is €23 million.

Alternative 3:

Mixed offer of €6.00 plus 0.40 shares of Adagio stock per share of Tantalus stock. A mixed offer will still result in some dilution, although not as much as a pure stock offer. Agrawal begins by calculating Adagio's post-merger value. Agrawal inserts €180 million for C because the company is paying €6 per share for 30 million shares: $V_{A^*} = V_A + V_T + S - C$ = €1,125 + 300 + 90 − 180 = €1,335 million.

Next, Agrawal determines that Adagio must issue 12 million shares to complete the transaction: 0.40 × 30 million = 12 million. Combined with the original 75 million shares outstanding, Adagio's post-merger number of shares outstanding will be 87 million. Agrawal divides €1,335 million by 87 million and finds that each share given to the Tantalus shareholders is worth €15.35.

The total value paid to Tantalus shareholders includes a cash component, €6.00 × 30 million = €180 million, and a stock component, 12 million shares issued with a value of €15.35 each equaling €184 million. Added together, the total value is €180 + €184 = €364 million, and the premium is therefore €364 million − €300 million = €64 million. The acquirer's gain is €26 million.

Conclusion:

Agrawal should recommend that the Tantalus Software Solutions management team opt for the all-stock offer because that alternative provides Tantalus shareholders the most value (the highest premium).

In Example 7, Adagio's gain ranged from €30 million in the pure cash offer to €26 million in the mixed offer and €23 million in the pure stock offer. If the dilution of a stock offer reduces the acquirer's gains from the transaction, why would an acquirer ever pay stock in a merger? The answer brings us back to the beginning of the section, where we pointed out that the price and form of payment in a merger determine the distribution of risks and benefits. The choice of payment method is influenced by both parties' confidence in the estimated synergies and the relative value of the acquirer's shares.

The more confident the managers are that the estimated synergies will be realized, the more the acquiring managers will prefer to pay with cash and the more the target managers will prefer to receive stock. And the more the merger is paid for with the acquirer's stock, the more that the risks and benefits of realizing synergies will be passed on to the target shareholders. For example, in the cash offer we analyzed in Example 7, if the synergies later turned out to be worth €60 million rather than the originally estimated €90 million, then the Tantalus shareholders' premium would be unaffected but Adagio's gain would completely evaporate. In contrast, if the synergies were greater than estimated, then Tantalus shareholders' premium would still be unchanged but Adagio's gain would increase.

When stock is used as payment, the target shareholders become part owners of the acquiring company. In the Adagio stock offer, Tantalus shareholders would receive 24 million shares and thus own 24/99 (24.2%) of the post-merger acquirer. Thus, Tantalus shareholders would participate by that proportion in any deviation of synergies from pre-merger estimates. If synergies were worth only €60 million, Adagio would lose its €23 million gain and Tantalus shareholders' gain from the transaction would fall by €7 million.

The other factor affecting the method of payment decision relates to the counterparties' confidence in the companies' relative values. The more confident managers are in estimates of the target company's value, the more the acquirer will prefer cash and the more the target will prefer stock. For example, what if Adagio estimates that Tantalus is worth more than €10 per share and consequently offers €12.50 per share in cash instead of €12.00? In that case, Tantalus shareholders would receive a premium that is €15 million higher and Adagio's gain from the transaction would be reduced by €15 million to €15 million.

10 BENEFITS FROM MERGERS

j describe characteristics of M&A transactions that create value;

What does the empirical evidence say about who actually gains in business combinations? Studies on the performance of mergers fall into two categories: short-term performance studies, which examine stock returns surrounding merger announcement dates, and long-term performance studies of post-merger companies. The empirical evidence suggests that merger transactions create value for target company shareholders in the short run. On average, target shareholders reap 30% premiums over the stock's pre-announcement market price, and the acquirer's stock price falls, on average, between 1% and 3%. Moreover, on average, both the acquirer and target tend to see higher stock returns surrounding cash acquisition offers than around share offers.

The high average premiums paid to target shareholders may be attributed, at least partly, to the winner's curse—the tendency for competitive bidding to result in overpayment. Even if the average bidding company accurately estimates the target company's value, some bidders will overestimate the target's value and other potential buyers will underestimate its value. Unless the winner can exploit some strong synergies that are unavailable to other bidders, the winning bidder is likely to be the one who most overestimates the value.

Roll argues that high takeover bids may stem from hubris, from "the overbearing presumption of bidders that their valuations are correct." Implied in this behavior is that these executives are somehow smarter than everyone else and can see value where others cannot. Even if there were no synergies from a merger, managerial hubris would still lead to higher-than-market bids and a transfer of wealth from the acquiring company's shareholders to the target's shareholders. The empirical evidence is consistent with Roll's hubris hypothesis.

When examining a longer period, empirical evidence shows that acquirers outperformed comparable non-acquirer companies over a five-year period from 2011–2016. This outperformance has been attributed to the reinvestment of capital via M&A into the business and growth, whereas the non-acquirers allocated a large portion of their capital to share buybacks and dividends, which ultimately shortchanged shareholders through poor growth. Total shareholder returns to acquiring companies over 2011–16 were about 16%, compared with 8% for non-acquirers.

Nevertheless, it is important that analysts thoroughly scrutinize estimates of synergy and post-merger value creation by deal. Analysts must attempt to distinguish those deals that create value and those that do not. Even though investment in quality growth via M&As produces outperformance for shareholders, companies with surplus cash but few new investment opportunities are prone to make acquisitions rather than distribute excess cash to shareholders. When distinguishing value-creating deals, analysts must examine the operational strengths possessed by the acquirer and the target to discern the likelihood that post-merger synergies will be achieved.

Based on past empirical results, the following are characteristics of M&A deals that create value:

- **The buyer is strong**. Acquirers whose earnings and share prices grow at a rate above the industry average for three years before the acquisition earn statistically significant positive returns on announcement.
- **The transaction premiums are relatively low**. Acquirers earn negative returns on announcement when paying a high premium.
- **The number of bidders is low**. Acquirer stock returns are negatively related to the number of bidders.
- **The initial market reaction is favorable**. Initial market reaction is an important barometer for the value investors place on the gains from merging as well as an indication of future returns. If the acquiring company's stock price falls when the deal is announced, investors are sending a message that the merger benefits are doubtful or that the acquirer is paying too much.

CORPORATE RESTRUCTURING 11

k compare and contrast equity carve-outs, spin-offs, split-offs, and liquidation;

l explain common reasons for restructuring.

Just as mergers and acquisitions are a means by which companies grow bigger, a corporate restructuring is usually used in reference to ways that companies get smaller—by selling, splitting off, or otherwise shedding operating assets. When a company decides to sell, liquidate, or spin off a division or a subsidiary, it is referred to as a **divestiture**.

Given that, as we have discussed, many companies have great difficulty actually achieving the planned synergies of a business combination, it is unsurprising that many companies seek to undo previous mergers. Indeed, periods of intense merger activity are often followed by periods of heightened restructuring activity. Of course, previous mergers that did not work out as planned are not the only reason companies may choose to divest assets. Some of the common reasons for restructuring follow:

- **Change in strategic focus.** Either through acquisitions or other investments over time, companies often become engaged in multiple markets. Management may hope to improve performance by eliminating divisions or subsidiaries that are outside the company's core strategic focus.
- **Poor fit.** Sometimes a company will decide that a particular division is a poor fit within the overall company. For example, the company may lack the expertise or resources to fully exploit opportunities pursued by the division and may decide to sell the segment to another company that does have the necessary resources. Or, the division might simply not be profitable enough to justify continued investment based on the company's cost of capital.
- **Reverse synergy.** Managers may feel that a segment of the company is undervalued by the market, sometimes because of poor performance of the overall company or because the division is not a good strategic fit. In these cases, it is possible that the division and the company will be worth more separately than combined.
- **Financial or cash flow needs.** If times are tough, managers may decide to sell off portions of the company as a means by which to raise cash or cut expenses.

Restructuring can take many forms, but the three basic ways that a company divests assets are a sale to another company, a spin-off to shareholders, or liquidation. As part of a sale to another company, a company might offer to sell the assets of a division or may offer an equity carve-out. An **equity carve-out** involves the creation of a new legal entity and sales of equity in it to outsiders.

In a **spin-off**, shareholders of the parent company receive a proportional number of shares in a new, separate entity. Whereas the sale of a division results in an inflow of cash to the parent company, a spin-off does not. A spin-off simply results in shareholders owning stock in two different companies where there used to be one. A similar type of transaction is called a **split-off**, whereby some of the parent company's shareholders are given shares in a newly created entity in exchange for their shares of the parent company. **Liquidation** involves breaking up a company, division, or subsidiary and selling off its assets piecemeal. For a company, liquidation is typically associated with bankruptcy.

SUMMARY

Mergers and acquisitions are complex transactions. The process often involves not only the acquiring and target companies but also a variety of other stakeholders, including competition law regulatory agencies. To fully evaluate a merger, analysts must ask two fundamental questions: First, will the transaction create value; and second, does the acquisition price outweigh the potential benefit? This reading has made the following important points.

- An acquisition is the purchase of some portion of one company by another. A merger represents the absorption of one company by another such that only one entity survives following the transaction.

- Mergers can be categorized by the form of integration. In a statutory merger, one company is merged into another; in a subsidiary merger, the target becomes a subsidiary of the acquirer; and in a consolidation, both the acquirer and target become part of a newly formed company.
- Horizontal mergers occur among peer companies engaged in the same kind of business. Vertical mergers occur among companies along a given value chain. Conglomerates are formed by companies in unrelated businesses.
- Merger activity tends to be concentrated in industries undergoing changes, such as deregulation or technological advancement.
- The motives for M&A activity include synergy, growth, market power, the acquisition of unique capabilities and resources, diversification, increased earnings, management's personal incentives, tax considerations, and the possibilities of uncovering hidden value. Cross-border motivations may involve technology transfer, product differentiation, government policy, and the opportunities to serve existing clients abroad.
- A merger transaction may take the form of a stock purchase (when the acquirer gives the target company's shareholders some combination of cash or securities in exchange for shares of the target company's stock) or an asset purchase (when the acquirer purchases the target company's assets and payment is made directly to the target company). The decision of which approach to take will affect other aspects of the transaction, such as how approval is obtained, which laws apply, how the liabilities are treated, and how the shareholders and the company are taxed.
- The method of payment for a merger can be cash, securities, or a mixed offering with some of both. The exchange ratio in a stock or mixed offering determines the number of shares that stockholders in the target company will receive in exchange for each of their shares in the target company.
- Hostile transactions are those opposed by target managers, whereas friendly transactions are endorsed by the target company's managers. There are a variety of both pre- and post-offer defenses a target can use to ward off an unwanted takeover bid.
- Examples of pre-offer defense mechanisms include poison pills and puts, incorporation in a jurisdiction with restrictive takeover laws, staggered boards of directors, restricted voting rights, supermajority voting provisions, fair price amendments, and golden parachutes.
- Examples of post-offer defenses include the "just say no" defense, litigation, greenmail, share repurchases, leveraged recapitalization, "crown jewel" defense, Pac-Man® defense, or finding a white knight or a white squire.
- Competition law prohibits mergers and acquisitions that impede competition.
- Three major tools for valuing a target company are discounted cash flow analysis (which involves discounting free cash flows estimated with pro forma financial statements), comparable company analysis (which estimates a company's intrinsic value based on relative valuation metrics for similar companies), and comparable transaction analysis (which derives valuation from details of recent takeover transactions for comparable companies).
- In a merger bid, the gain to target shareholders is measured as the control premium, which equals the price paid for the target company in excess of its value. The acquirer gains equal the value of any synergies created by the merger minus the premium paid to target shareholders. Together, the bid and the method of

payment determine the distribution of risks and returns among acquirer and target shareholders with regard to realization of synergies as well as correct estimation of the target company's value.

- The empirical evidence suggests that merger transactions create value for target company shareholders. Acquirers, in contrast, tend to accrue value in the years following a merger. This finding suggests that synergies are often overestimated or difficult to achieve.
- When a company decides to sell, liquidate, or spin off a division or a subsidiary, it is referred to as a divestiture. Companies may divest assets for a variety of reasons, including a change in strategic focus, poor fit of the asset within the corporation, reverse synergy, or cash flow needs.
- The three basic ways that a company divests assets are a sale to another company, a spin-off to shareholders, and liquidation.

REFERENCES

Bruner, Robert F. 2005. *Deals from Hell: M&A Lessons That Rise Above the Ashes*. New York: John Wiley & Sons.

Koller, T., M. Goedhart, and D. Wessels. 2005. *Valuation: Measuring and Managing the Value of Companies*. 4th ed. Hoboken, NJ: John Wiley & Sons.

Liu, Miao. 2016. "*A Comparative Study of Takeover Defenses in UK, US and Chinese Law.*" Doctoral thesis, Durham University.

Malmendier, Ulrike, Enrico Moretti, and Florian Peters. 2018. "Winning by Losing: Evidence on the Long-run Effects of Mergers." *Review of Financial Studies* 31 (8).

Roll, Richard. 1986. "The Hubris Hypothesis on Corporate Takeovers." *Journal of Business* 59 (April): 176–216.

Solnik, Bruno, and Dennis McLeavey. 2004. *International Investments*. 5th ed. Boston: Addison Wesley.

Suisse, Credit. 2016. "*Tying the Knot: M&A As a Path to Value Creation.*" Credit Suisse Corporate Insights (Q4):9–10.

Weston, J. Fred, and Samuel C. Weaver. 2001. *Mergers & Acquisitions*. New York: McGraw-Hill.

Weston, J. Fred, Kwang S. Chung, and Susan E. Hoag. 1990. *Mergers, Restructuring, and Corporate Control*. New York: Prentice Hall.

PRACTICE PROBLEMS

The following information relates to Questions 1–5

Modern Auto, an automobile parts supplier, has made an offer to acquire Sky Systems, creator of software for the airline industry. The offer is to pay Sky Systems' shareholders the current market value of their stock in Modern Auto's stock. The relevant information used in those calculations is as follows:

	Modern Auto	Sky Systems
Share price	$40	$25
Number of outstanding shares (millions)	40	15
Earnings (millions)	$100	$30

Although the combined company's total earnings will not increase and are estimated to be $130 million, Charles Wilhelm (treasurer of Modern Auto) argues that there are two attractive reasons to merge. First, Wilhelm says, "The merger of Modern Auto and Sky Systems will result in lower risk for our shareholders because of the diversification effect." Second, Wilhelm also says, "If our EPS increases, our stock price will increase in line with the EPS increase because our P/E will stay the same."

Sky Systems managers are not interested in the offer by Modern Auto. The managers, instead, approach HiFly, Inc., which is in the same industry as Sky Systems, to see if it would be interested in acquiring Sky Systems. HiFly is interested, and both companies believe there will be synergies from this acquisition. If HiFly were to acquire Sky Systems, it would do so by paying $400 million in cash.

1 The acquisition of Sky Systems by Modern Auto and the acquisition of Sky Systems by HiFly, respectively, would be examples of a:

A vertical merger and a horizontal merger.

B conglomerate merger and a vertical merger.

C conglomerate merger and a horizontal merger.

2 If Sky Systems were to be acquired by Modern Auto under the terms of the original offer, the post-merger EPS of the new company would be *closest* to:

A $2.00.

B $2.32.

C $2.63.

3 Are Wilhelm's two statements about his shareholders benefiting from the diversification effect of the merger and about the increase in the stock price, respectively, correct?

	The Merger Will Result in Lower Risk for Shareholders	Stock Price Will Increase in Line with the EPS Increase
A	No	No
B	No	Yes
C	Yes	No

© 2013 CFA Institute. All rights reserved.

4 Which of the following defenses *best* describes the role of HiFly in the acquisition scenario?

A Crown jewel

B Pac-Man®

C White knight

5 Suppose HiFly acquires Sky Systems for the stated terms. The gain to Sky Systems shareholders resulting from the merger transaction would be *closest* to:

A $25 million.

B $160 million.

C $375 million.

The following information relates to Question 6

Kinetic Corporation is considering acquiring High Tech Systems. Jim Smith, vice president of finance at Kinetic, is conducting the analysis of High Tech. Smith is aware of several approaches that could be used for this purpose. He plans to consider each of these approaches in his analysis and has collected or estimated the necessary financial data.

While discussing his analysis with a colleague, Smith makes two comments. His first comment is: "If there were a pre-announcement run-up in Quadrant's price because of speculation, the takeover premium should be computed based on the price prior to the run-up." Smith's second comment is: "Because the comparable transaction approach is based on the acquisition price, the takeover premium is implicitly recognized in this approach."

6 Are Smith's two comments about his analysis correct?

A Both of his comments are correct.

B Both of his comments are incorrect.

C His first comment is correct, and his second comment is incorrect.

The following information relates to Questions 7–12 and is based on "Corporate Governance" and this reading

Mark Zin and Stella Lee are CEO and CFO, respectively, of Moonbase Corporation. They are concerned that Moonbase is undervalued and subject to a hostile takeover bid. To assess the value of their own firm, they are reviewing current financial data for Jupiter PLC, Saturn Corporation, and Voyager Corporation, three firms they believe are comparable to Moonbase.

Relative Valuation Ratio	Jupiter	Saturn	Voyager
P/E	23.00	19.50	21.50
P/B	4.24	5.25	4.91
P/CF	12.60	11.40	13.30

Zin believes Moonbase should trade at similar multiples to these firms and that each valuation ratio measure is equally valid. Moonbase has a current stock price of \$34.00 per share, earnings of \$1.75 per share, book value of \$8.50 per share, and cash flow of \$3.20 per share. Using the average of each of the three multiples for the three comparable firms, Zin finds that Moonbase is undervalued.

Lee states that the low valuation reflects current poor performance of a subsidiary of Moonbase. She recommends that the board of directors consider divesting the subsidiary in a manner that would provide cash inflow to Moonbase.

Zin proposes that some action should be taken before a hostile takeover bid is made. He asks Lee if changes can be made to the corporate governance structure in order to make it more difficult for an unwanted suitor to succeed.

In response, Lee makes two comments of actions that would make a hostile takeover more difficult. Lee's first comment is, "Moonbase can institute a poison pill that allows our shareholders, other than the hostile bidder, to purchase shares at a substantial discount to current market value." Lee's second comment is, "Moonbase can instead institute a poison put. The put allows shareholders the opportunity to redeem their shares at a substantial premium to current market value."

Zin is also concerned about the general attitude of outside investors with the governance of Moonbase. He has read brokerage reports indicating that the Moonbase governance ratings are generally low. Zin believes the following statements describe characteristics that should provide Moonbase with a strong governance rating.

Statement 1 Moonbase's directors obtain advice from the corporate counsel to aid them in assessing the firm's compliance with regulatory requirements.

Statement 2 Five of the 10 members of the board of directors are not employed by Moonbase and are considered independent. Although not employed by the company, two of the independent directors are former executives of the company and thus can contribute useful expertise relevant for the business.

Statement 3 The board's audit committee is organized so as to have sufficient resources to carry out its task, with an internal staff that reports routinely and directly to the audit committee.

Zin is particularly proud of the fact that Moonbase has begun drafting a "Statement of Corporate Governance" (SCG) that will be available on the company website for viewing by shareholders, investment analysts, and any interested stakeholders. In particular, the SCG pays special attention to policies that ensure effective contributions from the board of directors. These policies include the following:

Policy 1 Training is provided to directors prior to joining the board and periodically thereafter.

Policy 2 Statements are provided of management's assessment of the board's performance of its fiduciary responsibilities.

Policy 3 Statements are provided of directors' responsibilities regarding oversight and monitoring of the firm's risk management and compliance functions.

Zin concludes the discussion by announcing that Johann Steris, a highly regarded ex-CFO of a major corporation, is under consideration as a member of an expanded board of directors. Zin states that Steris meets all the requirements as an independent director, including the fact that he will not violate the interlocking directorship requirement. Steris also will bring experience as a member of the compensation committee of the board of another firm. He also comments that Steris desires to serve on either the audit or compensation committee of the Moonbase board and that good governance practice suggests that Steris would not be prohibited from serving on either committee.

7 The value the CEO estimated based on comparable company analysis is *closest* to:
   A $37.33.
   B $39.30.
   C $40.80.
8 The divestiture technique that Lee is recommending is *most likely*:
   A a spin-off.
   B a split-off.
   C an equity carve-out.
9 With regard to poison pills and puts, Lee's comments are:
   A correct.
   B incorrect with regard to the poison put.
   C incorrect with regard to the poison pill.
10 Which statement by Zin provides the *most* support for a strong governance rating?
   A Statement 1
   B Statement 2
   C Statement 3
11 Which policy of the Statement of Corporate Governance is *least likely* to ensure effective contributions from the board of directors?
   A Policy 1
   B Policy 2
   C Policy 3
12 Is Zin's comment that good governance practice does not preclude Steris from serving on either of the two committees of the Moonbase board correct?
   A Yes.
   B No, good governance practice precludes Steris from serving on the audit committee.
   C No, good governance practice precludes Steris from serving on the compensation committee.

The following information relates to Questions 13–15

Josh Logan is a buy-side equity analyst who follows Durtech. Logan's supervisor believes that Durtech is a likely takeover candidate and has asked Logan to estimate the company's value per share in the event of an all-stock takeover bid. Logan plans to estimate Durtech's value per share using three approaches: discounted cash flow, comparable company analysis, and comparable transaction analysis.

Durtech has 1.2 million common shares outstanding and no outstanding long-term debt or preferred stock. Logan estimates that Durtech's free cash flows at the end of the next three years will be $5.0 million, $6.0 million, and $7.0 million, respectively. After Year 3, he projects that free cash flow will grow at 5% per year. He determines the appropriate discount rate for this free cash flow stream is 15% per year.

Applying discounted cash flow analysis to the preceding information, Logan determines that Durtech's fair enterprise value is $61.8 million. In a separate analysis based on ratios, Logan estimates that at the end of the third year, Durtech will be worth 10 times its Year 3 free cash flow.

Logan gathers data on two companies comparable to Durtech: Alphatech and Betatech. He believes that price-to-earnings, price-to-sales, and price-to-book-value per share of these companies should be used to value Durtech. The relevant data for the three companies are given in Exhibit 1.

Exhibit 1 Valuation Variables for Durtech and Comparable Companies

Valuation Variables	Alphatech	Betatech	Durtech
Current stock price ($)	72.00	45.00	24.00
Earnings per share ($)	2.00	1.50	1.00
Sales per share ($)	32.00	22.50	16.00
Book value per share ($)	18.00	10.00	8.00

Logan also identifies one recent takeover transaction and analyzes its takeover premium (the amount by which its takeover price per share exceeds its current stock price). Omegatech is comparable to the possible transaction on Durtech. Omegatech had a stock price of $44.40 per share prior to a newspaper report of a takeover rumor. After the takeover rumor was reported, the price rose immediately to $60.30 per share. Eventually, the takeover offer was accepted by Omegatech's shareholders for $55.00 per share. One-year trailing earnings per share for Omegatech immediately prior to the takeover were $1.25 per share.

13 Based on Exhibit 1 and the mean of each of the valuation ratios, Logan's estimate of Durtech's value per share should be *closest* to:

A $30.44.

B $33.67.

C $34.67.

14 Based on the premium on a recent comparable transaction, Logan's best estimate of the takeover premium for Durtech is *closest* to:

A 19.9%.

B 23.9%.

C 35.8%.

15 Using comparable transaction analysis, Logan's estimate of the fair acquisition value per share for Durtech is *closest* to:

A $35.52.

B $42.59.

C $44.00.

SOLUTIONS

1 C is correct. These are conglomerate and horizontal mergers, respectively.

2 C is correct. EPS is \$2.63.

Because Modern Auto's stock price is \$40 and Sky Systems' stock price is \$25, Modern Auto will acquire Sky Systems by exchanging 1 of its shares for 40/25 = 1.60 shares of Sky Systems. There are 15 million shares of Sky Systems. Their acquisition will take 15/1.60 = 9.375 million shares of Modern Auto. The total number of shares after the merger = 49.375 million. The EPS after the merger = 130/49.375 = \$2.63.

3 A is correct. Both of the statements by Wilhelm are wrong.

The first statement is wrong because diversification by itself does not lower risk for shareholders. Investors can diversify very cheaply on their own by purchasing stocks of different companies (for example, a Modern Auto shareholder could purchase stocks of Sky Systems).

The second statement is also wrong. The P/E will not necessarily remain the same following the merger and is more likely to decline. The pre-merger P/E for Modern Auto is 40/2.50 = 16. After the merger, the EPS would be \$130 million/49.375 million shares, or 2.6329. The post-merger P/E will probably fall to 40/2.6329 = 15.19.

4 C is correct. HiFly is a white knight.

5 A is correct.

$$\text{Target shareholders' gain} = \text{Premium} = P_T - V_T$$

P_T = Price paid for the target company = \$400 million as provided in the vignette

V_T = Pre-merger value of the target = \$25 share price × 15 million shares = \$375 million

\$400 million – \$375 million = \$25 million

6 A is correct. Both of Smith's statements are correct.

If there was a pre-announcement run-up in Quadrant's price because of speculation, the takeover premium should be computed based on the price prior to the run-up. Because the comparable transaction approach is based on the acquisition price, the takeover premium is implicitly recognized in this approach.

7 B is correct. Value is \$39.30.

Average P/E is 21.33 = (23.00 + 19.50 + 21.50)/3

Value based on P/E = 21.33 (1.75) = 37.33

Average P/B is 4.80 = (4.24 + 5.25 + 4.91)/3

Value based on P/B = 4.80 (8.50) = 40.80

Average P/CF is 12.43 = (12.60 + 11.40 + 13.30)/3

Value based on P/CF = 12.43 (3.20) = 39.79

Because Zin believes each valuation ratio is equally valid, value is a simple average of the three values.

Value = (37.33 + 40.80 + 39.79)/3 = 39.30

8 C is correct. An equity carve-out involves sale of equity in a new legal entity to outsiders and would thus result in a cash inflow for Moonbase. A spin-off or a split-off does not generate a cash flow to the firm.

9 B is correct. The first comment about the poison pill is correct, but the second comment is incorrect. Shareholders do not “put” their shares to the company; rather bondholders can exercise the put in the event of a hostile takeover. Bondholders have the right to sell their bonds back to the target at a redemption price that is pre-specified in the bond indenture, typically at or above par value.

10 C is correct. Statement 3 provides the most support for a strong governance rating. The statement describes the manner in which the audit committee should work. The other two statements do not support a strong governance rating as each casts doubt about the independence of the board from management’s control.

11 B is correct. The second policy is least likely to ensure effective contributions from the board. The board through self-assessment, and not management, should assess the board’s performance.

12 A is correct. As an independent director, without an interlocking relationship and with the expertise required, Steris would be eligible to serve on either of the two committees.

13 B is correct.

Step 1. Compute Valuation Ratios

Valuation Ratio	Alphatech	Betatech	Mean
P/E	36.00	30.00	33.00
P/S	2.25	2.00	2.125
P/BV	4.00	4.50	4.25

Step 2. Apply to Durtech’s Variables

Valuation Ratio	Durtech	Mean Multiple	Estimated Stock Price
Earnings per share	1.00	33.00	33.00
Sales per share	16.00	2.125	34.00
Book value per share	8.00	4.25	34.00

Step 3. Determine Mean Value: $(33 + 34 + 34)/3 = \$33.67$ per share

14 B is correct. A comparable transaction sells for premium of $55/44.4 - 1 = 23.9\%$.

15 C is correct. Omegatech’s transaction P/E: $55/1.25 = 44$. So estimated fair acquisition value per share is $44 \times 1 = \$44.00$.

READING

19

Capital Budgeting

by John D. Stowe, PhD, CFA, and Jacques R. Gagné, FSA, CFA, CIPM

John D. Stowe, PhD, CFA, is at Ohio University (USA). Jacques R. Gagné, FSA, CFA, CIPM, is at ENAP (Canada).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** calculate the yearly cash flows of expansion and replacement capital projects and evaluate how the choice of depreciation method affects those cash flows;
☐	**b.** explain how inflation affects capital budgeting analysis;
☐	**c.** evaluate capital projects and determine the optimal capital project in situations of 1) mutually exclusive projects with unequal lives, using either the least common multiple of lives approach or the equivalent annual annuity approach, and 2) capital rationing;
☐	**d.** explain how sensitivity analysis, scenario analysis, and Monte Carlo simulation can be used to assess the standalone risk of a capital project;
☐	**e.** describe types of real options relevant to a capital project;
☐	**f.** describe common capital budgeting pitfalls.

INTRODUCTION

1

Capital budgeting is the process that companies use for investing in capital projects—those projects with a life of a year or more. The Level I curriculum presented the basic principles of this process and the major discounted cash flow (DCF) investment criteria: net present value (NPV) and internal rate of return (IRR). The payback period and the accounting rate of return also are commonly used, although they have drawbacks compared with DCF methods. This reading expands on the basic principles by projecting the cash flows needed in these models and applying the models to a variety of more complicated situations.

This reading is organized as follows: Section 2 presents a crucial element of the capital budgeting process: organizing the cash flow information that is the raw material of the analysis. Sections 2.4 and 3 look further at cash flow analysis, considering

© 2020 CFA Institute. All rights reserved.

alternative depreciation methods, replacement projects (in contract to expansion projects), and the effects of inflation. Sections 4–7 demonstrate methods to extend the basic investment criteria to address economic alternatives and risk. These methods include mutually exclusive projects with differing lives, sensitivity analysis, scenario analysis, simulation models, and real options analysis. These sections also discuss several common pitfalls in capital budgeting analysis. Finally, Section 8 summarizes the reading.

2 CASH FLOW PROJECTIONS

a calculate the yearly cash flows of expansion and replacement capital projects and evaluate how the choice of depreciation method affects those cash flows;

In this section, we detail how cash flows are found for an "expansion" project. An expansion project is an independent investment that does not affect the cash flows for the rest of the company. In Section 2.4, we will deal with a "replacement" project, in which the cash flow analysis is more complicated. A replacement project must deal with the differences between the cash flows that occur with the new investment and the cash flows that would have occurred for the investment being replaced.

2.1 Table Format with Cash Flows Collected by Year

The cash flows for a conventional expansion project can be grouped into 1) the investment outlays, 2) after-tax operating cash flows over the project's life, and 3) terminal year after-tax non-operating cash flows. Exhibit 1 gives an example of the cash flows for a capital project in which all of the cash flows are collected by year.

Exhibit 1 Capital Budgeting Cash Flows Example (Cash Flows Collected by Year)

Year	0	1	2	3	4	5
Investment outlays						
Fixed capital	–200,000					
Net working capital	–30,000					
Total	–230,000					
Annual after-tax operating cash flows						
Sales		220,000	220,000	220,000	220,000	220,000
Cash operating expenses		90,000	90,000	90,000	90,000	90,000
Depreciation		35,000	35,000	35,000	35,000	35,000
Operating income before taxes		95,000	95,000	95,000	95,000	95,000
Taxes on operating income		38,000	38,000	38,000	38,000	38,000
Operating income after taxes		57,000	57,000	57,000	57,000	57,000
Add back: Depreciation		35,000	35,000	35,000	35,000	35,000
After-tax operating cash flow		92,000	92,000	92,000	92,000	92,000
Terminal year after-tax non-operating cash flows						
After-tax salvage value						40,000
Return of net working capital						30,000
Total						70,000

Exhibit 1 (Continued)

Year	0	1	2	3	4	5
Total after-tax cash flow	−230,000	92,000	92,000	92,000	92,000	162,000
Net present value at 10% required rate of return	162,217					
Internal rate of return	32.70%					

The investment outlays include a $200,000 outlay for fixed capital items. This outlay includes $25,000 for non-depreciable land, plus $175,000 for equipment that will be depreciated straight-line to zero over five years. The investment in net working capital is the net investment in short-term assets required for the investment. This amount is the investment in receivables and inventory needed, less the short-term payables generated by the project. In this case, the project requires $50,000 of current assets but generates $20,000 in current liabilities, resulting in a total investment in net working capital of $30,000. The total investment outlay at time zero is $230,000.

Each year, sales will be $220,000 and cash operating expenses will be $90,000. Annual depreciation for the $175,000 depreciable equipment is $35,000 (one-fifth of the cost). The result is an operating income before taxes of $95,000. Income taxes at a 40% rate are 0.40 × $95,000 = $38,000, which leaves operating income after taxes of $57,000. Adding back the depreciation charge of $35,000 gives the annual after-tax operating cash flow of $92,000.

At the end of Year 5, the company will sell off the fixed capital assets. In this case, the fixed capital assets (including the land) are sold for $50,000, which represents a gain of $25,000 over the remaining book value of $25,000. The gain of $25,000 is taxed at 40%, resulting in a tax of $10,000. This leaves $40,000 for the fixed capital assets after taxes. Additionally, the net working capital investment of $30,000 is recovered because the short-term assets (such as inventory and receivables) and short-term liabilities (such as payables) are no longer needed for the project. Total terminal year non-operating cash flows are then $70,000.

The investment project has a required rate of return of 10%. Discounting the future cash flows at 10% and subtracting the investment outlay gives an NPV of $162,217. The internal rate of return is 32.70%. Because the investment has a positive NPV, this project should be accepted. The project should also be accepted based on the IRR criterion because the IRR is greater than the required rate of return.

2.2 Table Format with Cash Flows Collected by Type

In the layout in Exhibit 1, we essentially collected the cash flows in the columns, by *year*, and then found the NPV by summing the present values of the annual cash flows (at the bottom of each column). There is another way of organizing the same information. We could also find the NPV by finding the present values of the cash flows in Exhibit 1 by rows, which are the *types* of cash flows. Exhibit 2 illustrates this approach:

Exhibit 2 Capital Budgeting Cash Flows Example (Cash Flows Collected by Type)

Time	Type of Cash Flow	Before-Tax Cash Flow	After-Tax Cash Flow	PV at 10%
0	Fixed capital	−200,000	−200,000	−200,000
0	Net working capital	−30,000	−30,000	−30,000

(continued)

Exhibit 2 (Continued)

Time	Type of Cash Flow	Before-Tax Cash Flow	After-Tax Cash Flow	PV at 10%
1–5	Sales minus cash expenses	220,000 – 90,000 = 130,000	130,000(1 – 0.40) = 78,000	295,681
1–5	Depreciation tax savings	None	0.40(35,000) = 14,000	53,071
5	After-tax salvage value	50,000	50,000 – 0.40(50,000 – 25,000) = 40,000	24,837
5	Return of net working capital	30,000	30,000	18,628
			NPV =	162,217

As Exhibit 2 shows, the outlays in fixed capital and in net working capital at time zero total \$230,000. For Years 1 through 5, the company realizes an after-tax cash flow for sales minus cash expenses of \$78,000, which has a present value of \$295,681. The depreciation charge results in a tax savings of \$14,000 per year, which has a present value of \$53,071. The present values of the after-tax salvage and of the return of net working capital are also shown in the Exhibit. The present value of all cash flows is an NPV of \$162,217. Obviously, collecting the after-tax cash flows by year, as in Exhibit 1, or by type, as in Exhibit 2, results in the same NPV.

2.3 Equation Format for Organizing Cash Flows

The capital budgeting cash flows in the foregoing example project were laid out in one of two alternative tabular formats. Analysts also may wish to take another approach. Instead of producing a table, you can also look at the cash flows using equations such as the following:

1 Initial outlay: For a new investment:

$$\text{Outlay} = \text{FCInv} + \text{NWCInv},$$

where

FCInv = investment in new fixed capital

NWCInv = investment in net working capital

This equation can be generalized for a replacement project (covered in Section 2.4.2), in which existing fixed capital is sold and provides some of the funding for the new fixed capital purchased. The outlay is then

$$\text{Outlay} = \text{FCInv} + \text{NWCInv} - \text{Sal}_0 + t(\text{Sal}_0 - B_0), \tag{1}$$

where

Sal_0 = cash proceeds (salvage value) from sale of old fixed capital

t = tax rate

B_0 = book value of old fixed capital

2 Annual after-tax operating cash flow:

$$CF = (S - C - D)(1 - t) + D, \text{ or} \tag{2}$$

$$CF = (S - C)(1 - t) + tD \tag{3}$$

where

S = sales
C = cash operating expenses
D = depreciation charge

3 Terminal year after-tax non-operating cash flow:

$$\text{TNOCF} = \text{Sal}_T + \text{NWCInv} - t(\text{Sal}_T - B_T) \tag{4}$$

where

Sal_T = cash proceeds (salvage value) from sale of fixed capital on termination date
B_T = book value of fixed capital on termination date

The outlay in the example is found with Equation 1:

Outlay = 200,000 + 30,000 – 0 + 0 = $230,000

For a replacement project, the old fixed capital would be sold for cash (Sal_0) and then there would be either taxes paid on the gain (if $\text{Sal}_0 - B_0$ were positive) or a tax saving (if $\text{Sal}_0 - B_0$ were negative). In this example, Sal_0 and $t(\text{Sal}_0 - B_0)$ are zero because no existing fixed capital is sold at time zero.

Using Equation 2, we find that the annual after-tax operating cash flow is

$$\begin{aligned}\text{CF} &= (S - C - D)(1 - t) + D\\ &= (220{,}000 - 90{,}000 - 35{,}000)(1 - 0.40) + 35{,}000\\ &= 95{,}000 \times (0.60) + 35{,}000\\ &= 57{,}000 + 35{,}000 = \$92{,}000\end{aligned}$$

Equation 2 is the project's net income plus depreciation. An identical cash flow results if we use Equation 3:

$$\begin{aligned}\text{CF} &= (S - C)(1 - t) + tD\\ &= (220{,}000 - 90{,}000)(1 - 0.40) + 0.40(35{,}000)\\ &= 130{,}000(0.60) + 0.40(35{,}000) = 78{,}000 + 14{,}000 = \$92{,}000\end{aligned}$$

Equation 3 is the after-tax sales and cash expenses plus the depreciation tax savings. The analyst can use either equation.

Equation 4 provides the terminal year non-operating cash flow:

$$\begin{aligned}\text{TNOCF} &= \text{Sal}_T + \text{NWCInv} - t(\text{Sal}_T - B_T)\\ &= 50{,}000 + 30{,}000 - 0.40(50{,}000 - 25{,}000)\\ &= 50{,}000 + 30{,}000 - 10{,}000 = \$70{,}000\end{aligned}$$

The old fixed capital (including land) is sold for $50,000, but $10,000 of taxes must be paid on the gain. Including the $30,000 return of net working capital gives a terminal year non-operating cash flow of $70,000.

The NPV of the project is the present value of the cash flows—an outlay of $230,000 at time zero, an annuity of $92,000 for five years, plus a single payment of $70,000 in five years:

$$\begin{aligned}\text{NPV} &= -230{,}000 + \sum_{t=1}^{5}\frac{92{,}000}{(1.10)^t} + \frac{70{,}000}{(1.10)^5}\\ &= -230{,}000 + 348{,}752 + 43{,}465 = \$162{,}217\end{aligned}$$

We obtain an identical NPV of $162,217 whether we use a tabular format collecting cash flows by year, a tabular format collecting cash flows by type, or an equation format using Equations 1 through 4. The analyst usually has some flexibility in choosing how to solve a problem. Furthermore, the analysis that an analyst receives from someone else could be in varying formats. The analyst must interpret this information correctly regardless of format. An analyst may need to present information in alternative formats, depending on what the client or user of the information wishes to see. All that is important is that the cash flows are complete (with no cash flows omitted and none double-counted), that their timing is recognized, and that the discounting is done correctly.

2.4 More on Cash Flow Projections

Cash flow analysis can become fairly complicated. Section 2.4 extends the analysis of the previous section to include more details on depreciation methods, replacement projects (as opposed to simple expansion projects), and the effects of inflation.

2.4.1 *Straight-Line and Accelerated Depreciation Methods*

Before going on to more-complicated investment decisions, we should mention the variety of depreciation methods that are in use. The example in Section 2.1 assumed straight-line depreciation down to a zero salvage value. The CFA accounting curriculum and most accounting texts give good descriptions of the straight-line method, the sum-of-years digits method, the double-declining balance method (and the 150% declining balance method), and the units-of-production and service hours method.

All countries specify the depreciation methods that are acceptable for tax purposes in their jurisdictions. The country specifies an asset class for each type of capital investment and a special depreciation schedule for each class. A few of the many examples of depreciation rate schedules appear in Exhibit 3.

Exhibit 3 Examples of Depreciation Rate Schedules

Year	5-Year SL	3-Year DDB	5-Year DDB	10-Year DDB	15-Year 150%DB
1	20.00%	33.33%	20.00%	10.00%	5.00%
2	20.00	44.45	32.00	18.00	9.50
3	20.00	14.81	19.20	14.40	8.55
4	20.00	7.41	11.52*	11.52	7.70
5	20.00		11.52	9.22	6.93
6			5.76	7.37	6.23
7				6.55*	5.90*
8				6.55	5.90
9				6.55	5.90
10				6.55	5.90
11				3.29	5.90
12					5.90
13					5.90
14					5.90

Exhibit 3 (Continued)

Year	5-Year SL	3-Year DDB	5-Year DDB	10-Year DDB	15-Year 150%DB
15					5.90
16					2.99

Notes: SL is straight-line depreciation. DDB is double-declining balance with a half-year convention. 150%DB is 150% declining balance with a half-year convention.
* Denotes a switch to the straight-line method of depreciation when optimal.

Take the five-year DDB property in Exhibit 3 as an example. With double-declining balance, the depreciation each year is 2/5 = 40% of the beginning-of-year book value. With a half-year convention, however, the asset is assumed to be in service for only six months during the first year, and only one-half of the depreciation, or 20%, is allowed the first year. After the first year, the depreciation rate is 40% of the beginning balance until Year 4, when straight-line depreciation would be at least as large, so we switch to straight-line. Yearly depreciation amounts and book values are shown in the following table.

Year	Depreciation (%)	Ending Book Value (%)
1	20.00	80.00
2	32.00 (= 40 × 80)	48.00
3	19.20 (= 40 × 48)	28.80
4	11.52* (= 40 × 28.8)	17.28
5	11.52 (= 40 × 28.8)	5.76
6	5.76 (= 20 × 28.8)	0.00

* Denotes a switch to the straight-line method of depreciation when optimal.

In Year 4, the beginning book value is 28.80% (100% minus the first three years' depreciation), and there are 2.5 years of remaining life. So with straight-line depreciation, 40% of the 28.80% (11.52%) is allocated to Year 4, 40% to Year 5 (11.52%), and 20% (5.76%) to Year 6. In Year 6, we have one-half of a year of the straight-line depreciation remaining because we assumed the asset was placed in service halfway through the first year.

Accelerated depreciation generally improves the NPV of a capital project compared with straight-line depreciation. For an example of this effect, we will assume the same capital project as in Exhibit 1, except that the depreciation is for the three-year property with double-declining balance with a half-year convention (3-Year DDB in Exhibit 3). When using straight-line, the depreciation was 20% per year ($35,000). Using the DDB depreciation percentages for the property shown in Exhibit 3, the first-year depreciation is 0.3333 × 175,000 = $58,327.50, second-year depreciation is 0.4445 × 175,000 = $77,787.50, third-year depreciation is 0.1481 × 175,000 = $25,917.50, fourth-year depreciation is 0.0741 × 175,000 = $12,967.50, and fifth-year depreciation is zero. The impact on the project's NPV and IRR is shown in Exhibit 4.

Exhibit 4 Capital Budgeting Example with Accelerated Depreciation

Year	0	1	2	3	4	5
Investment outlays:						
Fixed capital	–200,000					
Net working capital	–30,000					
Total	–230,000					
Annual after-tax operating cash flows:						
Sales		220,000	220,000	220,000	220,000	220,000
Cash operating expenses		90,000	90,000	90,000	90,000	90,000
Depreciation		58,328	77,788	25,918	12,968	0
Operating income before taxes		71,673	52,213	104,083	117,033	130,000
Taxes on operating income (40%)		28,669	20,885	41,633	46,813	52,000
Operating income after taxes		43,004	31,328	62,450	70,220	78,000
Add back: Depreciation		58,328	77,788	25,918	12,968	0
After-tax operating cash flow		101,331	109,115	88,367	83,187	78,000
Terminal year after-tax non-operating cash flows:						
After-tax salvage value						40,000
Return of net working capital						30,000
Total						70,000
Total after-tax cash flows	–230,000	101,331	109,115	88,367	83,187	148,000
Net present value at 10% required rate of return	$167,403					
Internal rate of return		34.74%				

As Exhibit 4 shows, the depreciation charges still sum to $175,000 (except for $2 of rounding), but they are larger in Years 1 and 2 and smaller in Years 3, 4, and 5. Although this method reduces operating income after taxes in Years 1 and 2 (and increases it in Years 3, 4, and 5), it reduces tax outflows in Years 1 and 2 and increases them later. Consequently, the after-tax operating cash flows (which were $92,000 per year) increase in the early years and decrease in later years. This dynamic increases the NPV from $162,217 to $167,403, a difference of $5,186. The IRR also increases from 32.70% to 34.74%.

The impact of accelerated depreciation can be seen without going through the complete analysis in Exhibit 4. We previously showed in Exhibit 2 that the present value of the depreciation tax savings (which was an annuity of 0.40 × $35,000 = $14,000 a year for five years) was $53,071. The present value of the tax savings from accelerated depreciation is shown in Exhibit 5.

Exhibit 5 Present Value of Tax Savings from Accelerated Depreciation

Year	Depreciation ($)	Tax Savings	PV at 10% ($)
1	58,327.50	0.40 × $58,327.5 = $23,331	21,210
2	77,787.50	0.40 × $77,787.5 = $31,115	25,715
3	25,917.50	0.40 × $25,917.5 = $10,367	7,789
4	12,967.50	0.40 × $12,967.5 = $5,187	3,543

Exhibit 5 (Continued)

Year	Depreciation ($)	Tax Savings	PV at 10% ($)
5	0	0.40 × $0 = $0	0
Total present value			58,257

By using the accelerated depreciation schedule, we increase the present value of the tax savings from $53,071 (from Exhibit 2) to $58,257, an increase of $5,186. The tax deferral associated with the accelerated depreciation (compared with straight-line) adds $5,186 to the NPV of the project.

There are a myriad of tax and depreciation schedules that apply to investment projects around the world. These tax and depreciation schedules are also subject to change from year to year. To accurately assess the profitability of a particular capital project, it is vital to identify and apply the schedules that are relevant to the capital budgeting decision at hand.

2.4.2 *Cash Flows for a Replacement Project*

In Section 2.1, we evaluated the cash flows for an expansion project, basing our after-tax cash flows on the outlays, annual operating cash flows after tax, and salvage value for the project by itself. In many cases, however, investing in a project will be more complicated. Investing could affect many of the company's cash flows. In principle, the cash flows relevant to an investing decision are the incremental cash flows: the cash flows the company realizes *with* the investment compared with the cash flows the company would realize *without* the investment. For example, suppose we are investing in a new project with an outlay of $100,000, and we sell off existing assets that the project replaces for $30,000. The incremental outlay is $70,000.

A very common investment decision is a replacement decision, in which old equipment is replaced with new equipment. This decision requires very careful analysis of the cash flows. The skills required to detail the replacement decision cash flows are also useful for other decisions in which an investment affects other cash flows in the company. We use the term "replacement" loosely, primarily to indicate that the cash flow analysis is more complicated than it was for the simpler expansion decision.

Assume we are considering the replacement of old equipment with new equipment that has more capacity and is less costly to operate. The characteristics of the old and new equipment are given in the following table:

Old Equipment		New Equipment	
Current book value	$400,000		
Current market value	$600,000	Acquisition cost	$1,000,000
Remaining life	10 years	Life	10 years
Annual sales	$300,000	Annual sales	$450,000
Cash operating expenses	$120,000	Cash operating expenses	$150,000
Annual depreciation	$40,000	Annual depreciation	$100,000
Accounting salvage value	$0	Accounting salvage value	$0
Expected salvage value	$100,000	Expected salvage value	$200,000

If the new equipment replaces the old equipment, an additional investment of $80,000 in net working capital will be required. The tax rate is 30%, and the required rate of return is 8%.

The cash flows can be found by carefully constructing tables like Exhibit 1 or by using Equations 1 through 4. The initial outlay is the investment in the new equipment plus the additional investment in net working capital less the after-tax proceeds from selling the old equipment:

$$\text{Outlay} = \text{FCInv} + \text{NWCInv} - \text{Sal}_0 + t(\text{Sal}_0 - B_0)$$
$$\text{Outlay} = 1{,}000{,}000 + 80{,}000 - 600{,}000 + 0.3(600{,}000 - 400{,}000) = \$540{,}000$$

In this case, the outlay of \$540,000 is \$1,080,000 for new equipment and net working capital minus the after-tax proceeds of \$540,000 the company receives from selling the old equipment. The incremental operating cash flows are

$$\begin{aligned} \text{CF} &= (S - C - D)(1 - t) + D \\ &= [(450{,}000 - 300{,}000) - (150{,}000 - 120{,}000) - (100{,}000 - 40{,}000)](1 - 0.30) \\ &\quad + (100{,}000 - 40{,}000) \\ &= (150{,}000 - 30{,}000 - 60{,}000)(1 - 0.30) + 60{,}000 = \$102{,}000. \end{aligned}$$

The incremental sales are \$150,000, incremental cash operating expenses are \$30,000, and incremental depreciation is \$60,000. The incremental after-tax operating cash flow is \$102,000 per year.

At the project termination, the new equipment is expected to be sold for \$200,000, which constitutes an incremental cash flow of \$100,000 over the \$100,000 expected salvage price of the old equipment. Because the accounting salvage values for both the new and old equipment were zero, this gain is taxable at 30%. The company also recaptures its investment in net working capital. The terminal year incremental after-tax non-operating cash flow is

$$\begin{aligned} \text{TNOCF} &= \text{Sal}_T + \text{NWCInv} - t(\text{Sal}_T - B_T) \\ &= (200{,}000 - 100{,}000) + 80{,}000 - 0.30[(200{,}000 - 100{,}000) - (0 - 0)] \\ &= \$150{,}000. \end{aligned}$$

Once the cash flows are identified, the NPV and IRR are readily found. The NPV, found by discounting the cash flows at the 8% required rate of return, is

$$\text{NPV} = -540{,}000 + \sum_{t=1}^{10} \frac{102{,}000}{1.08^t} + \frac{150{,}000}{1.08^{10}} = \$213{,}907$$

The IRR, found with a financial calculator, is 15.40%. Because the NPV is positive, this equipment replacement decision is attractive. The fact that the IRR exceeds the 8% required rate of return leads to the same conclusion.

The key to estimating the incremental cash flows for the replacement is to compare the cash flows that occur with the new investment to the cash flows that would have occurred without the new investment. The analyst is comparing the cash flows with a particular course of action to the cash flows with an alternative course of action.

3 EFFECTS OF INFLATION ON CAPITAL BUDGETING ANALYSIS

b explain how inflation affects capital budgeting analysis;

Inflation affects capital budgeting analysis in several ways. The first decision the analyst must make is whether to do the analysis in "nominal" terms or in "real" terms. Nominal cash flows include the effects of inflation, while real cash flows are adjusted downward to remove the effects of inflation. It is perfectly acceptable to do the analysis in either nominal or real terms, and sound decisions can be made either way. Inflation creates some issues, however, regardless of the approach.

The cash flows and discount rate used should both be nominal or both be real. In other words, nominal cash flows should be discounted at a nominal discount rate, and real cash flows should be discounted at a real rate. The real rate, just like real cash flows, has had the effect of inflation taken out. In general, the relationship between real and nominal rates is

$$(1 + \text{Nominal rate}) = (1 + \text{Real rate})(1 + \text{Inflation rate}).$$

Inflation reduces the value of depreciation tax savings (unless the tax system adjusts depreciation for inflation). The effect of expected inflation is captured in the discounted cash flow analysis. If inflation is higher than expected, the profitability of the investment is correspondingly lower than expected. Inflation essentially shifts wealth from the taxpayer to the government. Higher-than-expected inflation increases the corporation's real taxes because it reduces the value of the depreciation tax shelter. Conversely, lower-than-expected inflation reduces real taxes (the depreciation tax shelters are more valuable than expected).

Inflation also reduces the value of fixed payments to bondholders. When bonds are originally issued, bondholders pay a price for the bonds reflecting their inflationary expectations. If inflation is higher than expected, the real payments to bondholders are lower than expected. Higher-than-expected inflation shifts wealth from bondholders to the issuing corporations. Conversely, if inflation is lower than expected, the real interest expenses of the corporation increase, shifting wealth from the issuing corporation to its bondholders.

Finally, inflation does not affect all revenues and costs uniformly. The company's after-tax cash flows will be better or worse than expected depending on how particular sales outputs or cost inputs are affected. Furthermore, contracting with customers, suppliers, employees, and sources of capital can be complicated as inflation rises.

The capital budgeting model accommodates the effects of inflation, although inflation complicates the capital budgeting process (and the operations of a business, in general).

PROJECT ANALYSIS AND EVALUATION

4

c evaluate capital projects and determine the optimal capital project in situations of 1) mutually exclusive projects with unequal lives, using either the least common multiple of lives approach or the equivalent annual annuity approach, and 2) capital rationing;

Assessing the opportunity costs and analyzing the risks of capital investments becomes more complex and sophisticated as you examine real cases. The first project interaction we examine in this section is that of comparing **mutually exclusive projects** with unequal lives. We will briefly describe other project interactions but will not examine them in detail. We also examine the process of capital budgeting under capital rationing.

4.1 Mutually Exclusive Projects with Unequal Lives

We have previously looked at mutually exclusive projects and decided that the best project is the one with the greatest NPV. However, if the mutually exclusive projects have differing lives and the projects will be replaced (or replicated) repeatedly when they wear out, the analysis is more complicated. The analysis of a one-shot (one time only) investment differs from that of an investment chain (in which the asset is replaced regularly in the future).

For example, assume we have two projects with unequal lives of two and three years, with the following after-tax cash flows:

Project S

CF_t	–100	60	90	
t (time)	0	1	2	3

Project L

CF_t	–140	80	70	60
t (time)	0	1	2	3

Both projects have a 10% required rate of return. The NPV of Project S is \$28.93 and the NPV of Project L is \$35.66. Given that the two projects are mutually exclusive, Project L, with the greater NPV, should be chosen.

Let us now assume, however, that these are not one-shot investments but rather investments in assets that the company will need to replace when they wear out. Project S would be replaced every two years and Project L every three years. This situation is often referred to as a replacement chain. In this type of problem, one should examine the entire chain and not just the first link in the chain. If the projects are part of a replacement chain, examining the cash flows for only the initial investment for Projects S and L is improper because Project L provides cash flows during Year 3, when Project S provides none.

There are two logically equivalent ways of comparing mutually exclusive projects in a replacement chain. They are the "least common multiple of lives" approach and the "equivalent annual annuity" approach.

4.1.1 *Least Common Multiple of Lives Approach*

For the least common multiple of lives approach, the analyst extends the time horizon of analysis so that the lives of both projects will divide exactly into the horizon. For Projects S and L, the least common multiple of 2 and 3 is 6: The two-year project would be replicated three times over the six-year horizon and the three-year project would be replicated two times over the six-year horizon. The cash flows for replicating Projects S and L over a six-year horizon are shown on the following timelines:

Project S

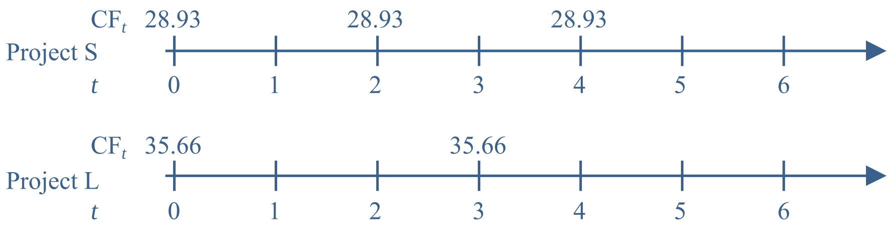
CF_t	–100	60	(–100+90)	60	(–100+90)	60	90
t	0	1	2	3	4	5	6

Project L

CF_t	–140	80	70	(–140+60)	80	70	60
t	0	1	2	3	4	5	6

Discounting the cash flows for the six-year horizon results in an NPV for Project S of \$72.59 and an NPV for Project L of \$62.45. Apparently, investing in Project S and replicating the investment over time has a greater NPV than choosing Project L and replicating it. This decision is the reverse of the one we made when looking solely at the NPVs of the initial investments!

Because the NPV of a single investment represents the present values of its cash flows, you can also visualize the NPV of a replacement chain as the present value of the NPVs of each investment (or link) in the chain. For Projects S and L, the NPVs of each investment are shown on the following timelines:



Investing in Project S is equivalent to receiving values of \$28.93 at times 0, 2, and 4, whereas investing in Project L is equivalent to receiving values of \$35.66 at times 0 and 3. The present values of these cash flow patterns are \$72.59 for Project S and \$62.45 for Project L. Discounting the NPVs of each investment in the chain is equivalent to discounting all of the individual cash flows in the chain.

4.1.2 *Equivalent Annual Annuity Approach*

The other method for properly evaluating a replacement chain is called the equivalent annual annuity (EAA) approach. The name for this approach is very descriptive. For an investment project with an outlay and variable cash flows in the future, the project NPV summarizes the equivalent value at time zero. For this same project, the EAA is the annuity payment (series of equal annual payments over the project's life) that is equivalent in value to the NPV.

Analysts can use a simple two-step procedure to find the EAA. The first step is to find the present value of all of the cash flows for an investment—the investment's NPV. The second step is to calculate an annuity payment that has a value equivalent to the NPV. For Project S above, we already calculated the NPV of the project over its two-year life to be \$28.93. The second step is to find an annuity payment for the two-year life that is equivalent. For a two-year life and a 10% discount rate, a payment of \$16.66 is the equivalent annuity.

The EAA for Project L is found by annuitizing its \$35.66 NPV over three years, so the EAA for Project L is \$14.34.

The decision rule for the EAA approach is to choose the investment chain that has the highest EAA, which in this case is Project S.

Given these two approaches to comparing replacement chains, which one should the analyst use? As a practical matter, the two approaches are logically equivalent and will result in the same decision. More specifically, the NPVs of a replacement chain over the least common multiple of lives (six years) were \$72.59 for Project S and \$62.45 for Project L. If we discount the EAA for Project S (\$16.66) and the EAA for Project L (\$14.34) for six years (treating each as a six-year annuity), we find the same NPVs. Hence, the least common multiple of lives and EAA approaches are consistent with each other. Consequently, the analyst can choose one approach over the other based on personal preference. Or, if the audience for the analyst's work prefers to see the analysis using one approach, the analyst can simply produce the analysis in that format.

4.2 Capital Rationing

When a company's capital budget has a size constraint, a **capital rationing** approach is needed. For example, the capital budget is a fixed money amount. A fixed capital budget can place the company in several interesting situations. To illustrate these, we will assume that the company has a fixed \$1,000 capital budget and has the opportunity to invest in four projects. The projects are of variable profitability.

In the first situation, the budget is adequate to invest in all profitable projects. Consider the four projects in Exhibit 6.

Exhibit 6 First Capital Rationing Example

	Investment Outlay	NPV	PI	IRR (%)
Project 1	600	220	1.37	15
Project 2	200	70	1.35	16
Project 3	200	−60	0.70	10
Project 4	400	−100	0.75	8

In this case, the company has two positive-NPV projects, Projects 1 and 2, which involve a total outlay of $800. Their total NPV is $290. The Profitability Index (PI) for each project – or ratio of the present value of future cash flows to initial investment – is 1.37 and 1.35 respectively. The company should choose these projects, and it will have $200 in its capital budget left over. These excess funds can be used elsewhere in the company (moved to someone else's budget, used to pay dividends or repurchase shares, or used to pay down debt). If a manager is afraid to return the excess funds and chooses to invest in Project 3, the manager will consume the whole capital budget but reduce the total NPV to $230, essentially destroying $60 of wealth for the company.

A second case exists in which the company has more profitable projects than it can choose, but it is able to invest in the most profitable ones available. Continuing with the $1,000 capital budget, this second case is illustrated in Exhibit 7.

Exhibit 7 Second Capital Rationing Example

	Investment Outlay	NPV	PI	IRR (%)
Project 5	600	300	1.50	16
Project 6	200	80	1.40	18
Project 7	200	60	1.30	12
Project 8	200	40	1.20	14

When the analyst has a fixed budget, the PI is especially useful because it shows the profitability of each investment per currency unit invested. If we rank these projects by their PIs, Projects 5, 6, and 7 are the best projects, and we are able to select them. This selection results in a total NPV of $440. The IRRs, shown in the last column, are not a reliable guide to choosing projects under capital rationing because a high-IRR project may have a low NPV. Wealth maximization is best guided by the NPV criterion.

A third case exists in which the company has more profitable projects than it can choose, but it is not able to invest in the most profitable ones available. Assume the company cannot invest in fractional projects: It must take all or none of each project it chooses. Continuing with the $1,000 capital budget, this case is illustrated in Exhibit 8.

Exhibit 8 Third Capital Rationing Example

	Investment Outlay	NPV	PI	IRR (%)
Project 9	600	300	1.50	15
Project 10	600	270	1.45	16

Exhibit 8 (Continued)

	Investment Outlay	NPV	PI	IRR (%)
Project 11	200	80	1.40	12
Project 12	400	100	1.25	11

In this example, an unlimited budget of $1,800 would generate a total NPV of $750. When the budget constraint is imposed, however, the highest NPV results from choosing Projects 9 and 12. The company is forced to choose its best project and its fourth-best project, as indicated by their relative PIs. Any other combination of projects either violates the budget constraint or has a lower total NPV.

Capital rationing has the potential to misallocate resources. Capital markets are supposed to allocate funds to their highest and best uses, with the opportunity cost of funds (used as the discount rate for NPVs or the hurdle rate for IRRs) guiding this allocation process. Capital rationing violates market efficiency if society's resources are not allocated where they will generate the best returns. Companies that use capital rationing may be doing either "hard" or "soft" capital rationing. Under hard capital rationing, the budget is fixed and managers cannot go beyond it. Under soft capital rationing, managers may be allowed to overspend their budgets if they argue effectively that the additional funds will be deployed profitably.

In the case of hard rationing, choosing the optimal projects that fit within the budget and maximize the company's NPV can be computationally intensive. Sometimes, managers use estimates and trial and error to find the optimal set of projects. The PI can be used as a guide in this trial and error process. Other times, the number of possibilities is so daunting that mathematical programming algorithms are used.

5 RISK ANALYSIS OF CAPITAL INVESTMENTS - STAND ALONE METHODS

d explain how sensitivity analysis, scenario analysis, and Monte Carlo simulation can be used to assess the standalone risk of a capital project;

So far, we have evaluated projects by calculating a single NPV to decide whether a project is profitable. We took a single value, or point estimate, of each input into the model and combined the values to calculate the NPV.

Risk is usually measured as a dispersion of outcomes. In the case of standalone risk, we typically measure the riskiness of a project by the dispersion of its NPVs or the dispersion of its IRRs. Sensitivity analysis, scenario analysis, and simulation analysis are very popular standalone risk analysis methods. These risk measures depend on the variation of the project's cash flows.

To illustrate the standalone risk tools, we will use the following "base case" capital project:

Unit price	$5.00
Annual unit sales	40,000
Variable cost per unit	$1.50
Investment in fixed capital	$300,000
Investment in working capital	$50,000

(continued)

Project life	6 years
Depreciation (straight-line)	$50,000
Expected salvage value	$60,000
Tax rate	40%
Required rate of return	12%

The outlay, from Equation 1, is $300,000 plus $50,000, or $350,000. The annual after-tax operating cash flow, from Equation 2, is

$$\begin{aligned} CF &= (S - C - D)(1 - t) + D \\ &= [(5 \times 40{,}000) - (1.50 \times 40{,}000) - (50{,}000)](1 - 0.40) + 50{,}000 \\ &= \$104{,}000 \end{aligned}$$

The terminal year after-tax non-operating cash flow, from Equation 4, is

$$\begin{aligned} TNOCF &= Sal_6 + NWCInv - t(Sal_6 - B_6) \\ &= 60{,}000 + 50{,}000 - 0.40(60{,}000 - 0) = \$86{,}000 \end{aligned}$$

The project NPV is

$$NPV = -350{,}000 + \sum_{t=1}^{6} \frac{104{,}000}{1.12^t} + \frac{86{,}000}{1.12^6} = -350{,}000 + 471{,}157 = \$121{,}157$$

5.1 Sensitivity Analysis

Sensitivity analysis calculates the effect on the NPV of changes in one input variable at a time. The foregoing base case has several input variables. If we wish to do a sensitivity analysis of several of them, we must specify the changes in each that we wish to evaluate. Suppose we want to consider the following:

	Base Value	Low Value	High Value
Unit price	$5.00	$4.50	$5.50
Annual unit sales	40,000	35,000	45,000
Variable cost per unit	$1.50	$1.40	$1.60
Expected salvage value	$60,000	$30,000	$80,000
Tax rate	40%	38%	42%
Required rate of return	12%	10%	14%

We have changed each of six input variables. Exhibit 9 shows the NPV calculated for the base case. Then the NPV is recalculated by changing one variable from its base case value to its high or low value.

Exhibit 9 Sensitivity of Project NPV to Changes in a Variable

	Project NPV			
Variable	Base Case ($)	With Low Estimate ($)	With High Estimate ($)	Range of Estimates ($)
Unit price	121,157	71,820	170,494	98,674
Annual unit sales	121,157	77,987	164,326	86,339
Cost per unit	121,157	131,024	111,289	19,735
Salvage value	121,157	112,037	127,236	15,199

Exhibit 9 (Continued)

Variable	Project NPV Base Case ($)	With Low Estimate ($)	With High Estimate ($)	Range of Estimates ($)
Tax rate	121,157	129,165	113,148	16,017
Required return	121,157	151,492	93,602	57,890

As Exhibit 9 shows, the project's NPV is most sensitive to changes in the unit price variable and least sensitive to changes in the salvage value. Roughly speaking, the project's NPV is most sensitive to changes in unit price and in unit sales. It is least affected by changes in cost per unit, salvage value, and the tax rate. Changes in the required rate of return also have a substantial effect but not as much as changes in price or unit sales.

In a sensitivity analysis, the manager can choose which variables to change and by how much. Many companies have access to software that can be instructed to change a particular variable by a certain amount—for example, to increase or decrease unit price, unit sales, and cost per unit by 10%. The software then produces the changes in NPV for each of these changes. Sensitivity analysis can be used to establish which variables are most influential on a project's success or failure.

5.2 Scenario Analysis

Sensitivity analysis calculates the effect on the NPV of changes in one variable at a time. In contrast, scenario analysis creates scenarios that consist of changes in several of the input variables and calculates the NPV for each scenario. Although corporations could construct a large number of scenarios, in practice they usually use only three. They can be labeled variously, but we will present an example with "pessimistic," "most likely," and "optimistic" scenarios. Continuing with the basic example from the previous section, the values of the input variables for the three scenarios appear in the following table.

Exhibit 10 Input Variables and NPV for Scenario Analysis

Variable	Scenario Pessimistic	Most Likely	Optimistic
Unit price	$4.50	$5.00	$5.50
Annual unit sales	35,000	40,000	45,000
Variable cost per unit	$1.60	$1.50	$1.40
Investment in fixed capital	$320,000	$300,000	$280,000
Investment in working capital	$50,000	$50,000	$50,000
Project life	6 years	6 years	6 years
Depreciation (straight-line)	$53,333	$50,000	$46,667
Salvage value	$40,000	$60,000	$80,000
Tax rate	40%	40%	40%
Required rate of return	13%	12%	11%

(continued)

Exhibit 10 (Continued)

	Scenario		
Variable	**Pessimistic**	**Most Likely**	**Optimistic**
NPV	−$5,725	$121,157	$269,685
IRR	12.49%	22.60%	34.24%

The most likely scenario is the same as the base case we used previously for sensitivity analysis, and the NPV for the most likely scenario is $121,157. To form the pessimistic and optimistic scenarios, managers change several of the assumptions for each scenario. For the pessimistic scenario, several of the input variables are changed to reflect higher costs, lower revenues, and a higher required rate of return. As the table shows, the result is a negative NPV for the pessimistic scenario and an IRR that is less than the pessimistic scenario's 13% required rate of return. For the optimistic scenario, the more favorable revenues, costs, and required rate of return result in very good NPV and IRR.

For this example, the scenario analysis reveals the possibility of an unprofitable investment, with a negative NPV and with an IRR less than the cost of capital. The range for the NPV is fairly large compared with the size of the initial investment, which indicates that the investment is fairly risky. This example included three scenarios for which management wants to know the investment's profitability for each set of assumptions. Other scenarios can be investigated if management chooses to do so.

5.3 Simulation (Monte Carlo) Analysis

Simulation analysis is a procedure for estimating a probability distribution of outcomes, such as for the NPV or IRR for a capital investment project. Instead of assuming a single value (a point estimate) for the input variables in a capital budgeting spreadsheet, the analyst can assume several variables to be stochastic, following their own probability distributions. By simulating the results hundreds or thousands of times, the analyst can build a good estimate of the distributions for the NPV or IRR. Because of the volume of computations, analysts and corporate managers rely heavily on their personal computers and specialized simulation software. Example 1 presents a simple simulation analysis.

EXAMPLE 1

Capital Budgeting Simulation

Gouhua Zhang has made the following assumptions for a capital budgeting project:

- Fixed capital investment is 20,000; no investment in net working capital is required.
- The project has an expected five-year life.
- The fixed capital is depreciated straight-line to zero over a five-year life. The salvage value is normally distributed with an expected value of 2,000 and a standard deviation of 500.
- Unit sales in Year 1 are normally distributed with a mean of 2,000 and a standard deviation of 200.

- Unit sales growth after Year 1 is normally distributed with a mean of 6% and standard deviation of 4%. Assume the same sales growth rate for Years 2–5.
- The sales price is 5.00 per unit, normally distributed with a standard deviation of 0.25 per unit. The same price holds for all five years.
- Cash operating expenses as a percentage of total revenue are normally distributed with a mean and standard deviation of 30% and 3%, respectively.
- The discount rate is 12% and the tax rate is 40%.

1 What are the NPV and IRR, using the expected values of all input variables?

2 Perform a simulation analysis and provide probability distributions for the NPV and IRR.

Solution to 1:

Exhibit 11 Expected Cash Flows for Simulation Example

Time	0	1	2	3	4	5
Fixed capital	−20,000					
After-tax salvage value						1,200
Price		5.00	5.00	5.00	5.00	5.00
Output		2,000	2,120	2,247	2,382	2,525
Revenue		10,000	10,600	11,236	11,910	12,625
Cash operating expenses		3,000	3,180	3,371	3,573	3,787
Depreciation		4,000	4,000	4,000	4,000	4,000
Operating income before taxes		3,000	3,420	3,865	4,337	4,837
Taxes on operating income		1,200	1,368	1,546	1,735	1,935
Operating income after taxes		1,800	2,052	2,319	2,602	2,902
Depreciation		4,000	4,000	4,000	4,000	4,000
Total after-tax cash flow	−20,000	5,800	6,052	6,319	6,602	8,102
NPV (at r = 12%)	3,294					
IRR	18.11%					

Based on the point estimates for each variable (the mean values for each), shown in Exhibit 11, Zhang should find the NPV to be 3,294 and the IRR to be 18.11%.

Solution to 2:

Zhang performs a simulation with 10,000 iterations. For each iteration, values for the five stochastic variables (price, output, output growth rate, cash expense percentage, and salvage value) are selected from their assumed distributions, and the NPV and IRR are calculated. After the 10,000 iterations, the resulting information about the probability distributions for the NPV and IRR is shown in Exhibit 12 and Exhibit 13.

Exhibit 12A Distribution for NPV

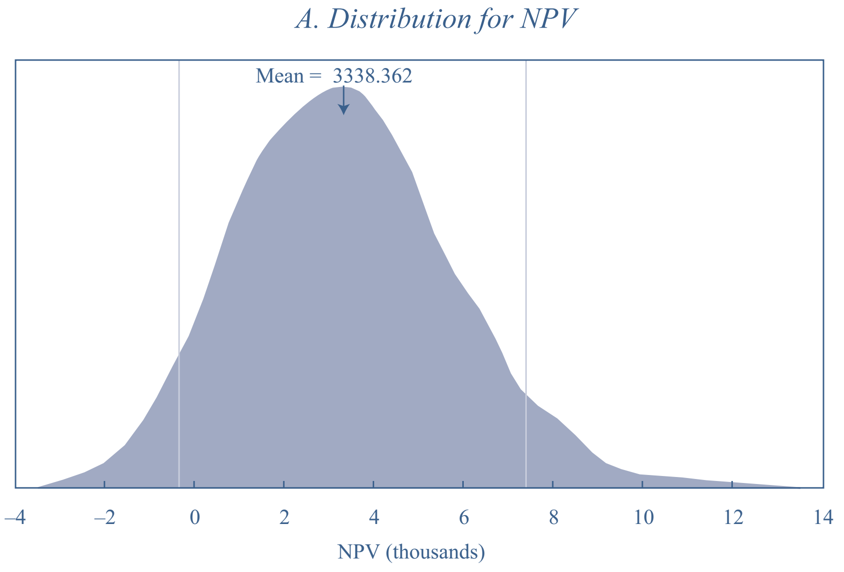


Exhibit 12B Distribution for IRR

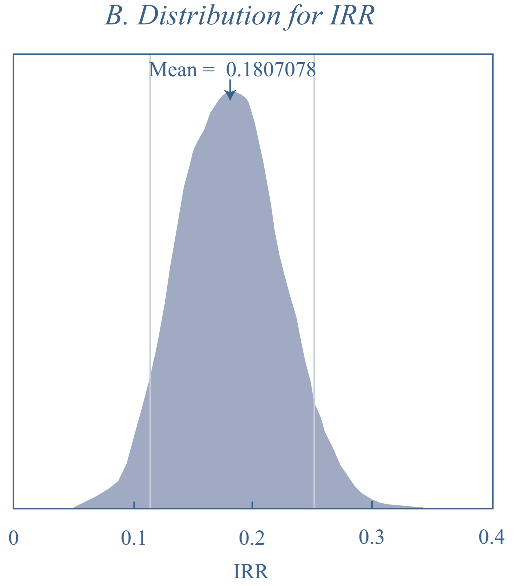


Exhibit 13 Summary Statistics for NPV and IRR

Statistic	NPV	IRR
Mean	3,338	18.07%
Standard deviation	2,364	4.18%
Skewness	0.2909	0.1130
Kurtosis	3.146	2.996

Exhibit 13 (Continued)

Statistic	NPV	IRR
Median	3,236	18.01%
90% confidence interval	−379 to 7,413	11.38% to 25.13%

Correlations between Input Variables and NPV and IRR

Input Variable	NPV	IRR
Output	0.71	0.72
Output growth rate	0.49	0.47
Price	0.34	0.34
Cash expense proportion	−0.28	−0.29
Salvage value	0.06	0.05

As the figure shows, the distributions for the NPV and IRR are somewhat normal looking. The means and standard deviations for each are given in Exhibit 13. Both distributions have a slight positive skewness, which means the distributions are skewed to the right. The two kurtosis values are fairly close to 3.0, which means that the distributions are not peaked or fat-tailed compared with the standard normal distribution. The median is the value at which 50% of the 10,000 outcomes fall on either side. The 90% confidence intervals show that 90% of the observations fall between −379 and 7,413 for the NPV and between 11.38% and 25.13% for the IRR. Although not shown in the table, 7.04% of the observations had a negative NPV and an IRR less than the 12% discount rate.

The means of the NPV and IRR from the simulation (in Exhibit 13) are fairly close to their values calculated using point estimates for all of the input variables (in Exhibit 11). This is not always the case, but it is here. The additional information from a simulation is the dispersions of the NPV and IRR. Given his assumptions and model, the simulation results show Zhang the distributions of NPV and IRR outcomes that should be expected. Managers and analysts often prefer to know these total distributions rather than just their mean values.

The correlations in Exhibit 13 can be interpreted as sensitivity measures. Changes in the "output" variable have the highest correlation with NPV and IRR outcomes. The salvage value has the lowest (absolute value) correlation.

This capital budgeting simulation example was not very complex, with only five stochastic variables. The example's five input variables were assumed to be normally distributed—in reality, many other distributions can be used. Finally, the randomly chosen values for each variable were assumed to be independent. They can be selected jointly instead of independently. Simulation techniques have proved to be a boon for addressing capital budgeting problems.

Sensitivity analysis, scenario analysis, and simulation analysis are well-developed standalone risk analysis methods. These risk measures depend on the variation of the project's cash flows.

6 REAL OPTIONS

e describe types of real options relevant to a capital project;

Real options are capital budgeting options that allow managers to make decisions in the future that alter the value of capital budgeting investment decisions made today. Instead of making all capital budgeting decisions now, at time zero, managers can wait and make additional decisions at future dates when these future decisions are contingent upon future economic events or information. These sequential decisions, in which future decisions depend on the decisions made today as well as on future economic events, are very realistic capital budgeting applications.

Real options are like financial options—they just deal with real assets instead of financial assets. A simple financial option could be a call option on a share of stock. Suppose the stock is selling for \$50, the exercise (strike) price is \$50, and the option expires in one year. If the stock goes up to \$60, you exercise the option and have a gain of \$10 in one year. If the stock goes down to \$40, you do not exercise, and you have no gain. However, zero gain is better than the \$10 loss you would have had if you had purchased the stock at the beginning of the year. Real options, like financial options, entail the right to make a decision but not the obligation. The corporation should exercise a real option only if it is value-enhancing.

Just as financial options are contingent on an underlying asset, real options are contingent on future events. The flexibility that real options give to managers can greatly enhance the NPV of the company's capital investments. The following are several types of these real options:

- *Timing Options.* Instead of investing now, the company can delay investing. Delaying an investment and basing the decision on hopefully improved information that you might have in, say, a year could help improve the NPV of the projects selected. **Project sequencing** options allow the firm to defer the decision to invest in a future project until the outcome of some or all of a current project is known. Projects are sequenced through time, so that investing in a project creates the option to invest in future projects.
- *Sizing Options.* If after investing, the company can abandon the project when the financial results are disappointing, it has an **abandonment option**. At some future date, if the cash flow from abandoning a project exceeds the present value of the cash flows from continuing the project, managers should exercise the abandonment option. Conversely, if the company can make additional investments when future financial results are strong, the company has a **growth option** or an expansion option. When estimating the cash flows from an expansion, the analyst must be wary of **cannibalization**, which occurs when an investment takes customers and sales away from another part of the company.
- *Flexibility Options.* Once an investment is made, other operational flexibilities may be available besides abandonment or expansion. For example, suppose demand exceeds capacity. Management may be able to exercise a **price-setting option**. By increasing prices, the company could benefit from the excess demand, which it cannot do by increasing production. There are also **production-flexibility options**, which offer the operational flexibility to alter production when demand varies from what is forecast. Even though it is expensive, the company can profit from working overtime or from adding additional shifts. The company can also work with customers and suppliers for

their mutual benefit whenever a demand–supply mismatch occurs. This type of option also includes the possibility of using different inputs or producing different outputs.

- *Fundamental Options*. In cases like the aforementioned, there are options embedded in a project that can raise its value. In other cases, the whole investment is essentially an option. The payoffs from the investment are contingent on an underlying asset, just like most financial options. For example, the value of an oil well or refinery investment is contingent on the price of oil. The value of a gold mine is contingent on the price of gold. If oil prices are low, you likely would not choose to drill a well. If oil prices are high, you go ahead and drill. Many R&D (research and development) projects also look like options.

There are several approaches to evaluating capital budgeting projects with real options. One of the difficulties with real options is that the analysis can be very complicated. Although some of the problems are simple and can be readily solved, many of them are so complex that they are expensive to evaluate or you may not have much confidence in the analysis. Four common sense approaches to real options analysis follow.

1 **Use DCF analysis without considering options.** If the NPV is positive without considering real options, and the project has real options that would simply add more value, it is unnecessary to evaluate the options. Just go ahead and make the investment.

2 **Consider the Project NPV = NPV (based on DCF alone) − Cost of options + Value of options.** Calculate the NPV based on expected cash flows. Then simply add the value associated with real options. For example, if a project has a negative NPV based on DCF alone of $50 million, will the options add at least that much to its value?

3 **Use decision trees.** Although they are not as conceptually sound as option pricing models, decision trees can capture the essence of many sequential-decision-making problems.

4 **Use option pricing models.** Except for simple options, the technical requirements for solving these models may require you to hire special consultants or "quants." Some large companies have their own specialists.

The analyst is confronted with 1) a variety of real options that investment projects may possess and 2) a decision about how to reasonably value these options. Example 2 deals with production flexibility; in this case, an additional investment outlay gives the company an option to use alternative fuel sources.

EXAMPLE 2

Production-Flexibility Option

Auvergne AquaFarms estimated the NPV of the expected cash flows from a new processing plant to be −€0.40 million. Auvergne is evaluating an incremental investment of €0.30 million that would give management the flexibility to switch among coal, natural gas, and oil as energy sources. The original plant relied only on coal. The option to switch to cheaper sources of energy when they are available has an estimated value of €1.20 million. What is the value of the new processing plant including this real option to use alternative energy sources?

Solution:

The NPV, including the real option, should be

$$\begin{aligned}\text{Project NPV} &= \text{NPV (based on DCF alone)} - \text{Cost of options} + \\ &\quad \text{Value of options}\\ \text{Project NPV} &= -0.40 \text{ million} - 0.30 \text{ million} + 1.20 \text{ million}\\ &= €0.50 \text{ million}\end{aligned}$$

Without the flexibility offered by the real option, the plant is unprofitable. The real option to adapt to cheaper energy sources adds enough to the value of this investment to give it a positive NPV.

7 COMMON CAPITAL BUDGETING PITFALLS

f describe common capital budgeting pitfalls.

Although the principles of capital budgeting may be easy to learn, applying the principles to real world investment opportunities can be challenging. Some of the common capital budgeting mistakes that managers make are listed here.

Not incorporating economic responses into the investment analysis. Economic responses to an investment often affect its profitability, and these responses have to be correctly anticipated. For example, in response to a successful investment, competitors can enter and reduce the investment's profitability. Similarly, vendors, suppliers, and employees may want to gain from a profitable enterprise. Companies that make highly profitable investments often find that a competitive marketplace eventually causes profitability to revert to normal levels.

Misusing capital budgeting templates. Because hundreds or even thousands of projects need to be analyzed over time, corporations provide standardized capital budgeting templates for managers to use in evaluating projects. This situation creates risks if the template model does not match the project or if employees input inappropriate information.

Pushing pet projects. Projects that influential managers want the corporation to invest in are known as **pet projects**. Ideally, such projects will receive the normal scrutiny that other investments receive and will be selected on the strength of their own merits. Often, unfortunately, pet projects are selected without undergoing normal capital budgeting analysis. Or, the pet project receives the analysis, but overly optimistic projections are used to inflate the project's profitability.

Basing investment decisions on EPS, net income, or return on equity. Managers sometimes have incentives to boost EPS, net income, or ROE. Many investments, even those with strong NPVs, do not boost these accounting numbers in the short run and may even reduce them. Paying attention to short-run accounting numbers can result in choosing projects that are not in the long-run economic interests of the business.

Using IRR to make investment decisions. The NPV criterion is economically sound. The IRR criterion is also sound for independent projects (with conventional cash flow patterns). If projects are mutually exclusive or competitive with each other, investing in projects based on the IRR will tend to result in choosing smaller, short-term projects with high IRRs at the expense of larger, longer-term, high-NPV projects. Basing decisions on paybacks or accounting rates of return is even more dangerous. These measures can be economically unsound.

Incorrectly accounting for cash flows. In analyzing a complicated project, it is easy to omit relevant cash flows, double-count cash flows, and mishandle taxes.

Over- or underestimating overhead costs. In large companies, the cost of a project must include the overhead it generates for such things as management time, information technology support, financial systems, and other support. Although these items are hard to estimate, over- or underestimating these overhead costs can lead to poor investment decisions.

Not using the appropriate risk-adjusted discount rate. The required rate of return for a project should be based on its risk. If a project is being financed with debt (or with equity), you should still use the project's required rate of return and not the cost of debt (or the cost of equity). Similarly, a high-risk project should not be discounted at the company's overall cost of capital, but at the project's required rate of return. Discount rate errors have a huge impact on the computed NPVs of long-lived projects.

Overspending and underspending the capital budget. Spending all of the investment budget just because it is available is another common mistake. Politically, many managers will spend all of their budget and argue that their budget is too small. In a well-run company, managers will return excess funds whenever their profitable projects cost less than their budget, and managers will make a sound case for extra funds if their budget is too small.

Failing to consider investment alternatives. Generating good investment ideas is the most basic step in the capital budgeting process, and many good alternatives are never even considered.

Incorrectly handling sunk costs and opportunity costs. Ignoring sunk costs is difficult for managers to do. Furthermore, not identifying the economic alternatives (real and financial) that are the opportunity costs is probably the biggest failure in much analysis. Only costs that change with the decision are relevant.

SUMMARY

Capital budgeting is the process that companies use for decision making on capital projects—those projects with a life of a year or more. This reading developed the principles behind the basic capital budgeting model, the cash flows that go into the model, and several extensions of the basic model.

- Analysts often organize the cash flows for capital budgeting in tables, summing all of the cash flows occurring at each point in time. These totals are then used to find an NPV or IRR. Alternatively, tables collecting cash flows by type can be used. Equations for the capital budgeting cash flows are as follows:

  Initial outlay:

  $$\text{Outlay} = \text{FCInv} + \text{NWCInv} - \text{Sal}_0 + \text{t}(\text{Sal}_0 - \text{B}_0)$$

  Annual after-tax operating cash flow:

  $$\text{CF} = (\text{S} - \text{C} - \text{D})(1 - t) + \text{D, or}$$

  $$\text{CF} = (\text{S} - \text{C})(1 - t) + t\text{D}$$

  Terminal year after-tax non-operating cash flow:

  $$\text{TNOCF} = \text{Sal}_\text{T} + \text{NWCInv} - t(\text{Sal}_\text{T} - \text{B}_\text{T})$$

- Depreciation schedules affect taxable income, taxes paid, and after-tax cash flows, and therefore capital budgeting valuations.
- When inflation exists, the analyst should perform capital budgeting analysis in "nominal" terms if cash flows are nominal and in "real" terms if cash flows are real.

- Inflation reduces the value of depreciation tax savings (unless the tax system adjusts depreciation for inflation). Inflation reduces the value of fixed payments to bondholders. Inflation usually does not affect all revenues and costs uniformly. Contracting with customers, suppliers, employees, and sources of capital can be complicated as inflation rises.
- Two ways of comparing mutually exclusive projects in a replacement chain are the "least common multiple of lives" approach and the "equivalent annual annuity" approach.
- For the least common multiple of lives approach, the analyst extends the time horizon of analysis so that the lives of both projects will divide exactly into the horizon. The projects are replicated over this horizon, and the NPV for the total cash flows over the least common multiple of lives is used to evaluate the investments.
- The equivalent annual annuity is the annuity payment (series of equal annual payments over the project's life) that is equivalent in value to the project's actual cash flows. Analysts find the present value of all of the cash flows for an investment (the NPV) and then calculate an annuity payment that has a value equivalent to the NPV.
- With capital rationing, the company's capital budget has a size constraint. Under "hard" capital rationing, the budget is fixed. In the case of hard rationing, managers use trial and error and sometimes mathematical programming to find the optimal set of projects. In that situation, it is best to use the NPV or PI valuation methods.
- Sensitivity analysis calculates the effect on the NPV of changes in one input variable at a time.
- Scenario analysis creates scenarios that consist of changes in several of the input variables and calculates the NPV for each scenario.
- Simulation (Monte Carlo) analysis is used to estimate probability distributions for the NPV or IRR of a capital project. Simulations randomly select values for stochastic input variables and then repeatedly calculate the project NPV and IRR to find their distributions.
- Real options can be classified as 1) timing options; 2) sizing options, which can be abandonment options or growth (expansion) options; 3) flexibility options, which can be price-setting options or production-flexibility options; and 4) fundamental options.

PRACTICE PROBLEMS

1 FITCO is considering the purchase of new equipment. The equipment costs \$350,000, and an additional \$110,000 is needed to install it. The equipment will be depreciated straight-line to zero over a five-year life. The equipment will generate additional annual revenues of \$265,000, and it will have annual cash operating expenses of \$83,000. The equipment will be sold for \$85,000 after five years. An inventory investment of \$73,000 is required during the life of the investment. FITCO is in the 40% tax bracket, and its cost of capital is 10%. What is the project NPV?

A \$52,122

B \$64,090

C \$97,449

2 After estimating a project's NPV, the analyst is advised that the fixed capital outlay will be revised upward by \$100,000. The fixed capital outlay is depreciated straight-line over an eight-year life. The tax rate is 40%, and the required rate of return is 10%. No changes in cash operating revenues, cash operating expenses, or salvage value are expected. What is the effect on the project NPV?

A \$100,000 decrease

B \$73,325 decrease

C \$59,988 decrease

3 When assembling the cash flows to calculate an NPV or IRR, the project's after-tax interest expenses should be subtracted from the cash flows for:

A the IRR calculation but not the NPV calculation.

B both the NPV calculation and the IRR calculation.

C neither the NPV calculation nor the IRR calculation.

4 Standard Corporation is investing \$400,000 of fixed capital in a project that will be depreciated straight-line to zero over its 10-year life. Annual sales are expected to be \$240,000, and annual cash operating expenses are expected to be \$110,000. An investment of \$40,000 in net working capital is required over the project's life. The corporate income tax rate is 30%. What is the after-tax operating cash flow expected in Year 1?

A \$63,000

B \$92,000

C \$103,000

5 Five years ago, Frater Zahn's Company invested £38 million—£30 million in fixed capital and another £8 million in working capital—in a bakery. Today, Frater Zahn's is selling the fixed assets for £21 million and liquidating the investment in working capital. The book value of the fixed assets is £15 million, and the marginal tax rate is 40%. The fifth year's after-tax non-operating cash flow to Frater Zahn's is *closest* to:

A £20.6 million.

B £23.0 million.

C £26.6 million.

© 2011 CFA Institute. All rights reserved.

The following information relates to Questions 6–8

Nanjing Techno Group is considering the purchase of a new 60-ton stamping press. The press costs CNY360,000, and an additional CNY40,000 is needed to install it. The press will be depreciated straight-line to zero over a five-year life. The press will generate no additional revenues, but it will reduce cash operating expenses by CNY140,000 annually. The press will be sold for CNY120,000 after five years. An inventory investment of CNY60,000 is required during the life of the investment. Nanjing Techno is in the 40% tax bracket.

6 What is the Nanjing Techno net investment outlay?

A CNY400,000

B CNY420,000

C CNY460,000

7 Nanjing Techno's incremental annual after-tax operating cash flow is *closest* to:

A CNY116,000.

B CNY124,000.

C CNY140,000.

8 What is the terminal year after-tax non-operating cash flow at the end of Year 5?

A CNY108,000.

B CNY132,000.

C CNY180,000.

The following information relates to Questions 9–12

Mei Yang is head of analyst recruiting for PPA Securities. She has been very frustrated by the number of job applicants who, in spite of their stellar pedigrees, seem to have little understanding of basic financial concepts. Yang has written a set of conceptual questions and simple problems for the human resources department to use to screen for the better candidates in the applicant pool. A few of her corporate finance questions and problems follow.

Concept 1 "A company invests in depreciable assets, financed partly by issuing fixed-rate bonds. If inflation is lower than expected, the value of the real tax savings from depreciation and the value of the real after-tax interest expense are both reduced."

Concept 2 "Sensitivity analysis and scenario analysis are useful tools for estimating the impact on a project's NPV of changing the value of one capital budgeting input variable at a time."

Concept 3 "When comparing two mutually exclusive projects with unequal lives, the IRR is a good approach for choosing the better project because it does not require equal lives."

Concept 4 "Project-specific betas should be used instead of company betas whenever the risk of the project differs from that of the company."

Problem	"Bo Gansu Electronics Co., Ltd. is investing CNY100 in a project that is being depreciated straight-line to zero over a two-year life with no salvage value. The project will generate earnings before interest and taxes of CNY50 each year for two years. Bo Gansu Company's weighted average cost of capital and required rate of return for the project are both 12%, and its tax rate is 30%."

9 For Concept 1, the statement is correct regarding the effects on:

A the real tax savings from depreciation but incorrect regarding the real after-tax interest expense.

B both the real tax savings from depreciation and the real after-tax interest expense.

C neither the real tax savings from depreciation nor the real after-tax interest expense.

10 For Concept 2, the statement is correct regarding:

A sensitivity analysis but incorrect regarding scenario analysis.

B scenario analysis but incorrect regarding sensitivity analysis.

C both sensitivity analysis and scenario analysis.

11 Are the statements identified as Concept 3 and Concept 4 correct?

A No for Concepts 3 and 4

B No for Concept 3 but yes for Concept 4

C Yes for Concept 3 but no for Concept 4

12 The after-tax operating cash flows for the Bo Gansu Company are:

A CNY50 in both years.

B CNY70 in both years.

C CNY85 in both years.

The following information relates to Questions 13–18

The capital budgeting committee for Laroache Industries is meeting. Laroache is a North American conglomerate that has several divisions. One of these divisions, Laroache Livery, operates a large fleet of vans. Laroache's management is evaluating whether it is optimal to operate new vans for two, three, or four years before replacing them. The managers have estimated the investment outlay, annual after-tax operating expenses, and after-tax salvage cash flows for each of the service lives. Because revenues and some operating costs are unaffected by the choice of service life, they were ignored in the analysis. Laroache Livery's opportunity cost of funds is 10%. The following table gives the cash flows in thousands of Canadian dollars (C$).

Service Life	Investment	Year 1	Year 2	Year 3	Year 4	Salvage
2 years	−40,000	−12,000	−15,000			20,000
3 years	−40,000	−12,000	−15,000	−20,000		17,000
4 years	−40,000	−12,000	−15,000	−20,000	−25,000	12,000

Schoeman Products, another division of Laroache, has evaluated several investment projects and now must choose the subset of them that fits within its C$40 million capital budget. The outlays and NPVs for the six projects are given below. Schoeman cannot buy fractional projects and must buy all or none of a project. The currency amounts are in millions of Canadian dollars.

Project	Outlay	PV of Future Cash Flows	NPV
1	31	44	13
2	15	21	6
3	12	16.5	4.5
4	10	13	3
5	8	11	3
6	6	8	2

Schoeman wants to determine which subset of the six projects is optimal.

A final proposal comes from the division Society Services, which has an investment opportunity with a real option to invest further if conditions warrant. The crucial details are as follows:

- The original project:
  - An outlay of C$190 million at time zero.
  - Cash flows of C$40 million per year for Years 1–10 if demand is "high."
  - Cash flows of C$20 million per year for Years 1–10 if demand is "low."
- Additional cash flows with the optional expansion project:
  - An outlay of C$190 million at time one.
  - Cash flows of C$40 million per year for Years 2–10 if demand is "high."
  - Cash flows of C$20 million per year for Years 2–10 if demand is "low."
- Whether demand is "high" or "low" in Years 1–10 will be revealed during the first year. The probability of "high" demand is 0.50, and the probability of "low" demand is 0.50.
- The option to make the expansion investment depends on making the initial investment. If the initial investment is not made, the option to expand does not exist.
- The required rate of return is 10%.

Society Services wants to evaluate its investment alternatives.

The internal auditor for Laroache Industries has made several suggestions for improving capital budgeting processes at the company. The internal auditor's suggestions are as follows:

Suggestion 1 "In order to put all capital budgeting proposals on an equal footing, the projects should all use the risk-free rate for the required rate of return."

Suggestion 2 "Because you cannot exercise both of them, you should not permit a given project to have both an abandonment option and an expansion/growth option."

Suggestion 3 "When rationing capital, it is better to choose the portfolio of investments that maximizes the company NPV than the portfolio that maximizes the company IRR."

Suggestion 4 "Project betas should be used for establishing the required rate of return whenever the project's beta is different from the company's beta."

13 What is the optimal service life for Laroache Livery's fleet of vans?

A Two years

B Three years

C Four years

14 The optimal subset of the six projects that Schoeman is considering consists of Projects:

A 1 and 5

B 2, 3, and 4

C 2, 4, 5, and 6

15 What is the NPV (C$ millions) of the original project for Society Services without considering the expansion option?

A −6.11

B −5.66

C 2.33

16 What is the NPV (C$ millions) of the optimal set of investment decisions for Society Services including the expansion option?

A 6.34

B 12.68

C 31.03

17 Should the capital budgeting committee accept the internal auditor's first and second suggestions, respectively?

A No for Suggestions 1 and 2

B No for Suggestion 1 and yes for Suggestion 2

C Yes for Suggestion 1 and no for Suggestion 2

18 Should the capital budgeting committee accept the internal auditor's third and fourth suggestions, respectively?

A No for Suggestions 3 and 4

B Yes for Suggestions 3 and 4

C No for Suggestion 3 and yes for Suggestion 4

The following information relates to Questions 19–24

Maximilian Böhm is reviewing several capital budgeting proposals from subsidiaries of his company. Although his reviews deal with several details that may seem like minutiae, the company places a premium on the care it exercises in making its investment decisions.

The first proposal is a project for Richie Express, which is investing $500,000, all in fixed capital, in a project that will have operating income after taxes of $20,000 and depreciation of $40,000 each year for the next three years. Richie Express will sell the asset in three years, paying 30% taxes on any excess of the selling price over book value. The proposal indicates that a $647,500 terminal selling price will enable the company to earn a 15% internal rate of return on the investment. Böhm doubts that this terminal value estimate is correct.

Another proposal concerns Gasup Company, which does natural gas exploration. A new investment has been identified by the Gasup finance department with the following projected cash flows:

- Investment outlays are $6 million immediately and $1 million at the end of the first year.
- After-tax operating cash flows are $0.5 million at the end of the first year and $4 million at the end of each of the second, third, fourth, and fifth years. In addition, an after-tax outflow occurs at the end of the five-year project that has not been included in the operating cash flows: $5 million required for environmental cleanup.
- The required rate of return on natural gas exploration is 18%.

The Gasup analyst is unsure about the calculation of the NPV and the IRR because the outlay is staged over two years.

Finally, Dominion Company is evaluating two mutually exclusive projects: The Pinto grinder involves an outlay of $100,000, annual after-tax operating cash flows of $45,000, an after-tax salvage value of $25,000, and a three-year life. The Bolten grinder has an outlay of $125,000, annual after-tax operating cash flows of $47,000, an after-tax salvage value of $20,000, and a four-year life. The required rate of return is 10%. The NPV and EAA of the Pinto grinder are $30,691 and $12,341, respectively. Whichever grinder is chosen, it will have to be replaced at the end of its service life. The analyst is unsure about which grinder should be chosen.

Böhm and his colleague Beth Goldberg have an extended conversation about capital budgeting issues, which includes the following comments. Goldberg makes two comments about real options:

Comment 1 "The abandonment option is valuable, but it should be exercised only when the abandonment value is above the amount of the original investment."

Comment 2 "If the cost of a real option is less than its value, this will increase the NPV of the investment project in which the real option is embedded."

Böhm also makes several comments about specific projects under consideration:

Comment A The land and building were purchased five years ago for $10 million. This is the amount that should now be included in the fixed capital investment."

Comment B "We can improve the project's NPV by using the after-tax cost of debt as the discount rate. If we finance the project with 100% debt, this discount rate would be appropriate."

Comment C "It is generally safer to use the NPV than the IRR in making capital budgeting decisions. When evaluating mutually exclusive projects, however, if the projects have conventional cash flow patterns and have the same investment outlays, it is acceptable to use either the NPV or IRR."

Comment D "You should not base a capital budgeting decision on its immediate impact on EPS."

19 What terminal selling price is required for a 15% internal rate of return on the Richie project?

A $588,028

B $593,771

C $625,839

20 The NPV and IRR, respectively, of the Gasup Company investment are *closest* to:

A $509,600 and 21.4%.

B $509,600 and 31.3%.

C $946,700 and 31.3%.

21 Of the two grinders that the Dominion Company is evaluating, Böhm should recommend the:

A Bolten grinder because its NPV is higher than the Pinto grinder NPV.

B Bolten grinder because its EAA is higher than the Pinto grinder EAA.

C Pinto grinder because its EAA is higher than the Bolten grinder EAA.

22 Are Goldberg's comments about real options correct?

A No for Comment 1 and Comment 2

B No for Comment 1 and yes for Comment 2

C Yes for Comment 1 and no for Comment 2

23 Is Böhm most likely correct regarding Comment A about the $10 million investment and Comment B about using the after-tax cost of debt?

A No for both comments

B Yes for both comments

C No for Comment A and yes for Comment B

24 Is Böhm most likely correct regarding Comment C that it is acceptable to use either NPV or IRR and Comment D about the immediate impact on EPS?

A No for both comments

B Yes for both comments

C No for Comment C and yes for Comment D

The following information relates to Questions 25–30

Geet Patel is a sell-side analyst who covers the pharmaceutical industry. One of the companies she follows, PRT-Assam Pharma, is evaluating a regional distribution center. The financial predictions for the capital project are as follows:

- Fixed capital outlay is INR1.50 billion.
- Investment in net working capital is INR0.40 billion.
- Straight-line depreciation is over a six-year period with zero salvage value.
- Project life is 12 years.
- Additional annual revenues are INR0.10 billion.

- Annual cash operating expenses are reduced by INR0.25 billion.
- The capital equipment is sold for INR0.50 billion in 12 years.
- Tax rate is 40%.
- Required rate of return is 12%.

Patel is evaluating this investment to see whether it has the potential to affect PRT-Assam Pharma's stock price. Patel estimates the NPV of the project to be INR0.41 billion, which should increase the company's value.

Patel is evaluating the effects of other changes to her capital budgeting assumptions. She wants to know the effect of a switch from straight-line to accelerated depreciation on the company's operating income and the project's NPV. She also believes that the initial outlay might be much smaller than initially assumed. Specifically, she thinks the outlay for fixed capital might be INR0.24 billion lower, with no change in salvage value.

When reviewing her work, Patel's supervisor provides the following comments. "I note that you are relying heavily on the NPV approach to valuing the investment decision. I don't think you should use an IRR because of the multiple IRR problem that is likely to arise with the PRT-Assam Pharma project. The equivalent annual annuity, however, would be a more appropriate measure to use for the project than the NPV. I suggest that you compute an EAA."

25 Patel should estimate the after-tax operating cash flow for Years 1–6 and 7–12, respectively, to be *closest* to:

A INR0.31 billion and INR0.21 billion.

B INR0.31 billion and INR0.25 billion.

C INR0.35 billion and INR0.25 billion.

26 Patel should estimate the initial outlay and the terminal year non-operating cash flow, respectively, to be *closest* to:

A INR1.50 billion and INR0.70 billion.

B INR1.90 billion and INR0.70 billion.

C INR1.90 billion and INR0.90 billion.

27 Is Patel's estimate of the NPV of the project correct?

A Yes

B No, the NPV is –INR0.01 billion

C No, the NPV is INR0.34 billion

28 A switch from straight-line to accelerated depreciation would:

A increase the NPV and decrease the first-year operating income after taxes.

B increase the first-year operating income after taxes and decrease the NPV.

C increase both the NPV and first-year operating income after taxes.

29 If the outlay is lower by the amount that Patel suggests, the project NPV should increase by an amount *closest* to:

A INR0.09 billion.

B INR0.14 billion.

C INR0.17 billion.

30 How should you evaluate the comments by Patel's supervisor about not using the IRR and about using the EAA? The supervisor is:

A incorrect about both.

B correct about IRR and incorrect about EAA.

C incorrect about IRR and correct about EAA.

The following information relates to Questions 31–36

Carlos Velasquez, CFA, is a financial analyst with Embelesado, S.A., a Spanish manufacturer of sailboats and sailing equipment. Velasquez is evaluating a proposal for Embelesado to build sailboats for a foreign competitor that lacks production capacity and sells in a different market. The sailboat project is perceived to have the same risk as Embelesado's other projects.

The proposal covers a limited time horizon—three years—after which the competitor expects to be situated in a new, larger production facility. The limited time horizon appeals to Embelesado, which currently has excess capacity but expects to begin its own product expansion in slightly more than three years.

Velasquez has collected much of the information necessary to evaluate this proposal in Exhibits 1 and 2.

Exhibit 1 Selected Data for Sailboat Proposal (currency amounts in € millions)

Initial fixed capital outlay	60
Annual contracted revenues	60
Annual operating costs	25
Initial working capital outlay (recovered at end of the project)	10
Annual depreciation expense (both book and tax accounting)	20
Economic life of facility (years)	3
Salvage (book) value of facility at end of project	0
Expected market value of facility at end of project	5

Exhibit 2 Selected Data for Embelesado, S.A.

Book value of long-term debt/total assets	28.6%
Book value of equity/total assets	71.4%
Market value of long-term debt/market value of company	23.1%
Market value of equity/market value of company	76.9%
Coupon rate on existing long-term debt	8.5%
Interest rate on new long-term debt	8.0%
Cost of equity	13.0%
Marginal tax rate	35.0%
Maximum acceptable payback period	2 years

Velasquez recognizes that Embelesado is currently financed at its target capital structure and expects that the capital structure will be maintained if the sailboat project is undertaken. Embelesado's managers disagree, however, about the method that should be used to evaluate capital budgeting proposals.

One of Embelesado's vice presidents asks Velasquez the following questions:

Question 1	Will projects that meet a corporation's payback criterion for acceptance necessarily have a positive NPV?
Question 2	For mutually exclusive projects, will the NPV and IRR methods necessarily agree on project ranking?
Question 3	For the sailboat project, what will be the effects of using accelerated depreciation (for both book and tax accounting) instead of straight-line depreciation on a) the NPV and b) the total net cash flow in the terminal year?
Question 4	Assuming a 13% discount rate, what will be the increase in the sailboat project's NPV if the expected market value of the facility at end of project is €15 million rather than €5 million?

31 The weighted average cost of capital for Embelesado is *closest* to:

A 10.78%.

B 11.20%.

C 11.85%.

32 The total net cash flow (in € millions) for the sailboat project in its terminal year is *closest* to:

A 33.00.

B 39.75.

C 43.00.

33 The IRR for the sailboat project is *closest* to:

A 18.5%.

B 19.7%.

C 20.3%.

34 The best responses that Velasquez can make to Question 1 and Question 2 are:

	Question 1	Question 2
A	No	No
B	No	Yes
C	Yes	No

35 In response to Question 3, what are the *most likely* effects on the NPV and the total net cash flow in the terminal year, respectively?

	NPV	Total Net Cash Flow in Terminal Year
A	Increase	Increase
B	Increase	Decrease
C	Decrease	Increase

36 In response to Question 4, the increase in the sailboat project's NPV (in € millions) is *closest* to:

A 4.50.

B 6.50.

C 6.76.

The following information relates to Questions 37–42

María Hernández is a sell-side analyst covering the electronics industry in Spain. One of the companies she follows, SG Electronics, S.A., has recently announced plans to begin producing and selling a new series of video cameras. Hernández estimates that this project will increase the value of the company, and consequently, she plans to change her research opinion on the company from a "hold" to a "buy." Her initial financial predictions for the project are as follows:

- Fixed capital equipment outlay is €2,750,000.
- At the beginning of the project, a required increase in current assets of €200,000 and a required increase in current liabilities of €125,000.
- Straight-line depreciation to zero over a five-year life.
- Project life of five years.
- Incremental annual unit sales of 3,000 at a unit price of €600.
- Annual fixed cash expenses of €125,000; variable cash expenses of €125 per unit.
- The capital equipment is expected to be sold for €450,000 at the end of Year 5. At the end of the project, the net working capital investment will be recovered.
- Tax rate of 40%.
- Based on the capital asset pricing model, the required rate of return is 12%.

Hernández estimates the project's expected NPV to be €975,538 and the IRR to be 24.6%. She also performs a sensitivity analysis by changing the input variable assumptions used in her initial analysis.

When reviewing Hernández's work, her supervisor, Arturo Costa, notes that she did not include changes in the depreciation method, initial fixed capital outlay, or inflation assumptions in her sensitivity analysis. As a result, Costa asks the following questions:

Question 1 "What would be the effect on the project's NPV if the initial fixed capital equipment outlay increased from €2,750,000 to €3,000,000, with everything else held constant?"

Question 2 "How would a higher-than-expected inflation rate affect the value of the real tax savings from depreciation and the value of the real after-tax interest expense, with everything else held constant?"

Question 3 "You are using a required rate of return of 12% when the company's weighted average cost of capital (WACC) is 10%. Why are you using a required rate of return for the project greater than the company's WACC?"

Before ending the meeting, Costa tells Hernández: “Last year the company produced a prototype at a cost of €500,000. Now management is having doubts about the market appeal of the product in its current design, and so they are considering delaying the start of the project for a year, until the prototype can be shown to industry experts.”

37 Using Hernández’s initial financial predictions, the estimated annual after-tax operating cash flow is *closest* to:

A €780,000.

B €1,000,000.

C €1,075,000.

38 Using Hernández’s initial financial predictions, the estimated terminal year after-tax non-operating cash flow is *closest* to:

A €195,000.

B €270,000.

C €345,000.

39 Hernández’s best response to Costa’s first question is that the project’s NPV would decrease by an amount *closest* to:

A €142,000.

B €178,000.

C €250,000.

40 Hernández’s *best* response to Costa’s second question is that:

A real tax savings from depreciation and real interest expense would be lower.

B real tax savings from depreciation would be higher and real interest expense would be lower.

C real tax savings from depreciation would be lower and real interest expense would be higher.

41 Hernández’s *best* response to Costa’s third question is: “Because:

A the project will plot above the security market line.”

B the project’s beta is greater than the company’s beta.”

C the project’s IRR is greater than the required rate of return.”

42 Should Costa’s end-of-meeting comments result in changes to Hernández’s capital budgeting analysis?

A No

B Yes, but only to incorporate the possible delay

C Yes, to incorporate both the possible delay and the cost of producing the prototype

SOLUTIONS

1 C is correct.

Outlay = FCInv + NWCInv − Sal_0 + t(Sal_0 − B_0)

= (350,000 + 110,000) + 73,000 − 0 + 0

= $533,000

The installed cost is $350,000 + $110,000 = $460,000, so the annual depreciation is $460,000/5 = $92,000. The annual after-tax operating cash flow for Years 1–5 is

$$\begin{aligned} CF &= (S - C - D)(1 - t) + D \\ &= (265{,}000 - 83{,}000 - 92{,}000)(1 - 0.40) + 92{,}000 \\ &= \$146{,}000. \end{aligned}$$

The terminal year after-tax non-operating cash flow in Year 5 is

$$\begin{aligned} TNOCF &= Sal_5 + NWCInv - t(Sal_5 - B_5) \\ &= 85{,}000 + 73{,}000 - 0.40(85{,}000 - 0) \\ &= \$124{,}000. \end{aligned}$$

The NPV is

$$NPV = -533{,}000 + \sum_{t=1}^{5} \frac{146{,}000}{1.10^t} + \frac{124{,}000}{1.10^5} = \$97{,}449$$

2 B is correct. The additional annual depreciation is $100,000/8 = $12,500. The depreciation tax savings is 0.40 ($12,500) = $5,000. The change in project NPV is

$$-100{,}000 + \sum_{t=1}^{8} \frac{5{,}000}{(1.10)^t} = -100{,}000 + 26{,}675 = -\$73{,}325$$

3 C is correct. Financing costs are not subtracted from the cash flows for either the NPV or the IRR. The effects of financing costs are captured in the discount rate used.

4 C is correct. The annual depreciation charge is $400,000/10 = $40,000. The after-tax operating cash flow in Year 1 should be

$$\begin{aligned} CF &= (S - C - D)(1 - t) + D \\ &= (240{,}000 - 110{,}000 - 40{,}000)(1 - 0.30) + 40{,}000 \\ &= 63{,}000 + 40{,}000 \\ &= \$103{,}000. \end{aligned}$$

5 C is correct. The terminal year after-tax non-operating cash flow is

$$\begin{aligned} TNOCF &= Sal_5 + NWCInv - t(Sal_5 - B_5) \\ &= 21 + 8 - 0.40(21 - 15) \\ &= £26.6.\ \text{million}. \end{aligned}$$

6 C is correct. The investment outlay is

$$\begin{aligned} Outlay &= FCInv + NWCInv - Sal_0 + t(Sal_0 - B_0) \\ &= (360{,}000 + 40{,}000) + 60{,}000 - 0 + 0 \\ &= CNY460{,}000. \end{aligned}$$

7 A is correct. Depreciation will be CNY400,000/5 = CNY80,000 per year. The annual after-tax operating cash flow is

$$\begin{aligned} CF &= (S - C - D)(1 - t) + D \\ &= [0 - (-140{,}000) - 80{,}000](1 - 0.40) + 80{,}000 \\ &= \text{CNY}116{,}000. \end{aligned}$$

8 B is correct. The terminal year non-operating cash flow is

$$\begin{aligned} TNOCF &= Sal_5 + NWCInv - t(Sal_5 - B_5) \\ &= 120{,}000 + 60{,}000 - 0.40(120{,}000 - 0) \\ &= \text{CNY}132{,}000. \end{aligned}$$

9 C is correct. The value of the depreciation tax savings is increased, and the value of the real after-tax interest expense is also increased. Because of the lower inflation, the value has increased (essentially discounting at a lower rate).

10 A is correct. The statement is correct for sensitivity analysis but not for scenario analysis (in which several input variables are changed for each scenario).

11 B is correct. Either the least-common multiple of lives or the equivalent annual annuity approach should be used (both use the NPV, not the IRR). Concept 4 is correct as given.

12 C is correct. The problem gives EBIT, not EBITDA.

$$CF = (S - C - D)(1 - t) + D$$

$$= 50(1 - 0.3) + 50$$

$$= \text{CNY}85 \text{ each year}$$

13 B is correct. The way to solve the problem is to calculate the equivalent annual annuity and choose the service life with the lowest annual cost. For a two-year service life, the NPV is

$$NPV = -40{,}000 + \frac{-12{,}000}{1.10^1} + \frac{-15{,}000}{1.10^2} + \frac{20{,}000}{1.10^2} = -46{,}776.86$$

The EAA (PV = −46,776.86, N = 2, and i = 10%) is −26,952.38.

For a three-year service life, the NPV is

$$\begin{aligned} NPV &= -40{,}000 + \frac{-12{,}000}{1.10^1} + \frac{-15{,}000}{1.10^2} + \frac{-20{,}000}{1.10^3} + \frac{17{,}000}{1.10^3} \\ &= -65{,}559.73 \end{aligned}$$

The EAA (PV = −65,559.73, N = 3, and i = 10%) is −26,362.54.

For a four-year service life, the NPV is

$$\begin{aligned} NPV &= -40{,}000 + \frac{-12{,}000}{1.10^1} + \frac{-15{,}000}{1.10^2} + \frac{-20{,}000}{1.10^3} + \frac{-25{,}000}{1.10^4} \\ &\quad + \frac{12{,}000}{1.10^4} = -87{,}211.26 \end{aligned}$$

The EAA (PV = −87,211.26, N = 4, and i = 10%) is −27,512.61.

The three-year service life has the lowest annual cost. Laroache should replace the vans every three years.

14 A is correct. To help the selection process, use the profitability index for each project, which shows the total present value per dollar invested.

Project	Outlay	PV of Future Cash Flows	NPV	PI	PI Rank
1	31	44	13	1.419	1
2	15	21	6	1.400	2
3	12	16.5	4.5	1.375	(tie) 3
4	10	13	3	1.300	6
5	8	11	3	1.375	(tie) 3
6	6	8	2	1.333	5

Try to incorporate the high-PI projects into the budget using trial and error. These trials include the following:

Set of Projects	Total Outlay	Total NPV
1 and 5	39	16
2, 3, and 4	37	13.5
2, 3, and 5	35	13.5
2, 4, 5, and 6	39	14

Among the sets of projects suggested, the optimal set is the one with the highest NPV, provided its total outlay does not exceed C$40 million. The set consisting of Projects 1 and 5 produces the highest NPV.

15 B is correct.

If demand is "high," the NPV is

$$\text{NPV} = -190 + \sum_{t=1}^{10} \frac{40}{1.10^t} = \text{C\$55.783 million}$$

If demand is "low," the NPV is

$$\text{NPV} = -190 + \sum_{t=1}^{10} \frac{20}{1.10^t} = -\text{C\$67.109 million}$$

The expected NPV is 0.50(55.783) + 0.50(–67.109) = –C$5.663 million.

16 B is correct. Assume we are at Time 1. The NPV of the expansion (at Time 1) if demand is "high" is

$$\text{NPV} = -190 + \sum_{t=1}^{9} \frac{40}{1.10^t} = \text{C\$40.361 million}$$

The NPV of the expansion (at Time 1) if demand is "low" is

$$\text{NPV} = -190 + \sum_{t=1}^{9} \frac{20}{1.10^t} = -\text{C\$74.820 million}$$

The optimal decision is to expand if demand is "high" and not expand if "low."

Because the expansion option is exercised only when its value is positive, which happens 50% of the time, the expected value of the expansion project, at Time 0, is

$$\text{NPV} = \frac{1}{1.10} 0.50(40.361) = \text{C\$18.346 million}$$

The total NPV of the initial project and the expansion project is

NPV = −C$5.663 million + C$18.346 million = C$12.683 million.

The optional expansion project, handled optimally, adds sufficient value to make this a positive NPV project.

17 A is correct. Both suggestions are bad. In valuing projects, expected cash flows should be discounted at required rates of return that reflect their risk, not at a risk-free rate that ignores risk. Even though both options cannot be simultaneously exercised, they can both add value. If demand is high, you can exercise the growth option, and if demand is low, you can exercise the abandonment option.

18 B is correct. Both suggestions are good. Choosing projects with high IRRs might cause the company to concentrate on short-term projects that reduce the company's NPV. Whenever the project risk differs from the company risk, a project-specific required rate of return should be used.

19 C is correct. The after-tax operating cash flow for each of the next three years is \$20,000 + \$40,000 = \$60,000. The book value in three years will be \$380,000 (the original cost less three years' depreciation). So the terminal year after-tax non-operating cash flow will be Sal_3 − 0.30(Sal_3 −\$380,000), where Sal_3 is the selling price. For a 15% return, the PV of future cash flows must equal the investment:

$$500{,}000 = \frac{60{,}000}{1.15} + \frac{60{,}000}{1.15^2} + \frac{60{,}000}{1.15^3} + \frac{\mathrm{Sal}_3 - 0.30\left(\mathrm{Sal}_3 - 380{,}000\right)}{1.15^3}$$

There are several paths to follow to solve for Sal_3.

$$363{,}006.5 = \frac{\mathrm{Sal}_3 - 0.30\left(\mathrm{Sal}_3 - 380{,}000\right)}{1.15^3}$$

$$\mathrm{Sal}_3 - 0.30\left(\mathrm{Sal}_3 - 380{,}000\right) = 552{,}087.5$$

$$0.70\ \mathrm{Sal}_3 = 438{,}087.5$$

$$\mathrm{Sal}_3 = \$625{,}839$$

20 A is correct. The cash flows (in \$ millions) for the five-year gas project are as follows:

Time	Outlays	After-Tax Operating Cash Flows	Total After-Tax Cash Flows
0	6.0	0.0	−6.0
1	1.0	0.5	−0.5
2	0.0	4.0	4.0
3	0.0	4.0	4.0
4	0.0	4.0	4.0
5	5.0	4.0	−1.0

Given the required rate of return of 18%, the NPV can be calculated with the following equation or with a financial calculator:

$$\mathrm{NPV} = -6.0 + \frac{-0.5}{1.18} + \frac{4.0}{1.18^2} + \frac{4.0}{1.18^3} + \frac{4.0}{1.18^4} + \frac{-1.0}{1.18^5}$$

$$\mathrm{NPV} = \$509{,}579$$

Similarly, the IRR can be calculated from the following equation:

$$-6.0 + \frac{-0.5}{1+r} + \frac{4.0}{(1+r)^2} + \frac{4.0}{(1+r)^3} + \frac{4.0}{(1+r)^4} + \frac{-1.0}{(1+r)^5} = 0$$

Solving for r with a financial calculator or spreadsheet software will yield 21.4% for the internal rate of return. Note that in spite of the fact that we are dealing with a non-conventional cash flow pattern, the IRR has a unique solution. The NPV profile declines as the required rate of return increases, and the NPV value crosses the x-axis (required rate of return) only one time, at 21.4%.

21 C is correct. Because the mutually exclusive projects have unequal lives, the EAA should be used instead of the NPV. The NPV and EAA for the Pinto grinder are correct. For the Bolten grinder, the NPV is

$$\text{NPV} = -125{,}000 + \sum_{t=1}^{4} \frac{47{,}000}{1.10^t} + \frac{20{,}000}{1.10^4} = 37{,}644$$

To find the Bolten EAA, take the NPV for Bolten and annualize it for four years (N = 4, PV = 37,644, and i = 10%). The Bolten EAA is \$11,876. Consequently, the Pinto grinder has the better EAA of \$12,341.

22 B is correct. Goldberg's first comment is wrong. A project should be abandoned in the future only when its abandonment value is more than the discounted value of the remaining cash flows. Goldberg's second comment is correct.

23 A is correct. The \$10 million original cost is a sunk cost and not relevant. The correct investment is today's opportunity cost, the market value today. The correct discount rate is the project required rate of return.

24 C is correct. Even if they are the same size, a short-term project with a high IRR can have a lower NPV than a longer-term project. The immediate impact on EPS does not capture the full effect of the cash flows over the project's entire life.

25 A is correct. The annual depreciation charge for Years 1–6 is 1.5/6 = 0.25. Annual after-tax operating cash flows for Years 1–6 are as follows:

$$\text{CF} = (\text{S} - \text{C} - \text{D})(1 - \text{t}) + \text{D}$$

$$= [0.10 - (-0.25) - 0.25](1 - 0.40) + 0.25$$

$$= 0.06 + 0.25 = \text{INR}0.31 \text{ billion}$$

Annual after-tax operating cash flows for Years 7–12 are as follows:

$$\text{CF} = (\text{S} - \text{C} - \text{D})(1 - \text{t}) + \text{D}$$

$$= [0.10 - (-0.25) - 0](1 - 0.40) + 0$$

$$= \text{INR}0.21 \text{ billion}$$

26 B is correct.

Outlay at time zero is as follows:

$$\text{Outlay} = \text{FCInv} + \text{NWCInv} - \text{Sal}_0 + \text{t}(\text{Sal}_0 - \text{B}_0)$$

$$= 1.50 + 0.40 - 0 + 0$$

$$= \text{INR}1.90 \text{ billion}$$

Terminal year after-tax non-operating cash flow is

$$\text{TNOCF} = \text{Sal}_{12} + \text{NWCInv} - \text{t}(\text{Sal}_{12} - \text{B}_{12})$$

$= 0.50 + 0.40 - 0.40(0.50 - 0)$

$= \text{INR}0.70$ billion

27 B is correct. The cash flows, computed in the first two questions, are as follows:

Time 0	–INR1.90 billion
Time 1–6	INR0.31 billion
Time 7–12	INR0.21 billion
Time 12	INR0.70 billion

The NPV is

$$\begin{aligned} \text{NPV} &= -1.90 + \sum_{t=1}^{6}\frac{0.31}{1.12^t} + \sum_{t=7}^{12}\frac{0.21}{1.12^t} + \frac{0.70}{1.12^{12}} \\ &= -1.90 + 1.2745 + 0.4374 + 0.1797 \\ &= -\text{INR}0.0084 \text{ billion} \approx -\text{INR}0.01 \text{ billion} \end{aligned}$$

28 A is correct. Accelerated depreciation shifts depreciation expense toward the earlier years so that first-year operating income after taxes will be lower. Because depreciation is a non-cash expense, however, it must be added back to operating income after taxes in order to obtain after-tax operating cash flow. This process shifts cash flows from later years to earlier years, increasing the NPV.

29 C is correct. The outlay is lower by INR0.24, which will decrease the annual depreciation by INR0.04 for the first six years. The annual additional taxes from the loss of the depreciation tax shelter are INR0.04(0.40) = INR0.016. The after-tax cash flows are higher by €0.24 at time zero (because of the smaller investment) and lower by €0.016 for the first six years. The NPV increases by

$$\text{NPV} = +0.24 - \sum_{t=1}^{6}\frac{0.016}{1.12^t} = 0.24 - 0.0658 = 0.1742 = \text{INR}0.17 \text{ billion}$$

30 A is correct. Both of the supervisor's comments are incorrect. Because the PRT-Assam Pharma project is a conventional project (an outflow followed by inflows), the multiple IRR problem cannot occur. The EAA is preferred over the NPV when dealing with mutually exclusive projects with differing lives, a scenario that is not relevant for this decision. The PRT-Assam Pharma project is freestanding, so the NPV approach is appropriate.

31 B is correct. The weighted average cost of capital for Embelesado is calculated as follows:

$$\begin{aligned} \text{WACC} &= (\text{Market weight of debt} \times \text{After-tax cost of debt}) + \\ &\quad\ (\text{Market weight of equity} \times \text{Cost of equity}) \\ &= w_d k_d(1 - \text{t}) + w_{cs}k_{cs} = 0.231(8.0\%)(1 - 0.35) + 0.769(13.0\%) \\ &= 1.201\% + 9.997\% \\ &= 11.198\%, \text{ rounded to } 11.20\% \end{aligned}$$

32 C is correct. The terminal year cash flow (in millions) is as follows:

Revenues	€60.00
Less operating costs	25.00
Less depreciation expenses	20.00
= Taxable income	15.00
Less taxes @ 35%	(5.25)

= Net income	9.75
Plus depreciation expenses	20.00
= After-tax operating CF	29.75
+ Recover WC	10.00
+ Ending market value	5.00
Less taxes on sale proceeds @ 35%	(1.75)*
= Terminal Year CF	€43.00

* The tax on the sale proceeds is 35% multiplied by the gain of €5.00 = €1.75.

33 C is correct. This is the IRR for a project with the following cash flows: (€70,000) in Year 0, €29,750 at Years 1 and 2, and €43,000 at Year 3.

	Years 1 & 2	Year 3
Revenues	€60,000	€60,000
Less operating costs	25,000	25,000
Less depreciation expense	20,000	20,000
= Taxable income	15,000	15,000
Less taxes @ 35%	5,250	5,250
= Net income	9,750	9,750
Plus depreciation expense	20,000	20,000
= After-tax operating CF	€29,750	29,750
+ Recover WC		10,000
+ Salvage value		5,000
– Less taxes on salvage value @ 35%		1,750
= Terminal year CF		€43,000

The IRR of 20.29% is readily found with a financial calculator:

$$70{,}000 = \frac{29{,}750}{(1 + IRR)^1} + \frac{29{,}750}{(1 + IRR)^2} + \frac{43{,}000}{(1 + IRR)^3}$$

You can also "reverse-engineer" the answer using the choices given in the question.

34 A is correct. Projects with shorter paybacks do not necessarily have a positive NPV. For mutually exclusive projects, the NPV and IRR criteria will not necessarily provide the same project ranking.

35 B is correct. Additional depreciation in earlier periods will shield Embelesado from additional taxes, thus increasing the net cash flows in earlier years of the project and increasing the project's NPV. This also means that there will be less depreciation expense in the terminal year of the project, however, thus shielding less income and increasing taxes. Terminal-year net cash flow will likely decrease.

36 A is correct. The entire €10 million will be subject to taxes, resulting in an additional €6.5 million after taxes. As indicated in the following calculation, when discounted at 13% for three years, this has a present value of €4.5048 (rounded to €4.50 million):

$$PV = \frac{10.0(1 - 0.35)}{(1.13)^3} = \frac{6.50}{(1.13)^3} = 4.50$$

37 B is correct. Using the equation CF = (S − C) × (1 − t) + tD, the numbers are as follows:

$$\begin{aligned}
\text{Sales} &= \text{P} \times \text{Q} = €600 \times 3{,}000 = €1{,}800{,}000 \\
\text{Costs} &= \text{Variable cost} \times \text{Q} + \text{Fixed costs} = (125 \times 3{,}000) + \\
&\quad €125{,}000 \\
&= 500{,}000 \\
\text{Depreciation expense} &= €2{,}750{,}000 \div 5 = €550{,}000 \\
\text{CF} &= (1{,}800{,}000 - 500{,}000) \times (1 - 0.40) + (550{,}000 \times \\
&\quad 0.40) \\
&= 780{,}000 + 220{,}000 \\
&= €1{,}000{,}000
\end{aligned}$$

38 C is correct. The terminal year non-operating cash flow includes the after-tax salvage value and the recovery of net working capital = €450,000 × (1 − 0.40) + €75,000 = €345,000.

(Note: Terminal year recovery of net working capital investment = Decrease in current assets − Decrease in current liabilities = €200,000 − €125,000 = €75,000.)

39 B is correct. Calculations: The outlay is higher by €250,000, which will increase annual depreciation by €50,000 over the five-year period. The annual additional tax savings from the higher depreciation expense is 50,000 × (0.40) = 20,000. Therefore NPV should decrease by

$$\text{NPV} = -250{,}000 + \sum_{t=1}^{5} \frac{20{,}000}{1.12^t} = -250{,}000 + 72{,}095.524 = -177{,}904$$

40 A is correct. Higher-than-expected inflation increases the corporation's real taxes because it reduces the value of the depreciation tax shelter; it also decreases the real interest expense because payments to bondholders in real terms are lower than expected.

41 B is correct. When a project is more or less risky than the company, project beta and not WACC should be used to establish the required rate of return for the capital project. In this case, the required rate of return is greater than the WACC, which means the project beta (risk) is greater than the company's beta.

42 B is correct. Timing options (e.g., delay investing) should be included in the NPV analysis, but sunk costs should not.

Equity Valuation

STUDY SESSIONS

Study Session 8	Equity Valuation (1)
Study Session 9	Equity Valuation (2)

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to analyze and evaluate equity securities using appropriate valuation concepts and techniques. The candidate should also be able to estimate risk and expected return of equities in global contexts.

Companies across the world differ widely in their operating and reporting models and risk–return considerations. A privately held, early stage financial technology startup with few physical assets or cash flows will look and operate differently than a mature auto manufacturer with complex operations across the globe. Fortunately, equity valuation methods exist that, based on the fundamental inputs available, can be applied to value the business, investment, or transaction in question. In each case, determining the most appropriate method to apply requires a sound understanding of the company and its industry.

© 2021 CFA Institute. All rights reserved.

EQUITY VALUATION

STUDY SESSION 8

Equity Valuation (1)

This study session introduces essential equity valuation concepts. The various definitions of value and the application of equity valuation techniques to solve everyday problems are first discussed. A five-step equity valuation process is then described with the three main categories of equity valuation models (absolute, relative, total entity) presented in step three. Key return measures including the equity risk premium and derivation of the equity required return using various models (CAPM, multifactor, build up) conclude the session.

READING ASSIGNMENTS

Reading 20	Equity Valuation: Applications and Processes by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA
Reading 21	Return Concepts by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA

© 2021 CFA Institute. All rights reserved.

READING

20

Equity Valuation: Applications and Processes

by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA

Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA). Thomas R. Robinson, PhD, CFA, CAIA, is at Robinson Global Investment Management LLC, (USA). John D. Stowe, PhD, CFA, is at Ohio University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** define valuation and intrinsic value and explain sources of perceived mispricing;
☐	**b.** explain the going concern assumption and contrast a going concern value to a liquidation value;
☐	**c.** describe definitions of value and justify which definition of value is most relevant to public company valuation;
☐	**d.** describe applications of equity valuation;
☐	**e.** describe questions that should be addressed in conducting an industry and competitive analysis;
☐	**f.** contrast absolute and relative valuation models and describe examples of each type of model;
☐	**g.** describe sum-of-the-parts valuation and conglomerate discounts;
☐	**h.** explain broad criteria for choosing an appropriate approach for valuing a given company.

INTRODUCTION AND VALUE DEFINITIONS 1

a define valuation and intrinsic value and explain sources of perceived mispricing;

© 2020 CFA Institute. All rights reserved.

b explain the going concern assumption and contrast a going concern value to a liquidation value;

c describe definitions of value and justify which definition of value is most relevant to public company valuation;

Every day, thousands of participants in the investment profession—investors, portfolio managers, regulators, researchers—face a common and often perplexing question: What is the value of a particular asset? The answers to this question usually influence success or failure in achieving investment objectives. For one group of those participants—equity analysts—the question and its potential answers are particularly critical because determining the value of an ownership stake is at the heart of their professional activities and decisions. **Valuation** is the estimation of an asset's value based on variables perceived to be related to future investment returns, on comparisons with similar assets, or, when relevant, on estimates of immediate liquidation proceeds. Skill in valuation is a very important element of success in investing.

We address some basic questions: What is value? Who uses equity valuations? What is the importance of industry knowledge? How can the analyst effectively communicate his analysis? We answer these and other questions and lay a foundation for the topics that follow.

The following section defines value and describes the various uses of equity valuation. The subsequent sections examine the steps in the valuation process, including the analyst's role and responsibilities, and discuss how valuation results are communicated. They also provide some guidance on the content and format of an effective research report.

1.1 Value Definitions and Valuation Applications

Before summarizing the various applications of equity valuation tools, it is helpful to define what is meant by "value" and to understand that the meaning can vary in different contexts. The context of a valuation, including its objective, generally determines the appropriate definition of value and thus affects the analyst's selection of a valuation approach.

1.1.1 *What Is Value?*

Several perspectives on value serve as the foundation for the variety of valuation models available to the equity analyst. Intrinsic value is the necessary starting point, but other concepts of value—going-concern value, liquidation value, and fair value—are also important.

1.1.1.1 Intrinsic Value A critical assumption in equity valuation, as applied to publicly traded securities, is that the market *price* of a security can differ from its intrinsic *value*. The **intrinsic value** of any asset is the value of the asset given a hypothetically complete understanding of the asset's investment characteristics. For any particular investor, an estimate of intrinsic value reflects his or her view of the "true" or "real" value of an asset. If one assumed that the market price of an equity security perfectly reflected its intrinsic value, "valuation" would simply require looking at the market price. Roughly, it is just such an assumption that underpins traditional efficient market theory, which suggests that an asset's market price is the best available estimate of its intrinsic value.

An important theoretical counter to the notion that market price and intrinsic value are identical can be found in the Grossman–Stiglitz paradox. If market prices, which are essentially freely obtainable, perfectly reflect a security's intrinsic value, then a rational investor would not incur the costs of obtaining and analyzing information to obtain a second estimate of the security's value. If no investor obtains and analyzes information about a security, however, then how can the market price reflect the security's intrinsic value? The **rational efficient markets formulation** (Grossman and Stiglitz 1980) recognizes that investors will not rationally incur the expenses of gathering information unless they expect to be rewarded by higher gross returns compared with the free alternative of accepting the market price. Furthermore, modern theorists recognize that when intrinsic value is difficult to determine, as is the case for common stock, and when trading costs exist, even further room exists for price to diverge from value (Lee, Myers, and Swaminathan 1999).

Thus, analysts often view market prices both with respect and with skepticism. They seek to identify mispricing, and at the same time, they often rely on price eventually converging to intrinsic value. They also recognize distinctions among the levels of **market efficiency** in different markets or tiers of markets (for example, stocks heavily followed by analysts and stocks neglected by analysts). Overall, equity valuation, when applied to market-traded securities, admits the possibility of mispricing. Throughout the discussion, then, we distinguish between the market price, P, and the intrinsic value ("value" for short), V.

For an active investment manager, valuation is an inherent part of the attempt to produce investment returns that exceed the returns commensurate with the investment's risk—that is, positive excess risk-adjusted returns. An excess risk-adjusted return is also called an **abnormal return** or **alpha**. (Return concepts will be more fully discussed later.) The active investment manager hopes to capture a positive alpha as a result of his or her efforts to estimate intrinsic value. Any departure of market price from the manager's estimate of intrinsic value is a perceived **mispricing** (i.e., a difference between the estimated intrinsic value and the market price of an asset).

These ideas can be illuminated through the following expression that identifies two possible sources of perceived mispricing:

$$V_E - P = (V - P) + (V_E - V),$$

where

V_E = estimated value
P = market price
V = intrinsic value

[Note: One can derive the above expression as $V_E - P = V_E - P + V - V = (V - P) + (V_E - V)$.]

This expression states that the difference between a valuation estimate and the prevailing market price is, by definition, equal to the sum of two components. The first component is the true mispricing—that is, the difference between the true but unobservable intrinsic value V and the observed market price P (this difference contributes to the abnormal return). The second component is the difference between the valuation estimate and the true but unobservable intrinsic value—that is, the error in the estimate of the intrinsic value.

To obtain a useful estimate of intrinsic value, an analyst must combine accurate forecasts with an appropriate valuation model. The quality of the analyst's forecasts, in particular the expectational inputs used in valuation models, is a key element in determining investment success. For active security selection to be consistently successful, the manager's expectations must differ from consensus expectations and be, on average, correct as well.

Uncertainty is constantly present in equity valuation. Confidence in one's expectations is always realistically partial. In applying any valuation approach, analysts can never be sure that they have accounted for all the sources of risk reflected in an asset's price. Because competing equity risk models will always exist, there is no obvious final resolution to this dilemma. Even if an analyst makes adequate risk adjustments, develops accurate forecasts, and employs appropriate valuation models, success is not assured. Temporal market conditions may prevent the investor from capturing the benefits of any perceived mispricing. Convergence of the market price to perceived intrinsic value may not happen within the investor's investment horizon, if at all. So, besides evidence of mispricing, some active investors look for the presence of a particular market or corporate event (**catalyst**) that will cause the marketplace to re-evaluate a company's prospects.

1.1.1.2 Going-Concern Value and Liquidation Value A company generally has one value if it is to be immediately dissolved and another value if it will continue in operation. In estimating value, a **going-concern assumption** is the assumption that the company will continue its business activities into the foreseeable future. In other words, the company will continue to produce and sell its goods and services, use its assets in a value-maximizing way for a relevant economic time frame, and access its optimal sources of financing. The **going-concern value** of a company is its value under a going-concern assumption. Models of going-concern value are our focus.

Nevertheless, a going-concern assumption may not be appropriate for a company in financial distress. An alternative to a company's going-concern value is its value if it were dissolved and its assets sold individually, known as its **liquidation value**. For many companies, the value added by assets working together and by human capital applied to managing those assets makes estimated going-concern value greater than liquidation value (although, a persistently unprofitable business may be worth more "dead" than "alive"). Beyond the value added by assets working together or by applying managerial skill to those assets, the value of a company's assets would likely differ depending on the time frame available for liquidating them. For example, the value of nonperishable inventory that had to be immediately liquidated would typically be lower than the value of inventory that could be sold during a longer period of time (i.e., in an "orderly" fashion). Thus, such concepts as **orderly liquidation value** are sometimes distinguished.

1.1.1.3 Fair Market Value and Investment Value For an analyst valuing public equities, intrinsic value is typically the relevant concept of value. In other contexts, however, other definitions of value are relevant. For example, a buy–sell agreement among the owners of a private business—specifying how and when the owners (e.g., shareholders or partners) can sell their ownership interest and at what price—might be primarily concerned with equitable treatment of both sellers and buyers. In that context, the relevant definition of value would likely be fair market value. **Fair market value** is the price at which an asset (or liability) would change hands between a willing buyer and a willing seller when the former is not under any compulsion to buy and the latter is not under any compulsion to sell. Furthermore, the concept of fair market value generally includes an assumption that both buyer and seller are informed of all material aspects of the underlying investment. Fair market value has often been used in valuation related to assessing taxes. In a financial reporting context—for example, in valuing an asset for the purpose of impairment testing—financial reporting standards reference **fair value**, a related (but not identical) concept and provide a specific definition: "Fair value is the amount for which an asset could be exchanged, a liability settled, or an equity instrument granted could be exchanged between knowledgeable, willing parties in an arm's length transaction."

Assuming the marketplace has confidence that the company's management is acting in the owners' best interests, market prices should tend, in the long run, to reflect fair market value. In some situations, however, an asset is worth more to a particular buyer (e.g., because of potential operating synergies). The concept of value to a specific buyer taking account of potential synergies and based on the investor's requirements and expectations is called **investment value**.

1.1.1.4 Definitions of Value: Summary Analysts valuing an asset need to be aware of the definition or definitions of value relevant to the assignment. For the valuation of public equities, an intrinsic value definition of values is generally relevant. Intrinsic value, estimated under a going-concern assumption, is the focus of these equity valuation sections.

APPLICATIONS OF EQUITY VALUATION

2

d describe applications of equity valuation;

Investment analysts work in a wide variety of organizations and positions. As a result, they apply the tools of equity valuation to address a range of practical problems. In particular, analysts use valuation concepts and models to accomplish the following:

- *Selecting stocks*. Stock selection is the primary use of the tools presented here. Equity analysts continually address the same question for every common stock that is either a current or prospective portfolio holding or for every stock that he or she is responsible for covering: Is this security fairly priced, overpriced, or underpriced relative to its current estimated intrinsic value and relative to the prices of comparable securities?
- *Inferring* (*extracting*) *market expectations*. Market prices reflect the expectations of investors about the future performance of companies. Analysts may ask: What expectations about a company's future performance are consistent with the current market price for that company's stock? What assumptions about the company's fundamentals would justify the current price? (**Fundamentals** are characteristics of a company related to profitability, financial strength, or risk.) These questions may be relevant to the analyst for several reasons:
  - The analyst can evaluate the reasonableness of the expectations implied by the market price by comparing the market's implied expectations to his own expectations.
  - The market's expectations for a fundamental characteristic of one company may be useful as a benchmark or comparison value of the same characteristic for another company.

  To extract or reverse-engineer a market expectation, the analyst selects a valuation model that relates value to expectations about fundamentals and is appropriate given the characteristics of the stock. Next, the analyst estimates values for all fundamentals in the model except the fundamental of interest. The analyst then solves for that value of the fundamental of interest that results in a model value equal to the current market price.
- *Evaluating corporate events*. Investment bankers, corporate analysts, and investment analysts use valuation tools to assess the impact of such corporate events as mergers, acquisitions, divestitures, spin-offs, and going-private transactions. Each of these events affects a company's future cash flows and thus the value of its equity. Furthermore, in mergers and acquisitions, the acquiring company's own common stock is often used as currency for the purchase. Investors then want to know whether the stock is fairly valued.
- *Rendering fairness opinions*. The parties to a merger may be required to seek a fairness opinion on the terms of the merger from a third party, such as an investment bank. Valuation is central to such opinions.
- *Evaluating business strategies and models*. Companies concerned with maximizing shareholder value evaluate the effect of alternative strategies on share value.

- *Communicating with analysts and shareholders.* Valuation concepts facilitate communication and discussion among company management, shareholders, and analysts on a range of corporate issues affecting company value.
- *Appraising private businesses.* Valuation of the equity of private businesses is important for transactional purposes (e.g., acquisitions of such companies or buy–sell agreements for the transfer of equity interests among owners when one of them dies or retires) and tax-reporting purposes (e.g., for the taxation of estates), among others. The absence of a market price imparts distinctive characteristics to such valuations, although the fundamental models are shared with public equity valuation. An analyst encounters these characteristics when evaluating initial public offerings, for example.
- *Share-based payment (compensation).* Share-based payments (e.g., restricted stock grants) are sometimes part of executive compensation. Estimation of their value frequently depends on using equity valuation tools.

Inferring Market Expectations

On 2 January 2019, Apple Inc. (AAPL) lowered its revenue guidance citing a variety of reasons, one of which was the weakening economies in some of its Asian markets. Apple's share price fell approximately 10%. When Biogen Inc. announced on 21 March 2019 that its experimental drug for Alzheimer's had failed in late-stage clinical trials, the company's share price dropped approximately 30%. What contributes to such large single-day price movements—changes in estimates of underlying intrinsic value, or market overreaction to negative news?

A rich stream of academic research probes overall market overreaction and underreaction based on large samples—for example, De Bondt and Thaler (1985), Abarbanell and Bernard (1992), and more recently, Bordalo et al. (2017) and Bouchaud et al. 2018). However, one classic research study addresses the topic with a case study of a single such dramatic price drop. This case study, shown in Exhibit 1, is useful for studying equity valuation.

Exhibit 1

Cornell's (2001) case study focuses on the 21 September 2000 press release by Intel Corporation containing information about its expected revenue growth for the third quarter of 2000. The announced growth fell short of the company's prior prediction by 2 to 4 percentage points and short of analysts' projections by 3 to 7 percentage points. In response to the announcement, Intel's stock price fell nearly 30% during the following five days—from $61.50 just prior to the press release to only $43.31 five days later.

To assess whether the information in Intel's announcement was sufficient to explain such a large loss of value, Cornell (2001) estimated the value of a company's equity as the present value of expected future cash flows from operations minus the expenditures needed to maintain the company's growth. (We will discuss such *free cash flow models* in detail at a later stage.)

Using a conservatively low discount rate, Cornell estimated that Intel's price before the announcement, $61.50, was consistent with a forecasted growth rate of 20% a year for the subsequent 10 years and then 6% per year thereafter. Intel's price after the announcement, $43.31, was consistent with a decline of the 10-year growth rate to well under 15% per year. In the final year of the

Exhibit 1 (Continued)

forecast horizon (2009), projected revenues with the lower growth rate would be $50 billion below the projected revenues based on the pre-announcement price. Because the press release did not obviously point to any changes in Intel's fundamental long-run business conditions (Intel attributed the quarterly revenue growth shortfall to a cyclical slowing of demand in Europe), Cornell's detailed analysis left him skeptical that the stock market's reaction could be explained in terms of fundamentals.

Assuming Cornell's methodology was sound, one interpretation is that investors' reaction to the press release was irrational. An alternative interpretation is that Intel's stock was overvalued prior to the press release and that the press release was "a kind of catalyst that caused movement toward a more rational price, even though the release itself did not contain sufficient long-run valuation information to justify that movement" (Cornell 2001, p. 134).

EXAMPLE 1

Referring to Exhibit 1 on Cornell's study of the Intel stock price reaction, explain how an analyst could evaluate the two possible interpretations.

Solution:

To evaluate whether the market reaction to Intel's announcement was an irrational reaction or a rational reduction of a previously overvalued price, one could compare the expected 20% growth implicit in the pre-announcement stock price to some benchmark—for example, the company's actual recent revenue growth, the industry's recent growth, and/or forecasts for the growth of the industry or the economy. Finding the growth rate implied in the company's stock price is an example of using a valuation model and a company's actual stock price to infer market expectations.

Note: Cornell (2001) observed that the 20% revenue growth rate implied by the pre-announcement stock price was much higher than Intel's average growth rate during the previous five years, which occurred when the company was much smaller. He concluded that Intel's stock was overvalued prior to the press release.

These examples illustrate the role of expectations in equity valuation and typical situations in which a given set of facts may be given various interpretations. These examples also illustrate that differences between market price and intrinsic value can occur suddenly, offering opportunities for astute investment managers to generate alpha.

3 THE VALUATION PROCESS, UNDERSTANDING THE BUSINESS AND INDUSTRY AND COMPETITIVE ANALYSIS

e describe questions that should be addressed in conducting an industry and competitive analysis;

In general, the valuation process involves the following five steps:

1 *Understanding the business*. Industry and competitive analysis, together with an analysis of financial statements and other company disclosures, provides a basis for forecasting company performance.
2 *Forecasting company performance*. Forecasts of sales, earnings, dividends, and financial position (pro forma analysis) provide the inputs for most valuation models.
3 *Selecting the appropriate valuation model*. Depending on the characteristics of the company and the context of valuation, some valuation models may be more appropriate than others.
4 *Converting forecasts to a valuation*. Beyond mechanically obtaining the "output" of valuation models, estimating value involves judgment.
5 *Applying the valuation conclusions*. Depending on the purpose, an analyst may use the valuation conclusions to make an investment recommendation about a particular stock, provide an opinion about the price of a transaction, or evaluate the economic merits of a potential strategic investment.

Most of these steps are addressed in detail later. Here, we provide an overview of each.

3.1 Understanding the Business

To forecast a company's financial performance as a basis for determining the value of an investment in the company or its securities, it is helpful to understand the economic and industry contexts in which the company operates, the company's strategy, and the company's previous financial performance. Industry and competitive analysis, together with an analysis of the company's financial reports, provides a basis for forecasting performance.

3.1.1 *Industry and Competitive Analysis*

Because similar economic and technological factors typically affect all companies in an industry, industry knowledge helps analysts understand the basic characteristics of the markets served by a company and the economics of the company. An airline industry analyst will know that labor costs and jet fuel costs are the two largest expenses of airlines and that in many markets airlines have difficulty passing through higher fuel prices by raising ticket prices. Using this knowledge, the analyst may inquire about the degree to which different airlines hedge the commodity price risk inherent in jet fuel costs. With such information in hand, the analyst is better able to evaluate risk and forecast future cash flows. In addition, the analyst would run sensitivity analyses to determine how different levels of fuel prices would affect valuation.

Various frameworks exist for industry and competitive analysis. The primary usefulness of such frameworks is that they can help ensure that an analysis gives appropriate attention to the most important economic drivers of a business. In other words, the objective is *not* to prepare some formal framework representing industry structure or corporate strategy, but rather to use a framework to organize thoughts about an industry and to better understand a company's prospects for success in competition with other companies in that industry. Further, although frameworks can provide a template, obviously the informational content added by an analyst makes the framework relevant to valuation. Ultimately, an industry and competitive analysis should highlight which aspects of a company's business present the greatest challenges and opportunities and should thus be the subject of further investigation and/or more extensive **sensitivity analysis** (an analysis to determine how changes in an assumed input would affect the outcome of an analysis). Frameworks may be useful as analysts focus on questions relevant to understanding a business.

- *How attractive are the industries in which the company operates in terms of offering prospects for sustained profitability?*

Inherent industry profitability is one important factor in determining a company's profitability. Analysts should try to understand **industry structure**—the industry's underlying economic and technical characteristics—and the trends affecting that structure. Basic economic factors—supply and demand—provide a fundamental framework for understanding an industry. Porter's (1985, 1998, 2008) five forces that characterize industry structure—explained in detail at a later stage and summarized in Exhibit 2— can help analysts assess industry profitability and prospects for companies.

Exhibit 2 Summary of Porter's Forces

Force	Features
Rivalry (intra-industry)	Lower rivalry, few competitors and/or good brand identification
Threat of new entrants	High costs to enter (& other barriers)
Threat of substitutes	Few substitutes exist, or cost to switch is high
Bargaining power of suppliers	Many suppliers exist
Bargaining power of buyers	Many customers for an industry' s product exist

Analysts must also stay current on facts and news concerning all the industries in which the company operates, including recent developments (e.g., management, technological, or financial). Particularly important to valuation are any factors likely to affect the industry's longer term profitability and growth prospects, such as demographic trends.

- *What is the company's relative competitive position within its industry, and what is its competitive strategy?*

The level and trend of the company's market share indicate its relative competitive position within an industry. In general, a company's value is higher to the extent that it can create and sustain an advantage relative to its competition. Porter identifies several generic corporate strategies for achieving above-average performance:

i. Cost leadership—being the lowest cost producer while offering products comparable to those of other companies so that products can be priced at or near the industry average

ii. Differentiation—offering unique products or services along some dimensions that are widely valued by buyers so that the company can command premium prices

iii. Focus—seeking a competitive advantage within a target segment or segments of the industry based on either cost leadership (cost focus) or differentiation (differentiation focus)

The term "business model" refers generally to how a company makes money: which customers it targets, what products or services it will sell to those customers, and how it delivers those products or services (including how it finances its activities). The term is broadly used and sometimes encompasses aspects of the generic strategies just described. For example, an airline with a generic cost leadership strategy might have a business model characterized as a low-cost carrier. Low-cost carriers offer a single class of service and use a single type of aircraft to minimize training costs and maintenance charges.

- *How well has the company executed its strategy, and what are its prospects for future execution?*

Competitive success requires both appropriate strategic choices and competent execution. Analyzing the company's financial reports provides a basis for evaluating a company's performance against its strategic objectives and for developing expectations about a company's likely future performance. A historical analysis means more than just reviewing, say, the 10-year historical record in the most recent annual report. It often means looking at the annual reports from 10 years prior, 5 years prior, and the most recent 2 years. Why? Because looking at annual reports from prior years often provides useful insights into how management has historically foreseen challenges and has adapted to changes in business conditions through time. (In general, the investor relations sections of most publicly traded companies' websites provide electronic copies of their annual reports from at least the most recent years.)

In examining financial and operational strategic execution, two caveats merit mention. First, the importance of qualitative—that is, non-numeric factors—must be considered. Such non-numeric factors include the company's ownership structure, its intellectual and physical property, the terms of its intangible assets (e.g., licenses and franchise agreements), and the potential consequences of legal disputes or other contingent liabilities. Second, it is important to avoid simply extrapolating past operating results when forecasting future performance. In general, economic and technological forces can contribute to the phenomenon of "regression toward the mean." Specifically, successful companies tend to draw more competitors into their industries and find that their ability to generate above-average profits comes under pressure. Conversely, poorly performing companies are often restructured in such a manner as to improve their long-term profitability. Thus, in many cases, analysts making long-term horizon growth forecasts for a company's earnings and profits (e.g., forecasts beyond the next 10 years) plausibly assume company convergence toward the forecasted average growth rate for the underlying economy.

4 ANALYSIS OF FINANCIAL REPORTS AND SOURCES OF INFORMATION

The aspects of a financial report that are most relevant for evaluating a company's success in implementing strategic choices vary across companies and industries. For established companies, financial ratio analysis is useful. Individual drivers of profitability for merchandising and manufacturing companies can be evaluated against the company's stated strategic objectives. For example, a manufacturing company aiming to create a sustainable competitive advantage by building strong brand recognition could be expected to have substantial expenditures for advertising but relatively higher prices for its goods. Compared with a company aiming to compete on cost, the branded company would be expected to have higher gross margins but also higher selling expenses as a percentage of sales.

EXAMPLE 2

Competitive Analysis

The following companies are among the largest publicly-traded providers of oilfield services, based on revenues in the most recent fiscal year:

- Schlumberger Ltd. (executive offices in Paris, Houston, London, and the Hague)
  - Revenue: $32.8 billion
  - Net income: $2.2 billion
- Halliburton (executive offices in Houston)
  - Revenue: $24.0 billion
  - Net income: $1.7 billion
- Baker Hughes, a GE Company (executive offices in Houston)
  - Revenue: $22.9 billion
  - Net income: $0.3 billion
- Saipem S.p.A. (executive offices in Milan)
  - Revenue (2017): €9.0 billion
  - Net income (loss) (2017): –€0.3 billion
- National Oilwell Varco Inc. (executive offices in Houston)
  - Revenue: $8.5 billion
  - Net income (loss): –$0.02 billion
- Weatherford International plc (executive offices in Baar, Switzerland)
  - Revenue: $5.7 billion
  - Net income (loss): –$2.8 billion

Note: Financial data are for fiscal 2018, except where noted.
Sources: Companies' 10-K, 20-F, or Investor Relations websites.

These companies provide tools and services—often of a very technical nature—to expedite the drilling activities of oil and gas producers and drilling companies.

1 Discuss the economic factors that may affect demand for the services provided by oilfield services companies, and explain a logical framework for analyzing and forecasting revenue for these companies.

2 Explain how comparing the level and trend in profit margin (net income/sales) and revenue per employee for the companies shown may help in evaluating whether one of these companies is the cost leader in the peer group.

Solution to 1:

Because the products and services of these companies relate to oil and gas exploration and production, the levels of exploration and production activities by oil and gas producers are probably the major factors that determine the demand for their services. In turn, the prices of natural gas and crude oil are important in determining the level of exploration and production activities. Therefore, among other economic factors, an analyst should research those relating to supply and demand for natural gas and crude oil.

- Supply factors in natural gas, such as natural gas inventory levels.

- Demand factors in natural gas, including household and commercial use of natural gas and the amount of new power generation equipment being fired by natural gas.
- Supply factors in crude oil, including capacity constraints and production levels in OPEC and other oil-producing countries, as well as new discoveries of off-shore and land-based oil reserves.
- Demand factors in crude oil, such as household and commercial use of oil and the amount of new power generation equipment using oil products as its primary fuel.
- For both crude oil and natural gas, projected economic growth rates could be examined as a demand factor and depletion rates as a supply-side factor.

Note: Energy analysts should be familiar with sources for researching supply and demand information, such as the International Energy Agency (IEA), the European Petroleum Industry Association (EUROPIA), the Energy Information Administration (EIA), the American Gas Association (AGA), and the American Petroleum Institute (API).

Solution to 2:

Profit margin reflects cost structure. In interpreting profit margin, however, analysts should evaluate any differences in companies' abilities to affect profit margin through power over price. A successfully executed cost leadership strategy will lower costs and raise profit margins. All else equal, we would also expect a cost leader to have relatively high sales per employee, reflecting efficient use of human resources.

With newer companies, or companies involved in creating new products or markets, nonfinancial measures may be critical to obtaining an accurate picture of corporate prospects. For example, a biotechnology company's clinical trial results or an internet company's unique visitors per day may provide information helpful for evaluating future revenue.

4.1 Sources of Information

Important perspectives on industry and competition are sometimes provided by companies themselves in regulator-mandated disclosures, regulatory filings, company press releases, investor relations materials, and contacts with analysts. Analysts can compare the information provided directly by companies to their own independent research.

Regulatory requirements concerning disclosures and filings vary internationally. In some markets, such as Canada and the United States, regulations require management to provide industry and competitive information and access to those filings is freely available (e.g., www.sedar.com for Canadian filings, www.sec.gov for US filings, and individual companies' Investor Relations websites). To take the case of the United States, in annual filings with the Securities and Exchange Commission made on Form 10-K for US companies and Form 20-F for non-US companies, companies provide industry and competitive information in the business description section and in the management discussion and analysis (MD&A). Interim filings (e.g., the quarterly SEC Form 10-Q for US companies and Form 6-K for non-US companies) provide interim financial statements but typically less-detailed coverage of industry and competition. In other jurisdictions, listed companies' financial disclosures can be found on individual companies' Investor Relations websites or centrally at government websites (e.g. Companies House in the UK at https://www.gov.uk/government/organisations/

companies-house), stock exchange websites (e.g. Shenzhen Stock Exchange disclosures at http://www.szse.cn), or central banks' websites (e.g., National Bank of Belgium at https://www.nbb.be/en/central-balance-sheet-office). Required disclosures concerning industry and competitive information differ across jurisdictions.

So far as analyst–management contacts are concerned, analysts must be aware when regulations (e.g., Regulation FD in the United States) prohibit companies from disclosing material nonpublic information to analysts without also disseminating that information to the public. General management insights based on public information, however, can still be useful to analysts, and many analysts consider in-person meetings with a company's management essential to understanding a company.

The CFA Institute Code of Ethics and Standards of Professional Conduct prohibit use of material inside information, and Regulation FD (and similar regulations in other countries) is designed to prohibit companies from selectively offering such information. These ethical and legal requirements assist analysts by clarifying their main role and purpose.

Company-provided sources of information in addition to regulatory filings include press releases and investor relations materials. The press releases of most relevance to analysts are the press releases that companies issue to announce their periodic earnings. Companies typically issue these earnings press releases several weeks after the end of an accounting period and several weeks before they file their interim financial statements. Earnings press releases summarize the company's performance for the period and usually include explanations for the performance and financial statements (often abbreviated versions). Following their earnings press releases, many companies host conference calls in which they further elaborate on their reported performance and typically allocate some time to answer questions posed by analysts. On their corporate websites, many companies post audio downloads and transcripts of conference calls and presentations made in analyst conferences. The audio files and transcripts of conference calls and conference presentations provide access not only to the company's reports but also to analysts' questions and the company's answers to those questions.

Apart from company-provided sources of information, analysts also obtain information from third-party sources, such as industry organizations, regulatory agencies, and commercial providers of market intelligence.

Sources of ESG Information: The Case of the US Auto Industry

The evaluation of environmental, social, and governance (ESG) factors can help analysts identify potential business risks and practices that may produce long-term competitive advantages relative to peers. In the following example, we discuss the sources of ESG-related information that an analyst following US-domiciled automakers might consider.

The automotive industry is among the most resource-intensive manufacturing industries in the world. New vehicles are subject to multiple governmental standards concerning safety, fuel efficiency and emissions control, vehicle recycling, and theft prevention, among others. Manufacturing and assembly facilities must conform to strict standards for air emissions, water discharge, and hazardous waste management.

Because an auto company's manufacturing process and vehicles can significantly affect the environment, the industry is heavily regulated. The global nature of the automotive industry requires careful consideration of different regulatory environments within countries and regions. Regulatory bodies in the United States, such as the Environmental Protection Agency, as well as non-US regulatory bodies, such as the European Commission, the European Environment Agency, and the UK-based Environment Agency, help develop and track environmental standards and legislation.

The potential for serious injuries from manufacturing increases the importance of automobile worker safety. In addition, labor relations are also very important for US automakers because of the sizable representation of employees in labor unions. Avoiding costly lawsuits, lost production from work stoppages, and negative publicity are primary concerns for automakers.

Information relevant to analyzing ESG considerations for US automakers can be found in many sources that are common to most industries. These sources include corporate filings, press releases, investor calls and webcasts, and trade publications. Sustainability reports (often called corporate sustainability reports, or CSRs) are also relevant to analysts when examining ESG considerations. These reports address the economic, environmental, and social effects resulting from an organization's everyday activities and the organization's values and governance (see https://www.globalreporting.org/information/sustainability-reporting/Pages/default.aspx). Although there is no uniform standard for their issuance or disclosure by companies, sustainability reports can provide analysts with a better understanding of a company's sustainable business practices and whether a company's resource management supports an economically sustainable business model.

For more specific ESG-related information, analysts following US automakers may consult labor union boycott lists and disclosures from the Occupational Safety and Health Administration (OSHA) and the US Equal Employment Opportunity Commission (EEOC). As the federal agency responsible for overseeing working conditions for most private sector employers in the United States, OSHA can help analysts identify auto manufacturers that have demonstrated a history of safety violations or an improvement in workplace safety. The EEOC's litigation database helps in the investigation of any notable workplace discrimination issues that have affected individual automakers.

Several not-for-profit organizations can be valuable ESG resources to analysts of US automakers (or other industries, for that matter). The Sustainable Accounting Standards Board (SASB) sets industry-specific ESG standards and can help analysts identify ESG considerations that have a quantitative impact on companies' financial performance. The Carbon Disclosure Project collects and synthesizes self-reported environmental data that can provide for important information regarding automakers' exposure to climate change and water scarcity. Finally, Ceres, an organization committed to driving sustainability research and advocacy, can provide analysts with access to sustainability research reports for the auto industry.

5 CONSIDERATIONS IN USING ACCOUNTING INFORMATION

In evaluating a company's historical performance and developing forecasts of future performance, analysts typically rely heavily on companies' accounting information and financial disclosures. Companies' reported results vary in their persistence (i.e., sustainability). In addition, the information that companies disclose can vary substantially with respect to the *accuracy* of reported accounting results as reflections of economic performance and the *detail* in which results are disclosed.

The term **quality of earnings analysis** broadly includes the scrutiny of *all* financial statements, including the balance sheet, to evaluate both the sustainability of a company's performance and how accurately the reported information reflects economic reality. Equity analysts will generally develop better insights into a company and improve forecast accuracy by developing an ability to assess a company's quality of earnings. With regard to sustainability of performance, an analyst aims to identify aspects of reported performance that are less likely to recur. For example, earnings with significant components of nonrecurring events—such as positive litigation settlements, nonpermanent tax reductions, or gains on sales of nonoperating assets—are considered to be of lower quality than earnings derived mainly from the company's core business operations.

In addition to identifying nonrecurring events, an analyst aims to identify reporting decisions that may result in a level of reported earnings that is unlikely to continue. A good starting point for this type of quality of earnings analysis is a comparison of a company's net income with its operating cash flow. As a simple hypothetical example, consider a company that generates revenues and net income but no operating cash flow because it makes all sales on account and never collects its receivables. One systematic way to make the comparison is to decompose net income into a cash component (combining operating and investing cash flows) and an accrual component (defined as net income minus the cash component). Capital markets research shows that the cash component is more persistent than the accrual component of earnings, with the result that a company with a relatively higher amount of current accruals will have a relatively lower ROA in the future (Sloan 1996). Here, greater persistency means that compared to accruals in the current period, the cash component in the current period is more predictive of future net income. A relatively higher proportion of accruals can be interpreted as lower earnings quality.

A quality of earnings analysis for a particular company requires careful scrutiny of accounting statements, footnotes, and other relevant disclosures. Sources for studying quality of earnings analysis and accounting risk factors include Mulford and Comiskey (2005) and Schilit and Perler (2010). Examples of a few of the many available indicators of possible problems with a company's quality of earnings are provided in Exhibit 3.

Exhibit 3 Selected Quality of Earnings Indicators

Category	Observation	Potential Interpretation
Revenues and gains	Recognizing revenue early, for example: ■ bill-and-hold sales, and ■ recording sales of equipment or software prior to installation and acceptance by customer.	Acceleration in the recognition of revenue boosts reported income, masking a decline in operating performance.
	Classification of nonoperating income or gains as part of operations.	Income or gains may be nonrecurring and may not relate to true operating performance, possibly masking declines in operating performance.

(continued)

Exhibit 3 (Continued)

Category	Observation	Potential Interpretation
Expenses and losses	Recognizing too much or too little reserves in the current year, such as: ■ restructuring reserves, ■ loan-loss or bad-debt reserves, and ■ valuation allowances against deferred tax assets.	May boost current income at the expense of future income, or alternatively, may decrease current year's earnings to boost future years' performance.
	Deferral of expenses by capitalizing expenditures as an asset, for example: ■ customer acquisition costs and ■ product development costs.	May boost current income at the expense of future income. May mask problems with underlying business performance.
	Use of aggressive estimates and assumptions, such as: ■ asset impairments, ■ long depreciable lives, ■ long periods of amortization, ■ high assumed discount rate for pension liabilities, ■ low assumed rate of compensation growth for pension liabilities, and ■ high expected return on assets for pension.	Aggressive estimates may indicate actions taken to boost current reported income. Changes in assumptions may indicate an attempt to mask problems with underlying performance in the current period.
Balance sheet issues (may also affect earnings)	Use of off-balance sheet financing (financing that does not appear on the balance sheet), such as securitizing receivables.	Assets and/or liabilities may not be properly reflected on the balance sheet.
Operating cash flow	Characterization of an increase in a bank overdraft as operating cash flow.	Operating cash flow may be artificially inflated.

The following example illustrates the importance of accounting practices in influencing reported financial results and the need for analysts to exercise judgment when using those results in any valuation model.

EXAMPLE 3

Historical Example

Quality of Earnings Warning Signs: Aggressive Estimates

In the section of his 2007 letter to the shareholders of Berkshire Hathaway titled "Fanciful Figures—How Public Companies Juice Earnings," Warren Buffett referred to the investment return assumption (the anticipated return on a defined benefit pension plan's current and future assets):

> "Decades of option-accounting nonsense have now been put to rest, but other accounting choices remain—important among these [is] the investment-return assumption a company uses in calculating pension expense. It will come as no surprise that many companies continue

> to choose an assumption that allows them to report less-than-solid 'earnings.' For the 363 companies in the S&P that have pension plans, this assumption in 2006 averaged 8%."
>
> (www.berkshirehathaway.com/letters/2007ltr.pdf. See pp.18–19.)

In his explanation, Buffett assumes a 5% return on cash and bonds, which averaged 28% of US pension fund assets. Therefore, this implies that the remaining 72% of pension fund assets—predominately invested in equities—must earn a return of 9.2%, after all fees, to achieve the 8% overall return on the pension fund assets. To illustrate one perspective on an average pension fund achieving that 9.2% return, he estimates that the Dow Jones Industrial Index would need to close at about 2,000,000 on 31 December 2099 (compared to a level under 13,000 at the time of his writing) for this century's returns on that US stock index to match just the 5.3% average annual compound return achieved in the 20th century.

1. How do aggressively optimistic estimates for returns on pension assets affect pension expense?
2. Where can information about a company's assumed returns on its pension assets be found?

Solution to 1:

The amount of "expected return on plan assets" associated with the return assumption is a deduction in calculating pension expense. An aggressively optimistic estimate for the rate of return that pension assets will earn means a larger-than-warranted deduction in calculating pension expense, and subtraction will lead to understating pension expense and overstating net income. In fact, pension expense could become pension income depending on the numbers involved.

Solution to 2:

Information about a company's assumed return on its pension assets can be found in the footnotes to the company's financial statements.

The next examples of poor earnings quality, in which management made choices going beyond making an aggressive estimate, are reminiscent of a humorous vignette from Benjamin Graham (1936) in which the chair of a company outlines plans for return to profitability, as follows: "Contrary to expectations, no changes will be made in the company's manufacturing or selling policies. Instead, the bookkeeping system is to be entirely revamped. By adopting and further improving a number of modern accounting and financial devices, the corporation's earning power will be amazingly transformed."

EXAMPLE 4

Quality of Earnings Warning Signs: Extreme Cases

CASE A.

In 2018, the Securities and Exchange Commission (SEC) charged Tangoe Inc., a formerly publicly-traded telecommunications expense management company, with fraudulent accounting practices that had allowed the company to improperly recognize revenues. Among the violations cited by the SEC were improperly recording revenue from customers who were unlikely to pay and understating the allowance for bad debts (*Sources:* US Securities and Exchange Commission press release 2018-175, issued 4 September 2018, and the related SEC complaint.)

1 Describe the financial statement impact of the accounting violations cited by the SEC.

2 How would a company's Accounts Receivable turnover (or days receivable) serve as an early warning sign of the revenue accounting violations cited by the SEC?

Solution to 1:

On the income statement, improperly recognizing revenue from customers unlikely to pay would inflate reported revenue and—all else equal—reported earnings. On the balance sheet, the improper practices would result in inflated receivables. On the statement of cash flows, if the amount of revenues included in net income exceeds the amount of cash collected from customers—all else equal—net income will exceed operating cash flow. (In actuality, this was not the case with Tangoe, where the company had other adjustments.)

Solution to 2:

Improperly recognizing revenue from customers who are unlikely to pay and understating the allowance for bad debts—all else equal—would result in a lower Accounts Receivable turnover (and higher days receivable).

Note: Analysis of Tangoe's last years of publicly-reported data actually shows the following (all $ in thousands):

- Revenues increased 12% from 2013 to 2014 (from $188,914 to $212,476), while average receivables increased by 32% (from $40,701 to $50,110).
- Accounts receivable turnover decreased from 4.6x (= $188,914/$40,701) to 4.2x (= $212,476/$50,110).
- Days receivable increased from 79 days (= 365/4.6) to 87 days (= 365/4.2)

The SEC also charged the company with other revenue recognition violations, including improperly recording a loan from a business partner as revenue, counting contingency-fee receipts as revenue, and recording customers' prepayments for future services as current revenue. Violations like these would result in understating such liabilities as loans payable and unearned revenue. The company, which paid penalties to settle the SEC's charges, was delisted from the NASDAQ stock exchange in 2017 and then was subsequently purchased by a private investment firm.

CASE B.

Livent, Inc., was a publicly traded theatrical production company that staged a number of smash hits, such as Tony-award winning productions of *Showboat* and *Fosse*. Livent capitalized preproduction costs, including expenses for preopening advertising, publicity and promotion, set construction, props, costumes, and salaries and fees paid to the cast, crew, and musicians during rehearsals. The company then amortized these capitalized costs over the expected life of the theatrical production based on anticipated revenues.

1 State the effect of Livent's accounting for preproduction costs on its reported earnings per share.

2 State the effect of Livent's accounting for preproduction costs on its balance sheet.

3 If an analyst calculated EBITDA/interest expense and debt/EBITDA based on Livent's accounting for preproduction costs without adjustment, how might the analyst be misled in assessing Livent's financial strength? (Recall that EBITDA is defined as earnings before interest, taxes, depreciation, and amortization. Such ratios as EBITDA/interest expense and debt/EBITDA indicate one aspect of a company's financial strength: debt-paying ability.)

Solution to 1:

Livent's accounting for preproduction costs immediately increased reported earnings per share because it deferred expenses.

Solution to 2:

Instead of immediately expensing costs, Livent reported the amounts on its balance sheet as an asset. The warning signal—the deferral of expenses—can indicate aggressive accounting; preproduction costs should have been expensed immediately because of the tremendous uncertainty about revenues from theatrical productions. There was no assurance that there would be revenues against which expenses could be matched.

Solution to 3:

Livent did not deduct preproduction costs from earnings as expenses. If the amortization of capitalized preproduction costs were then added back to earnings, the EBITDA/interest and debt/EBITDA would not reflect in any way the cash outflows associated with such items as paying pre-opening salaries; but cash outflows reduce funds available to meet debt obligations. The analyst who mechanically added back amortization of preproduction costs to calculate EBITDA would be misled into overestimating Livent's financial strength. Based on a closer look at the company's accounting, the analyst would properly not add back amortization of preproduction expenses in computing EBITDA. If preproduction expenses are not added back, a very different picture of Livent's financial health would emerge.

Note: In 1996, Livent's reported debt/EBITDA was 1.7, but the ratio without adding back amortization for preproduction costs was 5.5. In 1997, debt/EBITDA was 3.7 based on a positive EBITDA of $58.3 million; however, EBITDA without the add-back was *negative* $52.6 million. In November 1998, Livent declared bankruptcy and is now defunct. The criminal trial, in Canada, concluded in 2009 with the conviction of Livent's co-founders on charges of fraud and forgery.

In general, growth in an asset account (such as accounts receivable in the Tangoe example and deferred costs in the Livent example) at a much faster rate than the growth rate of sales may indicate aggressive accounting.

Far more serious than aggressive accounting is the deliberate misstatement of financial reports (i.e., fraudulent financial reporting). In general, publicly-traded companies' annual financial statements are audited by certified, professional auditors. The official standards used by auditors can provide useful insights to analysts about a variety of risk factors that may signal possible future negative surprises. For example, both international auditing standards issued by the IAASB and US auditing standards issued by the PCAOB include examples of fraud risk indicators (IAASB 2018, PCAOB

2017). Fraud risk indicators are typically categorized as relating to incentives to commit fraud, opportunity to commit fraud, or attitude toward committing fraud. A working selection of risk factors for misreporting or misappropriation include the following:

- Excessive pressure on company personnel to make revenue or earnings targets, particularly when combined with a dominant, aggressive management team or individual.
- Management and/or directors' compensation tied to profitability or stock price (through ownership or compensation plans). Although such arrangements are usually desirable, they can be a risk factor for aggressive financial reporting.
- Economic, industry, or company-specific pressures on profitability, such as loss of market share or declining margins.
- Management pressure to meet debt covenants or earnings expectations, including "a practice by management of committing to analysts, creditors, and other third parties to achieve aggressive or unrealistic forecasts" (PCAOB, 2017).
- Existence of related-party transactions.
- Complex organizational structure, creating difficulty in determining who controls the company.
- High turnover—of management, directors, or legal counsel.
- Reported (through regulatory filings) disputes with and/or changes in auditors.
- A history of securities law violations, reporting violations, or persistent late filings.

6 FORECASTING COMPANY PERFORMANCE, SELECTING THE APPROPRIATE VALUATION METHOD, ABSOLUTE AND RELATIVE VALUATION MODELS

f contrast absolute and relative valuation models and describe examples of each type of model;

The second step in the valuation process—forecasting company performance—can be viewed from two perspectives: the economic environment in which the company operates and the company's own operating and financial characteristics.

Companies do business within larger contexts of particular industries, national economies, and world trade. Viewing a company within those larger contexts, a top-down forecasting approach moves from international and national macroeconomic forecasts to industry forecasts and then to individual company and asset forecasts. For example, a revenue forecast for a major home appliance manufacturer could start with industry unit sales forecasts that are, in turn, based on GDP forecasts. Forecasted company unit sales would equal forecasted industry unit sales multiplied by the appliance manufacturer's forecasted market share. A revenue projection would be based on forecasted company unit sales and sales prices.

Alternatively, a bottom-up forecasting approach aggregates forecasts at a micro level to larger scale forecasts, under specific assumptions. For example, a clothing retailer may have several stores in operation with two new stores about to open. Using information based on the sales per square meter of the existing stores (perhaps during their initial period of operation), the analyst could forecast sales per square meter of the new stores that, added to forecasts of a similar type for existing stores, would give a sales forecast for the company as a whole. In making such a bottom-up sales forecast, the analyst would be making assumptions about selling prices and merchandise costs. Forecasts for individual retailers could be aggregated into forecasts for the group, continuing in a bottom-up fashion.

In general, analysts integrate insights from industry and competitive analysis with financial statement analysis to formulate specific forecasts of such items as a company's sales, earnings, and cash flow. Analysts generally consider qualitative as well as quantitative factors in financial forecasting and valuation. For example, an analyst might modify his or her forecasts and valuation judgments based on qualitative factors, such as the analyst's opinion about the business acumen and integrity of management and/or the transparency and quality of a company's accounting practices. Such qualitative factors are necessarily subjective.

6.1 Selecting the Appropriate Valuation Model

This section discusses the third step in the valuation process—selecting the appropriate model for the valuation task at hand. Detailed descriptions of the valuation models are presented later. Absolute valuation models and relative valuation models are the two broad types of valuation models that incorporate a going-concern assumption. Here, we describe absolute and relative valuation models in general terms and discuss a number of issues in model selection. In practice, analysts frequently use more than one approach to estimate the value of a company or its common stock (Pinto, Robinson, and Stowe 2019).

6.1.1 *Absolute Valuation Models*

An **absolute valuation model** is a model that specifies an asset's intrinsic value. Such models are used to produce an estimate of value that can be compared with the asset's market price. The most important type of absolute equity valuation models are present value models. In finance theory, present value models are considered the fundamental approach to equity valuation. The logic of such models is that the value of an asset to an investor must be related to the returns that investor expects to receive from holding that asset. Generally speaking, those returns can be referred to as the asset's cash flows, and present value models are also referred to as discounted cash flow models.

A **present value model** or **discounted cash flow model** applied to equity valuation derives the value of common stock as the present or discounted value of its expected future cash flows (such models are known as income models of valuation in private business appraisal). For common stock, one familiar type of cash flow is dividends, which are discretionary distributions to shareholders authorized by a corporation's board of directors. Dividends represent cash flows at the shareholder level in the sense that they are paid directly to shareholders. Present value models based on dividends are called **dividend discount models**. Rather than defining cash flows as dividends, analysts frequently define cash flows at the company level. Common shareholders in principle have an equity ownership claim on the balance of the cash flows generated by a company after payments have been made to claimants senior to common equity, such as bondholders and preferred stockholders (and the government as well, which takes taxes), whether such flows are distributed in the form of dividends.

The two main company-level definitions of cash flow in current use are free cash flow and residual income. Free cash flow is based on cash flow from operations but takes into account the reinvestment in fixed assets and working capital necessary for a going concern. The **free cash flow to equity model** defines cash flow net of payments to providers of debt, whereas the **free cash flow to the firm model** defines cash flows before those payments. We will define free cash flow and each model with more precision later. A residual income model is based on accrual accounting earnings in excess of the opportunity cost of generating those earnings.

Because the present value approach is the familiar technique for valuing bonds (here, the term "bonds" refers to all debt securities and loans), it is helpful to contrast the application of present value models to equity valuation with present value models as applied to bond valuation. The application of present value models to common stock typically involves greater uncertainty than is the case with bonds. That uncertainty centers on two critical inputs for present value models—the cash flows and the discount rate(s). Bond valuation discounts a stream of cash payments specified in a legal contract (the **bond indenture**). In contrast, in equity valuation an analyst must define the specific cash flow stream to be valued—dividends or free cash flow—and then forecast the amounts of those cash flows. Unlike bond valuation, no cash flow stream is contractually owed to common stockholders. Clearly, a company's total cash flows—and therefore the cash flows potentially available to common stockholders—will be affected by business, financial, technological, and other factors and are subject to greater variation than the contractual cash flow of a bond. Furthermore, the forecasts for common stock cash flows extend indefinitely into the future because common stock has no maturity date. In addition to the greater uncertainty involved in forecasting cash flows for equity valuation, significant uncertainty exists in estimating an appropriate rate at which to discount those cash flows. In contrast with bond valuation, in which a discount rate can usually be based on market interest rates and bond ratings, equity valuation typically involves a more subjective and uncertain assessment of the appropriate discount rate. Finally, in addition to the uncertainty associated with cash flows and discount rates, the equity analyst may need to address other issues, such as the value of corporate control or the value of unused assets.

The present value approach applied to stock valuation, therefore, presents a high order of complexity. Present value models are ambitious in what they attempt—an estimate of intrinsic value—and offer many challenges in application. Graham and Dodd (1934) suggested that the analyst consider stating a range of intrinsic values, and that suggestion remains valid. To that end, **sensitivity analysis** is an essential tool in applying discounted cash flow valuation. We discuss sensitivity analysis in more detail next.

Another type of absolute valuation is **asset-based valuation**, which values a company on the basis of the market value of the assets or resources it controls. For appropriate companies asset-based valuation can provide an independent estimate of value, and an analyst typically finds alternative, independent estimates of value to be useful. Exhibit 4 describes instances in which this approach to absolute valuation could be appropriate.

Exhibit 4

Asset-Based Valuation

Analysts often apply asset-based valuation to natural resource companies. For example, a crude oil producer, such as Petrobras, might be valued on the basis of the market value of its current proven reserves in barrels of oil, minus a discount for estimated extraction costs. A forest industry company, such as Weyerhaeuser, might be valued on the basis of the board meters (or board feet) of timber it controls. Today, however, fewer companies than in the past are involved only in natural resources extraction or production. For example, Occidental Petroleum features petroleum in its name but also has substantial chemical manufacturing operations. For such cases, the total company might be valued as the sum of its divisions, with the natural resource division valued on the basis of its proven resources.

6.1.2 *Relative Valuation Models*

Relative valuation models constitute the second broad type of going-concern valuation models. A **relative valuation model** estimates an asset's value relative to that of another asset. The idea underlying relative valuation is that similar assets should sell at similar prices. Relative valuation is typically implemented using price multiples (ratios of stock price to a fundamental, such as cash flow per share) or enterprise multiples (ratios of the total value of common stock and debt net of cash and short-term investments to certain of a company's operating assets to a fundamental, such as operating earnings).

Perhaps the most familiar price multiple is the price-to-earnings ratio (P/E), which is the ratio of a stock's market price to the company's earnings per share. A stock selling at a P/E that is low relative to the P/E of another closely comparable stock (in terms of anticipated earnings growth rates and risk, for example) is *relatively undervalued* (a good buy) relative to the comparison stock. For brevity, an analyst might state simply *undervalued*, but the analyst must realize that if the comparison stock is overvalued (in an absolute sense, in relation to intrinsic value), so might be the stock being called undervalued. Therefore, it is useful to maintain the distinction between *undervalued* and *relatively undervalued*. Investing to exploit perceived mispricing in either case (absolute or relative mispricing) relies on a basis of differential expectations—that is, investor expectations that differ from and are more accurate than those reflected in market prices, as discussed earlier.

The more conservative investing strategies based on relative valuation involve overweighting (underweighting) relatively undervalued (overvalued) assets, with reference to benchmark weights. The more aggressive strategies allow short selling of perceived overvalued assets. Such aggressive approaches are known as relative value investing (or relative spread investing, if using implied discount factors). A classic example is **pairs trading** that utilizes pairs of closely related stocks (e.g., two automotive stocks), buying the relatively undervalued stock and selling short the relatively overvalued stock. Regardless of which direction the overall stock market goes, the investor will be better off to the extent that the relatively undervalued stock ultimately rises more (falls less) than the relatively overvalued stock.

Frequently, relative valuation involves a group of comparison assets, such as an industry group, rather than a single comparison asset. The application of relative valuation to equity is often called the method of comparables (or just comparables) and is the subject of a later reading.

EXAMPLE 5

Relative Valuation Models

While researching Smithson Genomics, Inc., a (fictitious) healthcare information services company, you encounter a difference of opinions. One analyst's report claims that Smithson is at least 15% *overvalued*, based on a comparison of its P/E with the median P/E of peer companies in the healthcare information services industry and taking account of company and peer group fundamentals. A second analyst asserts that Smithson is *undervalued* by 10%, based on a comparison of Smithson's P/E with the median P/E of the Russell 3000 Index, a broad-based US equity index. Both analyses appear to be carefully executed and reported. Can both analysts be right?

Solution:

Yes. The assertions of both analysts concern *relative* valuations, and their benchmarks for comparisons differ. The first analyst compared Smithson to its peers in the healthcare information services industry and considers the company to be *relatively overvalued* compared to that group. The second analyst compared Smithson to the overall market as represented by the Russell 3000 and considers the company to be *relatively undervalued* compared to that group. If the entire healthcare information services industry is undervalued in relation to the Russell 3000, both analysts can be right because they are making relative valuations.

The investment implications of each analyst's valuation generally would depend on additional considerations, including whether the market price of the Russell 3000 fairly represents that index's intrinsic value and whether the market liquidity of an otherwise attractive investment would accommodate the intended position size. The analyst in many cases may want to supplement relative valuation with estimates of intrinsic value.

The method of comparables is characterized by a wide range of possible implementation choices; a later reading discusses various alternative price and enterprise multiples. Practitioners will often examine a number of price and enterprise multiples for the complementary information they can provide. In summary, the method of comparables does not specify intrinsic value without making the further assumption that the comparison asset is fairly valued. The method of comparables has the advantages of being simple, related to market prices, and grounded in a sound economic principle (that similar assets should sell at similar prices). Price and enterprise multiples are widely recognized by investors, so analysts can communicate the results of an absolute valuation in terms of a price or enterprise multiple.

7 VALUATION OF THE TOTAL ENTITY AND ITS COMPONENTS AND ISSUES IN MODEL SELECTION AND INTERPRETATION

g describe sum-of-the-parts valuation and conglomerate discounts;

h explain broad criteria for choosing an appropriate approach for valuing a given company.

A variation to valuing a company as a single entity is to estimate its value as the sum of the estimated values of its various businesses considered as independent, going-concern entities. A valuation that sums the estimated values of each of the company's businesses as if each business were an independent going concern is known as a **sum-of-the-parts valuation**. (The value derived using a sum-of-the-parts valuation is sometimes called the **breakup value** or **private market value**.)

Sum-of-the-parts analysis is most useful when valuing a company with segments in different industries that have different valuation characteristics. Sum-of-the-parts analysis is also frequently used to evaluate the value that might be unlocked in a restructuring through a spin-off, split-off, tracking stock, or equity (IPO) carve-out.

Example 6 shows a case in which a sum-of-the-parts valuation could be used to gain insight into a company's future prospects. In practice, a detailed breakdown of each business segment's contribution to earnings, cash flow, and value would be needed.

EXAMPLE 6

Sum-of-the-Parts Valuation

Donaldson Company, Inc., is one of the largest and most successful filtration manufacturers in the world. Consistent with FASB guidance related to segment reporting, the company has identified two reportable segments: Engine Products and Industrial Products. Segment selection was based on the internal organizational structure, management of operations, and performance evaluation by management and the company's board of directors. 2018 10-K data (in millions of US dollars) for the segments appear in the following table.

Descriptions of the segments from the company's 2018 10-K are as follows:

The Engine Products segment sells to original equipment manufacturers (OEMs) in the construction, mining, agriculture, aerospace, defense, and truck markets and to independent distributors, OEM dealer networks, private label accounts, and large equipment fleets. Products include replacement filters for both air and liquid filtration applications, air filtration systems, liquid filtration systems for fuel, lube and hydraulic applications, and exhaust and emissions systems.

The Industrial Products segment sells to various industrial end-users, OEMs of gas-fired turbines, and OEMs and end-users requiring clean air. Products include dust, fume, and mist collectors; compressed air purification systems; gas and liquid filtration for food; beverage and industrial processes; air filtration systems for gas turbines; and specialized air and gas filtration systems for such applications as membrane-based products as well as specialized air and gas filtration systems for such applications as hard disk drives and semi-conductor manufacturing.

	Engine Products	Industrial Products	Total Company*
Fiscal 2018			
Net sales	$1,849.0	$885.2	$2,734.2
Earnings (loss) before income taxes	261.3	137.1	363.6
Assets	1,110.3	631.9	1,976.6
Capital expenditures	64.6	31.4	97.5
Fiscal 2017			
Net sales	$1,553.3	$818.6	$2,371.9
Earnings (loss) before income taxes	219.7	129.1	322.0
Assets	849.6	638.3	1,979.7
Capital expenditures	29.7	23.4	65.9
Fiscal 2016			
Net sales	$1,391.3	$829.0	$2,220.3
Earnings (loss) before income taxes	163.5	119.0	257.4

(continued)

	Engine Products	Industrial Products	Total Company*
Assets	841.4	646.9	1,787.0
Capital expenditures	37.5	27.3	72.9

* Total company results differ from the sum of the two divisions by allocated corporate and unallocated amounts.

1 Why might an analyst use a sum-of-the-parts approach to value Donaldson?

2 How might an analyst use the provided information in an analysis and valuation?

Solution to 1:

On the one hand, the Engine Products segment is already significantly larger than the Industrial Products segment and is growing at a much faster rate in terms of sales, income, assets, and capital expenditures. On the other hand, profit margins appear to be higher for Industrial Products. In 2018, the EBIT-to-sales ratio was 15.5% for the Industrial Products segment versus 14.1% for the Engine Products segment.

An investor presentation by Donaldson's management in May 2013 indicated that they expected Industrial Products to become 48% of the company's product portfolio by 2021. However, the recent results noted show that the Engine Products segment has become a larger and larger part of Donaldson's total business despite its lower margins. Whether or not the company will ultimately be successful in changing their product mix is fundamental to an analyst forming an opinion on Donaldson's share price.

Solution to 2:

An analyst might use the information from Example 6 to develop separate valuations for each of the segments based on forecasts for each segment's sales and profitability. The value of the company in total would be the sum of the value of each of the segments, adjusted for corporate items—such as taxes, overhead expenses, and assets/liabilities not directly attributable to the separate operating systems.

The concept of a conglomerate discount often arises in connection with situations warranting a sum-of-the parts valuation. **Conglomerate discount** refers to the concept that the market applies a discount to the stock of a company operating in multiple, unrelated businesses compared to the stock of companies with narrower focuses. Alternative explanations for the conglomerate discount include 1) inefficiency of internal capital markets (i.e., companies' allocation of investment capital among divisions does not maximize overall shareholder value); 2) endogenous factors (i.e., poorly performing companies tend to expand by making acquisitions in unrelated businesses); and 3) research measurement errors (i.e., conglomerate discounts do not actually exist, and evidence suggesting that they do is a result of flawed measurement). Examples in which conglomerate discounts appear most observable occur when companies divest parts of the company that have limited synergies with their core businesses.

Note that a break-up value in excess of a company's unadjusted going-concern value may prompt strategic actions, such as a divestiture or spin-off.

7.1 Issues in Model Selection and Interpretation

How does one select a valuation model? The broad criteria for model selection are that the valuation model be:

- consistent with the characteristics of the company being valued;
- appropriate given the availability and quality of data; and
- consistent with the purpose of valuation, including the analyst's perspective.

Note that using more than one model can yield incremental insights.

Selection of a model consistent with the characteristics of the company being valued is facilitated by having a good understanding of the business, which is the first step in the valuation process. Part of understanding a company is understanding the nature of its assets and how it uses those assets to create value. For example, a bank is composed largely of marketable or potentially marketable assets and securities. Thus, for a bank, a relative valuation based on assets (as recognized in accounting) has more relevance than a similar exercise for a service company with few marketable assets.

In selecting a model, data availability and quality can be limiting factors. For example, a dividend discount model is the simplest discounted cash flow model; but if a company has never paid dividends and no other information exists to assess a company's future dividend policy, an analyst may have more confidence applying an apparently more complex present value model. Similar considerations also apply in selecting a specific relative valuation approach. For example, meaningful comparisons using P/Es may be hard to make for a company with highly volatile or persistently negative earnings.

Model selection can also be influenced by the purpose of the valuation or the perspective of the analyst. For example, an investor seeking a controlling equity position in a company may elect to value the company based on forecasted free cash flows rather than forecasted dividends because such flows might potentially be redirected by such an acquirer without affecting the value of the acquisition (this valuation approach will be discussed in detail in another reading). When an analyst reads valuations and research reports prepared by others, the analyst should consider how the writer's perspective (and potential biases) may have affected the choice of a particular valuation approach and/or valuation inputs. Specific guidance on model selection will be offered later when discussing present value models and price multiples.

As a final note to this introduction of model selection, it is important to emphasize that professionals frequently use multiple valuation models or factors in common stock selection. According to the *Merrill Lynch Institutional Factor Survey* (2018), respondent institutional investors report using an average of approximately 17 valuation factors in selecting stocks. (*Note:* In this report, the term "factor" covers market-based metrics, such as price multiples, as well as accounting-based metrics, such as return on equity.) There are a variety of ways in which multiple factors can be used in stock selection. One prominent way, stock screens, will be discussed in a later reading. As another example, analysts can rank each security in a given investment universe by relative attractiveness according to a particular valuation factor. The rankings for individual securities could be combined into a single composite ranking by assigning weights to the individual factors. Analysts may use a quantitative model to assign those weights.

8 CONVERTING FORECASTS TO A VALUATION AND APPLYING THE VALUATION CONCLUSION: THE ANALYST'S ROLE AND RESPONSIBILITIES

Converting forecasts to valuation involves more than inputting the forecast amounts to a model to obtain an estimate of the value of a company or its securities. Two important aspects of converting forecasts to valuation are sensitivity analysis and situational adjustments.

Sensitivity analysis is an analysis to determine how changes in an assumed input would affect the outcome. Some sensitivity analyses are common to most valuations. For example, a sensitivity analysis can be used to assess how a change in assumptions about a company's future growth—for example, decomposed by sales growth forecasts and margin forecasts—and/or a change in discount rates would affect the estimated value. Other sensitivity analyses depend on the context. For example, assume an analyst is aware that a competitor to the target company plans to introduce a competing product. Given uncertainty about the target company's competitive response—whether it will lower prices to retain market share, offer discounts to its distributors, increase advertising, or change a product feature—the analyst could create a baseline forecast and then analyze how different competitive responses would affect the forecasted financials and, in turn, the estimated valuation.

Situational adjustments may be required to incorporate the valuation impact of specific issues. Three such issues that could affect value estimates are control premiums, lack of marketability discounts, and illiquidity discounts. A controlling ownership position in a company (e.g., more than 50% of outstanding shares, although a far smaller percentage often affords an investor the ability to significantly influence a company) carries with it control of the board of directors and the valuable options of redeploying the company's assets or changing the company's capital structure. The value of a stock investment that would give an investor a controlling position will generally reflect a **control premium**; that is, it will be higher than a valuation produced by a generic quantitative valuation expression that did not explicitly model such a premium. A second issue generally not explicitly modeled is that investors require an extra return to compensate for lack of a public market or lack of marketability. The value of non-publicly traded stocks generally reflects a **lack of marketability discount**. Among publicly traded (i.e., marketable) stocks, the prices of shares with less depth to their markets (less liquidity) often reflect an **illiquidity discount**. An illiquidity discount would also apply if an investor wishes to sell an amount of stock that is large relative to that stock's trading volume (assuming it is not large enough to constitute a controlling ownership). The price that could be realized for that block of shares would generally be lower than the market price for a smaller amount of stock, a so-called **blockage factor**.

8.1 Applying the Valuation Conclusion: The Analyst's Role and Responsibilities

As noted earlier, the purposes of valuation and the intended consumer of the valuation vary:

- Analysts associated with investment firms' brokerage operations are perhaps the most visible group of analysts offering valuation judgments. Their research reports are widely distributed to current and prospective retail and institutional brokerage clients. The term brokerage typically means the business of acting as agents for buyers and sellers. Analysts who work at brokerage firms are known as **sell-side analysts** because brokerage firms sell investments and services to such institutions as investment management firms.

- In investment management firms, trusts and bank trust departments, and similar institutions, an analyst may report valuation judgments to a portfolio manager or to an investment committee as input to an investment decision. Such analysts are widely known as **buy-side analysts**. The analyst's valuation expertise is important not only in investment disciplines involving security selection based on detailed company analysis but also in highly quantitative investment disciplines. Quantitative analysts work in developing, testing, and updating security selection methodologies. Ranking stocks by some measure(s) of relative attractiveness (subject to a risk control discipline), as we will discuss in more detail later, forms one key part of quantitative equity investment disciplines.
- Analysts at corporations may perform some valuation tasks similar to those of analysts at money management firms (e.g., when the corporation manages in-house a sponsored pension plan). Both corporate analysts and investment bank analysts may also identify and value companies that could become acquisition targets.
- Analysts at independent vendors of financial information usually offer valuation information and opinions in publicly distributed research reports, although some focus solely on organizing and analyzing corporate information.

In conducting their valuation activities, investment analysts play a critical role in collecting, organizing, analyzing, and communicating corporate information, and in some contexts, recommending appropriate investment actions based on sound analysis. When they do those tasks well, analysts help their clients, the capital markets, and the suppliers of capital:

- Analysts help their clients achieve their investment objectives by enabling those clients to make better buy and sell decisions.
- Analysts contribute to the efficient functioning of capital markets by providing analysis that leads to informed buy and sell decisions and thus to asset prices that better reflect underlying values. When asset prices accurately reflect underlying values, capital flows more easily to its highest-value uses.
- Analysts benefit the suppliers of capital, including shareholders, when they are effective monitors of management's performance. This monitoring can serve to keep managers' actions more closely aligned with shareholders' best interests [see Jensen and Meckling (1976) for classic analysis of the costs of stockholder–manager conflicts].

What Are Analysts Expected to Do?

When analysts at brokerage firms recommend a stock to the public that later performs very poorly, or when they fail to uncover negative corporate activities, they can sometimes come under public scrutiny. Industry leaders may then be asked to respond to such criticism and to comment on expectations about the role and responsibilities of analysts. One such instance occurred in the United States as a consequence of the late 2001 collapse of Enron Corporation, an energy, utility, trading, and telecommunication company. In testimony before the US Senate (excerpted below), the President and CEO of AIMR (predecessor organization of CFA Institute) offered a summary of the working conditions and responsibilities of brokerage analysts. In the following passage, **due diligence** refers to investigation and analysis in support of a recommendation; the failure to exercise due diligence may sometimes result in liability according to various securities laws. "Wall Street analysts" refers to analysts working in the US brokerage industry (sell-side analysts).

> What are Wall Street analysts expected to do? These analysts are assigned companies and industries to follow, are expected to research fully these companies and the industries in which they operate, and to forecast their future prospects. Based on this analysis, and using appropriate valuation models, they must then determine an appropriate fair price for the company's securities. After comparing this fair price to the current market price, the analyst is able to make a recommendation. If the analyst's "fair price" is significantly above the current market price, it would be expected that the stock be rated a "buy" or "market outperform."
>
> How do Wall Street analysts get their information? Through hard work and due diligence. They must study and try to comprehend the information in numerous public disclosure documents, such as the annual report to shareholders and regulatory filings . . . and gather the necessary quantitative and qualitative inputs to their valuation models.
>
> This due diligence isn't simply reading and analyzing annual reports. It also involves talking to company management, other company employees, competitors, and others, to get answers to questions that arise from their review of public documents. Talking to management must go beyond participation in regular conference calls. Not all questions can be voiced in those calls because of time constraints, for example, and because analysts, like journalists, rightly might not wish to "show their cards," and reveal the insights they have gotten through their hard work, by asking a particularly probing question in the presence of their competitors.
>
> Wall Street analysts are also expected to understand the dynamics of the industry and general economic conditions before finalizing a research report and making a recommendation. Therefore, in order for their firm to justify their continued employment, Wall Street analysts must issue research reports on their assigned companies and must make recommendations based on their reports to clients who purchase their firm's research
>
> (*Source:* Thomas A. Bowman, CFA. Testimony to the Committee on Governmental Affairs (excerpted) US Senate, 27 February 2002).

From the beginnings of the movement to organize financial analysis as a profession rather than as a commercial trade, one guiding principle has been that the analyst must hold himself accountable to both standards of competence and standards of conduct. Competence in investment analysis requires a high degree of training, experience, and discipline (as reflected in the examination and work experience requirements that are prerequisites for obtaining the CFA designation). Additionally, the investment professional is in a position of trust, requiring ethical conduct toward the public, clients, prospects, employers, employees, and fellow analysts. For CFA Institute members, this position of trust is reflected in the Code of Ethics and Standards of Professional Conduct, as well as in the Professional Conduct Statement that they submit annually. The Code and Standards guide the analyst to independent, well-researched, and well-documented analysis and are described in the following sections.

COMMUNICATING VALUATION RESULTS

9

Writing is an important part of an analyst's job. Whether for review by an investment committee or a portfolio manager in an investment management firm or for distribution to the retail or institutional clients of a brokerage firm, research reports share several common elements. In this section, we briefly discuss the content and format of an effective research report and the analyst's responsibilities for preparing a report.

9.1 Contents of a Research Report

A primary determinant of a research report's contents is what the intended readers seek to gain from reading the report. From a sell-side analyst's report, an intended reader would be interested in the investment recommendation. In evaluating how much attention and weight to give to a recommendation, the reader will look for persuasive supporting arguments. A key element supporting any recommendation is the intrinsic value of the security.

Given the importance of the estimated intrinsic value of the security, most research reports provide the reader with information about the key assumptions and expectations underlying that estimated intrinsic value. The information typically includes an update on the company's financial and operating results, a description of relevant aspects of the current macroeconomic and industry context, and an analysis and forecast for the industry and company. Because some readers of research reports are interested in background information, some reports contain detailed historical descriptive statistics about the industry and company.

A report can include specific forecasts, key valuation inputs (e.g., the estimated cost of capital), a description of the valuation model, and a discussion of qualitative factors and other considerations that affect valuation. Superior research reports also objectively address the uncertainty associated with investing in the security and/or the valuation inputs involving the greatest amount of uncertainty. By converting forecasts into estimated intrinsic value, a comparison between intrinsic value and market price provides the basis for an investment recommendation. When a research report states a target price for a stock (based on its intrinsic value) in its investment recommendation, the report should clarify the basis for computing the target, a time frame for reaching the target, and information on the uncertainty of reaching the target. An investment recommendation may be accompanied by an explanation of the underlying rationale (i.e., investment thesis), which summarizes why a particular investment offer would provide a way to profit from the analyst's outlook.

Although a well-written report cannot compensate for a poor analysis, a poorly written report can detract from the credibility of an excellent analysis. Writing an effective research report is a challenging task. In summary, an effective research report:

- contains timely information;
- is written in clear, incisive language;
- is objective and well researched, with key assumptions clearly identified;
- distinguishes clearly between facts and opinions;
- contains analysis, forecasts, valuation, and a recommendation that are internally consistent;
- presents sufficient information to allow a reader to critique the valuation;
- states the key risk factors involved in an investment in the company; and
- discloses any potential conflicts of interests faced by the analyst.

Although these general characteristics are all desirable attributes of a useful and respected report, in some situations the requirements are more specific. For example, regulations governing disclosures of conflicts and potential conflicts of interest vary across countries, so an analyst must remain up-to-date on relevant disclosure requirements. In some situations, investment recommendations are affected by policies of the firm employing an analyst; for example, a policy might require that a security's price must be *X*% below its estimated intrinsic value to be considered a "buy." Even in the absence of such a policy, an analyst needs to maintain a conceptual distinction between a "good company" and a "good investment" because returns on a common stock investment always depend on the price paid for the stock, whether the business prospects of the issuing company are good, bad, or indifferent. Exhibit 5 provides a small sample of possible research report content.

Exhibit 5

Research Reports

The following two passages are closely based on the valuation discussions of actual companies in two actual short research notes. The dates and company names used in the passages, however, are fictional.

A At a recent multiple of 6.5, our earnings per share multiple for 2020, the shares were at a discount to our projection of 14% growth for the period … MXI has two operating segments … In valuing the segments separately, employing relative acquisition multiples and peer mean values, we found fair value to be above recent market value. In addition, the shares trade at a discount to book value (0.76). Based on the value indicated by these two valuation metrics, we view the shares as worth holding. However, in light of a weaker economy over the near term, dampening demand for MXI's services, our enthusiasm is tempered. [*Elsewhere in the report, MXI is evaluated as being in the firm's top category of investment attractiveness*.]

B Although TXI outperformed the overall stock market by 20% since the start of the year, it definitely looks undervalued, as shown by its low multiples … [*the values of the P/E and another multiple are stated*]. According to our dividend discount model valuation, we get to a valuation of €3.08, implying an upside potential of 36.8% based on current prices. The market outperform recommendation is reiterated. [*In a parenthetical expression, the current dividend, assumed dividend growth rates, and their time horizons are given. The analyst also briefly explains and calculates the discount rate. Elsewhere in the report the current price of TXI is given as €2.25*.]

Although some of the concepts mentioned in the two passages may not yet be familiar, you can begin to assess the two reporting efforts.

Passage A communicates the analysis awkwardly. The meaning of "the shares were at a discount to our projection of 14% growth for the period" is not completely clear. Presumably, the analyst is projecting the earnings growth rate for 2020 and stating that the P/E is low in relation to that expected growth rate. The analyst next discusses valuing MXI as the sum of its divisions. In describing the method as "employing relative acquisition multiples and peer mean values," the analyst does not convey a clear picture of what was done. It is probable that companies similar to each of MXI's divisions were identified; then, the mean or average value of some unidentified multiple for those comparison companies was calculated and used as the basis for valuing MXI. The writer is vague, however, on the extent of MXI's undervaluation. The analyst states that MXI's price is

Exhibit 5 (Continued)

below its book value (an accounting measure of shareholders' investment) but draws no comparison with the average price-to-book value ratio for stocks similar to MXI, for example. (The price-to-book ratio is discussed in a later reading.) Finally, the verbal summation is feeble and hedged. Although filled with technical verbiage, Passage A does not communicate a coherent valuation of MXI.

In the second sentence of Passage B, by contrast, the analyst gives an explicit valuation of TXI and the information needed to critique it. The reader can also see that €3.08, which is elsewhere stated in the research note as the target price for TXI, implies the stated price appreciation potential for TXI [(€3.08/€2.25) – 1, approximately 37%]. In the first sentence in Passage B, the analyst gives information that might support the conclusion that TXI is undervalued, although the statement lacks strength because the analyst does not explain why the P/E is "low." Nevertheless, the verbal summary is clear. Using less space than the analyst in Passage A, the analyst in Passage B has done a better job of communicating the results of his valuation.

9.2 Format of a Research Report

Equity research reports may be logically presented in several ways. The firm in which the analyst works sometimes specifies a fixed format for consistency and quality control purposes. Without claiming superiority to other ways to organize a report, we offer Exhibit 6 as an adaptable format by which the analyst can communicate research and valuation findings in detail. (Shorter research reports and research notes obviously may employ a more compact format.)

Exhibit 6 A Format for Research Reports

Section	Purpose	Content	Comments
Table of Contents	■ Show report's organization	■ Consistent with narrative in sequence and language	This is typically used only in very long research reports.
Summary and Investment Conclusion	■ Communicate the large picture ■ Communicate major specific conclusions of the analysis ■ Recommend an investment course of action	■ Capsule description of the company ■ Major recent developments ■ Earnings projections ■ Other major conclusions ■ Valuation summary ■ Investment action	An executive summary; may be called simply "Summary."
Business Summary	■ Present the company in more detail ■ Communicate a detailed understanding of the company's economics and current situation ■ Provide and explain specific forecasts[a]	■ Company description to the divisional level ■ Industry analysis ■ Competitive analysis ■ Historical performance ■ Financial forecasts	Reflects the first and second steps of the valuation process. Financial forecasts should be explained adequately and reflect quality of earnings analysis.

(continued)

Exhibit 6 (Continued)

Section	Purpose	Content	Comments
Risks	■ Alert readers to the risk factors in investing in the security	■ Possible negative industry developments ■ Possible negative regulatory and legal developments ■ Possible negative company developments ■ Risks in the forecasts ■ Other risks	Readers should have enough information to determine how the analyst is defining and assessing the risks specific to investing in the security.
Valuation	■ Communicate a clear and careful valuation	■ Description of model(s) used ■ Recapitulation of inputs ■ Statement of conclusions	Readers should have enough information to critique the analysis.
Historical and Pro Forma Tables	■ Organize and present data to support the analysis in the Business Summary		This is generally a separate section only in longer research reports. Many reports fold all or some of this information into the Business Summary section.

[a] Actual outcomes can and generally will differ from forecasts. A discussion of key random factors and an examination of the sensitivity of outcomes to the outcomes of those factors are useful.

9.3 Research Reporting Responsibilities

All analysts have an obligation to provide substantive and meaningful content in a clear and comprehensive report format. Analysts who are CFA Institute members, however, have an additional and overriding responsibility to adhere to the Code of Ethics and the Standards of Professional Conduct in all activities pertaining to their research reports. The CFA Institute Code of Ethics states:

> Members of CFA Institute must . . . use reasonable care and exercise independent professional judgment when conducting investment analysis, making investment recommendations, taking investment actions, and engaging in other professional activities.

Going beyond this general statement of responsibility, some specific Standards of Professional Conduct particularly relevant to an analyst writing a research report are shown in Exhibit 7.

Exhibit 7 Selected CFA Institute Standards of Professional Conduct Pertaining to Research Reports*

Standard of Professional Conduct	Responsibility
I(B)	Members and Candidates must use reasonable care and judgment to achieve and maintain independence and objectivity in their professional activities. Members and Candidates must not offer, solicit, or accept any gift, benefit, compensation, or consideration that reasonably could be expected to compromise their own or another's independence and objectivity.
I(C)	Members and Candidates must not knowingly make any misrepresentations relating to investment analysis, recommendations, actions, or other professional activities.
V(A)1	Members and Candidates must exercise diligence, independence, and thoroughness in analyzing investments, making investment recommendations, and taking investment actions.
V(A)2	Members and Candidates must have a reasonable and adequate basis, supported by appropriate research and investigation, for any investment analysis, recommendation, or action.
V(B)1	Members and Candidates must disclose to clients and prospective clients the basic format and general principles of the investment processes used to analyze investments, select securities, and construct portfolios and must promptly disclose any changes that might materially affect those processes.
V(B)2	Members and Candidates must disclose to clients and prospective clients significant limitations and risks associated with the investment process.
V(B)3	Members and Candidates must use reasonable judgment in identifying which factors are important to their investment analyses, recommendations, or actions and include those factors in communications with clients and prospective clients.
V(B)4	Members and Candidates must distinguish between fact and opinion in the presentation of investment analysis and recommendations.
V(C)	Members and Candidates must develop and maintain appropriate records to support their investment analysis, recommendations, actions, and other investment-related communications with clients and prospective clients.

* See the most recent edition of the CFA Institute *Standards of Practice Handbook* (www.cfainstitute.org).

SUMMARY

In this reading, we have discussed the scope of equity valuation, outlined the valuation process, introduced valuation concepts and models, discussed the analyst's role and responsibilities in conducting valuation, and described the elements of an effective research report in which analysts communicate their valuation analysis.

- Valuation is the estimation of an asset's value based on either variables perceived to be related to future investment returns or comparisons with closely similar assets.
- The intrinsic value of an asset is its value given a hypothetically complete understanding of the asset's investment characteristics.
- The assumption that the market price of a security can diverge from its intrinsic value—as suggested by the rational efficient markets formulation of efficient market theory—underpins active investing.
- Intrinsic value incorporates the going-concern assumption, that is, the assumption that a company will continue operating for the foreseeable future. In contrast, liquidation value is the company's value if it were dissolved and its assets sold individually.
- Fair value is the price at which an asset (or liability) would change hands if neither buyer nor seller were under compulsion to buy/sell and both were informed about material underlying facts.
- In addition to stock selection by active traders, valuation is also used for:
  - inferring (extracting) market expectations;
  - evaluating corporate events;
  - issuing fairness opinions;
  - evaluating business strategies and models; and
  - appraising private businesses.
- The valuation process has five steps:
  1. Understanding the business.
  2. Forecasting company performance.
  3. Selecting the appropriate valuation model.
  4. Converting forecasts to a valuation.
  5. Applying the analytical results in the form of recommendations and conclusions.
- Understanding the business includes evaluating industry prospects, competitive position, and corporate strategies—all of which contribute to making more accurate forecasts. Understanding the business also involves analysis of financial reports, including evaluating the quality of a company's earnings.
- In forecasting company performance, a top-down forecasting approach moves from macroeconomic forecasts to industry forecasts and then to individual company and asset forecasts. A bottom-up forecasting approach aggregates individual company forecasts to industry forecasts, which in turn may be aggregated to macroeconomic forecasts.
- Selecting the appropriate valuation approach means choosing an approach that is:
  - consistent with the characteristics of the company being valued;
  - appropriate given the availability and quality of the data; and

- consistent with the analyst's valuation purpose and perspective.

- Two broad categories of valuation models are absolute valuation models and relative valuation models.
  - Absolute valuation models specify an asset's intrinsic value, supplying a point estimate of value that can be compared with market price. Present value models of common stock (also called discounted cash flow models) are the most important type of absolute valuation model.
  - Relative valuation models specify an asset's value relative to the value of another asset. As applied to equity valuation, relative valuation is also known as the method of comparables, which involves comparison of a stock's price multiple to a benchmark price multiple. The benchmark price multiple can be based on a similar stock or on the average price multiple of some group of stocks.
- Two important aspects of converting forecasts to valuation are sensitivity analysis and situational adjustments.
  - Sensitivity analysis is an analysis to determine how changes in an assumed input would affect the outcome of an analysis.
  - Situational adjustments include control premiums (premiums for a controlling interest in the company), discounts for lack of marketability (discounts reflecting the lack of a public market for the company's shares), and illiquidity discounts (discounts reflecting the lack of a liquid market for the company's shares).
- Applying valuation conclusions depends on the purpose of the valuation.
- In performing valuations, analysts must hold themselves accountable to both standards of competence and standards of conduct.
- An effective research report:
  - contains timely information;
  - is written in clear, incisive language;
  - is objective and well researched, with key assumptions clearly identified;
  - distinguishes clearly between facts and opinions;
  - contains analysis, forecasts, valuation, and a recommendation that are internally consistent;
  - presents sufficient information that the reader can critique the valuation;
  - states the risk factors for an investment in the company; and
  - discloses any potential conflicts of interest faced by the analyst.
- Analysts have an obligation to provide substantive and meaningful content. CFA Institute members have an additional overriding responsibility to adhere to the CFA Institute Code of Ethics and relevant specific Standards of Professional Conduct.

REFERENCES

Abarbanell, Jeffery S., and Victor L. Bernard. 1992. “Tests of Analysts’ Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior.” *Journal of Finance* 47 (3): 1181–207.

BofA Merrill Lynch & Co.“2018 Institutional Factor Survey.” In US Equity Strategy Year Ahead, BofA Merrill Lynch & Co. (20 November).

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2017. “*Diagnostic Expectations and Stock Returns.*” NBER Working Paper 23863.

Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar. 2018. “Sticky Expectations and the Profitability Anomaly.” *Journal of Finance* 74 (2): 639–74.

Cornell, Bradford. 2001. “Is the Response of Analysts to Information Consistent with Fundamental Valuation? The Case of Intel.” *Financial Management* 30 (1): 113–36.

De Bondt, Werner F.M., and Richard Thaler. 1985. “Does the Stock Market Overreact?” *Journal of Finance* 40 (3): 793–805.

Graham, Benjamin. 1936. “*U.S. Steel Announces Sweeping Modernization Scheme.*” Unpublished satire by Ben Graham, written in 1936 and given by the author to Warren Buffett in 1954. Presented in Warren Buffet’s “Chairman’s Letter to the Shareholders of Berkshire Hathaway Inc” (1 March 1991). www.berkshirehathaway.com/letters/1990.html.

Graham, Benjamin, and David L. Dodd. 1934. *Security Analysis*. McGraw-Hill Professional Publishing.

Grossman, Sanford, and Joseph Stiglitz. 1980. “On the Impossibility of Informationally Efficient Markets.” *American Economic Review* 70 (3): 393–408.

International Auditing and Assurance Standards Board (IAASB). 2018. International Standard on Auditing (ISA) 240, The Auditor’s Responsibility to Consider Fraud and Error in an Audit of Financial Statements. In 2018 Handbook of International Quality Control, Auditing, Review, Other Assurance, and Related Services Pronouncements. https://www.iaasb.org/.

Jensen, Michael C., and William H. Meckling. 1976. “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure.” *Journal of Financial Economics* 3 (4): 305–60.

Lee, Charles M.C., James Myers, and Bhaskaran Swaminathan. 1999. “What Is the Intrinsic Value of the Dow?” *Journal of Finance* 54 (5): 1693–741.

Moody’s. 2000. *Putting EBITDA in Perspective*. Moody’s Investors Service Global Credit Research.

Mulford, Charles W., and Eugene F. Comiskey. 2005. *Creative Cash Flow Reporting: Uncovering Sustainable Financial Performance*. Hoboken, NJ: John Wiley & Sons.

Pinto, Jerald E., Thomas R. Robinson, and John D. Stowe. 2019. “Equity Valuation: A Survey of Professional Practice.” *Review of Financial Economics* 37 (2): 219–33.

Porter, Michael E.1985. The Competitive Advantage: Creating and Sustaining Superior Performance. New York: Free Press. (Republished with new introduction in 1998.)

Porter, Michael E. 2008. “The Five Competitive Forces That Shape Strategy.” *Harvard Business Review* 86 (1): 78–93.

Public Company Accounting Oversight Board (PCAOB). 2017. Auditing Standard (AS) 2401: Consideration of Fraud in a Financial Statement Audit. In Auditing Standards of the Public Company Accounting Oversight Board (15 December). www.pcaob.org.

Schilit, Howard, and Jeremy Perler. 2010. *Financial Shenanigans: How to Detect Accounting Gimmicks & Fraud in Financial Reports*, 3rd edition. New York: McGraw-Hill.

Sloan, Richard G. 1996. “Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?” *Accounting Review* 71 (3): 289–315.

PRACTICE PROBLEMS

1 Critique the statement: "No equity investor needs to understand valuation models because real-time market prices for equities are easy to obtain online."

2 The reading defined intrinsic value as "the value of an asset given a hypothetically complete understanding of the asset's investment characteristics." Discuss why "hypothetically" is included in the definition and the practical implication(s).

3 A Explain why liquidation value is generally not relevant to estimating intrinsic value for profitable companies.

B Explain whether making a going-concern assumption would affect the value placed on a company's inventory.

4 Explain how the procedure for using a valuation model to infer market expectations about a company's future growth differs from using the same model to obtain an independent estimate of value.

5 Exhibit 1, based on a study of Intel Corporation that used a present value model (Cornell 2001), examined what future revenue growth rates were consistent with Intel's stock price of $61.50 just prior to its earnings announcement and $43.31 only five days later. "Using a conservatively low discount rate, Cornell estimated that Intel's price before the announcement, $61.50, was consistent with a forecasted growth rate of 20% a year for the subsequent 10 years and then 6% per year thereafter." Discuss the implications of using a higher discount rate than Cornell did.

6 Discuss how understanding a company's business (the first step in equity valuation) might be useful in performing a sensitivity analysis related to a valuation of the company.

7 In a research note on the ordinary shares of the (fictitious) company Milan Fashion Group (MFG) dated early July 2019 when the price was €7.73 and projected annual dividends were €0.05, an analyst stated a target price of €9.20. The research note did not discuss how the target price was obtained or how it should be interpreted. Assume the target price represents the expected price of MFG. What further specific pieces of information would you need to form an opinion on whether MFG was fairly valued, overvalued, or undervalued?

8 You are researching (hypothetical) company XMI Corporation (XMI). XMI has shown steady earnings per share growth (18% a year during the last seven years) and trades at a very high multiple to earnings (its P/E is currently 40% above the average P/E for a group of the most comparable stocks). XMI has generally grown through acquisition, using XMI stock to purchase other companies whose stock traded at lower P/Es. In investigating the financial disclosures of these acquired companies and talking to industry contacts, you conclude that XMI has been forcing the companies it acquires to accelerate the payment of expenses before the acquisition deals are closed. As one example, XMI asks acquired companies to immediately pay all pending accounts payable, whether or not they are due. Subsequent to the acquisition, XMI reinstitutes normal expense payment patterns.

© 2016 CFA Institute. All rights reserved.

A What are the effects of XMI's pre-acquisition expensing policies?

B The statement is made that XMI's "P/E is currently 40% above the average P/E for a group of the most comparable stocks." What type of valuation model is implicit in that statement?

The following information relates to Questions 9–16

Global-Guardian Capital is a rapidly growing international investment firm. The firm's research team is responsible for identifying undervalued and overvalued publicly-traded equities that have a market capitalization greater than $500 million.

Due to the rapid growth of assets under management, Global-Guardian Capital recently hired a new analyst, Jack Richardson, to support the research process. At the new analyst orientation meeting, the director of research made the following statements about equity valuation at the firm:

Statement 1 "Analysts at Global-Guardian Capital seek to identify mispricing, relying on price eventually converging to intrinsic value. However, convergence of the market price to an analyst's estimate of intrinsic value may not happen within the portfolio manager's investment time horizon. So, besides evidence of mispricing, analysts should look for the presence of a particular market or corporate event—that is, a catalyst—that will cause the marketplace to re-evaluate the subject firm's prospects."

Statement 2 "An active investment manager attempts to capture positive alpha. But mispricing of assets is not directly observable. It is therefore important that you understand the possible sources of perceived mispricing."

Statement 3 "For its distressed securities fund, Global-Guardian Capital screens its investable universe of securities for companies in financial distress."

Statement 4 "For its core equity fund, Global-Guardian Capital selects financially sound companies that are expected to generate significant positive free cash flow from core business operations within a multiyear forecast horizon."

Statement 5 "Global-Guardian Capital's research process requires analysts to evaluate the reasonableness of the expectations implied by the market price by comparing the market's implied expectations to his or her own expectations."

After the orientation meeting, the director of research asks Richardson to evaluate three companies that are retailers of men's clothing: Diamond Co., Renaissance Clothing, and Deluxe Men's Wear.

Richardson starts his analysis by evaluating the characteristics of the men's retail clothing industry. He finds few barriers to new retail entrants, high intra-industry rivalry among retailers, low product substitution costs for customers, and a large number of wholesale clothing suppliers.

While conducting his analysis, Richardson discovers that Renaissance Clothing included three non-recurring items in their most recent earnings release: a positive litigation settlement, a one-time tax credit, and the gain on the sale of a non-operating asset.

To estimate each firm's intrinsic value, Richardson applies appropriate discount rates to each firm's estimated free cash flows over a ten-year time horizon and to the estimated value of the firm at the end of the ten-year horizon.

Michelle Lee, a junior technology analyst at Global-Guardian, asks the director of research for advice as to which valuation model to use for VEGA, a fast growing semiconductor company that is rapidly gaining market share.

The director of research states that "the valuation model selected must be consistent with the characteristics of the company being valued."

Lee tells the director of research that VEGA is not expected to be profitable for several more years. According to management guidance, when the company turns profitable, it will invest in new product development; as a result, it does not expect to initiate a dividend for an extended period of time. Lee also notes that she expects that certain larger competitors will become interested in acquiring VEGA because of its excellent growth prospects. The director of research advises Lee to consider that in her valuation.

9 Based on Statement 2, which of the following sources of perceived mispricing do active investment managers attempt to identify? The difference between:

- **A** intrinsic value and market price.
- **B** estimated intrinsic value and market price.
- **C** intrinsic value and estimated intrinsic value.

10 With respect to Statements 3 and 4, which of the following measures of value would the distressed securities fund's analyst consider that a core equity fund analyst might ignore?

- **A** Fair value
- **B** Liquidation value
- **C** Fair market value

11 With respect to Statement 4, which measure of value is *most* relevant for the analyst of the fund described?

- **A** Liquidation value
- **B** Investment value
- **C** Going-concern value

12 According to Statement 5, analysts are expected to use valuation concepts and models to:

- **A** value private businesses.
- **B** render fairness opinions.
- **C** extract market expectations.

13 Based on Richardson's industry analysis, which of the following characteristics of men's retail clothing retailing would *positively* affect its profitability? That industry's:

- **A** entry costs.
- **B** substitution costs.
- **C** number of suppliers.

14 Which of the following statements about the reported earnings of Renaissance Clothing is *most accurate*? Relative to sustainable earnings, reported earnings are likely:

- **A** unbiased.
- **B** upward biased.
- **C** downward biased.

15 Which valuation model is Richardson applying in his analysis of the retailers?

A Relative value

B Absolute value

C Sum-of-the-parts

16 Which valuation model would the director of research *most likely* recommend Lee use to estimate the value of VEGA?

A Free cash flow

B Dividend discount

C P/E relative valuation

The following information relates to Questions 17–20

Bruno Santos is an equity analyst with a regional investment bank. Santos reviews the growth prospects and quality of earnings for Phoenix Enterprises, one of the companies he follows. He has developed a stock valuation model for this firm based on its forecasted fundamentals. His revenue growth rate estimate is less than that implied by the market price.

Phoenix's financial statements over the past five years show strong performance, with above average growth. Santos has decided to use a lower forecasted growth rate in his models, reflecting the effect of "regression to the mean" over time. He notes two reasons for his lower growth rate forecast:

Reason 1 Successful companies tend to draw more competition, putting their high profits under pressure.

Reason 2 Phoenix's intellectual property and franchise agreements will be weakening over time.

Santos meets with Walter Hartmann, a newly hired associate in his department. In their conversation, Hartmann states, "Security analysts forecast company performance using both top-down and bottom-up analysis. I can think of three examples:

1 A restaurant chain forecasts its sales to be its market share times forecast industry sales.

2 An electric utility company forecasts that its sales will grow proportional to increases in GDP.

3 A retail furniture company forecasts next year's sales by assuming that the sales in its newly built stores will have similar sales per square meter to that of its existing stores."

Hartmann is reviewing some possible trades for three stocks in the health care industry based on a pairs-trading strategy. Hartmann's evaluations are as follows:

- HG Health is 15% overvalued.
- Corgent Cell Sciences is 10% overvalued.
- Johnson Labs is 15% undervalued.

17 Based on Santos's revenue growth rate estimate, the shares of Phoenix are *most likely*:

A undervalued.

B fairly valued.

C overvalued.

18 Which of the reasons given by Santos *most likely* justifies a reduction in Phoenix's forecasted growth rate?

A Reason 1 only

B Reason 2 only

C Both Reason 1 and Reason 2

19 Which of Hartmann's examples of company performance forecasting *best* describes an example of bottom-up forecasting?

A Restaurant chain

B Electric utility company

C Retail furniture company

20 Based on his trading strategy, which of the following should Hartmann recommend?

A Short HG Health and Corgent Cell Sciences

B Buy Johnson Labs and Corgent Cell Sciences

C Buy Johnson Labs and short Corgent Cell Sciences

The following information relates to questions 21–24

Abby Dormier is a sell-side analyst for a small Wall Street brokerage firm; she covers publicly and actively traded companies with listed equity shares. Dormier is responsible for issuing either a buy, hold, or sell rating for the shares of Company A and Company B. The appropriate valuation model for each company was chosen based on the following characteristics of each company:

> **Company A** is an employment services firm with no debt and has fixed assets consisting primarily of computers, servers, and commercially available software. Many of the assets are intangible, including human capital. The company has a history of occasionally paying a special cash dividend.
>
> **Company B** operates in three unrelated industries with differing rates of growth: tobacco (60% of earnings), shipbuilding (30% of earnings), and aerospace consulting (10% of earnings). The company pays a regular dividend that is solely derived from the earnings produced by the tobacco division.

Dormier considers the following development in making any necessary adjustments to the models before assigning ratings:

> Company B has finalized the terms to acquire 70% of the outstanding shares of Company X, an actively traded tobacco company, in an all-stock deal.

Dormier assigns ratings to each of the companies and provides a rationale for each rating. The director of research asks Dormier: "How did you arrive at these recommendations? Describe how you used a top-down approach, which is the policy at our company."

Dormier replies, "I arrived at my recommendations through my due diligence process. I have studied all of the public disclosure documents; I have participated in the company conference calls, being careful with my questions in such a public forum; and I have studied the dynamics of the underlying industries. The valuation models are robust and use an extensive set of company-specific quantitative and qualitative inputs."

21 Based on Company A's characteristics, which of the following absolute valuation models is *most* appropriate for valuing that company?

A Asset based

B Dividend discount

C Free cash flow to the firm

22 Based on Company B's characteristics, which of the following valuation models is *most* appropriate for valuing that company?

A Asset based

B Sum of the parts

C Dividend discount

23 Which of the following is *most likely* to be appropriate to consider in Company B's valuation of Company X?

A Blockage factor

B Control premium

C Lack of marketability discount

24 Based on Dormier's response to the director of research, Dormier's process could have been more consistent with the firm's policy by:

A incorporating additional micro-level inputs into her valuation models.

B evaluating the impact of general economic conditions on each company.

C asking more probing questions during publicly available company conference calls.

SOLUTIONS

1 The statement is flawed in at least two ways. First, active investors believe that stock prices do not always accurately reflect all relevant information on the security; for such investors, knowledge of equity valuation models is important for identifying investment opportunities because they represent a way to translate the investor's forecasts into value estimates for comparison with market prices. Thus, the "all" in "all investors" is misleading. Second, not all equities are publicly traded and have market prices, and the most recent market price can be stale for the many public equities that trade only infrequently.

2 No matter how diligent the analyst, some uncertainty always exists concerning 1) the accuracy of the analyst's forecasts and 2) whether an intrinsic value estimate accounts for all sources of risk reflected in market price. Thus, knowledge of a stock's investment characteristics is always incomplete. The practical consequences are that an investor can only estimate intrinsic value and active security selection carries the risk of making mistakes in estimating value.

3 **A** Liquidation value is typically not relevant to estimating intrinsic value for profitable companies because, in general, value would be destroyed by selling such a company's assets individually. Stated another way, the value added by being a going concern is a relevant investment characteristic that an intrinsic value estimate would recognize.

B A going-concern assumption generally increases the value placed on a company's inventory relative to not making that assumption. Usually, inventory that can be sold in the company's regular distribution channels would realize higher amounts than inventory that must be sold immediately because a company is being liquidated.

4 The key difference is that for inferring investor expectations the market price is used as the model input for value, whereas for obtaining an independent estimate of value, value is left as the unknown in the model. In the latter case, value is estimated based on the analyst's estimates for the variables that determine value.

5 Consider the present value of a single cash flow. If one increased the discount rate, one would also need to increase the cash flow if a constant present value were to be maintained. By a similar argument, if Cornell had used a higher discount rate, he would have needed to project a higher level of assumed future cash flows than he did for their present value to have been consistent with the given pre-announcement price of \$61.50. Thus, the implied growth rate consistent with a price of \$61.50 would have been higher than the 20% growth rate estimated by Cornell.

6 An understanding of the company's business facilitates a focus on the key business aspects that affect value, and from a practical perspective, it highlights the critical inputs to a forecast that should be tested using sensitivity analysis.

7 You need to know 1) the time horizon for the price target and 2) the required rate of return on MFG. The price target of €9.20 represents a potential 20% return from investing in the stock if the time horizon is one year, calculated as (€9.20 + €0.05)/€7.73 – 1.0 = 0.197; without a time frame, however, you cannot evaluate the attractiveness of that return. Given that the time frame for the return is established, you need to have an estimate of the required rate of return over the same time horizon.

If the expected return of 19.7% exceeds the security's required return for the same horizon—in other words, if the share's expected alpha is positive—then MFG would appear to be undervalued.

8 **A** Accelerating the payment of expenses reduces the acquired companies' last reported pre-acquisition cash flow. Accelerating expense recognition reduces the acquired companies' last reported pre-acquisition earnings. XMI's cash flow and earnings growth rates following the acquisitions would be expected to be biased upwards because of the depressed levels for the acquirees.

B That is an example of a relative valuation model (or the method of comparables), which compares a company's market multiple to the multiples of similar companies.

9 A is correct. The difference between the true (real) but unobservable intrinsic value and the observed market price contributes to the abnormal return or alpha, which is the concern of active investment managers.

10 B is correct. The measure of value the distressed securities fund's analyst would consider that the core equity fund analyst might ignore is liquidation value. The liquidation value of a company is its value if it were dissolved and its assets sold individually.

11 C is correct. For its core equity fund, Global-Guardian Capital screens its investable universe of securities for well-capitalized companies that are expected to generate significant future free cash flow from core business operations. The concern with future free cash flows implies that going-concern value is relevant.

12 C is correct. Market prices reflect the expectations of investors about the future performance of companies. The analyst can evaluate the reasonableness of the expectations implied by the market price by comparing the market's implied expectations to his own expectations. This process assumes a valuation model, as discussed in the text.

13 C is correct. The men's retail clothing industry is characterized by a large number of wholesale clothing suppliers. When many suppliers of the products needed by industry participants exist, competition among suppliers should limit their ability to raise input prices. Thus, the large number of suppliers is a factor that should positively affect industry profitability.

14 B is correct. The effects of favorable nonrecurring events in reported earnings would tend to bias reported earnings upward relative to sustainable earnings because non-recurring items are by definition not expected to repeat. Renaissance Clothing included three non-recurring items in their most recent earnings release that all led to higher earnings for the current period: a positive litigation settlement, a one-time tax credit, and the gain on the sale of a non-operating asset.

15 B is correct. An absolute valuation model is a model that specifies an asset's intrinsic value. The most important type of absolute equity valuation models are present value models (also referred to as discounted cash flow models), and the model described by Richardson is of that type.

16 A is correct. The broad criteria for model selection are that a valuation model be consistent with the characteristics of the company being valued—that it be appropriate given the availability and quality of the data and consistent with the purpose of the valuation. VEGA currently has negative earnings, making the use of P/E relative valuation difficult if not impossible. As VEGA does not pay a dividend and is not expected to for the foreseeable future, the application of a dividend discount model is problematic. However, the lack of a dividend would not be an obstacle to free cash flow valuation. Furthermore, the director of research has advised that the possibility that competitors may seek to acquire VEGA be taken in to account in valuing VEGA. The reading states that free cash flow valuation can be appropriate in such circumstances. Thus, the director of research would be most likely to recommend free cash flow valuation.

17 C is correct. If the revenue growth rate inferred by the market price exceeds the growth rate that the firm could reasonably expect, Santos should conclude that the market price is too high and thus that the firm is overvalued.

18 C is correct. Increased competition for successful firms can cause a regression to the mean of a company's growth rate. Expiring and weakening intellectual property and franchise agreements can also reduce potential growth.

19 C is correct. The retail furniture company forecasting sales based on sales per square meter is an example of bottom-up forecasting because it aggregates forecasts at a micro level to larger-scale forecasts.

20 C is correct. Pairs trading involves buying an undervalued stock and shorting an overvalued stock in the same industry. Hartmann should buy Johnson Labs (15% undervalued) and short Corgent Cell Sciences (10% overvalued).

21 C is correct. The free cash flow to the firm model is the most appropriate of the choices because it can be used whether the company has significant marketable assets or consistently pays a cash dividend. Much of Company A's assets are intangible, and although the company has a history of paying a dividend, it has been only occasionally and in the form of a special dividend (i.e., not a consistent cash dividend).

22 B is correct. This valuation model would be consistent with the characteristics of the company. Company B is a conglomerate operating in three unrelated industries with significantly different expected revenue growth rates. The sum-of-the-parts valuation model sums the estimated values of each of the company's businesses as if each business were an independent going concern. Sum-of-the-parts analysis is most useful when valuing a company with segments in different industries that have different valuation characteristics.

23 B is correct. A control premium may be reflected in the value of a stock investment that would give an investor a controlling position. Company B acquired 70% of the outstanding stock of Company X; more than 50% is considered a controlling ownership position.

24 B is correct. A top-down forecasting approach moves from macroeconomic forecasts to industry forecasts and then to individual company and asset forecasts. Analysts are expected to understand the general economic conditions before finalizing a research report and making a recommendation. According to Dormier's response, she did not comment on the general economic conditions—although such considerations would be consistent with the firm's policy of using a top-down approach.

READING

21

Return Concepts

by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA

Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA). Thomas R. Robinson, PhD, CFA, CAIA is at Robinson Global Investment Management LLC, (USA). John D. Stowe, PhD, CFA, is at Ohio University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** contrast realized holding period return, expected holding period return, required return, return from convergence of price to intrinsic value, discount rate, and internal rate of return;
☐	**b.** calculate and interpret an equity risk premium using historical and forward-looking estimation approaches;
☐	**c.** determine the required return on an equity investment using the capital asset pricing model, the Fama–French model, the Pastor–Stambaugh model, macroeconomic multifactor models, and the build-up method (e.g., bond yield plus risk premium);
☐	**d.** explain beta estimation for public companies, thinly traded public companies, and non-public companies;
☐	**e.** describe strengths and weaknesses of methods used to estimate the required return on an equity investment;
☐	**f.** explain international considerations in required return estimation;
☐	**g.** explain and calculate the weighted average cost of capital for a company;
☐	**h.** evaluate the appropriateness of using a particular rate of return as a discount rate, given a description of the cash flow to be discounted and other relevant facts.

© 2020 CFA Institute. All rights reserved.

1 RETURN CONCEPTS

a contrast realized holding period return, expected holding period return, required return, return from convergence of price to intrinsic value, discount rate, and internal rate of return;

The return on an investment is a fundamental element in evaluating an investment:

- Investors evaluate an investment in terms of the return they expect to earn on it compared with a level of return viewed as fair given everything they know about the investment, including its risk.
- Analysts need to specify the appropriate rate or rates with which to discount expected future cash flows when using present value models of stock value.

We present and illustrate key return measures relevant to valuation, starting with an overview of return concepts. The subsequent section presents the main approaches to estimating the equity risk premium, a key input in determining the required rate of return on equity in several important models. With a means to estimate the equity risk premium in hand, the next sections discuss and illustrate the major models for estimating the required return on equity. We then present the weighted average cost of capital, a discount rate used when finding the present value of cash flows to all providers of capital, and also explain certain facts concerning discount rate selection. We then conclude the coverage of return concepts with a summary and practice problems.

1.1 Return Concepts

A sound investment decision depends critically on the correct use and evaluation of rate of return measures. The following sections explain the major return concepts most relevant to valuation. Please note that this is by no means an exhaustive list of return concepts used in finance. Also, references to *return* in this context refer to *rate of return*, not a money amount of return.

1.1.1 *Holding Period Return*

The holding period rate of return (for short, the **holding period return**) is the return earned from investing in an asset over a specified period. The specified period is the holding period under examination, whether it is one day, two weeks, four years, or any other length of time. To use a hypothetical return figure of 0.8% for a one-day holding period, we would say that "the one-day holding period return is 0.8%" (or equivalently, "the one-day return is 0.8%" or "the return is 0.8% over one day"). Such returns can be separated into investment income and price appreciation (also referred to as "capital gain") components. If the asset is a share purchased now (at $t = 0$, with t denoting time) and sold at $t = H$, the holding period is $t = 0$ to $t = H$ and the holding period return is

$$\begin{aligned} r &= \frac{D_H + P_H}{P_0} - 1 \\ &= \frac{D_H}{P_0} + \frac{P_H - P_0}{P_0} \qquad (1) \\ &= \text{Dividend yield} + \text{Price appreciation return} \end{aligned}$$

where D_t and P_t are per-share dividends and share price at time t. Equation 1 shows that the holding period return is the sum of two components: dividend yield (D_H/P_0) and price appreciation return ($[P_H - P_0]/P_0$), also known as the capital gains yield.

Equation 1 assumes, for simplicity, that any dividend is received at the end of the holding period. More generally, the holding period return would be calculated based on reinvesting any dividend received between $t = 0$ and $t = H$ in additional shares on the date the dividend was received at the price then available. Holding period returns are sometimes annualized—e.g., the return for a specific holding period may be converted to an annualized return, usually based on compounding at the holding period rate. For example, $(1.008)^{365} - 1 = 17.3271$, or 1,732.71%, is one way to annualize a one-day 0.80% return. As the example shows, however, annualizing holding period returns, when the holding period is a fraction of a year, are unrealistic when the reinvestment rate is not an actual, available reinvestment rate.

1.1.2 *Realized and Expected (Holding Period) Return*

In the expression for the holding period return, the selling price, P_H, and in general, the dividend, D_H, are not known as of $t = 0$. For a holding period in the past, the selling price and the dividend are known, and the return is called a realized holding period return or, more simply, a realized return. For example, with a beginning price of €50.00, an ending or selling price of €52.00 six months later, and a dividend equal to €1.00 (all amounts referring to the past), the realized return is €1.00/€50.00 + (€52.00 − €50.00)/€50.00 = 0.02 + 0.04 = 0.06, or 6% over six months. In forward-looking contexts, holding-period returns are random variables because future selling prices and dividends may both take on a range of values. Nevertheless, an investor can form an expectation concerning the dividend and selling price and thereby have an **expected holding-period return**, or simply expected return, for the stock that consists of the expected dividend yield and the expected price appreciation return.

Although professional investors often formulate expected returns based on explicit valuation models, a return expectation does not have to be based on a model or on specific valuation knowledge. Any investor can have a personal viewpoint on the future returns on an asset. In fact, because investors formulate expectations in varying ways and on the basis of different information, different investors generally have different expected returns for an asset. The comparison point for interpreting the investment implication of the expected return for an asset is its required return, the subject of the next section.

1.1.3 *Required Return*

A **required rate of return** (for short, required return) is the minimum level of expected return that an investor requires in order to invest in the asset over a specified period, given the asset's riskiness. It represents the opportunity cost for investing in the asset—the highest level of expected return available elsewhere from investments of similar risk. As the opportunity cost for investing in the asset, the required return represents a threshold value for being fairly compensated for the asset's risk. If the investor's expected return exceeds the required return, the asset will appear to be undervalued because it is expected to return more-than-fair compensation for the asset's risk. By contrast, if the expected return on the asset falls short of the required rate of return, the asset will appear to be overvalued.

The valuation examples presented in our coverage will illustrate the use of required return estimates grounded in market data (such as observed asset returns) and explicit models for required return. We will refer to any such estimate of the required return used in an example as *the* required return on the asset for the sake of simplicity, although other estimates are usually defensible. For example, under the capital asset pricing model (discussed in more detail later), the required return for an asset is equal to the risk-free rate of return plus a premium (or discount) related to the asset's sensitivity to market returns. That sensitivity can be estimated based on returns for an observed market portfolio and the asset. That is one example of a required return estimate grounded in a formal model based on marketplace variables (rather than a single investor's return requirements). Market variables should contain information about investors' asset risk perceptions and their level of risk aversion, both of which are important in determining fair compensation for risk.

The concept of risk-free rate mentioned in the previous paragraphs is important in valuation. "Risk-free rate" in financial theory is the rate of return on an asset that produces the same, known rate of return in all future economic states. In investment practice, "risk-free rate" typically refers to a rate of return on an investment with assured (or nearly assured) payments. This risk-free rate then serves as a reference rate for practical purposes such as valuing other investments. In a given market, the yield on a sovereign debt instrument (e.g., Treasury bills or Treasury bonds in the United States and German Treasury bills [Schätze] in the eurozone) is typically used by practitioners to represent the risk-free rate for valuation purposes. Note that the concept of a risk-free rate for purposes of equity valuation differs from the concept of risk-free rates used as official benchmarks in pricing rate-linked products and contracts. For example, the recommended risk-free reference rate for rate-linked products and contracts in the eurozone is the euro short-term rate €STR, as of 2018, as a replacement for the previously used euro overnight index average EONIA.

We use the notation r for the required rate of return on the asset being discussed. The required rate of return on common stock and debt are also known as the **cost of equity** and **cost of debt**, respectively, taking the perspective of the issuer. To raise new capital, the issuer would have to price the security to offer a level of expected return that is competitive with the expected returns being offered by similarly risky securities. The required return on a security is therefore the issuer's marginal cost for raising additional capital of the same type.

The difference between the expected return and the required rate of return on an asset is the asset's expected alpha (or *ex ante* alpha) or expected abnormal return:

$$\text{Expected alpha} = \text{Expected return} - \text{Required return} \tag{2a}$$

When an asset is efficiently priced (its price equals its intrinsic value), expected return should equal required return and the expected alpha is zero. In investment decision-making and valuation, the focus is on expected alpha. To evaluate the actual results of an investment discipline, however, the analyst would examine realized alpha. Realized alpha (or *ex post* alpha) over a given holding period is

$$\begin{aligned}\text{Realized alpha} = {} & \text{Actual holding-period return} \\ & - \text{Contemporaneous required return}\end{aligned} \tag{2b}$$

Estimates of required returns are essential for using present value models of value. Present value models require the analyst to establish appropriate discount rates for determining the present values of expected future cash flows.

Expected return and *required rate of return* are sometimes used interchangeably in conversation and writing. As discussed, doing so is not necessarily correct. When current price equals perceived value, expected return should be the same as the required rate of return. When price is below (above) the perceived value, however, expected return will exceed (be less than) the required return as long as the investor expects price to converge to value over her time horizon.

Given an investor's expected holding-period return, we defined expected alpha in relation to a required return estimate. In the next section, we show the conversion of a value estimate into an estimate of expected holding-period return.

1.1.4 *Expected Return Estimates from Intrinsic Value Estimates*

When an asset is mispriced, one of several outcomes is possible. Take the case of an asset that an investor believes is 25% undervalued in the marketplace. Over the investment time horizon, the mispricing may

- increase (the asset may become more undervalued);

- stay the same (the asset may remain 25% undervalued);
- be partially corrected (e.g., the asset may become undervalued by 15%);
- be corrected (the asset price changes to exactly reflect value); or
- reverse or be overcorrected (the asset may become overvalued).

Generally, convergence of price to value is the equilibrium and anticipated outcome when the investor's value estimate is more accurate than the market's, as reflected in the market price. In that case, the investor's expected rate of return has two components: the required return (earned on the asset's current market price) and a return from convergence of price to value.

We can illustrate how expected return may be estimated when an investor's value estimate, V_0, is different from the market price. Suppose the investor expects price to fully converge to value over τ years. $(V_0 - P_0)/P_0$ is an estimate of the return from convergence over the period of that length, essentially the expected alpha for the asset stated on a per-period basis. With r_τ being the required return on a periodic (not annualized) basis and $E(R_\tau)$ the expected holding-period return on the same basis, then:

$$E(R_\tau) \approx r_\tau + \frac{V_0 - P_0}{P_0}$$

Although only an approximation, the expression does illustrate that an expected return can be viewed as the sum of two returns: the required return and a return from convergence of price to intrinsic value. (Note that the expression assumes that the required rate of return and intrinsic value are static over the holding period and that convergence happens either smoothly over the holding period or all at once at its end.)

To illustrate, assume that an investor estimates the required return for an investment in the shares of the (fictitious) Mixama Motors Company is 6.3%. The investor estimates that the intrinsic value of the company's shares is $176.30 per share, and the current market price is $127.97. Thus, in the investor's view, Mixama Motors's shares are undervalued by $V_0 - P_0$ = \$176.30 − \$127.97 = \$48.33, or 37.77% as a fraction of the market price (\$48.33/\$127.97). If price were expected to converge to value in exactly one year, an investor would earn 37.77% + 6.3% = 44.07%. Alternatively, if the investor expected the undervaluation to disappear by the end of nine months, then the investor would earn approximately 42.5% over the nine-month period which is the sum of the 37.77% price appreciation and the required return on a nine-month basis. The required return on a nine-month basis (τ = 9/12 = 0.75) is $(1.063)^{0.75} - 1 = 0.0469$, or 4.69%. The total expected return is

$$\begin{aligned} E(R_\tau) &\approx r_\tau + \frac{V_0 - P_0}{P_0} \\ &= 4.69\% + 37.77\% \\ &= 42.46\% \end{aligned}$$

In other words, the sum of the price appreciation of 37.77% on a nine-month basis, when added to the required return of 4.69% on a nine-month basis, gives an estimate of the nine-month holding period return of approximately 42.5%. Another possibility is that price converges to value in two years. The expected two-year holding-period return would be 13.00% + 37.77% = 50.77%, in which the required return component is calculated as $(1.063)^2 - 1 = 0.1300$. This expected return based on two-year convergence could be compared to the expected return based on one-year convergence of 44.07% by annualizing it: $(1.5077)^{1/2} - 1 = 0.2279$, or 22.79% per year.

Active investors essentially "second-guess" the market price. The risks of that activity include the risks that 1) their value estimates are not more accurate than the market price and 2) even if they are more accurate, the value discrepancy may not narrow over the investors' time horizon. Clearly, the convergence component of expected return can be quite risky.

EXAMPLE 1

Required Return

Thomas Weeramantry is a co-manager of a diversified global equity portfolio. He is researching Microsoft Corporation, one of the largest US-headquartered technology companies. Weeramantry gathered a number of research reports on Microsoft and began his analysis of the company in February 2019, when the current share price for Microsoft was \$112.17. In one research report, the analyst offered the following facts, opinions, and estimates concerning Microsoft:

- The most recent quarterly dividend was \$0.46 per share.
- Microsoft's required return on equity is estimated to be 7.0%.
- A one-year target price for Microsoft shares is estimated at \$126.40.

An analyst's target price is the price at which the analyst believes the security should sell at a stated future point in time. Based only on the information given, answer the following questions concerning Microsoft. For both questions, ignore returns from reinvesting the quarterly dividends.

1 What is the analyst's one-year expected return?

2 What is a target price that is *most* consistent with Microsoft being fairly valued?

Solution to 1:

Over one year, the analyst expects Microsoft to pay \$1.84 in dividends (4 quarters × \$0.46). Using the target price of \$126.40 and dividends of \$1.84, the analyst's expected return is (\$1.84/\$112.17) + (\$126.40 − \$112.17)/\$ 112.17 = 0.016 + 0.127 = 0.143, or 14.3%.

Solution to 2:

If Microsoft is fairly valued, it should return its cost of equity (required return), which is 7.0%. Under that assumption, Target price = Current price × (1 + Required return) − Dividend = \$112.17 × 1.07 − \$1.84 = \$118.18; the dividend is subtracted to isolate the return from price appreciation. Another solution approach involves subtracting the dividend yield from the required return to isolate the anticipated price appreciation return: 7.0% − 1.6% = 5.4%. Thus, 1.054 × \$112.17 = \$118.23.

1.1.5 *Discount Rate*

Discount rate is a general term for any rate used in finding the present value of a future cash flow. A discount rate reflects the compensation required by investors for delaying consumption—generally assumed to equal the risk-free rate—and their required compensation for the risk of the cash flow. Generally, the discount rate used to determine intrinsic value depends on the characteristics of the investment rather than on the characteristics of the purchaser. That is, *for the purposes of estimating intrinsic value*, a required return based on marketplace variables is used rather than a personal required return influenced by such factors as whether the investor is diversified in his personal portfolio. On the other hand, some investors will make judgmental adjustments to such required return estimates, knowing the limitations of the finance models used to estimate such returns.

In principle, because of varying expected future inflation rates and the possibly varying risk of expected future cash flows, a distinct discount rate could be applicable to each distinct expected future cash flow. Similarly, when expected future annual cash flows are grouped into multi-year subperiods—each with a different assumed growth rate—analysts sometimes apply different required returns to discount the different subperiods' expected cash flows. In practice, a single required return is generally used to discount all expected future cash flows.

Sometimes an internal rate of return is used as a required return estimate, as discussed in the next section.

1.1.6 *Internal Rate of Return*

The **internal rate of return** (IRR) on an investment is the discount rate that equates the present value of the asset's expected future cash flows to the asset's price—that is, the amount of money needed today to purchase a right to those cash flows.

In a model that views the intrinsic value of a common equity share as the present value of expected future cash flows, if price is equal to current intrinsic value—the condition of market informational efficiency—then, generally, a discount rate can be found (usually by iteration) that equates that present value to the market price. An IRR computed under the assumption of market efficiency has been used to estimate the required return on equity. An example is the historical practice of many US state regulators of estimating the cost of equity for regulated utilities using the model illustrated in Equation 3b below. The issue of cost of equity arises because regulators set prices sufficient for utilities to earn their cost of capital (see Fernandez 2019 for a description of methodology for required return used by various European utilities regulators).

To illustrate, the simplest version of a present value model results from defining cash flows as dividends and assuming a stable dividend growth rate for the indefinite future. The stable growth rate assumption reduces the sum of results in a very simple expression for intrinsic value:

$$\text{Intrinsic value} = \frac{\text{Year-ahead dividend}}{\text{Required return} - \text{Expected dividend growth rate}} \tag{3a}$$

If the asset is correctly valued now (Market price = Intrinsic value), given consensus estimates of the year-ahead dividend and future dividend growth rate (which are estimates of the dividend expectations built into the price), we can solve for a required return—an IRR implied by the market price:

$$\text{Required return estimate} = \frac{\text{Year-ahead dividend}}{\text{Market price}} + \text{Expected dividend growth rate} \tag{3b}$$

The use of such an IRR as a required return estimate assumes not only market efficiency but also the correctness of the particular present value model (in the foregoing example, the stable growth rate assumption is critical) and the estimated inputs to the selected model. In Equation 3b and similar cases, although the asset's risk is incorporated indirectly into the required return estimate via the market price, the adjustment for risk is not explicit as it is in many competing models that will be presented.

Finally, obtaining an IRR from a present value model should not be confused with the somewhat similar-looking exercise that involves inferring what the market price implies about future growth rates of cash flows, given an independent estimate of required return: That exercise has the purpose of assessing the reasonableness of the market price.

2 EQUITY RISK PREMIUM: HISTORICAL AND FORWARD-LOOKING ESTIMATES

b calculate and interpret an equity risk premium using historical and forward-looking estimation approaches;

The equity risk premium is the incremental return (*premium*) that investors require for holding equities rather than a risk-free asset. Thus, it is the difference between the required return on equities and a specified expected risk-free rate of return. The equity risk premium, like the required return, depends strictly on expectations for the future because the investor's returns depend only on the investment's future cash flows. Possibly confusingly, *equity risk premium* is also commonly used to refer to the realized excess return of stocks over a risk-free asset over a given past period. The realized excess return could be very different from the premium that, based on available information, was contemporaneously being expected by investors.

Using the equity risk premium, the required return on the broad equity market or an average-systematic-risk equity security is

Required return on equity = Current expected risk-free return + Equity risk premium

where, for consistency, the definition of risk-free asset (e.g., government bills or government bonds) used in estimating the equity risk premium should correspond to the one used in specifying the current expected risk-free return.

The importance of the equity risk premium in valuation is that, in perhaps a majority of cases in practice, analysts estimate the required return on a common equity issue as either

$$\begin{aligned}\text{Required return on share } i &= \text{Current expected risk-free return} \\ &\quad + \beta_i(\text{Equity risk premium})\end{aligned} \tag{4}$$

or

$$\begin{aligned}\text{Required return on share } i &= \text{Current expected risk-free return} \\ &\quad + \text{Equity risk premium} \\ &\quad \pm \text{Other risk premia (or discounts)} \\ &\qquad \text{appropriate for } i\end{aligned} \tag{5}$$

- Equation 4 adjusts the equity risk premium for the share's particular level of systematic risk as measured by beta (β_i)—an average systematic risk security has a beta of 1, whereas beta values above and below 1 indicate greater-than-average and smaller-than-average systematic risk. Equation 4 will be explained later as the capital asset pricing model.
- Equation 5 does not make a beta adjustment to the equity risk premium but adds premia/discounts required to develop an overall equity risk adjustment. Equation 5 will be explained later as the build-up method for estimating the required return. It is particularly useful in the valuation of private businesses.

In practice, analysts may use more than one approach to develop the cost of equity. A survey of analysts shows that the CAPM approach is used by 68% of respondents, while a build-up approach (bond yield plus a premium) is used by 43% of respondents (Pinto, Robinson, and Stowe 2019). Typically, analysts estimate the equity risk premium for the national equity market of the issues being analyzed (but if a global CAPM is being used, a world equity premium is estimated that takes into account the totality of equity markets).

Even for the longest-established developed markets, the magnitude of the equity risk premium is difficult to estimate and can be a reason for differing investment conclusions among analysts. Therefore, we will introduce the topic of estimation in some detail. Whatever estimates analysts decide to use, when an equity risk premium estimate enters into a valuation, analysts should be sensitive to how their value conclusions could be affected by estimation error.

Two broad approaches are available for estimating the equity risk premium. One is based on historical average differences between equity market return and government debt returns, and the other is based on current expectational data. Survey data in Pinto, Robinson, and Stowe (2019) indicate that the approaches are almost equally widespread among analysts, with around 36% of respondents using historical estimates and 35% using forward-looking estimates. The following sections describe the two approaches.

2.1 Historical Estimates

A historical equity risk premium estimate is usually calculated as the mean value of the differences between broad-based equity-market-index returns and government debt returns over some selected sample period. When reliable long-term records of equity returns are available, historical estimates have been a familiar and popular choice of estimation. If investors do not make systematic errors in forming expectations, then, over the long term, average returns should be an unbiased estimate of what investors expected. The fact that historical estimates are based on data also gives them an objective quality.

In using a historical estimate to represent the equity risk premium going forward, the analyst is assuming that returns are stationary—that is, the parameters that describe the return-generating process are constant over the past and into the future.

The analyst's major decisions in developing a historical equity risk premium estimate include the selection of:

- the equity index to represent equity market returns;
- the period for computing the estimate;
- the type of mean calculated; and
- the proxy for the risk-free return.

Analysts try to select an equity index that accurately represents the average returns earned by equity investors in the market being examined. Broad-based, market-value-weighted indexes are typically selected.

Specifying the length of the sample period typically involves trade-offs. Dividing a data period of a given length into smaller subperiods does not increase precision in estimating the mean—only extending the length of the dataset can increase precision. (Note that this result contrasts with the estimation of variance and covariance, in which higher frequency of estimation for a given time span *does* increase the precision in estimating variance and covariance.) Thus, a common choice is to use the longest reliable returns series available. The assumption of stationarity is usually more difficult to maintain, however, as the series starting point is extended to the distant past. The specifics of the type of non-stationarity are also important. For a number of equity markets, research has brought forth abundant evidence of non-constant underlying return volatility. Non-stationarity—in which the equity risk premium has fluctuated in the short term but around a central value—is a less serious impediment to using a long data series than the case in which the risk premium has shifted to a permanently different level (Cornell 1999). Empirically, the expected equity risk premium is countercyclical in the United States—that is, the expected premium is high during bad times but low during good times (Fama and French 1989, Ferson and Harvey

1991). This property leads to some interesting challenges: For example, when a series of strong market returns has increased enthusiasm for equities and raised historical-mean equity risk premium estimates, the forward-looking equity risk premium may have actually declined.

Practitioners taking a historical approach to equity premium estimation often focus on the type of mean calculated and the proxy for the risk-free return. There are two choices for computing the mean and two broad choices for the proxy for the risk-free return.

The mean return of a historical set of annual return differences between equities and government debt securities can be calculated using either a geometric mean or an arithmetic mean:

- A geometric mean equity risk premium estimate equal to the compound annual excess return of equities over the risk-free return, or
- An arithmetic mean equity risk premium estimate equal to the sum of the annual return differences divided by the number of observations in the sample.

The risk-free rate can also be represented in two ways:

- as a long-term government bond return, or
- as a short-term government debt instrument (Treasury bill) return.

Dimson, Marsh, and Staunton (2008) presented evidence on realized excess returns of stocks over government debt ("historical equity risk premia") using survivorship-bias-free return datasets for 17 developed markets for the 108 years extending from 1900 through 2007. (Note that in a given year, the excess return of stocks over government debt is calculated as [(1 + Equity market return)/(1 + Risk-free rate of return)] − 1 ≈ Equity market return − Risk-free rate of return, where a specified government debt instrument return represents the risk-free rate of return.) In their 2018 update, the authors provided the risk premiums for 21 markets (Dimson, Marsh, and Staunton 2018, published as *Credit Suisse Global Investment Returns Sourcebook*, 2018). Exhibit 1 excerpts their findings, showing results for the four combinations of mean computation and risk-free return representation (two mean return choices × two risk-free return choices). In the table, *standard error* and *standard deviation* are those of the annual excess return series.

Exhibit 1 Historical Equity Risk Premia: Twenty-One Major Markets, 1900–2017

Panel A: Historical Equity Risk Premia Relative to Bonds, 1900–2017

Country	Geometric Mean	Arithmetic Mean	Standard Error	Standard Deviation
Australia	5.0%	6.6%	1.7%	18.1%
Austria	2.9	21.5	14.1	151.5
Belgium	2.2	4.3	1.9	20.8
Canada	3.5	5.1	1.7	18.2
Denmark	2.2	3.8	1.7	18.0
Finland	5.2	8.7	2.7	29.7
France	3.1	5.4	2.1	22.5
Germany	5.1	8.4	2.6	28.2
Ireland	2.7	4.7	1.8	19.7
Italy	3.2	6.5	2.7	29.1
Japan	5.1	9.1	3.0	32.2

Exhibit 1 (Continued)

Panel A: Historical Equity Risk Premia Relative to Bonds, 1900–2017

Country	Geometric Mean	Arithmetic Mean	Standard Error	Standard Deviation
Netherlands	3.3	5.6	2.0	22.1
New Zealand	4.0	5.6	1.6	17.7
Norway	2.5	5.4	2.5	27.4
Portugal	5.3	9.4	2.9	31.4
South Africa	5.3	7.1	1.8	19.4
Spain	1.8	3.8	1.9	20.5
Sweden	3.1	5.3	2.0	21.2
Switzerland	2.2	3.7	1.6	17.4
United Kingdom	3.7	5.0	1.6	17.0
United States	4.4	6.5	1.9	20.7
Europe	3.0	4.3	1.4	15.7
World ex US	2.8	3.8	1.3	14.4
World	3.2	4.4	1.4	15.3

Panel B: Historical Equity Risk Premia Relative to Bills, 1900–2017

Country	Geometric Mean	Arithmetic Mean	Standard Error	Standard Deviation
Australia	6.1%	7.4%	1.5%	16.3%
Austria	5.8	10.6	3.4	37.0
Belgium	3.0	5.4	2.2	23.5
Canada	4.2	5.6	1.5	16.8
Denmark	3.4	5.3	1.9	20.5
Finland	6.0	9.5	2.7	29.5
France	5.6	8.1	2.2	23.9
Germany	6.2	9.9	2.9	31.1
Ireland	3.7	6.0	2.0	21.2
Italy	5.8	9.6	2.9	31.2
Japan	6.3	9.4	2.5	27.3
Netherlands	4.6	6.7	2.0	22.2
New Zealand	4.6	6.1	1.7	18.0
Norway	3.3	6.1	2.4	25.8
Portugal	4.7	9.3	3.1	33.5
South Africa	6.2	8.2	2.0	21.5
Spain	3.4	5.5	2.0	21.4
Sweden	4.1	6.0	1.9	20.3
Switzerland	3.8	5.4	1.7	18.6
United Kingdom	4.5	6.2	1.8	19.5
United States	5.6	7.5	1.8	19.5
Europe	3.5	5.2	1.8	19.1

(continued)

Exhibit 1 (Continued)

Panel B: Historical Equity Risk Premia Relative to Bills, 1900–2017				
Country	**Geometric Mean**	**Arithmetic Mean**	**Standard Error**	**Standard Deviation**
World ex US	3.6	5.2	1.7	18.4
World	4.3	5.7	1.6	16.9

Note: "World" represents a market-capitalization-weighted (in early decades, GDP-weighted) average of country results in USD terms. Statistics for Austria exclude 1921–22, and statistics for Germany exclude 1922–23.

The following excerpt from Exhibit 1 presents a comparison of historical equity risk premium estimates for the United States and Japan. This comparison highlights some of the issues that can arise in using historical estimates. As background to the discussion, note that as a mathematical fact, the geometric mean is always less than (or equal to) the arithmetic mean; furthermore, the yield curve is typically upward sloping (long-term bond yields are typically higher than short-term yields).

Exhibit 1 Historical Equity Risk Premia: 1900–2017 (excerpted)

	United States		Japan	
	Geometric Mean	**Arithmetic Mean**	**Geometric Mean**	**Arithmetic Mean**
Premium relative to bills	5.6%	7.5%	6.3%	9.4%
Premium relative to bonds	4.4	6.5	5.1	9.1

For the United States, estimates of the equity risk premium relative to long-term government bonds run from 4.4% (geometric mean relative to bonds) to 7.5% (arithmetic mean relative to bills). The United States illustrates the typical case in which realized values relative to bills, for any definition of mean, are higher than those relative to bonds.

The premium estimates for Japan are notably higher than for the United States. Because the promised yield on long-term bonds is usually higher than that on short-term bills, the nearly equal arithmetic mean premium relative to bonds compared with bills in the case of Japan is atypical and suggests a historically flatter yield curve. The analyst would need to consider whether similar conditions would apply to the future before using the estimate as a forecast. In all markets, the geometric mean premium relative to long-term bonds gives the smallest risk premium estimate (an exception is Portugal). Note the following:

- For each market, the variation in year-to-year results is very large as shown by the standard deviations. As a result, the sample mean estimates the true mean with potentially substantial error. For example, the standard error relative to bills is 1.6% for the World index and ranges from 1.5% for Australia and Canada to 3.4% for Austria. So, for the World index, a two standard deviation interval for the underlying mean (an interval within which the underlying mean is expected to lie with a 0.95 probability) is a wide 2.5% to 8.9% (i.e., 5.7% arithmetic mean ± 3.2% standard deviation) even with 118 years of data. This problem of sampling error becomes more acute, the shorter the series on which the mean estimate is based.

- The variation in the historical equity risk premium estimates across countries is substantial. Based on geometric mean data in Panel A of Exhibit 1, the histogram in Exhibit 2 lists markets within 1% return intervals. For more than half of the markets, the realized equity risk premium was between 2% and 4%. As Exhibit 1, Panel A shows, the mean ("World") value is 3.2%. Approximately one-third of the values fall between the two extreme intervals, however.

Exhibit 2 Distribution of Realized Premium (Geometric Mean, Relative to Bonds)

		Interval for Realized Premium x (%)				
		$1 \le x < 2$	$2 \le x < 3$	$3 \le x < 4$	$4 \le x < 5$	$5 \le x < 6$
	6		Belgium	France		Australia
	5		Denmark	Sweden		Germany
Number of markets	4		Switzerland	Italy		Japan
	3		Norway	Netherlands		Finland
	2		Ireland	Canada	New Zealand	Portugal
	1	Spain	Austria	United Kingdom	United States	South Africa

The next two sections discuss choices related to the calculation of a historical equity risk premium estimate.

2.1.1 *Arithmetic Mean or Geometric Mean*

A decision with an important impact on the risk premium estimate is the choice between an arithmetic mean and a geometric mean. The geometric mean is smaller by an amount equal to about one half the variance of returns, so it is always smaller than the arithmetic mean given any variability in returns. (The geometric mean equals the arithmetic mean when the returns for all periods are equal).

In actual professional practice, both geometric and arithmetic means have been used in equity risk premium estimation.

The arithmetic mean return as the average one-period return best represents the mean return in a single period. There are two traditional arguments in favor of using the arithmetic mean in equity risk premium estimation, one relating to the type of model in which the estimates are used and the second relating to a statistical property. The major finance models for estimating required return—in particular, the CAPM and multifactor models—are single-period models; so the arithmetic mean, with its focus on single-period returns, appears to be a model-consistent choice. A statistical argument has also been made for the arithmetic mean: With serially uncorrelated returns and a *known* underlying arithmetic mean, the unbiased estimate of the expected terminal value of an investment is found by compounding forward at the arithmetic mean. For example, if the arithmetic mean is 8%, an unbiased estimate of the expected terminal value of a €1 million investment in five years is €1(1.08)^5 = €1.47 million. In practice, however, the underlying mean is not known. It has been established that compounding forward using the *sample* arithmetic mean, whether or not returns are serially uncorrelated, overestimates the expected terminal value of wealth (see Hughson, Stutzer, and Yung 2006 for a proof). Note that even when returns are not serially uncorrelated, using the arithmetic mean (even a known value) tends to overestimate the expected value of terminal wealth. Returns that revert to the mean are one example of serial correlation of practical concern. In the example, if 8% is merely the sample arithmetic mean (used as an estimate of the unknown underlying mean), we would expect terminal wealth to be less than €1.47 million. Practically, only the first traditional argument still has force.

The geometric mean return of a sample represents the compound rate of growth that equates the beginning value to the ending value of one unit of money initially invested in an asset. Present value models involve the discounting over multiple periods. Discounting is just the reverse side of compounding in terms of finding amounts of equivalent worth at different points in time; because the geometric mean is a compound growth rate, it appears to be a logical choice for estimating a required return in a multiperiod context, even when using a single-period required return model. In contrast to the sample arithmetic mean, using the sample geometric mean does not introduce bias in the calculated expected terminal value of an investment (Hughson, Stutzer, and Yung 2006). Equity risk premium estimates based on the geometric mean have tended to be closer to supply-side and demand-side estimates from economic theory than arithmetic mean estimates. For the foregoing reasons, the geometric mean is increasingly preferred for use in historical estimates of the equity risk premium.

2.1.2 *Long-Term Government Bonds or Short-Term Government Bills*

The choices for the risk-free rate are a short-term government debt rate, such as a 30-day T-bill rate, or a long-term government bond yield to maturity (YTM). Government bonds are preferred to even the highest-rated corporate bonds because they typically have less (near zero) default and equity market risk.

A bond-based equity risk premium estimate in almost all cases is smaller than a bill-based estimate (see Exhibit 1). But a normal upward-sloping yield curve tends to offset the effect of the risk-free rate choice on a required return estimate, because the current expected risk-free rate based on a bond will be larger than the expectation based on a bill. With an inverted yield curve, however, the short-term yields exceed long-term yields and the required return estimate based on using a risk-free rate based on a bill can be much higher.

Industry practice has tended to favor use of a long-term government bond rate in premium estimates despite the fact that such estimates are often used in one-period models such as the CAPM. A risk premium based on a bill rate may produce a better estimate of the required rate of return for discounting a one-year-ahead cash flow, but a premium relative to bonds should produce a more plausible required return/discount rate in a multiperiod context of valuation.

To illustrate a reason for the preference, take the case of bill-relative and bond-relative premia estimates of 5.5% and 4.5%, respectively, for a given market. Assume the yield curve is inverted: The current rates are 9% and 6% on the bill and bond, respectively. The required return on average-risk equity based on bills is 14.5% (9% + 5.5%), compared with 10.5% based on bonds (6% + 4.5%). That 14.5% rate may be appropriate for discounting a one-year-ahead cash flow in a current high interest and inflation environment. The inverted yield curve, however, predicts a downward path for short-rates and inflation. Most of the cash flows lie in the future, and the premium for expected average inflation rates built into the long-bond rate is more plausible. A practical principle is that for the purpose of valuation, the analyst should try to match the duration of the risk-free-rate measure to the duration of the asset being valued. If the analyst has adopted a short-term risk-free rate definition, nevertheless, a practical approach to dealing with the situation just presented would be to use an expected average short-term bill rate rather than the current 9% rate. Advocates of using short-term rates point out that long-term government bonds are subject to risks, such as interest rate risk, that complicate their interpretation.

In practice, many analysts use the current YTM on a long-term government bond as an approximation for the expected return on it. The analyst needs to be clear that she is using a current yield observation, reflecting current inflation expectations. The yield on a recently issued ("on the run") bond mitigates distortions related to liquidity and discounts/premiums relative to face value. The available maturities of liquid government bonds change over time and differ among national markets. If a 20-year maturity is available and trades in a liquid market, however, its yield is a reasonable choice as an estimate of the risk-free rate for equity valuation. (Note that the Ibbotson US long-term government bond yield is based on a portfolio of 20-year average maturity T-bonds. We use that series in the suggested historical estimate of the US equity risk premium.) In many international markets, only bonds of shorter maturity are available or have a liquid market. A 10-year government bond yield is another common choice.

Valuation requires definite estimates of required returns. The data in Exhibit 1 provide one practical starting point for an estimate of equity risk premium for the markets given. As discussed, one mainstream choice among alternative estimates of the historical equity risk premium is the geometric mean historical equity risk premium relative to government bonds.

2.1.3 *Adjusted Historical Estimates*

A historical risk premium estimate may be adjusted in several ways to neutralize the effect of biases that may be present in the underlying equity market return series. One type of adjustment is made to offset the effect of biases in the data series being used to estimate the equity risk premium. A second type of adjustment is made to take account of an independent estimate of the equity risk premium. In both cases, the adjustment could be upward or downward.

One issue is **survivorship bias** in equity market data series. This bias arises when poorly performing or defunct companies are removed from membership in an index, so that only relative winners remain. Survivorship bias tends to inflate historical estimates of the equity risk premium. For many developed markets, equity returns series are now available that are free or nearly free of survivorship bias. When using a series that has such bias, however, the historical risk premium estimate should be adjusted downward. Guidance for such adjustment based on research is sometimes available but is beyond the scope of this discussion. A conceptually related issue with historical estimates can arise when a market has experienced a string of unexpectedly positive or negative events and the surprises do not balance out over the period of sampled data. For example, a string of positive inflation and productivity surprises may result in a series of high returns that increase the historical mean estimate of the equity risk premium. In such cases, a forward-looking model estimate may suggest a much lower value of the equity risk premium. To mitigate that concern, the analyst may adjust the historical estimate downward based on an independent forward-looking estimate (or upward, in the case of a string of negative surprises). Many experts believe that the historical record for various major world markets has benefited from a majority of favorable circumstances that cannot be expected to be duplicated in the future; their recommended adjustment to historical mean estimates is downward. Dimson, Marsh, and Staunton (2002) have argued that historical returns have been advantaged by re-pricings as increasing scope for diversification has led to a lower level of market risk. In the case of the United States, Ibbotson and Chen (2003) recommended a 1.25 percentage point downward adjustment to the Morningstar (Ibbotson) historical mean US equity risk premium estimate based on a lower estimate from a supply-side analysis of the equity risk premium.

Example 2 illustrates difficulties in historical data that could lead to a preference for an adjusted historical or forward-looking estimate.

EXAMPLE 2

The Indian Equity Risk Premium: Historical Estimates of the Equity Risk Premium in a Developing Market

Historical estimates of the equity risk premium in developing markets are often attended by a range of concerns. The case of India can serve as an example. A number of equity indexes are available, and each has possible limitations. Although not as broad-based as the alternatives, the S&P BSE Sensex Index (Sensex), a market-capitalization-weighted index of the shares of 30 leading companies, has the longest available record: Compiled since 1986, returns go back to 1979. Note the following facts concerning this index and other issues relevant to estimating the equity risk premium:

- The backfilled returns from 1979 to 1985 are based on the initial 30 issues selected in 1986, which were among the largest market cap as of 1986.
- The Sensex is a price index; a total return version of the index incorporating dividends is available from 1997 forward.
- Interest rates in India were suppressed by regulation prior to 1991 and were much more volatile thereafter. The benchmark interest rate reported by the Reserve Bank of India averaged 6.57% from 2000 until 2013, reaching an all-time high of 14.50% in August 2000 and a record low of 4.25% in April 2009. In April 2019, the rate was 6.00%. The post-regulation period appears to be associated with higher stock market volatility.
- In 2000, the exchange used this index to open its derivatives market via the trading of Sensex futures contracts. The development of Sensex options followed in 2001.
- Valuation levels have changed significantly in the last two decades. The price-to-earnings ratio for the Sensex was 45.5 in 1994 and reached a low of 13.0 in 1998. In April 2019, the price-to-earnings ratio was 28.4.
- Objective estimates of the extent of any bias can be developed.

(Sources: for institutional background: Varma and Barua 2006; https://tradingeconomics.com/india/interest-rate; andhttps://www.bseindia.com/markets/keystatics/Keystat_index.aspx)

Based only on the information given, address the following.

1 What factors could bias an unadjusted historical risk premium estimate upward?

2 What factors could bias an unadjusted historical risk premium estimate downward?

3 State and explain two indications that the historical time series is non-stationary.

4 Recommend and justify a preference for a historical or an adjusted historical equity risk premium estimate.

Solution to 1:

The backfilling of returns from 1979 to 1985 based on companies selected in 1986 could bias the estimate upward because of survivorship bias. The companies selected in 1986 are likely to have been among the most successful of the companies on the exchange as of 1979. Less clearly, another factor is the suppression of interest rates prior to 1991. An artificially low risk-free rate would bias the equity risk premium estimate upward unless the required return on equity were smaller by an equal amount. Finally, derivative contracts on the Sensex have been available only since the early 2000s. The ability to hedge market risk may result in future equity risk premiums being lower.

Solution to 2:

The failure to incorporate the return from dividends biases the equity risk premium estimate downward. Risk premiums are also generally believed to be inversely related to valuation ratios. The price-to-earnings ratio is currently well below its 1994 level, suggesting that the historical premium estimate may be biased downward.

Solution to 3:

The different levels of interest rates before and after the lifting of regulation in 1991 is one indication that the equity risk premium pre- and post-1991 could be different and that the overall series is non-stationary. A second is the higher level of stock market volatility pre- and post-regulation.

Solution to 4:

Given that objective estimates of the extent of biases can be developed, an adjusted historical estimate would be preferred because such an estimate is more likely to be unbiased and accurate.

In Example 2, one criticism that could be raised concerning any historical estimate is the shortness of the period in the dataset—the post-1991 reform period—that is definitely relevant to the present. Sampling error in any mean estimate—even one based on clean data—would be a major concern for this dataset; recall that the earlier discussion of a two standard deviation interval for the US equity risk premium was based on 113 years of data. The analyst might address specific concerns through an adjusted historical estimate. The analyst may also decide to investigate one or more forward-looking estimates. Forward-looking estimates are the subject of the next section. A later section on international issues will have more information on equity risk premium estimation for emerging markets such as India.

2.2 Forward-Looking Estimates

Because the equity risk premium is based only on expectations for economic and financial variables from the present going forward, it is logical to estimate the premium directly based on current information and expectations concerning such variables. Such estimates are often called forward-looking or *ex ante* estimates. In principle, such estimates may agree with, be higher than, or be lower than historical equity risk premium estimates. *Ex ante* estimates are likely to be less subject to an issue such as non-stationarity or data biases than historical estimates. Such estimates are often subject to other potential errors, however, related to financial and economic models and potential behavioral biases in forecasting.

2.2.1 *Gordon Growth Model Estimates*

Probably the most frequently encountered forward-looking estimate of the equity risk premium is based on a very simple form of a present value model called the constant growth dividend discount model or Gordon growth model, already shown as Equation 3a. For mature developed equity markets such as the Eurozone, United Kingdom, and North American markets, this model's assumptions are often met, at least approximately. Broad-based equity indexes are nearly always associated with a dividend yield, and year-ahead dividend payment may be fairly predictable. The expected dividend growth rate may be inferred based on published analyst or economic expectations, such as consensus analyst expectations of the earnings growth rate for an equity market index (which may be based on forecasts for the constituent companies or a top-down forecast). Specifically, the Gordon growth model (GGM) equity risk premium estimate is as follows:

GGM equity risk premium estimate
= Dividend yield on the index based on year-ahead aggregate forecasted dividends and aggregate market value
+ Consensus long-term earnings growth rate
– Current long-term government bond yield (6)

We can illustrate with the case of the United States. As of July 2019, the dividend yield on the S&P 500 Index as defined in Equation 6 was approximately 1.9% based on a price level of the S&P 500 of 3,020. The consensus analyst view was that earnings on the S&P 500 would actually fall over the next year (see www.standardandpoors.com). We will use the 7% long-term average growth rate as the long-term earnings growth forecast. Note that dividend growth should track earnings growth over the long term. The 30-year US government bond yield was 2.6%. Therefore, according to Equation 6, the Gordon growth model estimate of the US equity risk premium was 1.9% + 7.0% – 2.6%, or 6.3%. Like historical estimates, Gordon growth model estimates generally change through time. For example, this calculation carried out in 2001 and 2009 produced a GGM estimate of 2.4% (computed as 1.2% + 7.0% – 5.8%) and 3.9% (computed as 1.9% + 7.0% – 5.0%), respectively.

Equation 6 is based on an assumption of earnings growth at a stable rate. An assumption of multiple earnings growth stages is more appropriate for very rapidly growing economies. Taking an equity index in such an economy, the analyst may forecast a fast growth stage for the aggregate of companies included in the index, followed by a transition stage in which growth rates decline and a mature growth stage characterized by growth at a moderate, sustainable rate. The discount rate r that equates the sum of the present values of the expected cash flows of the three stages to the current market price of the equity index defines an IRR. Letting PVFastGrowthStage(r) stand for the present value of the cash flows of the fast earnings growth stage with the present value shown as a function of the discount rate r, and using a self-explanatory notation for the present values of the other phases, the equation for IRR is as follows:

$$\begin{aligned}\text{Equity index price} &= \text{PVFastGrowthStage}(r) + \text{PVTransition}(r)\\ &\quad + \text{PVMatureGrowthStage}(r)\end{aligned}$$

The IRR is computable using a spreadsheet's IRR function. Using the IRR as an estimate of the required return on equities (as described earlier), subtracting a government bond yield gives an equity risk premium estimate.

A consequence of the model underlying Equation 6, making assumptions of a constant dividend payout ratio and efficient markets, is that earnings, dividends, and prices are expected to grow at dividend growth rate, so that the P/E is constant. The analyst may believe, however, that the P/E will expand or contract. Some analysts make an adjustment to the estimate in Equation 6 to reflect P/E multiple expansion or contraction. From a given starting market level associated with a given level of earnings and a given P/E, the return from capital appreciation cannot be greater than the earnings growth rate unless the P/E multiple expands. P/E multiple expansion can result from an increase in the earnings growth rate and/or a decrease in risk.

2.2.2 *Macroeconomic Model Estimates*

Using relationships between macroeconomic variables and the financial variables that figure in equity valuation models, analysts can develop equity risk premium estimates. Such models may be more reliable when public equities represent a relatively large share of the economy, as in many developed markets. Many such analyses focus on the supply-side variables that fuel GDP growth (and are thus known as supply-side estimates). The Gordon growth model estimate, when based on a top-down economic analysis rather than using consensus analyst estimates, can be viewed as a supply-side estimate.

To illustrate a supply-side analysis, the total return to equity can be analyzed into four components as explained by Ibbotson and Chen (2003):

- expected inflation: EINFL;
- expected growth rate in real earnings per share: EGREPS;
- expected growth rate in the P/E (the ratio of share price to earnings per share): EGPE; and
- expected income component (including return from reinvestment of income): EINC.

The growth in P/E arises as a factor from a decomposition of the capital appreciation portion of returns—that is, $(P_t/P_t - 1) - 1.0 = [(P_t/E_t)/(P_t - 1/E_t - 1)](E_t/E_t - 1) - 1.0 = (1 + \text{EGPE})(1 + \text{EGREPS}) - 1.0$. So,

$$\text{Equity risk premium} = \left\{\left[(1 + \text{EINFL})(1 + \text{EGREPS})(1 + \text{EGPE}) - 1.0\right] + \text{EINC}\right\} - \text{Expected risk-free return} \tag{7}$$

In the following we illustrate this type of analysis using data for US equity markets as represented by the S&P 500.

- *Expected inflation.* A market forecast is available from the US Treasury and US Treasury inflation-protected securities (TIPS) yield curve:

$$\begin{aligned}\text{Implicit inflation forecast} &\approx \frac{1 + \text{YTM of 20-year maturity T-bonds}}{1 + \text{YTM of 20-year maturity TIPS}} - 1\\ &= \frac{1.023}{1.0066} - 1\\ &= 0.016 \text{ or } 1.6 \text{ percent.}\end{aligned}$$

  We will use an estimate of 1.6% per year, consistent with the TIPS analysis and other long-term forecasts. So, 1 + EINFL = 1.016.

- *Expected growth in real earnings per share.* This quantity should approximately track the real GDP growth rate. An adjustment upward or downward to the real GDP growth rate can be made for any expected differential growth between the companies represented in the equity index being used to represent the stock market and the overall economy.

  According to economic theory, the real GDP growth rate should equal the sum of labor productivity growth and the labor supply growth rate (which can be estimated as the sum of the population growth rate and the increase in the labor force participation rate). A forecasted 2% per year US labor productivity growth rate and 1% per year labor supply growth rate produces a 3% overall real GDP growth rate estimate of 3%. So, 1 + EGREPS = 1.03.

- *Expected growth in the P/E*. The baseline value for this factor is zero, reflecting an efficient markets view. When the analyst views a current P/E level as reflecting overvaluation or undervaluation, however, a negative or positive value, respectively, can be used, reflecting the analyst's investment time horizon. So, without presenting a case for misevaluation, 1 + EGPE = 1.
- *Expected income component*. Historically, for US markets the long-term value has been close to 4.5%, including reinvestment return of 20 bps (see Ibbotson and Chen (2003), p. 90). As of July 2019, however, the S&P 500 dividend yield is below the long-term average. A forward-looking estimate based on the forward expected dividend yield of 2.1% and 10 bps reinvestment return is 2.2%. So, EINC = 0.022.

Using the Ibbotson–Chen format and a risk-free rate of 3%, an estimate of the US equity risk premium estimate is

$$\left\{\left[(1.016)(1.03)(1) - 1.0\right] + 0.022\right\} - 0.025 = 0.0685 - 0.025 = 4.35\%$$

The supply side estimate of 4.35% is very close to the historical geometric mean estimate of 4.4% (see Exhibit 1).

2.2.3 *Survey Estimates*

One way to gauge expectations is to ask people what they expect. Survey estimates of the equity risk premium involve asking a sample of people—frequently, experts—about their expectations for it, or for capital market expectations from which the premium can be inferred.

For example, every year, Fernandez and colleagues conduct a global survey of finance professors, analysts, and managers of companies concerning what value they use for the equity risk premium. The 2018 survey revealed that the average was around 5.4% in the United States, Germany, and many other developed markets but was in excess of 10% for countries such as Argentina, Egypt, Greece, Iran, Pakistan, and Venezuela (Fernandez, Pershin, and Acin 2018).

The monograph *Rethinking the Equity Risk Premium* (Hammond, Leibowitz, and Siegel 2011) included the opinions of many financial experts regarding the future US equity risk premium. The range of equity risk premium estimates was 2.5% to 6.0%. Siegel (2017) describes the 2011 results as having converged around 4% for the equity risk premium, compared with an average of 3.7% estimated by roughly the same group of experts convened in 2002 by the Association of Investment Management and Research (later renamed as CFA Institute). Another source that surveys institutional investors reported risk premia estimates of 4.0% in 2012 and 4.6% in 2014 (Damodaran, 2016). One issue with survey risk premia estimates is that they are sensitive to recent market returns.

3 CAPITAL ASSET PRICING MODEL (CAPM)

c determine the required return on an equity investment using the capital asset pricing model, the Fama–French model, the Pastor–Stambaugh model, macroeconomic multifactor models, and the build-up method (e.g., bond yield plus risk premium);

d explain beta estimation for public companies, thinly traded public companies, and non-public companies;

e describe strengths and weaknesses of methods used to estimate the required return on an equity investment;

With means to estimate the equity risk premium in hand, the analyst can estimate the required return on the equity of a particular issuer. The choices include the following:

- the CAPM;
- a multifactor model such as the Fama–French or related models; and
- a build-up method, such as the bond yield plus risk premium method.

3.1 The Capital Asset Pricing Model

The CAPM is an equation for required return that should hold in **equilibrium** (the condition in which supply equals demand) if the model's assumptions are met; among the key assumptions are that investors are risk averse and that they make investment decisions based on the mean return and variance of returns of their total portfolio. The model's chief insight is that investors evaluate the risk of an asset in terms of the asset's contribution to the systematic risk of their total portfolio (systematic risk is risk that cannot be shed by portfolio diversification). Because the CAPM provides an economically grounded and relatively objective procedure for required return estimation, it has been widely used in valuation.

The expression for the CAPM that is used in practice was given earlier as Equation 4:

$$\text{Required return on share } i = \text{Current expected risk-free return} + \beta_1(\text{Equity risk premium})$$

For example, if the current expected risk-free return is 3%, the asset's beta is 1.20, and the equity risk premium is 4.5%, then the asset's required return is

$$\text{Required return on share } i = 0.030 + 1.20(0.045) = 0.084, \text{ or } 8.4\%$$

The asset's beta measures its market or systematic risk, which in theory is the sensitivity of its returns to the returns on the "market portfolio" of risky assets. Concretely, beta equals the covariance of returns with the returns on the market portfolio divided by the market portfolio's variance of returns. In typical practice for equity valuation, the market portfolio is represented by a broad value-weighted equity market index. The asset's beta is estimated by a least squares regression of the asset's returns on the index's returns and is available also from many vendors. In effect, in Equation 4 the analyst is adjusting the equity risk premium up or down for the asset's level of systematic risk by multiplying it by the asset's beta, adding that asset-specific risk premium to the current expected risk-free return to obtain a required return estimate.

In the typical case in which the equity risk premium is based on a national equity market index and estimated beta is based on sensitivity to that index, the assumption is being made implicitly that equity prices are largely determined by *local* investors. When equities markets are *segmented* in that sense (i.e., local market prices are largely determined by local investors rather than by investors worldwide), two issues with the same risk characteristics can have different required returns if they trade in different markets.

The opposite assumption is that all investors worldwide participate equally in setting prices (perfectly integrated markets). That assumption results in the international CAPM (or world CAPM), in which the risk premium is relative to a world market portfolio. Taking an equity view of the market portfolio, the world equity risk premium can be estimated historically based on the MSCI World index (returns available from 1970), for example, or indirectly as (US equity risk premium estimate)/(Beta of US stocks relative to MSCI World) = 4.5%/0.81 = 5.6%. Computing beta relative to MSCI World and using a national risk-free interest rate, the analyst can obtain international CAPM estimates of required return. In practice, the international CAPM is not commonly relied on for required return on equity estimation.

3.1.1 *Beta Estimation for a Public Company*

The simplest estimate of beta results from an ordinary least squares regression of the return on the stock on the return on the market. The result is often called an unadjusted or "raw" historical beta. The actual values of beta estimates are influenced by several choices:

- *The choice of the index used to represent the market portfolio.* For a number of markets, there are traditional choices. For US equities, the S&P 500 and NYSE Composite have been traditional choices.
- *The length of data period and the frequency of observations.* The most common choice is five years of monthly data, yielding 60 observations

  (a number of vendors including Morningstar and Compustat make that choice). Value Line uses five years of weekly observations. The Bloomberg default is two years of weekly observations, which can be changed at the user's option. One study of US stocks found support for five years of monthly data over alternatives (Bartholdy and Peare 2003). An argument can be made that the Bloomberg default can be especially appropriate in fast-growing markets.

The beta value in a future period has been found to be on average closer to the mean value of 1.0, the beta of an average-systematic-risk security, than to the value of the raw beta. Because valuation is forward looking, it is logical to adjust the raw beta so it more accurately predicts a future beta. The most commonly used adjustment was introduced by Blume (1971):

$$\text{Adjusted beta} = (2/3)(\text{Unadjusted beta}) + (1/3)(1.0) \quad (8)$$

For example, if the beta from a regression of an asset's returns on the market return is 1.30, adjusted beta is (2/3)(1.30) + (1/3)(1.0) = 1.20. Vendors of financial information often report raw and adjusted beta estimates together. Although most vendors use the Blume adjustment, some do not. For example, Morningstar (Ibbotson) adjusts raw beta toward the peer mean value (rather than toward the overall mean value of 1.0). The analyst of course needs to understand the basis behind the presentation of any data that he uses.

The following examples apply the CAPM to estimate the required return on equity.

EXAMPLE 3

Required Return with Adjusted Beta

Daphne Delacour, an equity analyst, is investigating the required return on a (fictitious) Indian construction company Treedle & Weer Ltd. Delacour's data source uses the Sensex (Indian stock market index) as the equity index for estimating beta, which generates a raw beta of 1.537. (The raw beta is the slope of the regression line running through the scatterplot of data points denoting the return on TWLT [*y*-axis] for different returns on the Sensex [*x*-axis].) Delacour's data source also displays a figure for R^2, which in this scenario indicates that beta explains more than 73% of variation in LT returns—an exceptionally good fit. The data source also displays a standard error of beta at 0.121—relatively small compared with the magnitude of the raw estimate, 1.537, and indicative that the beta has been estimated with accuracy.

Delacour also decides to use the CAPM to estimate LT stock's required return. As inputs, she has decided to use her own adjusted historical estimate of 4.0% for the Indian equity risk premium and the 10-year Indian government bond yield of 6.4% as the risk-free rate.

Based only on the information given, address the following:

1 Demonstrate the calculation of adjusted beta using the Blume method.

2 Estimate the required return on LT using the CAPM with an adjusted beta.

Solution to 1:

The calculation for adjusted beta is (2/3)(1.537) + (1/3)(1.0) = 1.358.

Solution to 2:

r = 6.4% + 1.358(4%) = 11.8%.

EXAMPLE 4

Calculating the Required Return on Equity Using the CAPM (1)

Exxon Mobil Corporation, BP p.l.c., and Total S.A. are three "super major" integrated oil and gas companies headquartered, respectively, in the United States, the United Kingdom, and France. An analyst estimates that the equity risk premium in the United States, the United Kingdom, and the Eurozone are, respectively, 4.4%, 5.5%, and 5.9%. Other information is summarized in Exhibit 3 (source: Bloomberg; Fernandez 2018 survey).

Exhibit 3 ExxonMobil, BP, and Total

Company	Beta	Estimated Equity Risk Premium	Risk-Free Rate
Exxon Mobil Corporation	0.90	4.4%	2.8%
BP p.l.c.	0.78	5.5	2.0
Total S.A.	0.71	5.9	1.7

Using the capital asset pricing model, calculate the required return on equity for:

1 Exxon Mobil Corporation.

2 BP p.l.c.

3 Total S.A.

Solution to 1:

The required return on ExxonMobil according to the CAPM is 2.8% + 0.90(4.4%) = 6.76%.

Solution to 2:

The required return on BP according to the CAPM is 2.0% + 0.78(5.5) = 6.29%.

Solution to 3:

The required return on Total stock according to the CAPM is 1.7% + 0.71(5.9) = 5.89%.

EXAMPLE 5

Calculating the Required Return on Equity Using the CAPM (2): Non-Traded Asset Case

Jill Adams is an analyst at a hedge fund that has been offered an equity stake in a privately held US property and liability insurer. Adams identifies Alleghany Corporation as a publicly traded comparable company and intends to use information about Alleghany in evaluating the offer. One sell-side analyst that Adams contacts puts Alleghany's required return on equity at 8.0%. Researching the required return herself, Adams determines that Alleghany has the historical betas shown in Exhibit 4 as of early 2019:

Exhibit 4 Alleghany Corporation: Historical Adjusted Betas

5-Year Adjusted Beta	10-Year Adjusted Beta
0.84	0.79

Source: Bloomberg.

The estimated US equity risk premium (relative to bonds) is 4.4%. The YTM for:

- 91-day US Treasury bills is 2.3%.
- 2-year US government bonds is 2.6%.
- 10-year US government bonds is 2.8%.

Adams follows the most common industry practices concerning the period for estimating beta and adjustments to beta.

1 Estimate Alleghany Corporation's required return based on the CAPM.
2 Is the sell-side analyst's estimate of 10% for Alleghany's cost of equity *most* consistent with Alleghany shares having above-average or below-average systematic risk?

Solution to 1:

Using a five-year horizon for calculating beta is the most common practice. Consistent with the definition of the equity risk premium, a long-bond yield is used in the CAPM: 2.80% + 0.84(4.40%) = 6.496% or 6.5%, approximately.

Solution to 2:

The analyst's estimate implies above-average systematic risk. A beta of 1 by definition represents the beta of the market and so shares of average systematic risk. A beta of 1 implies a required return of 2.80% + 1.0(4.40%) = 7.2%.

When a share issue trades infrequently, the most recent transaction price may be stale and not reflect underlying changes in value. If beta is estimated based on, for example, a monthly data series in which missing values are filled with the most recent transaction price, the estimated beta will be too small and the required return on equity will be underestimated. Several econometric techniques can be used to estimate the beta of infrequently traded securities (see Elton, Gruber, Brown, and Goetzmann 2014 for a summary). A practical alternative is to base the beta estimate on the beta of a comparable security.

3.1.2 *Beta Estimation for Thinly Traded Stocks and Non-public Companies*

Analysts do not have access to a series of market price observations for non-public companies with which to calculate a regression estimate of beta. However, using an industry classification system such as the MSCI/Standard & Poor's Global Industry Classification Standard (GICS) or the FTSE Industry Classification Benchmark (ICB) to identify publicly traded peer companies, the analyst can estimate indirectly the beta of the non-public company on the basis of the public peer's beta.

The procedure must take into account the effect on beta of differences in financial leverage between the non-public company and the benchmark. First, the benchmark beta is unlevered to estimate the beta of the benchmark's assets—reflecting just the systematic risk arising from the economics of the industry. Then, the asset beta is re-levered to reflect the financial leverage of the non-public company.

Let β_E be the equity beta before removing the effects of leverage, if any. This is the benchmark beta. If the debt of the benchmark is high quality (so an assumption that the debt's beta is zero should be approximately true), analysts can use the following expression for unleveraging the beta:

$$\beta_U \approx \left[\frac{1}{1+(1-t)\left(\frac{D}{E}\right)}\right]\beta_E \tag{9a}$$

where β_U is unlevered, also known as asset beta. Note that Equation 9a comes from the expression $\beta_U \approx [1 + (1 - t)(D/E)]^{-1} \times [\beta_E + (1 - t)(D/E)\beta_D]$, making the assumption that $\beta_D = 0$. This expression can be used when the debt's beta is known to be definitely non-zero.Then, if the subject company has debt and equity levels D' and E', respectively, and assuming the subject company's debt is high grade, the subject company's equity beta, β'_E ,is estimated as follows:

$$\beta'_E \approx \left[1+(1-t)\left(\frac{D'}{E'}\right)\right]\beta_U \tag{9b}$$

Expressions 9a and 9b hold under the assumption that the level of debt adjusts to the target capital structure weight as total firm value changes, consistent with the definition for the weighted average cost of capital that will be presented later. Exhibit 5 summarizes the steps.

Exhibit 5 Estimating a Beta for a Non-Traded Company

Step 1
Select the benchmark (comparable)
⬇
Step 2
Estimate benchmark's beta
⬇
Step 3
Unlever the benchmark's beta
⬇
Step 4
Lever the beta to reflect the subject company's financial leverage

To illustrate, suppose that a benchmark company is identified (Step 1) that is 40% funded by debt. By contrast, the weight of debt in the subject company's capital structure is only 20%. The benchmark's beta is estimated at 1.2 (Step 2). The 40% weight of debt in the benchmark implies that the weight of equity is 100% – 40% = 60%. Assume that the marginal rate of tax is 19%. Unlevering the benchmark beta (Step 3):

$$\beta_U \approx \left[\frac{1}{1+(1-t)\left(\frac{D}{E}\right)}\right]\beta_E = \left[\frac{1}{1+(1-0.19)\left(\frac{40}{60}\right)}\right]1.2 = 0.65 \times 1.2 = 0.78$$

Next, the unlevered beta of 0.78 is re-levered according to the financial leverage of the subject company, which uses 20% debt and 80% equity:

$$\beta'_E \approx \left[1+(1-t)\left(\frac{D'}{E'}\right)\right]\beta_U = \left[1+(1-0.19)\left(\frac{20}{80}\right)\right]0.72 = 1.2025 \times 0.72 = 0.87$$

$$\beta'_E \approx \left[1+(1-t)\left(\frac{D'}{E'}\right)\right]\beta_U = \left[1+(1-0.1)\left(\frac{20}{80}\right)\right]0.77 = 1.213 \times 0.77 = 0.93$$

Sometimes, instead of using an individual company as a benchmark, the required return will be benchmarked on a median or average industry beta. A process of unlevering and re-levering can be applied to such a beta based on the median or average industry capital structure.

EXAMPLE 6

Calculating the Required Return on Equity Using the CAPM (3)

Adams turns to determining a beta for use in evaluating the offer of an equity stake in a private insurer and rounds her beta estimate of Alleghany, the public comparable, to 0.85. Assume that as of the valuation date, Alleghany Corporation has 10% debt and 90% equity in its capital structure. The private insurer is 20% funded by debt.

If a beta of 0.85 is assumed for the comparable, what is the estimated beta of the private insurer? Assume marginal tax rate of 20%.

Solution:

The unlevered beta for Alleghany is calculated as [1/(1 + (1 − 0.2)(10/90)] × 0.85 = 0.918 × 0.85 = 0.78. For the private insurer, if debt is 20% of capital then equity is 80% of capital and D′/E′ = 20/80 = 0.25. Therefore, the estimate of the private insurer's equity beta is [1 + (1 – 0.2) × 0.25] × 0.78 = 1.2 × 0.78, or 0.936.

The CAPM is a simple, widely accepted, theory-based method of estimating the cost of equity. Beta, its measure of risk, is readily obtainable for a wide range of securities from a variety of sources and can be estimated easily when not available from a vendor. In portfolios, the idiosyncratic risk of individual securities tends to offset against each other leaving largely beta (market) risk. For individual securities, idiosyncratic risk can overwhelm market risk and, in that case, beta may be a poor predictor of future average return. Thus, the analyst needs to have multiple tools available.

MULTIFACTOR MODELS FOR EQUITY RETURN

4

c determine the required return on an equity investment using the capital asset pricing model, the Fama–French model, the Pastor–Stambaugh model, macroeconomic multifactor models, and the build-up method (e.g., bond yield plus risk premium);

d explain beta estimation for public companies, thinly traded public companies, and non-public companies;

e describe strengths and weaknesses of methods used to estimate the required return on an equity investment;

f explain international considerations in required return estimation;

A substantial amount of evidence has accumulated that the CAPM beta describes risk incompletely. In practice, coefficients of determination (R-squared) for individual stocks' beta regressions may range from 2% to 40%, with many under 10%. For many markets, evidence suggests that multiple factors drive returns. At the cost of greater complexity and expense, the analyst can consider a model for required return based on multiple factors. Greater complexity does not ensure greater explanatory power, however, and any selected multifactor model should be examined for the value it is adding.

Whereas the CAPM adds a single risk premium to the risk-free rate, arbitrage pricing theory (APT) models add a set of risk premia. APT models are based on a multifactor representation of the drivers of return. Formally, APT models express the required return on an asset as follows:

$$r = R_F + (\text{Risk premium})_1 + (\text{Risk premium})_2 + \ldots + (\text{Risk premium})_K \quad (10)$$

where $(\text{Risk premium})_i = (\text{Factor sensitivity})_i \times (\text{Factor risk premium})_i$. **Factor sensitivity** or **factor beta** is the asset's sensitivity to a particular factor (holding all other factors constant). In general, the **factor risk premium** for factor i is the expected return in excess of the risk-free rate accruing to an asset with unit sensitivity to factor i and zero sensitivity to all other factors.

One of the best-known models based on multiple factors expands on the CAPM with two additional factors. That model, the Fama–French model, is discussed next.

4.1 The Fama–French Model

By the end of the 1980s, empirical evidence had accumulated that, at least over certain long periods, in the United States and several other equity markets, investment strategies biased toward small-market-capitalization securities and/or value might generate higher returns over the long run than the CAPM predicts.

In 1993, researchers Eugene Fama and Kenneth French addressed these perceived weaknesses of the CAPM in a model with three factors, known as the Fama–French model (FFM). The FFM is among the most widely known non-proprietary multifactor models. The factors are as follows:

- RMRF, standing for $R_M - R_F$, the return on a market value-weighted equity index in excess of the one-month T-bill rate—this is one way the equity risk premium can be represented and is the factor shared with the CAPM.
- SMB (small minus big), a size (market capitalization) factor. SMB is the average return on three small-cap portfolios minus the average return on three large-cap portfolios. Thus SMB represents a small-cap return premium.
- HML (high minus low), the average return on two high book-to-market portfolios minus the average return on two low book-to-market portfolios. With high book-to-market (equivalently, low price-to-book) shares representing a value bias and low book-to-market representing a growth bias, in general, HML represents a value return premium.

Each of the factors can be viewed as the mean return to a zero-net investment, long–short portfolio. SMB represents the mean return to shorting large-cap shares and investing the proceeds in small-cap shares; HML is the mean return from shorting low book-to-market (high P/B) shares and investing the proceeds in high book-to-market shares. The FFM estimate of the required return is as follows:

$$r_i = R_F + \beta_i^{\text{mkt}}\text{RMRF} + \beta_i^{\text{size}}\text{SMB} + \beta_i^{\text{value}}\text{HML} \tag{11}$$

Historical data on the factors are publicly available for several countries. The historical approach is frequently used in estimating the risk premia of this model. The definitions of RMRF, SMB, and HML have a specificity that lends itself to such estimation. Nevertheless, the range of estimation approaches discussed earlier could also be applied to estimating the FFM factors. Note the definition of RMRF in terms of a short-term rate; available historical series are in terms of a premium over a short-term government debt rate. In using Equation 11, we would take a current short-term risk-free rate. Note as well that because other factors besides the market factor are included in Equation 11, the beta on the market in Equation 11 is generally not exactly the same value as the CAPM beta for a given stock.

We can illustrate the FFM using the case of the US equity market. Assume the short-term interest rate is 1.80%. Historical market, size, and value premiums based on Fama–French data from 1926 to 2018 are 8.3%, 3.2%, and 4.7%, respectively. In a more recent subperiod (1998 through 2018), however, the realized RMRF has averaged 6.6%, the SMB premium has averaged 2.1%, and the realized HML premium has averaged 1.2%. Thus, based on risk premiums for the last 20 years, one estimate of the FFM expected return for the US market as of year-end 2018 is

$$r_i = 0.018 + \beta_i^{\text{mkt}}0.066 + \beta_i^{\text{size}}0.021 + \beta_i^{\text{value}}0.012$$

Consider the case of a small-cap issue with value characteristics and above-average market risk—assume its FFM market beta is 1.20. If the issue's market capitalization is small, we expect it to have a positive size beta; for example, $\beta_i^{\text{size}} = 0.5$. If the shares sell cheaply in relation to book equity (i.e., they have a high book-to-market ratio), the value beta is also expected to be positive; for example, $\beta_i^{\text{value}} = 0.8$. For both the size and value betas, zero is the neutral value, in contrast with the market beta, where the neutral value is 1. Thus, according to the FFM, the shares' required return is close to 12%:

$$r_i = 0.0180 + 1.20(0.066) + 0.5(0.021) + 0.8(0.012) = 0.1173$$

The FFM market beta of 1.2 could be above or below the CAPM beta, but for this comparison, suppose the CAPM beta is 1.20. The CAPM estimate would be 0.0180 + 1.20(0.066) = 0.0972 or less by about 11.73 − 9.72, or 2.01 percentage points. In this case, positive size and value exposures help account for the different estimates in the two models.

Returning to the specification of the FFM to discuss its interpretation, note that the FFM factors are of two types:

- an equity market factor, which is identified with systematic risk as in the CAPM; and
- two factors related to company characteristics and valuation, size (SMB) and value (HML).

The FFM views the size and value factors as representing ("proxying for") a set of underlying risk factors. For example, small market-cap companies may be subject to risk factors such as less ready access to private and public credit markets and competitive disadvantages. High book-to-market may represent shares with depressed prices because of exposure to financial distress. The FFM views the return premiums to small size and value as compensation for bearing types of systematic risk. Many practitioners and researchers believe, however, that those return premiums arise from market inefficiencies rather than compensation for systematic risk.

EXAMPLE 7

Required Return: CAPM and FFM

Charlene Wei is an analyst researching a large, growing US-based technology company. The next task in her valuation is to estimate a required return on equity (which is also the required return on total capital, because this technology company has no long-term debt). Wei's approach is to use an equally weighted average of the CAPM and FFM estimates unless one method appears to be superior as judged by more than a five-point difference in adjusted R^2; in that case, only the estimate with superior explanatory power is used. Exhibit 6 shows the cost of equity information for the company. All the beta estimates in Exhibit 6 are significant at the 5% level.

Exhibit 6 CAPM and FFM Required Return Estimates

	Model A	Model B
1) Current risk-free rate	1.80%	1.80%
2) Beta with respect to market	0.93	0.96
3) Market (equity) risk premium	6.60%	6.60%
Premium for stock: (2) × (3) =	6.14%	6.34%
4) Size beta	—	−0.17
5) Size Premium (SMB)	—	2.10%
Premium for stock: (4) × (5) =	—	−0.36%
6) Value beta	—	−0.15

(continued)

Exhibit 6 (Continued)

	Model A	Model B
7) Value Premium	—	1.20%
Premium for stock: (6) × (7) =	—	−0.18%
R^2	0.54	0.55
Adjusted R^2	0.53	0.52

Sources: This example uses assumed values for the hypothetical company's size and value betas.

Wei is apprised that her firm's economic unit expects that the marketplace will favor growth-oriented equities over the coming year. Reviewing all the information, Wei makes the following statements:

- "This company's cost of equity benefits from the company's above-average market capitalization."
- "If our economic unit's analysis is correct, growth-oriented portfolios are expected to outperform value-oriented portfolios over the next year. As a consequence, we should favor the CAPM required return estimate over the Fama–French estimate."

Using only the foregoing information, address the following.

1 Estimate Microsoft's cost of equity using the:

A CAPM.

B Fama–French model.

2 Judge whether Wei's first statement, concerning the company's cost of equity, is accurate.

3 Judge whether Wei's second statement, concerning the expected relative performance of growth-oriented portfolios and the use of the CAPM and FFM required return estimates, is correct.

Solution to 1:

A The required return according to the CAPM is 1.80% + 0.93(6.60%) = 1.80% + 6.14% = 7.94%

B The required return according to the FFM is 1.80% + 0.96(6.60%) + (−0.17)(2.10%) + (−0.15)(1.20%) = 1.80% + 6.34% + (−0.36%) + (−0.18%) = 7.60%.

Solution to 2:

The statement is accurate. The SMB premium is positive, and the company has negative exposure to it, resulting in the required return estimate being lower by 44 bps.

Solution to 3:

The statement is incorrect. It suggests that computing a required return using a positive value premium is questionable when the investor short-term forecast is for growth to outperform value. Required return estimates should reflect the expected or long-run compensation for risk. The positive value of the value premium in the FFM reflects expected compensation for bearing risk over the long run, consistent with the company's cash flows extending out to the indefinite future. The economic unit's prediction for a short-term time horizon does not invalidate the use of a positive value premium for the Fama–French model.

The regression fit statistics for both the CAPM and FFM in Example 7 were relatively high. There is more to learn about the relative merits of the CAPM and FFM in practice, but the FFM appears to have the potential for being a practical addition to the analyst's toolkit. One study contrasting the CAPM and FFM for US markets found that whereas differences in the CAPM beta explained on average 3% of the cross-sectional differences in returns of the stocks over the next year, the FFM betas explained on average 5% of the differences (Bartholdy and Peare 2003). Neither performance appears impressive, but keep in mind that equity returns are subject to a very high degree of randomness over short horizons.

Researchers have also examined the relative usefulness of domestic versus global versions of the FFM. One study, using data from the United States, Japan, Canada, and the United Kingdom, concludes that domestic (i.e., home-country) versions of the Fama–French three-factor model are more useful than global versions for explaining time series variation in market returns of individual stocks or portfolios (Griffin 2002).

4.2 Extensions to the Fama–French Model

Other researchers have extended the thought process behind the FFM of extending the CAPM to capture observed patterns in equity returns that differences in the CAPM beta appear not to explain. One well-established relationship is that investors demand a return premium for assets that are relatively illiquid—assets that cannot be quickly sold in quantity without high explicit or implicit transaction costs. Pastor and Stambaugh (2003) extended the FFM to encompass compensation for the degree of liquidity of an equity investment.

This model has been applied to public security investment as well as certain private security investments (Metrick 2007). The Pastor–Stambaugh model (PSM) adds to the FFM a fourth factor, LIQ, representing the excess returns to a portfolio that invests the proceeds from shorting high-liquidity stocks in a portfolio of low-liquidity stocks:

$$r_i = R_F + \beta_i^{\text{mkt}}\text{RMRF} + \beta_i^{\text{size}}\text{SMB} + \beta_i^{\text{value}}\text{HML} + \beta_i^{\text{liq}}\text{LIQ} \quad (12)$$

An estimate of the liquidity premium for US equity markets is 4.5% (Metrick 2007). An estimate of the PSM model for US markets is as follows:

$$r_i = 0.003 + \beta_i^{\text{mkt}}0.059 + \beta_i^{\text{size}}0.003 + \beta_i^{\text{value}}0.027 + \beta_i^{\text{liq}}0.045$$

An average-liquidity equity should have a liquidity beta of 0, with no impact on required return. But below-average liquidity (positive liquidity beta) and above-average liquidity (negative liquidity beta) will tend to increase and decrease required return, respectively.

EXAMPLE 8

The Required Return for a Common Stock Investment

A common stock has the following characteristics:

Market beta	1.50
Size beta	0.15
Value beta	–0.52
Liquidity beta	0.20

Based only on the information given, infer the style characteristics of the above common stock issue.

Solution:

The issue appears to be small cap and have a growth orientation. The positive size beta indicates sensitivity to small-cap returns, as would characterize small-cap stocks. (A positive liquidity beta, as shown, would also be typical for small-cap stocks because they usually trade in less liquid markets than do large-cap stocks.) The negative value beta indicates a growth orientation.

The concept of liquidity may be distinguished from marketability. With reference to equities, liquidity relates to the ease and potential price impact of the sale of an equity interest into the market. Liquidity is a function of several factors including the size of the interest and the depth and breadth of the market and its ability to absorb a block (i.e., a large position) without an adverse price impact. In the strictest sense, marketability relates to the right to sell an asset.

Barring securities law or other contractual restrictions, all equity interests are potentially marketable—that is, they can be potentially marketed for sale in the sense of the existence of a market into which the security can be sold. In private business valuation, however, the two terms are often used interchangeably. The typical treatment in that context is to take a discount for lack of marketability (liquidity) from the value estimate, where justified, rather than incorporate the effect in the discount rate, as in the PSM (see Hitchner 2006).

4.3 Macroeconomic and Statistical Multifactor Models

The FFM and PSM are examples of one type of a range of models for required return that are based on multiple fundamental factors (factors that are attributes of the stocks or companies themselves, e.g., the P/E for a share or the company's financial leverage); the group includes several proprietary models as well. Models for required return have also been based on macroeconomic and statistical factors.

- In macroeconomic factor models, the factors are economic variables that affect the expected future cash flows of companies and/or the discount rate that is appropriate to determining their present values.
- In statistical factor models, statistical methods are applied to historical returns to determine portfolios of securities (serving as factors) that explain those returns in various senses.

A specific example of macroeconomic factor models is the five-factor BIRR model, presented in Burmeister, Roll, and Ross (1994), with factor definitions as follows:

1. Confidence risk: the unanticipated change in the return difference between risky corporate bonds and government bonds, both with maturities of 20 years. To explain the factor's name, when their confidence is high, investors are willing to accept a smaller reward for bearing the added risk of corporate bonds.
2. Time horizon risk: the unanticipated change in the return difference between 20-year government bonds and 30-day Treasury bills. This factor reflects investors' willingness to invest for the long term.
3. Inflation risk: the unexpected change in the inflation rate. Nearly all stocks have negative exposure to this factor, as their returns decline with positive surprises in inflation.

4 Business cycle risk: the unexpected change in the level of real business activity. A positive surprise or unanticipated change indicates that the expected growth rate of the economy, measured in constant dollars, has increased.
5 Market timing risk: The portion of the total return of an equity market proxy (e.g., the S&P 500 for the United States) that remains unexplained by the first four risk factors. Almost all stocks have positive sensitivity to this factor.

The fifth factor acknowledges the uncertainty surrounding the correct set of underlying variables for asset pricing; this factor captures influences on the returns to the market proxy not explained by the first four factors. For example, using such a model, the required return for a security could have the following form:

r_i = T-bill rate + (Sensitivity to confidence risk × 2.59%) − (Sensitivity to time horizon risk × 0.66%) − (Sensitivity to inflation risk × 4.32%) + (Sensitivity to business-cycle risk × 1.49%) + (Sensitivity to market-timing risk × 3.61%)

where the risk premia estimates are developed using econometric techniques referenced in Burmeister et al. (1994). Similar to models based on fundamental factors, models based on macroeconomic and statistical factors have various proprietary implementations.

4.4 Build-Up Method Estimates of the Required Return on Equity

Widely used by valuators of closely held businesses, the build-up method estimates the required return on an equity investment as the sum of the risk-free rate and a set of risk premia:

r_i = Risk-free rate + Equity risk premium ± One or more premia (discounts)

The build-up method parallels the risk premium approach embodied in multifactor models, with the difference that specific beta adjustments are not applied to factor risk premiums.

4.5 Build-Up Approaches for Private Business Valuation

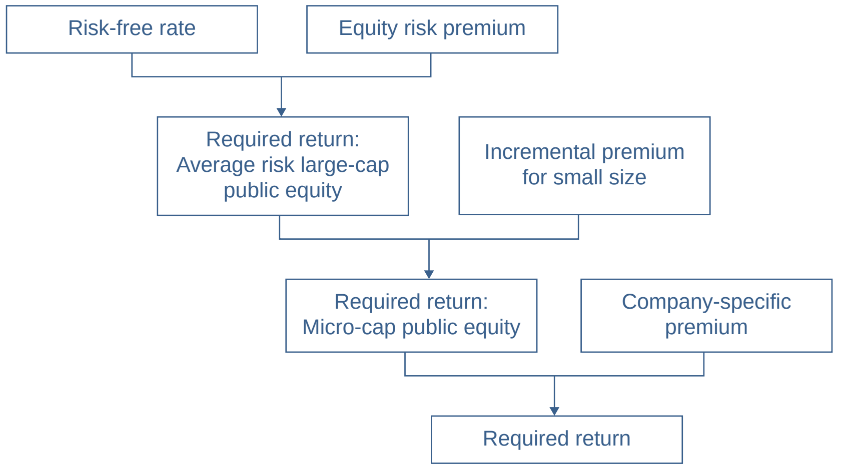
The need for estimates of the required return on the equity of a private business arises when present value models—known in such contexts as income models—are used in the process of valuing business interests. Because the valuation of such interests takes place not only for completely private investment purposes but where courts and tax authorities may play a role—such as in the valuation of a business included in an estate or the valuation of an equity interest for a legal dispute—the valuator may need to research which methods such authorities have found to be acceptable.

Standard approaches to estimating the required return on equity for publicly traded companies, such as the CAPM and the FFM, are adaptable for estimating the required rate of return for non-publicly traded companies. However, valuators often use an approach to valuation that relies on building up the required rate of return as a set of premia added to the risk-free rate. The premia include the equity risk premium and one or more additional premia, often based on factors such as size and perceived company-specific risk, depending on the facts of the exercise and the valuator's analysis of them. An expression for the build-up approach was presented in Equation 5. A traditional specific implementation is as follows (Hitchner 2006):

$$r_i = \text{Risk-free rate} + \text{Equity risk premium} + \text{Size premium}_i + \text{Specific-company premium}_i$$

Exhibit 7 explains the logic for a typical case. The equity risk premium is often estimated with reference to equity indexes of publicly traded companies. The market's largest market-capitalization companies typically constitute a large fraction of such indexes' value. With a beta of 1.0 implicitly multiplying the equity risk premium, the sum of the risk-free rate and equity risk premium is effectively the required return on an average-systematic-risk large-cap public equity issue. In the great majority of cases, private business valuation concerns companies much smaller in size than public large-cap issues. Valuators often add a premium related to the excess returns of small stocks over large stocks, reflecting an incremental return for small size. (The premium is typically after adjustment for the differences in the betas of small- and large-cap stocks to isolate the effect of size—a beta-adjusted size premium.) The level of the size premium is typically assumed to be inversely related to the size of the company being valued. When the size premium estimate is appropriately based on the lowest market-cap decile—frequently the case because many private businesses are small relative to publicly traded companies—the result corresponds to the return on an average-systematic-risk micro-cap public equity issue. An analysis of risk factors that are incremental to those captured by the previously included premia may lead the valuator to add a specific-company premium. This risk premium sometimes includes a premium for unsystematic risk of the subject company under the premise that such risk related to a privately held company may be less easily diversified away.

Exhibit 7 Required Return Estimate for a Privately Held Business



Two additional issues related to required return estimation for private companies are 1) consideration of the relative values of controlling versus minority interests in share value and 2) the effect on share value of the lack of ready marketability for a small equity interest in a private company. Lack of marketability is the inability to immediately sell shares resulting from lack of access to public equity markets because the shares are not registered for public trading. (Marketability may also be restricted by contractual or other reasons.)

With respect to the potential adjustment for the relative control associated with an equity interest in a private company, any adjustments related to the type of interest (controlling or minority) are traditionally made not in the required return but, if appropriate, directly to the preliminary value estimate. The issues involved in such adjustments are complex with some diversity of viewpoints among practitioners. Given these considerations, a detailed discussion is outside the scope of our coverage. Similarly, adjustments for lack of marketability are traditionally taken as an adjustment to the estimated value for an equity interest after any adjustment for the degree of control of the equity interest.

To illustrate, suppose an analyst is valuing a private integrated document management solutions company. The risk-free rate is 3.00%; the analyst's estimate of the equity risk premium is 4.20%; and based on assets and revenues, the company appears to correspond to the top half of the 10th decile of US public companies—decile 10a in Exhibit 8 with market capitalizations of equity ranging from over $129 million to about $207 million.

Exhibit 8 Estimates of US Beta Adjusted Size Premia

Market Cap Decile	Market Cap of Largest Company (in thousands)	Size Premium
6	$1,620,860	1.75%
7	1,090,515	1.77
8	682,750	2.51
9	422,811	2.80
10	206,795	6.10
	Breakdown of the 10th Decile	
10a	206,795	4.34%
10b	128,672	9.81

Source: SBBI Valuation Yearbook (2012), pp. 86–90.

Thus, ignoring any appropriate specific-company premium, an estimate of the required return on equity is 3.00% + 4.20% + 4.34% = 11.54%. A caution is that the size premium for the smallest decile (and especially the 10b component) may reflect not only the premium for healthy small-cap companies but also former large-cap companies that are in financial distress. If that is the case, the historical estimate may not be applicable without a downward adjustment for estimating the required return for a small but financially healthy private company.

A so-called modified CAPM formulation would seek to capture departures from average systematic risk. For example, if the analyst estimated that the company would have a beta of 1.2 if publicly traded, based on its publicly traded peer group, the required return estimate would be

Risk-free rate + Beta × Equity risk premium + Size premium,

or 3.00% + 1.2 × 4.20% + 4.34% = 12.38%. This result could be reconciled to a simple build-up estimate by including a differential return of (1.2 − 1.0)(4.20%) = 0.84% in the specific-company premium.

4.6 Bond Yield Plus Risk Premium

For companies with publicly traded debt, the **bond yield plus risk premium method** provides a quick estimate of the cost of equity. The estimate is

$$\text{BYPRP cost of equity} = \text{YTM on the company's long-term debt} + \text{Risk premium} \quad (13)$$

The YTM on the company's long-term debt includes

- a real interest rate and a premium for expected inflation, which are also factors embodied in a government bond yield; and
- a default risk premium.

The default risk premium captures factors such as profitability, the sensitivity of profitability to the business cycle, and leverage (operating and financial) that also affect the returns to equity. The risk premium in Equation 13 is the premium that compensates for the additional risk of the equity issue compared with the debt issue (recognizing that debt has a prior claim on the cash flows of the company). In US markets, the typical risk premium added is 3%–4%, based on experience.

EXAMPLE 9

The Cost of Equity of Vodafone from Two Perspectives

You are valuing the stock of Vodafone Group Plc as of early 2019, and you have gathered the following information:

UK gilt 30-year yield	1.70%
Vodafone Group Plc 7.875% 02/15/2030	4.64%

The Vodafone bonds, you note, are rated BBB+ by Standard & Poor's and Baa2 by Moody's Investors Service. The beta on Vodafone's stock is 0.70. As a matter of judgment, you have decided to use a risk premium of 3% in the bond yield plus risk premium approach.

1. Calculate the cost of equity using the CAPM. Assume that the equity risk premium is 5.20%.
2. Estimate the cost of equity using the bond yield plus risk premium approach, with a risk premium of 3.0%.
3. Suppose you found that Vodafone's stock, which closed at £140.52 on 8 April 2019, was slightly undervalued based on a DCF valuation using the CAPM cost of equity from Question 1. Does the alternative estimate of the cost of equity from Question 2 support the conclusion based on Question 1?

Solution to 1:

1.70% + 0.70(5.20%) = 5.34%.

Solution to 2:

Add 3.0% to the Vodafone bond YTM: 4.64% + 3.0% = 7.64%. Note that the difference between the Vodafone bond YTM and the long gilt YTM is 2.94%. This amount plus 3.0% is the total estimated risk premium versus UK treasury debt, 2.94% + 3.0% = 5.94%.

Solution to 3:

Not necessarily; *undervalued* means that the value of a security is greater than market price. All else equal, the lower the discount rate, the higher the estimate of value. The inverse relationship between discount rate and value, holding all else constant, is a basic relationship in valuation. If Vodafone appears to be undervalued using the CAPM cost of equity estimate of 5.34%, that does not necessarily mean it will also appear to be undervalued using a 7.64% cost of equity based on the bond yield plus risk premium method.

The bond yield plus risk premium method can be viewed as a build-up method applying to companies with publicly traded debt. The estimate provided can be a useful check when the explanatory power of more-rigorous models is low. Given that a company's shares have positive systematic risk, the yield on its long-term debt is revealing as a check on cost of equity estimate. For example, the 8.375% bonds of Koninklijke KPN N.V. due 1 October 2030 (rated BBB by Standard & Poor's and Baa3 by Moody's) were priced to yield 5.2% as of early April 2019, so an estimated required return for its stock not greater than 5.20% would be suspect.

4.7 The Required Return on Equity: International Issues

Among the issues that concern analysts estimating the required return of equities in a global context are

- exchange rates, and
- data and model issues in emerging markets.

An investor is ultimately concerned with returns and volatility stated in terms of her own currency. Historical returns are often available or can be constructed in local currency and home currency terms. Equity risk premium estimates in home currency terms can be higher or lower than estimates in local currency terms because exchange rate gains and losses from the equity component are generally not exactly offset by gains and losses from the government security component of the equity risk premium. For example, Dimson, Marsh, and Staunton (2018) report that the real (inflation-adjusted) return on Swiss equities from 1900 to 2017, using the geometric mean, was approximately 4.5% in Swiss franc terms, but for a US investor it was approximately 5.3% because the franc appreciated in value relative to the dollar. The US dollar estimate more accurately reflects a US investor's historical experience. A sound approach for any investor is to focus on the local currency record, incorporating any exchange rate forecasts.

The difficulty of required return and risk premium estimation in emerging markets has been previously mentioned. Of the numerous approaches that have been proposed to supplement or replace traditional historical and forward-looking methods, we can mention two.

- The country spread model for the equity risk premium. For an emerging equity market, this model states the following:

$$\text{Equity risk premium estimate} = \text{Equity risk premium for a developed market} + \text{Country premium}$$

The country premium represents a premium associated with the expected greater risk of the emerging market compared with the benchmark developed market. Typically, analysts hope that a sovereign bond yield spread is adequate for approximating this premium. Thus, the country premium is often estimated as the yield on emerging market bonds (denominated in the currency of the developed market) minus the yield on developed market government bonds.

To illustrate, the yield premium on Indian companies' dollar bonds over US Treasuries decreased to 3.19% in March 2013. Taking this premium as the country premium for India and using an estimate of 4.20% for the US equity risk premium, the Indian equity risk premium equals 4.20% + 3.19% = 7.39%.

- The country risk rating model provides a regression-based estimate of the equity risk premium based on the empirical relationship between developed equity market returns and Institutional Investor's semi-annual risk ratings for those markets. The estimated regression equation is then used with the risk ratings for less developed markets to predict the required return for those markets. This model has been recommended by Morningstar (Ibbotson).

5 WEIGHTED AVERAGE COST OF CAPITAL (WACC)

g explain and calculate the weighted average cost of capital for a company;

h evaluate the appropriateness of using a particular rate of return as a discount rate, given a description of the cash flow to be discounted and other relevant facts.

The cost of capital is most commonly estimated using the company's after-tax weighted average cost of capital, WACC for short: a weighted average of required rates of return for the component sources of capital.

The cost of capital is relevant to equity valuation when an analyst takes an indirect, total firm value approach using a present value model. Using the cost of capital to discount expected future cash flows available to debt and equity, the total value of these claims is estimated. The balance of this value after subtracting off the market value of debt is the estimate of the value of equity.

In many jurisdictions, corporations may deduct net interest expense from income in calculating taxes owed, but they cannot deduct payments to shareholders, such as dividends.

If the suppliers of capital are creditors and common stockholders, the expression for WACC is

$$\text{WACC} = \frac{\text{MVD}}{\text{MVD} + \text{MVCE}} r_d (1 - \text{Tax rate}) + \frac{\text{MVCE}}{\text{MVD} + \text{MVCE}} r, \quad (14)$$

where MVD and MVCE are the current market values of debt and (common) equity, not their book or accounting values. Dividing MVD or MVCE by the total market value of the firm, which is MVD + MVCE, gives the proportions of the company's total capital from debt or equity, respectively. These weights will sum to 1.0. Multiplying the before-tax required return on debt (r_d) by 1 minus the marginal corporate tax rate (1 − Tax rate) adjusts the pretax rate r_d downward to reflect the tax deductibility of corporate interest payments that is being assumed. Because distributions to equity are assumed not to be deductible by the corporations, a corporation's before and after-tax costs of equity are the same; no adjustment to r involving the corporate tax rate is appropriate. Generally speaking, it is appropriate to use a company's marginal tax rate rather than its current effective tax rate (reported taxes divided by pretax income) because the effective tax rate can reflect non-recurring items. A cost of capital based on the marginal tax rate usually better reflects a company's future costs in raising funds.

Because the company's capital structure (the proportions of debt and equity financing) can change over time, analysts often use *target* weights instead of the current market-value weights when calculating WACC. Target weights provide a good approximation of the WACC for cases in which the current weights misrepresent the company's normal capital structure.

The before-tax required return on debt is typically estimated using the expected YTM of the company's debt based on current market values. Analysts can choose from any of the methods presented for estimating the required return on equity, *r*. No tax adjustment is appropriate for the cost of equity, assuming payments to shareholders such as dividends are not tax deductible by companies.

EXAMPLE 10

The Weighted Average Cost of Capital for Vodafone

Taking an indirect, total firm value approach to valuing equity, suppose you have the inputs for estimating the cost of capital shown in Exhibit 9. Based only on the information given, estimate Vodafone's WACC.

Exhibit 9 Cost of Capital Data: Vodafone

Panel A: Capital Structure	Value
Long-term debt as a percentage of total capital, at market value	43%
Tax rate	19%

Panel B: Component Costs of Capital	Value
Cost of equity: CAPM estimate	5.34%
YTM of Vodafone long bond	4.64%

Solution:

Long-term debt as a percentage of total capital stated at market value is the weight to be applied to Vodafone's after-tax cost of debt in the WACC calculation. Therefore, Vodafone's WACC based on the data provided is approximately 5.73%, calculated as follows:

$$\begin{aligned}\text{WACC} &= 0.43(4.64\%)(1 - 0.19) + 0.57(5.34\%)\\ &= 1.616\% + 3.044\% = 4.66\%\end{aligned}$$

5.1 Discount Rate Selection in Relation to Cash Flows

When used as discount rates in valuation, required returns need to be defined appropriately relative to the cash flows to be discounted.

A cash flow after more-senior claims (e.g., promised payments on debt and taxes) have been fulfilled is a cash flow to equity. When a cash flow to equity is discounted, the required return on equity is an appropriate discount rate. When a cash flow is available to meet the claims of all of a company's capital providers—usually called a cash flow to the firm—the firm's weighted-average cost of capital is the appropriate discount rate.

Cash flows may be stated in nominal or real terms. When cash flows are stated in real terms, amounts reflect offsets made for actual or anticipated changes in the purchasing power of money. Nominal discount rates must be used with nominal cash flows, and real discount rates must be used with real cash flows. In valuing equity, we will use only nominal cash flows, and therefore we will make use of nominal discount rates. Because the tax rates applying to corporate earnings are generally stated in nominal money terms—such and such tax rates applying at stated levels of nominal pretax earnings—using nominal quantities is an exact approach because it reflects taxes accurately.

Equation 14 presents an after-tax weighted average cost of capital using the after-tax cost of debt. At a later stage, we will present cash flow to the firm definitions for which it is appropriate to use that definition of the cost of capital as the discount rate (i.e., rather than a pretax cost of capital reflecting a pretax cost of debt).

In short, in later sections we will be able to illustrate present value models of stock value using only two discount rates: the nominal required return on equity when the cash flows are those available to common shareholders, and the nominal after-tax weighted average cost of capital when the cash flows are those available to all the company's capital providers.

SUMMARY

We introduced several important return concepts. Required returns are important because they are used as discount rates in determining the present value of expected future cash flows. When an investor's intrinsic value estimate for an asset differs from its market price, the investor generally expects to earn the required return plus a return from the convergence of price to value. When an asset's intrinsic value equals price, however, the investor expects to earn only the required return.

For two important approaches to estimating a company's required return, the CAPM and the build-up model, the analyst needs an estimate of the equity risk premium. We examined realized equity risk premia for a group of major world equity markets and also explained forward-looking estimation methods. For determining the required return on equity, the analyst may choose from the CAPM and various multifactor models such as the Fama–French model and its extensions, examining regression fit statistics to assess the reliability of these methods. For private companies, the analyst can either adapt public equity valuation models for required return using public company comparables or use a build-up model, which starts with the risk-free rate and the estimated equity risk premium and adds additional appropriate risk premia.

When the analyst approaches the valuation of equity indirectly, by first valuing the total firm as the present value of expected future cash flows to all sources of capital, the appropriate discount rate is a weighted average cost of capital based on all sources of capital. Discount rates must be on a nominal (real) basis if cash flows are on a nominal (real) basis.

Among the major points are the following:

- The return from investing in an asset over a specified period is called the *holding period return*. *Realized return* refers to a return achieved in the past, and *expected return* refers to an anticipated return over a future period. A *required return* is the minimum level of expected return that an investor requires to invest in the asset over a specified period, given the asset's riskiness. The (*market*) *required return*, a required rate of return on an asset that is inferred using market prices or returns, is typically used as the *discount rate* in finding the present values of expected future cash flows. If an asset is perceived (is not perceived) as fairly priced in the marketplace, the required return should (should not) equal the investor's expected return. When an asset is believed to be mispriced, investors should earn a *return from convergence of price to intrinsic value.*

- An estimate of the equity risk premium—the incremental return that investors require for holding equities rather than a risk-free asset—is used in the CAPM and in the build-up approach to required return estimation.
- Approaches to equity risk premium estimation include historical, adjusted historical, and forward-looking approaches.
- In historical estimation, the analyst must decide whether to use a short-term or a long-term government bond rate to represent the risk-free rate and whether to calculate a geometric or arithmetic mean for the equity risk premium estimate. Forward-looking estimates include Gordon growth model estimates, supply-side models, and survey estimates. Adjusted historical estimates can involve an adjustment for biases in data series and an adjustment to incorporate an independent estimate of the equity risk premium.
- The CAPM is a widely used model for required return estimation that uses beta relative to a market portfolio proxy to adjust for risk. The Fama–French model is a three-factor model that incorporates the market factor, a size factor, and a value factor. The Pastor–Stambaugh extension to the FFM adds a liquidity factor. The bond yield plus risk premium approach finds a required return estimate as the sum of the YTM of the subject company's debt plus a subjective risk premium (often 3% to 4%).
- When a stock is thinly traded or not publicly traded, its beta may be estimated on the basis of a peer company's beta. The procedure involves unlevering the peer company's beta and then re-levering it to reflect the subject company's use of financial leverage. The procedure adjusts for the effect of differences of financial leverage between the peer and subject company.
- Emerging markets pose special challenges to required return estimation. The country spread model estimates the equity risk premium as the equity risk premium for a developed market plus a country premium. The country risk rating model approach uses risk ratings for developed markets to infer risk ratings and equity risk premiums for emerging markets.
- The weighted average cost of capital is used when valuing the total firm and is generally understood as the nominal after-tax weighted average cost of capital, which is later used in discounting nominal cash flows to the firm. The nominal required return on equity is used in discounting cash flows to equity.

REFERENCES

Bartholdy, Jan, and Paula Peare. 2003. "Unbiased Estimation of Expected Return Using CAPM." *International Review of Financial Analysis* 12 (1): 69–81.

Blume, Marshall. 1971. "On the Assessment of Risk." *Journal of Finance* 26 (1): 1–10.

Burmeister, Edwin, Richard Roll, and Stephen A. Ross. 1994. "A Practitioner's Guide to Arbitrage Pricing Theory." In *A Practitioner's Guide to Factor Models*. Charlottesville, VA: The Research Foundation of the Institute of Chartered Financial Analysts.

Copeland, Tom, Tim Koller, and Jack Murrin. 2000. *Valuation: Measuring and Managing the Value of Companies*. 3rd ed. Hoboken, NJ: John Wiley & Sons.

Cornell, Bradford. 1999. *The Equity Risk Premium*. Hoboken, NJ: John Wiley & Sons.

Damodaran, Aswath. 2016. "*Equity Risk Premiums (ERP): Determinants, Estimation and Implications—The 2016 Edition*" 5 March. Available at SSRN: http://ssrn.com/abstract=2742186.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2002. *Triumphs of the Optimists: 101 Years of Global Investment Returns*. Princeton, NJ: Princeton University Press.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2008. *Global Investment Returns Yearbook 2008.* ABN-AMRO, Royal Bank of Scotland, and London Business School.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2011. "Equity Premiums Around the World." In *Rethinking the Equity Risk Premium*, ed. Hammond, P. Brett, Jr, Martin L. Leibowitz, and Laurence B. Siegel.Charlottesville, VA: Research Foundation of CFA Institute.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2018. *Global Investment Returns Yearbook 2018.* Credit Suisse. London Business School.

Elton, Edwin J., Martin J. Gruber, Stephen J. Brown, and William N. Goetzmann. 2014. *Modern Portfolio Theory and Investment Analysis*. 9th ed. Hoboken, NJ: John Wiley & Sons.

Fama, Eugene F., and Kenneth R. French. 1989. "Business Conditions and Expected Returns on Stocks and Bonds." *Journal of Financial Economics* 25 (1): 23–50.

Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56.

Fama, Eugene F., and Kenneth R. French. 2001. "Disappearing Dividends: Changing Firm Characteristics of Lower Propensity to Pay?" *Journal of Financial Economics* 60 (1): 3–43.

Fernandez, Pablo. 2019. "*WACC and CAPM According to Utilities Regulators: Confusions, Errors and Inconsistencies.*" Available at SSRN: https://ssrn.com/abstract=3327206 or

Fernandez, Pablo, Vitaly Pershin, and Isabel Fernández Acin. 2018. "*Market Risk Premium and Risk-Free Rate Used for 59 Countries in 2018: A Survey.*" Available at SSRN: https://ssrn.com/abstract=3155709 or .

Ferson, Wayne E., and Campbell R. Harvey. 1991. "The Variation of Economic Risk Premiums." *Journal of Political Economy* 99 (2): 385–415.

Griffin, John M. 2002. "Are the Fama and French Factors Global or Country Specific?" *Review of Financial Studies* 15 (3): 783–803.

Hammond, P. Brett, Martin L. Leibowitz, and Laurence B. Siegel. 2011. *Rethinking the Equity Risk Premium*. Charlottesville, VA: Research Foundation of CFA Institute.

Hitchner, James R. 2006. *Financial Valuation: Applications and Models*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Hughson, Eric, Michael Stutzer, and Chris Yung. 2006. "The Misuse of Expected Returns." *Financial Analysts Journal* 62 (6): 88–96.

Ibbotson, Roger, and Peng Chen. 2003. "Long-Run Stock Returns: Participating in the Real Economy." *Financial Analysts Journal* 58 (1): 88–98.

Karolyi, G. Andrew, and René M. Stulz. 2003. "*Are Financial Assets Priced Locally or Globally?*" Handbook of the Economics of Finance, vol. 1: 975–1020.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15 (2): 145–61.

Merton, Robert C. 1980. "On Estimating the Expected Return on the Market: An Exploratory Investigation." *Journal of Financial Economics* 8 (4): 323–61.

Metrick, Andrew. 2007. *Venture Capital and the Finance of Innovation*. Hoboken, NJ: John Wiley & Sons, Inc.

Morningstar. 2007. *"Stocks, Bonds, Bills, and Inflation." 2007 Valuation Edition Yearbook*. Chicago: Morningstar, Inc.

Pastor, Lubos, and Robert F. Stambaugh. 2003. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy* 111 (3): 642–85.

Pinto, Jerald E., Thomas R. Robinson, and John D. Stowe. 2019. "Equity Valuation: A Survey of Professional Practice." *Review of Financial Economics* 37 (2): 219–33.

Siegel, Laurence B. 2017. "*The Equity Risk Premium: A Contextual Literature Review.*" Research Foundation Literature Reviews, vol. 12, no. 1 (November).

Varma, Jayanth R., and Samir K. Barua. 2006. "*A First Cut Estimate of the Equity Risk Premium in India.*" Indian Institute of Management Ahmedabad. Working Paper No. 2006-06-04.

PRACTICE PROBLEMS

1 A Canada-based investor buys shares of a company for C$72.08 on 15 October 20X0 with the intent of holding them for a year. The dividend rate was C$2.11 per year. The investor actually sells the shares on 5 November 20X0 for C$69.52. The investor notes the following additional facts:

- No dividends were paid between 15 October and 5 November.
- The required return on the company's equity was 8.7% on an annual basis and 0.161% on a weekly basis.

A State the lengths of the expected and actual holding periods.

B Given that the company's stock was fairly priced, calculate the price appreciation return (capital gains yield) anticipated by the investor given his initial expectations and initial expected holding period.

C Calculate the investor's realized return.

D Calculate the realized alpha.

2 At the time of valuation, the estimated betas for JPMorgan Chase & Co. and The Boeing Company were 1.50 and 0.80, respectively. The risk-free rate of return was 4.35%, and the equity risk premium was 8.04%. Based on these data, calculate the required rates of return for these two stocks using the CAPM.

3 The estimated factor sensitivities of Clearful Energy (a fictitious company) to Fama–French factors and the risk premia associated with those factors appear in the following table:

	Factor Sensitivity	Risk Premium (%)
Market factor	1.20	4.5
Size factor	−0.50	2.7
Value factor	−0.15	4.3

A Based on the Fama–French model, calculate the required return for Clearful Energy using these estimates. Assume that the Treasury bill rate is 4.7%.

B Describe the expected style characteristics of Clearful Energy based on its factor sensitivities.

4 An analyst's data source shows that Newmont Mining (NEM) has an estimated beta of −0.2. The risk-free rate of return is 2.5%, and the equity risk premium is estimated to be 4.5%.

A Using the CAPM, calculate the required rate of return for investors in NEM.

B The analyst notes that the current yield to maturity on corporate bonds with a credit rating similar to NEM is approximately 3.9%. How should this information affect the analyst's estimate?

5 An analyst wants to account for financial distress and market capitalization as well as market risk in his cost of equity estimate for a particular traded company. Which of the following models is *most appropriate* for achieving that objective?

A The CAPM.

B The Fama–French model.

C A macroeconomic factor model.

© 2021 CFA Institute. All rights reserved.

6 An analyst assembles the following facts concerning an Indian company's component costs of capital and capital structure. Based on the information given, calculate the company's WACC.

Component Costs of Capital	(%)
Cost of equity based on the CAPM	15.6
Pretax cost of debt	8.28
Tax rate	30
Target weight in capital structure	Equity 80, Debt 20

The following information relates to Questions 7–12

An equity index is established in 2001 for a country that has relatively recently established a market economy. The index vendor constructed returns for the five years prior to 2001 based on the initial group of companies constituting the index in 2001. From 2012 to 2016, a series of military confrontations concerning a disputed border disrupted the economy and financial markets. The dispute is conclusively arbitrated at the end of 2016. In total, 20 years of equity market return history is available as of the beginning of 2017. The geometric mean return relative to 10-year government bond returns over 20 years is 2% per year. The forward dividend yield on the index is 1%. Stock returns over 2012 to 2016 reflect the setbacks, but economists predict the country will be on a path of a 4% real GDP growth rate by 2019. Earnings in the public corporate sector are expected to grow at a 5% per year real growth rate. Consistent with that, the market P/E is expected to grow at 1% per year. Although inflation is currently high at 6% per year, the long-term forecast is for an inflation rate of 4% per year. The yield curve has usually been upward sloping, but currently the government yield curve is inverted; at the short end, yields are 9% and at 10-year maturities, yields are 7%.

7 The inclusion of index returns prior to 2001 would be expected to:

A bias the historical equity risk premium estimate upward.

B bias the historical equity risk premium estimate downward.

C have no effect on the historical equity risk premium estimate.

8 The events of 2012 to 2016 would be expected to:

A bias the historical equity risk premium estimate upward.

B bias the historical equity risk premium estimate downward.

C have no effect on the historical equity risk premium estimate.

9 In the current interest rate environment, using a required return estimate based on the short-term government bond rate and a historical equity risk premium defined in terms of a short-term government bond rate would be expected to:

A bias long-term required return on equity estimates upward.

B bias long-term required return on equity estimates downward.

C have no effect on long-term required return on equity estimates.

10 A supply side estimate of the equity risk premium as presented by the Ibbotson–Chen earnings model is *closest* to:

A 3.2%.

B 4.0%.

C 4.3%.

11 Common stock issues in the aforementioned market with average systematic risk are *most likely* to have required rates of return:

A between 2% and 7%.

B between 7% and 9%.

C of 9% or greater.

12 Which of the following statements is *most accurate*? If two equity issues have the same market risk but the first issue has higher leverage, greater liquidity, and a higher required return, the higher required return *is most likely* the result of the first issue's:

A greater liquidity.

B higher leverage.

C higher leverage and greater liquidity.

Questions 13 through 19 relate to Horizon Asset Management

Judy Chen is the primary portfolio manager of the global equities portfolio at Horizon Asset Management. Lars Johansson, a recently hired equity analyst, has been assigned to Chen to assist her with the portfolio.

Chen recently sold shares of Novo-Gemini, Inc. (a fictitious company) from the portfolio. Chen tasks Johansson with assessing Novo-Gemini's return performance, with specific trade information provided in Exhibit 1.

Exhibit 1 Novo-Gemini, Inc. Trade Details

1. Novo-Gemini shares were purchased for $20.75 per share.
2. At the time of purchase, research by Chen suggested that Novo-Gemini shares were expected to sell for $29.00 per share at the end of a three-year holding period.
3. At the time of purchase, the required return for Novo-Gemini based on the CAPM was estimated to be 12.6% on an annual basis.
4. Exactly three years after the purchase date, the shares were sold for $30.05 per share.
5. No dividends were paid by Novo-Gemini over the three-year holding period.

Chen explains to Johansson that, at the time of purchase, the CAPM used to estimate a required return for Novo-Gemini incorporated an unadjusted historical equity risk premium estimate for the US equity market. Chen notes that the US equities market has experienced a meaningful string of favorable inflation and productivity surprises in the past. She asks Johansson whether the historical equity risk premium should have been adjusted before estimating the required return for Novo-Gemini.

For another perspective on the reward for bearing risk, Chen asks Johansson to calculate a forward-looking equity risk premium for the US equity market using data on the S&P 500 in Exhibit 2.

Exhibit 2 S&P 500 Index Data	
Dividend yield, based on year-ahead aggregate forecasted dividends	1.2%
Consensus long-term earnings growth rate	4%
20-year US government bond yield	3%

Chen is now considering adding shares of Bezak, Inc. (a fictitious company) to the portfolio. Chen asks Johansson to calculate Bezak's weighted average cost of capital using the CAPM with the information provided in Exhibit 3.

Exhibit 3 Bezak, Inc.	
Pretax cost of debt	4.9%
Long-term debt as a percentage of total capital, at market value	25%
Marginal tax rate	30%
Bezak, Inc. beta	2.00
Estimated equity risk premium	5.5%
Risk-free rate	3.0%

Lastly, Chen asks Johansson to evaluate Twin Industries (a fictitious company), a privately owned US company that may initiate a public stock offering. Johansson decides to adapt the CAPM to estimate the required return on equity for Twin Industries. Using the MSCI/Standard & Poor's Global Industry Classification Standard (GICS), Johansson identifies a publicly traded peer company with an estimated beta of 1.09 that is much larger but otherwise similar to Twin Industries. Twin Industries is funded 49% by debt, whereas the publicly traded peer company is funded 60% by debt.

13 Based on Exhibit 1, the expected three-year holding period return for Novo-Gemini Inc. at the time of purchase was *closest* to:

A 39.76%.

B 42.76%.

C 44.82%.

14 Based on Exhibit 1, the realized three-year holding period return for Novo-Gemini Inc. was *closest* to:

A 39.76%.

B 42.76%.

C 44.82%.

15 Based on the historical record of surprises in inflation and productivity, the historical equity risk premium for the US equity market, if it is used as an estimate of the forward-looking equity risk premium, should *most likely* be:

A left unchanged.

B adjusted upward.

C adjusted downward.

16 Based on Exhibit 2, the forward-looking estimate for the US equity risk premium is *closest* to:

A 2.2%.

B 5.8%.

C 8.2%.

17 Based on Exhibit 3, and assuming interest on debt is tax-deductible, the WACC for Bezak, Inc. is *closest* to:

A 10.87%.

B 11.36%.

C 13.61%.

18 The estimate of beta for Twin Industries is *closest* to:

A 0.44.

B 0.85.

C 0.89.

19 A potential weakness of Johansson's approach to estimating the required return on equity for Twin Industries is that the return estimate:

A does not include a size premium.

B may overstate potential returns over the long term.

C does not consider systematic risk arising from the economics of the industry.

The following information relates to Questions 20–24

Erica Silverstein is evaluating a spinoff by Acme Manufacturing, Inc., a US industrial company. Acme will spin off one of its subsidiaries, LED Light Company, in a public offering. Silverstein uses the following information and assumptions in her analysis:

- The risk-free rate is 2%.
- LED Light's marginal tax rate is 30%.
- LED Light's debt-to-equity ratio is 0.5.
- The estimated equity risk premium is 6%.
- The yield to maturity on LED Light's long-term debt is forecasted to be 4%.
- The beta of the closest peer firm to LED Light is 0.8, and the firm has a debt-to-equity ratio of 0.4.

As part of her evaluation, Silverstein estimates the required return for LED Light using several approaches.

Approach 1 The CAPM using the beta of a peer firm of LED Light. The peer firm's beta is 0.8, and the firm has a debt-to-equity ratio of 0.4.

Approach 2 The Fama–French three-factor model using the estimated factor sensitivities and factor risk premiums presented in Exhibit 1.

Exhibit 1 Estimated Fama–French Factor Sensitivities and Premiums

Factor	Factor Sensitivity	Factor Premium (%)
Excess return on the market	1.5	6.0
Size (small minus big, SMB)	–0.1	5.0
Value (high minus low, HML)	0.5	4.0

Approach 3 A build-up method using an estimated risk premium of 6.5% that compensates for the additional risk of the equity issue compared with LED Light's long-term debt.

Silverstein is concerned that the required rates of return using these three approaches do not consider the relative liquidity anticipated for LED Light. So, she also estimates the required return using the Pastor–Stambaugh model, which yields an estimate of 14%.

Silverstein discusses her set of required return estimates for LED Light with a colleague. The colleague notes that Silverstein estimated the equity risk premium of 6% using historical data and suggests that Silverstein consider estimating a forwarding-looking estimate to use in her analysis.

Silverstein proceeds with estimating a forward-looking equity risk premium using the Gordon Growth Model approach, given the following additional assumptions:

- Dividend yield on the market index based on one-year-ahead aggregate forecasted dividends is 2.3%.
- Consensus long-term earnings growth rate is 5.9%.
- Expected growth rate in the P/E ratio is 0.1%.
- The 30-year Treasury yield is 2.8%.

20 Using Approach 1, LED Light's estimated required rate of return is *closest* to:

A 6.55%.

B 7.06%.

C 8.48%.

21 Using Approach 2, LED Light's estimated required rate of return is *closest* to:

A 10.5%.

B 12.5%.

C 13.5%.

22 Using Approach 3, LED Light's estimated required rate of return is *closest* to:

A 8.5%.

B 10.5%.

C 12.8%.

23 Using Silverstein's required return estimate based on the Pastor–Stambaugh model as the cost of equity, the weighted average cost of capital (WACC) for LED Light is *closest* to:

A 8.4%.

B 10.3%.

C 10.7%.

24 Silverstein should estimate the Gordon Growth Model (GGM) equity risk premium as:

A 4.3%.

B 5.4%.

C 9.3%.

SOLUTIONS

1 A The expected holding period was one year. The actual holding period was from 15 October 2017 to 5 November 2017, which is three weeks.

B Given fair pricing, the expected return equals the required return, 8.7%. The expected return from price appreciation over the initial anticipated one-year holding period must be equal to the required return minus the dividend yield, 2.11/72.08 = 0.0293, or 2.93%. Thus, expected price appreciation return was 8.7% – 2.93% = 5.77%.

C The realized return was ($69.52 – $72.08)/$72.08 = –0.03552, or –3.55% over three weeks. There was no dividend yield return over the actual holding period.

D The required return over a three-week holding period was $(1.00161)^3 - 1$ = 0.484%. Using the answer to C, the realized alpha was –3.552 – 0.484 = –4.036, or –4.04%.

2 For JPMorgan Chase, the required return is

$$r = R_F + \beta\left[E\left(R_M\right) - R_F\right] = 4.35\% + 1.50(8.04\%) = 4.35\% + 12.06\%$$
$$= 16.41\%$$

For Boeing, the required return is

$$r = R_F + \beta\left[E\left(R_M\right) - R_F\right] = 4.35\% + 0.80(8.04\%) = 4.35\% + 6.43\%$$
$$= 10.78\%$$

3 A The Fama–French model gives the required return as

$$\begin{aligned} r = {} & \text{T-bill rate} \\ & + (\text{Sensitivity to equity market factor} \times \text{Equity risk premium}) \\ & + (\text{Sensitivity to size factor} \times \text{Size risk premium}) \\ & + (\text{Sensitivity to value factor} \times \text{Value risk premium}) \end{aligned}$$

For Clearful Energy, the required return is

$$\begin{aligned} r &= 4.7\% + (1.20 \times 4.5\%) + (-0.50 \times 2.7\%) + (-0.15 \times 4.3\%) \\ &= 4.7\% + 5.4\% - 1.35\% - 0.645\% \\ &= 8.1\% \end{aligned}$$

B Clearful Energy appears to be a large-cap, growth-oriented, high-market-risk stock as indicated by its negative size beta, negative value beta, and market beta above 1.0.

4 A The required return is given by

$$r = 0.025 + (-0.2)(0.045) = 2.5\% - 0.9\% = 1.6\%$$

This example indicates that Newmont Mining has a required return of 1.6%. When beta is negative, the CAPM calculation yields a required rate of return that is below the risk-free rate, which is arguably not meaningful. Cases of equities with negative betas are relatively rare.

B The fact that the NEM's cost of debt is higher than the calculated required return on equity is another indicator that the return estimated using CAPM is not useful for valuing the company's equity.

5 B is correct. The Fama–French model incorporates market, size, and value risk factors. One possible interpretation of the value risk factor is that it relates to financial distress.

6 The company's WACC is 13.64% calculated as follows:

	Equity		Debt	WACC
Weight	0.80		0.20	
After-Tax Cost	15.6%		(1 – 0.30)8.28%	
Weight × Cost	12.48%	+	1.16%	= 13.64%

7 A is correct. The backfilling of index returns using companies that have survived to the index construction date is expected to introduce a positive survivorship bias into returns.

8 B is correct. The events of 2012 to 2016 depressed share returns but 1) are not a persistent feature of the stock market environment, 2) were not offset by other positive events within the historical record, and 3) have led to relatively low valuation levels, which are expected to rebound.

9 A is correct. The required return reflects the magnitude of the historical equity risk premium, which is generally higher when based on a short-term interest rate (as a result of the normal upward-sloping yield curve), and the current value of the rate being used to represent the risk-free rate. The short-term rate is currently higher than the long-term rate, which will also increase the required return estimate. The short-term interest rate, however, overstates the long-term expected inflation rate. Using the short-term interest rate, estimates of the long-term required return on equity will be biased upward.

10 C is correct. According to this model, the equity risk premium is

$$\text{Equity risk premium} = \left\{\left[(1 + \text{EINFL})(1 + \text{EGREPS})(1 + \text{EGPE}) - 1.0\right] + \text{EINC}\right\} - \text{Expected risk-free return}$$

Here,

$$\begin{aligned}
\text{EINFL} &= 4\% \text{ per year (long-term forecast of inflation)}\\
\text{EGREPS} &= 5\% \text{ per year (growth in real earnings)}\\
\text{EGPE} &= 1\% \text{ per year (growth in market P/E)}\\
\text{EINC} &= 1\% \text{ per year (dividend yield or the income portion)}\\
\text{Risk-free return} &= 7\% \text{ per year (for 10-year maturities)}
\end{aligned}$$

By substitution, we find:

$\{[(1.04)(1.05)(1.01)\text{-}1.0] + 0.01\} - 0.07 = 0.113 - 0.07 = 0.043$ or 4.3%.

11 C is correct. Based on a long-term government bond yield of 7%, a beta of 1, and any of the risk premium estimates that can be calculated from the givens (e.g., a 2% historical risk premium estimate or 4.3% supply-side equity risk premium estimate), the required rate of return would be at least 9%. Based on using a short-term rate of 9%, C is the correct choice.

12 B is correct. All else equal, the first issue's greater liquidity would tend to make its required return lower than the second issue's. However, the required return on equity increases as leverage increases. The first issue's higher required return must result from its higher leverage, more than offsetting the effect of its greater liquidity, given that both issues have the same market risk.

13 A is correct. This is the expected three-year holding period return, calculated as

$$\begin{aligned}\text{3-year expected return} &= (V_0 - P_0)/P_0 = (\$29.00 - \$20.75)/\$20.75\\ &= 39.76\%.\end{aligned}$$

14 C is correct. The realized holding period return (note that no dividends were paid during the three-year holding period) is 44.82%. Specifically, the realized three-year holding period is calculated as

$$\begin{aligned}\text{3-Year realized return} &= (P_H - P_0)/(P_0) = (30.05 - 20.75)/20.75\\ &= 44.82\%.\end{aligned}$$

15 C is correct. A string of favorable inflation and productivity surprises may result in a series of high returns that increase the historical mean estimate of the equity risk premium. To mitigate that concern, the analyst may adjust the historical estimate downward based on an independent forward-looking estimate.

16 A is correct. Given the data presented, the equity risk premium can be estimated as

Equity risk premium = Dividend yield on the index based on year-ahead aggregate forecasted dividends and aggregate market value + Consensus long-term earnings growth rate – Current long-term government bond yield.

The equity risk premium is 1.2% + 4.0% – 3.0% = 2.2%.

17 B is correct. The weighted average cost of capital is taking the sum product of each component of capital multiplied by the component's after-tax cost.

First, estimate the cost of equity using the CAPM:

Cost of equity = Risk-free rate + [Equity risk premium × Beta]

Cost of equity = 3.0% + [5.5% × 2.00] = 14%

Now, calculate Bezak's WACC:

	Equity		Debt	WACC
Weight	0.75		0.25	
After-tax cost	14%		(1 – 0.30) × 4.9%	
Weight × After-tax cost	10.5%	+	0.8575%	= 11.36%

18 B is correct. The steps to estimating a beta for a non-traded company are as follows:

Step 1 Select the comparable benchmark

Step 2 Estimate benchmark's beta

Step 3 Unlever the benchmark's beta

Step 4 Lever the beta to reflect the subject company's financial leverage

The beta of the benchmark peer company data is given as 1.09. Next, this beta needs to be unlevered, calculated as follows:

$$\beta_u = \left[\frac{1}{1 + (1 - t)\left(\frac{D}{E}\right)}\right]\beta_l$$

$$\beta_u\left[\frac{1}{1+(1-0.3)\left(\frac{0.60}{0.40}\right)}\right](1.09)$$

$$\beta_u = 0.532, \text{ or } 0.53$$

Then, the unlevered beta needs to be levered up to reflect the financial leverage of Twin Industries, calculated as follows:

$$\beta'_E \approx \left[1+(1-t)\left(\frac{D'}{E'}\right)\right]\beta_u$$

$$\beta'_E \approx \left[1+(1-t)\left(\frac{0.49}{0.51}\right)\right](0.532)$$

$$\beta_u = 0.889, \text{ or } 0.89$$

19 A is correct. Johansson intends to estimate a required return on equity using a modified CAPM approach. Twin Industries is stated to be smaller than the chosen proxy benchmark being used, and there is no size premium adjustment in the CAPM framework. The framework adjusts the beta for leverage differences but not for firm size differences. The build-up method may be more appropriate because it includes the equity risk premium and one or more additional premia, often based on factors such as size and perceived company-specific risk.

20 B is correct. The required return for LED Light using Approach 1 (CAPM using the beta of a peer firm) is calculated as follows:

First, compute the unlevered beta from the peer firm beta:

$$\beta_U = \left[\frac{1}{1+\left((1-t)\frac{D}{E}\right)}\right]\beta_E$$

$$\beta_U = \left[\frac{1}{1+((1-0.30)0.4)}\right]0.8 = \left[\frac{1}{1+0.28}\right]0.8 = 0.625$$

Second, compute the levered beta given the debt-to-equity of LED Light of 0.5:

$$\beta'_E = \left[1+\left((1-t)\frac{D'}{E'}\right)\right]\beta_U$$

$$\beta'_E = \left[1+((1-0.3)0.5)\right]0.625 = 0.84375$$

Finally, compute the required return with the levered beta:

Required rate of return = Current expected risk – free return + β_i (Equity risk premium)

Required rate of return = 0.02 + 0.84375 (0.06) = 0.070625 ≈ 7.06%

A is incorrect because the wrong debt-to-equity ratio is used in the unlevering and levering calculations:

$$\beta_U = \left[\frac{1}{1 + ((1 - 0.30)0.5)}\right]0.8 = \left[\frac{1}{1 + 0.35}\right]0.8 = 0.59259$$

β_E = [1 + ((1 – 0.3)0.4)]0.59259 = 0.7585152

Required rate of return on share i = 0.02 + 0.7585152 (0.06) = 0.065511 ≈ 6.55%

C is incorrect because the beta of the peer firm is not unlevered before applying the levering by LED Light's debt-to-equity ratio.

β_E = [1 + ((1 – 0.3) 0.5)] 0.8 = 1.08

Required rate of return on share i = 0.02 + 1.08 (0.06) = 0.0848 = 8.48%

21 B is correct. The required return for LED Light using Approach 2 (Fama–French three-factor model) is calculated as follows:

$$r_{LED} = R_F + \left[\beta_i^{mkt}(RMRF)\right] + \left[\beta_i^{size}(SMB)\right] + \left[\beta_i^{value}(HML)\right]$$

r_{LED} = 0.02 + [1.5 (0.06)] + [–0.1 (0.05)] + [0.5 (0.04)]

r_{LED} = 0.02 + 0.09 – 0.005 + 0.02 = 0.125 = 12.5%

A is incorrect because the risk-free rate is omitted from the calculation:

r_{LED} = [15 (0.06)] + [–0.1 (0.05)] + [0.5 (0.04)] = 0.09 – 0.005 + 0.02 = 0.105 = 10.5%

C is incorrect because the expected return on the market is used (that is, 0.02 + 0.06 = 0.08) instead of the market risk premium of 0.06, and the risk-free rate is omitted from the calculation:

r_{LED} = [1.5 (0.08)] + [–0.1 (0.05)] + [0.5 (0.04)] = 0.12 – 0.005 + 0.02 = 0.135 = 13.5%

22 B is correct. The required return for LED Light using Approach 3 (bond yield plus risk premium approach) is calculated as follows:

Required rate of return = Expected long-term yield + Equity risk premium

Required rate of return = 0.04 + 0.065 = 0.105 = 10.5%

A is incorrect because the risk-free rate is added to the equity risk premium. The premium should be relative to the company's long-term debt, not the risk-free rate.

Required rate of return on share i = 0.02 + 0.065 = 0.085 = 8.5%

C is incorrect because the beta has been incorrectly multiplied by the equity risk premium before being added to LED Light's long-term bond yield:

Required rate of return on share i = 0.04 + (1.35 × 0.065) = 0.12775 ≈ 12.8%

23 B is correct. The after-tax WACC for LED Light is calculated as:

$$WACC = \frac{MVD}{MVD + MVCE} r_d\left(1 - t) + \frac{MVCE}{MVD + MVCE}\right) r_e$$

Given LED Light's debt-to-equity ratio of 0.5, the weight of debt capital is 0.5/1.5, and the weight of equity common equity capital is 1.0/1.5.

$$WACC = \left[\left(\frac{0.5}{0.5+1.0}\right)0.04(1-0.30)\right] + \left[\left(\frac{1.0}{0.5+1.0}\right)0.14\right]$$

$$WACC = 0.009333 + 0.09333 = 0.102666 \approx 10.3\%$$

A is incorrect because the wrong capital weights are used (the given debt-to-equity ratio of 0.5 is treated as equal weights):

$$WACC = [(0.5)\ 0.04\ (1 - 0.30)] + [(0.5)\ 0.14]$$

$$WACC = 0.014 + 0.07 = 0.084000 = 8.4\%$$

C is incorrect because the adjustment for the after-tax cost of debt is omitted, resulting in an overstatement of the cost of debt and the weighted average cost of capital:

$$WACC = \left[\left(\frac{0.5}{0.5+1.0}\right)0.04\right] + \left[\left(\frac{1.0}{0.5+1.0}\right)0.14\right]$$

$$WACC = 0.01333 + 0.09333 = 0.10666 \approx 10.7\%$$

24 B is correct. The equity risk premium estimate based on the GGM equity risk premium approach is calculated as:

GGM equity risk premium estimate = Dividend yield on the market index based on one-year-ahead aggregate forecasted dividends and aggregate market value + consensus long-term earnings growth rate – current long-term government bond yield

GGM equity risk premium estimate = 2.3% + 5.9% – 2.8% = 5.4%

A is incorrect because the expected growth rate in the P/E ratio of 0.1% (inserted in the equation as 1.1% based on its treatment in the Ibbotson and Chen forward-looking, supply-side estimate equation) is incorrectly subtracted:

GGM equity risk premium estimate = 2.3% + 5.9% – 1.1% – 2.8% = 4.3%

C is incorrect because the expected growth rate in the P/E ratio of 0.1% (inserted in the equation as 1.1% based on its treatment in the Ibbotson and Chen forward-looking, supply-side estimate equation) is incorrectly added, and the long-term government bond yield of 2.8% is incorrectly omitted:

GGM equity risk premium estimate = 2.3% + 5.9% + 1.1% = 9.3%

EQUITY VALUATION

STUDY SESSION

9

Equity Valuation (2)

This study session focuses on financial modeling including the development of forecast model inputs using available industry and corporate information. Approaches for analyzing key balance sheet, income, and cash flow statement items are presented. Other factors affecting financial forecasts such as competition, inflation, deflation, and technology are considered. An example using pro forma financial statements to build a financial model is shown. The session ends with coverage of discounted cash flow (DCF) valuation models and an emphasis on the dividend discount model (DDM).

READING ASSIGNMENTS

Reading 22	Industry and Company Analysis by Matthew L. Coffina, CFA, Anthony M. Fiore, CFA, and Antonius J. van Ooijen, MSc, CFA
Reading 23	Discounted Dividend Valuation by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA

© 2021 CFA Institute. All rights reserved.

READING

22

Industry and Company Analysis

by Matthew L. Coffina, CFA, Anthony M. Fiore, CFA, and Antonius J. van Ooijen, MSc, CFA

Matthew L. Coffina, CFA, is at Morningstar Investment Management LLC (USA). Anthony M. Fiore, CFA, is at Silvercrest Asset Management (USA). Antonius J. van Ooijen, MSc, CFA, is at APG Asset Management (Netherlands).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** compare top-down, bottom-up, and hybrid approaches for developing inputs to equity valuation models;
☐	**b.** compare "growth relative to GDP growth" and "market growth and market share" approaches to forecasting revenue;
☐	**c.** evaluate whether economies of scale are present in an industry by analyzing operating margins and sales levels;
☐	**d.** demonstrate methods to forecast the following costs: cost of goods sold, selling general and administrative costs, financing costs, and income taxes;
☐	**e.** describe approaches to balance sheet modeling;
☐	**f.** describe the relationship between return on invested capital and competitive advantage;
☐	**g.** explain how competitive factors affect prices and costs;
☐	**h.** evaluate the competitive position of a company based on a Porter's five forces analysis;
☐	**i.** explain how to forecast industry and company sales and costs when they are subject to price inflation or deflation;
☐	**j.** evaluate the effects of technological developments on demand, selling prices, costs, and margins;
☐	**k.** explain considerations in the choice of an explicit forecast horizon;
☐	**l.** explain an analyst's choices in developing projections beyond the short-term forecast horizon;
☐	**m.** demonstrate the development of a sales-based pro forma company model.

© 2020 CFA Institute. All rights reserved.

1 INTRODUCTION AND INCOME STATEMENT MODELING: REVENUE

a compare top-down, bottom-up, and hybrid approaches for developing inputs to equity valuation models;

b compare “growth relative to GDP growth” and “market growth and market share” approaches to forecasting revenue;

Industry and company analysis is a key step in the process of valuing companies and the securities they have issued. We will focus on how analysts use industry information and corporate disclosures to forecast a company’s future financial results.

Financial forecasts are the basis for fundamental equity valuation based on discounted cash flows and/or market multiples. An effective forecast model must be based on a thorough understanding of a company’s business, management, strategy, external environment, and historical results. Thus, an analyst begins with a review of the company and its environment—its industry, key products, strategic position, management, competitors, suppliers, and customers. Using this information, an analyst identifies key revenue and cost drivers and assesses the likely impact of relevant trends, such as economic conditions and technological developments. An analyst’s understanding of the fundamental drivers of the business and an assessment of future events provide a basis for developing inputs to the forecast model.

We begin our discussion with an overview of developing a forecast model. We then describe the general approach to forecasting each component of the income statement (revenue, operating costs, and non-operating costs), balance sheet, and cash flow statement. We then turn to special topics: the impact of competitive factors on prices and costs; the effects of inflation and deflation; technological developments; long-term forecasting; and building a company model. We then conclude and present a summary.

1.1 Financial Modeling: An Overview

For most companies, financial modeling begins with the income statement. The income statement is a logical starting point because most companies derive the majority of their value from future cash flow generation, which is primarily determined by the amount of net income generated by the business. Exceptions include banks and insurance companies, for which the value of existing assets and liabilities on the balance sheet may be more relevant to the companies’ overall value than projected future income. The income statement also provides a useful starting point for modeling a company’s balance sheet and cash flow statement.

1.1.1 *Income Statement Modeling: Revenue*

Most companies receive revenue from multiple sources. For analyzing revenue, segment disclosures in companies’ financial reports are often the richest source of information. Both International Financial Reporting Standards (IFRS) and US Generally Accepted Accounting Principles (US GAAP) require companies to disclose certain information about business segments. This information includes how segments are defined; segment revenues, expenses, assets, and liabilities; analysis of revenue by geographical area; and reconciliation of segment accounts to the consolidated financial statement. According to accounting standards, separate financial information must be provided for any segment whose revenue, operating income, or assets account for 10% or more of the revenue, operating income, or assets of the combined company. In addition to the interim and annual financial reports issued by the company, important information can often be found in other disclosures, such as regulatory filings, management presentations, and conference calls, as well as in external data sources.

Revenue can be analyzed by geographical source, business segment, or product line. In a geographic analysis, the analyst places a company's revenue into various geographic "buckets" (groupings). These buckets may be narrowly defined, such as by individual countries, or more broadly defined, such as by region of the world. A geographic analysis can be particularly useful for global companies operating in multiple countries with different underlying growth rates or competitive dynamics. For example, a company may be experiencing relatively slow growth in one region of the world and relatively fast growth in other regions. By examining each region of the world separately, analysts can enhance their understanding of overall growth.

In a breakdown by segment, the analyst classifies a company's revenue into various business segments. Many companies operate in more than one industry or market niche with widely differing economics. Although information is often available for the different business segments, analysts should make an independent judgment about whether a company's chosen segmentation of its business is relevant and material. Sometimes analysts can regroup reported information in a manner that helps make important points.

Finally, a product line analysis provides the most granular level of detail. A product line analysis is most relevant for a company with a manageably small number of products that behave differently, but when combined, they account for most of the company's sales.

Example 1 introduces the first of many examples and exhibits that we use. Please note that many numbers have been rounded; so, in replicating results based on the numbers given in the text and exhibits, small apparent discrepancies may reflect the rounding error.

EXAMPLE 1

Analysis of Revenue (1)

Novo Nordisk is a Denmark-based biopharmaceutical company with a focus on diabetes drugs. The company provides detailed disclosure of revenue along geographic, business segment, and product lines. All figures are in millions of Danish krone (DKK).

In its 2018 annual report, Novo Nordisk provided the following geographic breakdown of sales for the previous three years.

Exhibit 1 Novo Nordisk's Sales by Geographic Region (DKK millions)

	2016	2017	2018
North America	59,242	58,009	56,908
Europe	20,682	21,189	21,679
Latin America	3,593	3,699	4,009
Japan and Korea	6,225	6,072	5,797
China	10,458	10,709	11,285

(continued)

Exhibit 1 (Continued)

	2016	2017	2018
AAMEO (Africa, Asia, Middle East, and Oceania)	11,580	12,018	12,153
Total Sales	111,780	111,696	111,831

The company also classified revenue into business segments it defined: diabetes care (including obesity products) and biopharmaceuticals. Within each segment, disclosure on several individual product lines was also provided.

Exhibit 2 Novo Nordisk's Sales by Business Segment (DKK millions)

	2016	2017	2018
Modern insulins	51,969	53,047	50,391
Human insulins	10,745	9,793	9,265
GLP-1	20,046	23,173	26,129
Other diabetes products	4,612	4,302	4,250
Obesity products	1,577	2,562	3,869
Total diabetes & obesity care	88,949	92,877	93,904
Haemophilia products	10,472	10,469	9,576
Growth disorder products	8,770	6,655	6,834
Other biopharmaceuticals	3,589	1,695	1,517
Total biopharmaceuticals	22,831	18,819	17,927
Total Sales	111,780	111,696	111,831

The company also reported that in local currencies (excluding the impact of exchange rate changes on the figures reported in DKK terms), overall sales rose 5% and the total diabetes and obesity care segment grew 6%.

Use the data in Exhibit 2 to answer the following questions:

1 Modern insulins provide certain advantages over the more traditional human insulins, such as having a faster or longer-lasting effect on blood sugar levels. GLP-1 analogs are even newer products that help the human body produce more insulin. Compare Novo Nordisk's recent sales growth rate of GLP-1 products with those of modern insulins and human insulins.

2 How did Novo Nordisk's sales breakdown change over the last two years? How did the relative contributions from diabetes care and biopharmaceuticals change?

Solution to 1:

Between 2016 and 2018, Novo Nordisk's sales of modern insulins dropped slightly: 50,391/51,969 – 1 ≈ -0.03 or 3%.. Human insulin sales dropped more substantially: 9,265/10,745 – 1 ≈ 0.138 or 13.8%. In contrast, sales of GLP-1 analogs grew by 26,129/20,046 -1 ≈ 30.3% between 2016 and 2018. To calculate a compound annual growth rate, take $(26,129/20,046)^{1/2} - 1 = 0.142$ or 14.2%.

Solution to 2:

In the past two years, Novo Nordisk's total sales remained broadly unchanged (in the reporting currency Danish krone, DKK). Diabetes care and obesity segment increased as a percentage of the total from 80% (88.9 bn/111.8 bn) to 84% (93.9 bn/111.8 bn), while biopharmaceuticals' share of revenue declined from 20% (22.8 bn/111.8 bn) to 16% (17.9 bn/111.8 bn).

Once the analyst has an understanding of the important components of a company's revenue, he must decide whether to use a top-down, bottom-up, or hybrid approach to projecting future revenue. A **top-down approach** usually begins at the level of the overall economy. Forecasts can then be made at more narrowly defined levels, such as sector, industry, and market for a specific product, to arrive at a revenue projection for the individual company. In contrast, a **bottom-up approach** begins at the level of the individual company or a unit within the company, such as individual product lines, locations, or business segments. Analysts then aggregate their projections for the individual products or segments to arrive at a forecast of total revenue for the company. Moreover, analysts also aggregate their revenue projections for individual companies to develop forecasts for a product market, industry, or the overall economy. A **hybrid approach** combines elements of both top-down and bottom-up analysis and can be useful for uncovering implicit assumptions or errors that may arise from using a single approach.

1.1.1.1 Top-Down Approaches to Modeling Revenue Two common top-down approaches to modeling revenue are "growth relative to GDP growth" and "market growth and market share."

In a growth relative to GDP growth approach, the analyst first forecasts the growth rate of nominal gross domestic product. The analyst then considers how the growth rate of the specific company being examined will compare with nominal GDP growth. The analyst may use a forecast for real GDP growth to project volumes and a forecast for inflation to project prices. Analysts often think in terms of percentage point premiums or discounts derived from a company's position in the industrial life cycle (e.g., embryonic, growth, shakeout, mature, or decline) or business cycle sensitivity. Thus, an analyst's conclusion may be that a health care company's revenue will grow at a rate of 200 bps above the nominal GDP growth rate. The forecast may also be in relative terms. Thus, if GDP is forecast to grow at 4% and the company's revenue is forecast to grow at a 15% faster rate, the forecast percent change in revenue would be 4% × (1 + 0.15) = 4.6%, or 60 bps higher in absolute terms.

In a market growth and market share approach, the analyst first forecasts growth in a particular market. The analyst then considers the company's current market share and how that share is likely to change over time. For example, if a company is expected to maintain an 8% market share of a given product market and the product market is forecast to grow from €18.75 billion to €20 billion in annual revenue, the forecast growth in company revenue is from a level of 8% × €18.75 billion = €1.5 billion to a level of 8% × €20 billion = €1.6 billion (considering this product market alone). If the product market revenue has a predictable relationship with GDP, regression analysis might be used to estimate the relationship.

1.1.1.2 Bottom-Up Approaches to Modeling Revenue Examples of bottom-up approaches to modeling revenue include the following:

- *Time series*: forecasts based on historical growth rates or time-series analysis.
- *Return on capital*: forecasts based on balance sheet accounts. For example, interest revenue for a bank may be calculated as loans multiplied by the average interest rate.
- *Capacity-based measure:* forecasts, for example, in retailing, based on same-store sales growth (for stores that have been open for at least 12 months) and sales related to new stores.

Time-series forecasts are among the simplest. For example, analysts may fit a trend line to historical data and then project sales over the desired time frame (e.g., using Excel's TREND formula). In such a case, analysts would be projecting historical growth rates to continue, but they might also use different assumptions—for example, they may project growth to decline linearly from current rates to some long-run rate. Note that time-series methods may also be used as tools in executing a top-down analysis, such as projecting GDP growth in a growth relative to GDP growth approach.

1.1.1.3 Hybrid Approaches to Modeling Revenue Hybrid approaches combine elements of both top-down and bottom-up analysis, and in practice they are the most commonly used approaches. For example, the analyst may use a market growth and market share approach to model individual product lines or business segments. Then, the analyst may aggregate the individual projections to arrive at a forecast for the overall company because the sum of forecast segment revenue equals the segment market size multiplied by the market share for all segments.

In a volume and price approach, the analyst makes separate projections for volumes (e.g., the number of products sold or the number of customers served) and average selling price. Depending on how these elements are forecast, this approach can be classified as top-down, bottom-up, or hybrid.

EXAMPLE 2

Analysis of Revenue (2)

Use the data in Example 1 on Novo Nordisk to answer the following questions:

1 Xiaoping Wu is an equity analyst covering European pharmaceutical companies for his clients in China. Wu projects that global nominal GDP will grow 3% annually over the long run, based on 2% real growth and 1% inflation. The incidence of diabetes is increasing globally because of increasingly unhealthy diets and sedentary lifestyles. As a result, Wu believes global sales of diabetes drugs will grow 100 bps faster than nominal GDP over the long run. Wu believes the revenue growth rate of Novo Nordisk's diabetes care segment will match the projected long-run growth rate of the diabetes drug market.

   A Is Wu using a top-down, bottom-up, or hybrid approach to modeling Novo Nordisk's revenue?

   B Based on Wu's projections for revenue growth, calculate the estimated revenue growth rate for the diabetes care segment in 2020. Assume no impact from exchange rate changes.

2 Helga Hansen is a buy-side analyst in Denmark. In 2019, Hansen was investigating Ozempic, a compound launched in 2018, in a class of diabetes drugs called GLP-1 analogs. As of 2019, Ozempic is one of several

products on the market, offered alongside Novo Nordisk's existing GLP-1 drug Victoza and competing with dulaglutide (Trulicity) by Eli Lilly and exenatide (brands Byetta & Bydureon) by AstraZeneca.

Eli Lilly and AstraZeneca reported global sales of their products in US dollars. Hansen converted the reported figures to euros using the annual average USD/EUR, GBP/EUR, and DKK/EUR exchange rates and compiled the following table comparing sales of the different GLP-1 products measured in millions of euros.

Compound/brand	Company	2016	2017	2018
Exenatide (Byetta & Bydureon)	AstraZeneca	751	664	601
Victoza	Novo Nordisk	2,692	3,115	3,264
Ozempic	Novo Nordisk	0	0	241
Dulaglutide (Trulicity)	Eli Lilly	836	1,796	2,709
Albiglutide (Tanzeum)	GSK	99	76	0

A What was the growth rate in total GLP-1 analog sales in 2018?

B What percentage of GLP-1 analog sales growth in 2018 was caused by Ozempic?

C A year earlier, Hansen projected that the growth rate of the GLP-1 analog market would slow to 18% in 2018. She also expected Trulicity to improve its market share by 10 percentage points. What was Hansen's estimate of 2018 Trulicity sales? How close was she to the actual result?

D Is Hansen's approach to modeling Novo Nordisk's sales best described as bottom-up, top-down, or hybrid?

Solution to 1A:

Wu's long-run revenue projections are based on Novo Nordisk's growth relative to nominal GDP growth, which is a top-down approach. However, his estimated growth rate is applied to only one of Novo Nordisk's segments (diabetes care), indicating a hybrid approach. Wu's four-year forecasts are also based in part on the historical growth rate of the diabetes care segment, which is a bottom-up approach. Wu is thus using a hybrid approach.

Solution to 1B:

The data in Example 1 indicate that Novo Nordisk's diabetes and obesity care segment grew approximately 1% in 2018 in DKK terms (= 93.90/92.88 − 1 ≈ 0.010) and, importantly, 6% in local currencies. Wu projects the long-run growth rate to be in line with the diabetes drug market growth at 4% (100 bps faster than GDP growth of 3%). The difference between the 2018 growth rate and projected long-run growth rate is 2% (= 6% − 4%), and Wu expects the modest deceleration in growth to occur linearly over four years, implying a reduction of 50 bps per year in the growth rate. The estimated growth rates by year are thus

2019 = 5.5%

2020 = 5%

2021 = 4.5%

2015 = 4%

Thereafter, 4%

The estimated revenue growth rate for 2020 is 5%.

Solution to 2A:

Total sales of GLP-1 analogs in 2018 were €6,815million (= 601 + 3,264 + 241 + 2,709), compared with total 2017 sales of €5,651million (= 664 + 3,115 +1,796 + 76). The growth rate was thus around 20.6% (= 6,815/5,651 − 1 ≈ 0.206).

Solution to 2B:

Total GLP-1 analog sales increased by €1,164 million (from €5,651 million to €6,815 million). The newly-launched Ozempic achieved sales of €241 million (starting from zero). So, Ozempic accounted for approximately 21% of the growth in sales of this drug class (= 241/1,164 ≈ 0.207, or 20.7%). Note that the figure for Trulicity was 78% (914/1,164 = 0.784).

Solution to 2C:

Based on 2017 sales of €5,651 million and a projected growth rate of 18%, Hansen projected the total GLP-1 analog market to be worth about €6,668 million in 2018 (= 5,651 × 1.18 ≈ 6,668). Trulicity's market share in 2017 was around 32%, which Hansen projected to improve by 10 percentage points, resulting in a 42% market share in 2018. Hansen thus projected 2018 Trulicity sales to be around €2,800 million (= 6,668 × 0.42 ≈ 2,800). Actual Trulicity sales in 2018 were €2,709 million, so Hansen's estimate was too high by €91 million (= 2,800 − 2,709 = 91).

Solution to 2D:

Hansen bases her estimates on market growth and market share, which would normally imply a top-down approach. The analysis, however, is applied to an individual product line, implying a bottom-up approach. Therefore, Hansen is using a hybrid approach.

2 INCOME STATEMENT MODELING: OPERATING COSTS

c evaluate whether economies of scale are present in an industry by analyzing operating margins and sales levels;

Disclosure about operating costs is frequently less detailed than disclosure about revenue. If relevant information is available, analysts may consider matching the cost analysis to the revenue analysis. For example, they might model costs separately for different geographic regions, business segments, or product lines. More frequently, analysts will be forced to consider costs at a more aggregated level than the level used to analyze revenue. Analysts should still keep in mind their revenue analysis when deriving cost assumptions. For instance, if a relatively low-margin product is expected to grow faster than a relatively high-margin product, analysts should project some level of overall margin deterioration, even if they are not certain about the precise margins earned on each product.

Once again, analysts can take a top-down, bottom-up, or hybrid view of costs. In a top-down approach, analysts may consider such factors as the overall level of inflation or industry-specific costs before making assumptions about the individual company. In contrast, in a bottom-up approach analysts would start at the company level, considering such factors as segment-level margins, historical cost growth rates, historical margin levels, or the costs of delivering specific products. A hybrid approach would incorporate both top-down and bottom-up elements.

When estimating costs, analysts should pay particular attention to fixed costs. Variable costs are directly linked to revenue growth, and they may be best modeled as a percentage of revenue or as projected unit volume multiplied by unit variable costs.

By contrast, increases in fixed costs are not directly related to revenue; rather, they are related to future investment in property, plant, and equipment (PP&E) and to total capacity growth. Practically, fixed costs may be assumed to grow at their own rate, based on an analysis of future PP&E growth. Analysts should determine whether, at its current level of output, the subject company has **economies of scale**, a situation in which average costs per unit of a good or service produced fall as volume rises. Factors that can lead to economies of scale include, at higher levels of production, greater bargaining power with suppliers, lower cost of capital, and lower per unit advertising expenses. Gross and operating margins tend to be positively correlated with sales levels in an industry that enjoys economies of scale.

Analysts must also be aware of any uncertainty surrounding estimates of costs. For example, banks and insurance companies create reserves against estimated future losses, while companies with large pension plans have long-duration liabilities, the true costs of which may not be known for many years. A review of disclosures about reserving practices related to future obligations and pensions can be helpful in assessing whether cost estimates are reasonable. But most of the time it is difficult for the external analyst to anticipate future revisions to cost estimates. Other aspects affecting the uncertainty of cost estimates include competitive factors and technological developments. This impact will be discussed in later sections.

EXAMPLE 3

Approaches to Modeling Operating Costs

Walgreens (full name Walgreens Boots Alliance Inc) and Rite Aid are two of the largest retail drugstore chains in the United States. For both companies, around two-thirds of their sales are from prescription pharmaceuticals, with the remaining third coming from front-of-store categories, such as beauty and wellness products and services, over-the-counter drugs, convenience foods, and administration of immunizations.

Although they are in the same industry and both operate predominantly in the United States, Walgreens and Rite Aid have very different operating margins. There is reason to believe that economies of scale exist in the drugstore business. For example, larger drugstore companies have greater bargaining leverage with suppliers and the ability to negotiate better reimbursement rates with third-party payers. Some relevant data are presented in Exhibit 3. The last column on the right includes results from the companies' 2018 fiscal year (ended in August 2018 for Walgreens and March 2018 for Rite Aid).

Exhibit 3 Financial Results for Walgreens and Rite Aid

	2016	2017	2018
Walgreens			
Revenue ($ millions)	117,351	118,214	131,537
Cost of goods sold ($ millions)	87,477	89,052	100,745

(continued)

Exhibit 3 (Continued)

	2016	2017	2018
Selling, general, and administrative ($ millions)	23,661	23,751	24,764
Operating income ($ millions)	6,001	5,557	6,414
End of period selling area square footage (millions sq. ft.)	112	112	150
Average selling area (millions sq. ft)	112	112	131
Same-store pharmacy sales growth	3.2%	3.1%	3.4%
Rite Aid			
Revenue ($ millions)	20,770	22,928	21,529
Cost of goods sold ($ millions)	15,778	17,863	16,749
Selling, general, and administrative ($ millions)	4,581	4,777	4,651
Operating income ($ millions)	337	242	133
End of period selling area square footage (millions sq. ft.)	45.2	44.9	26.8
Average selling area square footage (millions sq. ft.)	45.5	45.1	35.9
Same-store pharmacy sales growth	1.3%	-2.2%	-2.9%

(*Sources:* Rite Aid Form 10-K, Walgreens Alliance Boots 10-K, and www.ft.com.)

Customer service may be one driver of revenue for the retail drug business. Retail analysts commonly use a combination of qualitative and quantitative evidence to assess customer service. Qualitative evidence might come from personal store visits or customer surveys. Quantitative evidence may be based on such metrics as selling, general, and administrative (SG&A) expense per square foot. Too little spending on SG&A might indicate that stores are understaffed. Relatedly, same-store sales growth may be an indicator of customer satisfaction.

Use the data given to answer the following questions:

1 On the basis of the 2018 operating margins for Walgreens and Rite Aid, is there evidence suggesting that economies of scale exist in the retail drugstore business? If so, are economies of scale realized in cost of goods sold or SG&A expenses?

2 Marco Benitez is a US-based equity analyst with an independent research firm. Benitez is researching service levels in the US drugstore industry.

   A Calculate and interpret Walgreens' and Rite Aid's SG&A per average square foot for 2018 and 2016.

   B Assuming that customer satisfaction is a driver of sales growth, which company appears to have a more satisfied customer base over the period examined?

   C Benitez projects that Rite Aid's average selling area square footage will increase 2% annually over the next three years. He believes SG&A per average square foot will decrease 1% annually during this time. What is Benitez's projection for total SG&A expense in 2021?

3 Jason Lewis is another US-based equity analyst covering the retail drugstore industry. He is considering several approaches to forecasting Walgreens' and Rite Aid's' future costs. Classify each of the following as a bottom-up, top-down, or hybrid approach.

A Lewis believes government insurance programs in the United States will face budgetary pressures in the future, which will result in lower reimbursements across the retail drugstore industry. Lewis thinks this will lower all drugstores' gross margins.

B Lewis projects that Walgreens' historical rate of growth in SG&A expenses will continue for the next five years. But in the long-run, he projects SG&A to grow at the rate of inflation.

C To estimate Rite Aid's future lease expense, Lewis makes assumptions about square footage growth and average rent per square foot, based on past experience.

Solution to 1:

Walgreens 2018 operating margin (operating income divided by revenue) was 4.9% (= 6,414/131,537 ≈ 0.049). Rite Aid's operating margin in the comparable year was 0.6% (= 133/21,529 ≈ 0.006). Walgreens' much larger size ($131.5 bn in sales versus $21.5 bn for Rite Aid) combined with its much higher profitability provides evidence suggesting that there are economies of scale in the drugstore industry.

To answer the second part of the question, divide both companies' expense lines by revenue for 2018. Walgreens' cost of goods sold consumed 76.6% (≈ 100,745/131,537) of revenue, whereas Rite Aid's cost of goods sold consumed 77.8% (≈ 16,749/21,529). Walgreens' SG&A consumed 18.8% (≈ 24,764/131,537) of revenue, whereas Rite Aid's SG&A consumed 21.6% (≈ 4,651/21,529). The results indicate that there are economies of scale in both cost of goods sold and SG&A.

Solution to 2A:

Walgreens' average SG&A per square foot in 2018 and 2016 were $189 (≈ 24,764/131) and $211 (≈ 23,661/112), respectively. That is a decrease of approximately 10%. The same figures for Rite Aid were $129.6 (≈ 4,651/35.9) in 2018 and $100.7 (≈ 4,581/45.5) in 2016, an increase of 28.7%. The difference in the levels of SG&A per sq. ft. might be evidence of higher service levels at Walgreens compared to Rite Aid. Note that the change between 2016 and 2018 can be mostly attributed to acquisitions and disposals that the companies have been involved in.

Solution to 2B:

Walgreens appears to have the more satisfied customer base, considering just same-store sales. Comparing same-store sales growth for the two companies, Walgreens has consistently outperformed Rite Aid in the past three years. This result supports the hypothesis that Rite Aid's customers may be less satisfied than Walgreens' customers over the time period examined.

Solution to 2C:

Benitez projects Walgreens' selling area will be around 159 million square feet [= 150×1.02^3] by the end of fiscal 2021. He projects SG&A per average square foot will be $183.4 [= $189 \times (0.99)^3$]. He thus estimates total SG&A to be $29,160 million (159 × 183.4) in 2021.

Solution to 3A:

This case describes a top-down approach because Lewis considers the overall industry environment before individual companies.

Solution to 3B:

In this case, Lewis combines a bottom-up approach (projecting the historical rate of growth to continue) with a top-down approach (basing his long-run assumptions on the overall rate of inflation). Therefore, this is a hybrid approach.

Solution to 3C:

This case describes a bottom-up approach because Lewis bases his forecasts on Rite Aid's historical experience.

3 MODELING OPERATING COSTS: COGS AND SG&A

d demonstrate methods to forecast the following costs: cost of goods sold, selling general and administrative costs, financing costs, and income taxes;

The cost of goods sold (COGS) is typically the single largest cost for manufacturing and merchandising companies. For a manufacturer, COGS include raw materials along with the direct labor and overhead used in producing the goods.

Because sales minus COGS equals gross margin, COGS and gross margin vary inversely. Forecasting COGS as a percentage of sales and forecasting gross margin percentage are equivalent in that a value for one implies a value for the other.

Because COGS has a direct link with sales, forecasting this item as a percentage of sales is usually a good approach. Historical data on a company's COGS as a percentage of sales usually provide a useful starting point for estimates. For example, if a company is losing market share in a market in which the emergence of new substitute products are also putting the overall sector under pricing pressure, gross margins are likely to decline. But if the company is gaining market share because it has introduced new competitive and innovative products, especially if it has done so in combination with achieving cost advantages, gross margins are likely to improve.

Because cost of goods sold is relatively a large cost, a small error in this item can have a material impact on the forecasted operating profit. Thus, analysts should consider whether an analysis of these costs (e.g., by segment, by product category, or by volume and price components), when such an analysis is possible, can improve forecasting accuracy. For example, some companies face fluctuating input costs that can be passed on to customers only with a time lag. Particularly for companies that have low gross margins, sudden shocks in input costs can affect operating profit significantly. A good example is the sensitivity of airlines' profits to unhedged changes in jet fuel costs. In these cases, a breakdown of both costs and sales into volume and price components is essential for developing short-term forecasts, even if analysts use the overall relationship between sales and input cost for developing longer-term forecasts.

Analysts should also consider the impact of a company's hedging strategy. For example, commodity-driven companies' gross margins almost automatically decline if input prices increase significantly because of variable costs increasing at a faster rate than revenue growth. Assume a company's cost of goods sold as a percentage of sales equals 25%. If the input costs double and the company is able to pass the entire increase on to its clients through a 25% price increase, cost of goods sold as a percentage of sales will increase (to 40%) because an equal absolute amount has been added to the numerator and to the denominator. Thus, although the absolute amount of gross profit will remain constant, the gross margin will decrease (from 75% to 60%). Through various hedging strategies a company can mitigate the impact on profitability. For example, brewers often hedge the cost of barley, a key raw material needed for brewing beer, one year in advance. Although companies usually do not disclose their hedging positions, their general strategy is often revealed in the footnotes of the annual report. Further, the negative impact of increasing sales prices on sales volume can be mitigated by a policy of gradual sales price increases. For example, if the brewer expects higher barley prices because of a bad harvest, the brewer can slowly increase prices to avoid a strong price jump next year.

Competitors' gross margins can also provide a useful cross check for estimating a realistic gross margin. Gross margin differences among companies within a sector should logically relate to differences in their business operations. For example, in the Netherlands, supermarket chain Albert Heijn has a higher gross margin in the very competitive grocery sector because it can leverage its dominant 35% market share to achieve savings in purchases; it also has an ability to make higher margin private label products. All of these competitive advantages contribute to its structurally higher gross margin within the grocery sector. But if a new large competitor emerges (e.g., through consolidation of the fragmented market), Albert Heijn's above average gross margin could come under pressure. Also note that differences in competitors' gross margins does not always indicate a superior competitive position but instead could simply reflect differences in business models. For example, some companies in the grocery segment own and operate their own retail stores whereas other companies operate as wholesalers with franchised retail operations. In the franchised retailing business model, most of the operating costs are incurred by the franchisee; the wholesaler offers products with only a small markup to these franchisees. Compared with a grocer with its own stores, a supermarket wholesaler will have a much lower gross margin. The grocer with its own stores, however, will have much higher operating costs. Even though differences in business models can complicate direct comparisons, competitors' gross margins can nonetheless offer potentially useful insights.

3.1 Selling, General, and Administrative Expenses

Selling, general, and administrative (SG&A) expenses are the other main type of operating costs. In contrast to COGS, SG&A expenses have less of a direct relationship with the revenue of a company. As an illustration of the profit impact of COGS and SG&A, consider Thai cement and building materials company Siam Cement Group. A summary of its key income statement items is shown in Exhibit 4.

Exhibit 4 Siam Cement Group Financials

	2017 (Baht billions)	2018 (Baht billions)	YoY%	Percent of Sales 2017	Percent of Sales 2018
Net sales	450.92	478.44	6.1	100.0	100.0
Cost of goods sold	349.31	383.46	9.8	77.5	80.1
Gross profit	101.61	94.98	-6.5	22.5	19.9
SG&A	52.58	55.09	4.8	11.7	11.5
Selected SG&A items:					
Salary & personnel expenses	24.24	23.98	-1.0	5.4	5.0

(continued)

Exhibit 4 (Continued)

	2017 (Baht billions)	2018 (Baht billions)	YoY%	Percent of Sales 2017	Percent of Sales 2018
Freight costs	11.63	11.55	-0.6	2.6	2.4
Research & development	4.18	4.67	11.9	0.9	1.0
Promotion and advertising	2.62	2.58	-1.3	0.6	0.5
Operating income	62.4	51.7	-17.1	13.8	10.8

Note: "YoY%" means year-over-year percentage change.
Sources: Based on information in Siam Cement Group's annual reports.

As shown in the exhibit, Siam Cement was affected in 2018 by higher input costs that could not be fully passed on to customers. Consequently, despite reporting sales growth of 6.1%, gross profit fell by 6.5% and gross margin declined. The company was able to limit its other operating costs; SG&A expenses grew 4.8%, declining slightly as a percentage of revenue. Operating income fell 17.1%. This contrasted with the company's experience in 2016, when lower input costs resulted in widening of the gross margin to 24.7% from 22.3% in 2015 (not shown in the exhibit above).

Siam Cement's income statement illustrates that companies often disclose the different components of SG&A expenses. Siam Cement, for example, shows separate line items for a number of distribution cost items, such as freight and rental expenses, and a number of general and administrative expenses, such as depreciation, IT fees, professional fees, and research and development. Although SG&A expenses overall are generally less closely linked to revenue than COGS, certain expenses within SG&A are more variable than others. Specifically, selling and distribution expenses often have a large variable component and can be estimated, like COGS, as a percentage of sales. The largest component of selling expenses is often wages and salaries linked to sales. Therefore, selling expenses will usually increase with additional sales people and/or an overall increase in wages and benefits for the sales force.

Other general and administrative expenses are less variable. Overhead costs for employees, for example, are more related to the number of employees at the head office and supporting IT and administrative operations than to short-term changes in the level of sales. Research and development expense is another example of an expense that tends to fluctuate less than sales. Consequently, these expenses are more fixed in nature and tend to increase and decrease gradually over time than do corresponding changes in the company's revenue.

In addition to analyzing the historical relationship between a company's operating expenses and sales, benchmarking a company against its competitors can also be useful. By analyzing the cost structure of a company's competitors, the efficiency potential and margin potential of a specific company can be estimated. As a final measure, performing certain cross checks within a forecast model can be useful too. For example, in the supermarket sector the projected floor square footage (or metric equivalent) underlying the revenue projections should match with the floor space projections underlying the unit selling expense forecasts. Both sales and expense projections may be enhanced if the company provides a breakout of the product and/or geographical segments in the footnotes of the annual report.

EXAMPLE 4

L'Oréal's Operational Cost Structure vs. Competitors

As shown in Exhibit 5, L'Oréal reported an operating margin of 17.9% in 2018, which makes it the most profitable company among a selection of beauty companies. However, the average operating margin of 19.4% for home and personal goods companies operating in mass markets is even greater than that of L'Oréal. Luxury goods companies tend to have higher gross margin than mass market companies, but it is offset by high "go to market" costs for advertising and promotion (A&P) expenditures. With the exception of Avon, the business model of which is based on direct selling, A&P is substantially greater at the beauty companies than at the mass market producers.

L'Oréal is often considered to be a pure beauty company. But if the underlying business is considered in detail, the company's operations can be split 50/50 between a luxury beauty high-end part and a general consumer part. In the general consumer part, L'Oréal's products compete with such players as Colgate, Procter & Gamble, and Henkel in the mass market. Exhibit 5 presents relevant data.

Exhibit 5 European and US Home and Personal Care Companies, Beauty vs. Mass Market Companies: Simplified and Common Size Income Statement (2018)

Company	Sales (in millions for currencies given)	Sales	COGS	Gross Margin	A&P	SG&A/ Other	EBIT
Beauty							
L'Oréal	€26,937	100%	27.2%	72.8%	30.2%	24.7%	17.9%
Estée Lauder	$13,683	100	20.8	79.2	25.6	38.6	15.0
Beiersdorf	€7,233	100	42.5	57.5	21.2	21.0	15.3
Avon	$5,571	100	42.4	57.6	2.3	51.1	4.2
Average beauty		100%	33.2%	66.8%	19.8%	33.9%	13.1%
Mass market							
Colgate	$15,544	100%	40.6%	59.4%	10.2%	25.4%	23.8%
Reckitt Group	£12,597	100	39.4	60.6	12	24.4	24.2
Procter & Gamble	$66,832	100	51.3	48.7	10.6	17.6	20.5
Clorox	$6,124	100	56.3	43.7	9.3	15.9	18.5
Kimberly-Clark	$18,486	100	69.7	30.3	3.5	14.7	12.1
Henkel	€19,899	100	53.5	46.5	22.7	6.2	17.6
Average mass market		100%	51.8%	48.2%	9.4%	19.4%	19.4%

Notes: The COGS and SG&A expense percentages for some of the companies listed in Exhibit 5 have been adjusted to reflect differences in accounting choices. For example, some of the consumer product companies include shipping and handling expenses in cost of sales, whereas others include these costs as a component of SG&A expenses.
Sources: Based on information in company reports.

1 Assuming the following information, what will L'Oréal's new operating margin be?

- L'Oréal's beauty and mass market operations each represent half of revenues.
- L'Oréal will be able to bring the overall cost structure of its mass market operations in line with the average of mass market companies (EBIT = 19.4%).
- The cost structure of L'Oréal's beauty operations will remain stable (assumed EBIT = 17.9%).

2 What will happen to L'Oréal's operating margin if the company is able to adjust the operating cost structure of its mass market segment (50% of revenues) partly toward the average of its mass market peers but keep its high gross margin? Assume the following:

- The cost structure of half of the business, the beauty operations, will remain stable (EBIT = 17.9%).
- L'Oréal's mass market operations will have a gross margin of 60.3% (the average of the current gross margin of 72.8% and the 48.2% reported by its mass market peers).
- L'Oréal's A&P costs will fall by half from 30.2% of sales to 15.1% of sales, and other costs will remain stable.

Solution to 1:

Operating margin will increase from 17.9% to 18.7%, which is 50% of 19.4% (mass market EBIT) plus 50% of 17.9% (L'Oréal EBIT).

Solution to 2:

As shown in Exhibit 6, operating margin will increase from 17.9% to 19.3%. The operating margin of the mass market operations will improve by 170 bps to 20.7% because a 1,230 bps decline in gross margin (from 72.8% to 60.5%) will be more than offset by the 1,510 bps of decline in A&P expenditures (from 30.2% of sales to 15.1% of sales). The average of the EBIT for beauty (17.9%) and the new EBIT for mass market operations (20.7%) is 19.3%.

Exhibit 6 EBIT for L'Oréal Divisions

	L'Oréal	Beauty 50%	Mass Market 50%	Average
Sales	100%	100%	100%	100%
Cost of goods sold	27.2%	27.2%	40.7%	33.4%
Gross margin	72.8%	72.8%	60.5%	66.7%
A&P	30.2%	30.2%	15.1%	22.7%
SG&A/Other	24.7%	24.7%	24.7%	24.7%
EBIT	17.9%	17.9%	20.7%	19.3%

EXAMPLE 5

Analysis of the Consumer Goods Company Unilever

The consumer goods company Unilever reported an overall underlying operating margin of 18.4% in 2018. As shown in Exhibit 7, the operating margin is lowest in the fastest growing product category, home care products. The other parts of the business, personal care and foods categories, enjoy higher margins but are growing more slowly.

Exhibit 7 Unilever Revenue and Profit from Product Categories (€ millions, unless noted)

	2017	2018	18/17 YoY	Avg 2016–2018
Personal Care	20,697	20,624	3.1%	3.0%
Foods	22,444	20,227	2.0	2.4
Home Care	10,574	10,131	4.2	4.3
Total revenues	53,715	50,982	2.9	3.0
Underlying operating profit				
Personal Care	4,375	4,508	3.04%	
Foods	3,737	3,534	−5.4	
Home Care	1,288	1,317	2.25	
Total underlying operating profit	9,400	9,359	−0.4	
Underlying operating profit margin				
Personal Care	121.1%	21.9%		
Foods	16.7	17.5		
Home Care	12.2	13.0		
Total underlying operating profit margin	17.5%	18.4%		

Notes: USG is “underlying sales growth” (i.e., sales growth adjusted for currency, disposals, and acquisitions). USG is the organic sales growth based on volume, price, and mix changes. Underlying profit is operating profit adjusted for exceptional items, like restructuring costs

Source: Based on Unilever’s 2018 full year and fourth quarter results.

1 Determine the estimated sales, operating profit, and operating profit margin by using the following two approaches: (A) Assume consolidated sales growth of 3.0% (the average of the years 2017 and 2018) and overall stable operating margin of 18.4% for the next five years; and (B) assume each individual product’s sales growth and operating margin continue at the same rate reported in 2018. Which approach will result in a higher estimated operating profit after five years?

2 Compare and explain the results under the two alternative approaches described in Question 1 (A and B) with reference to the yearly growth rate in estimated total sales, the yearly growth rate in total operating profit, and the yearly profit margin.

3 Assume Unilever is able to grow revenues the next five years in each category in line with 2018 (Beauty & Personal Care 3.1%; Foods & Refreshments 2.0%; and Home Care 4.2%). But operating profit margins in Beauty and Personal Care will fall 20 bps annually for the next five years (as a result of high competition, limited growth, and costs resulting from the adoption of sustainable packaging) and operating profit margins in Foods & Refreshments and Home Care segments will increase by 15 and 50 bps, respectively, each year for the next five years (helped by increasing demand for the company's products and better utilization of its factories). Using approach (B), calculate the overall operating profit margin.

Solution to 1:

Exhibit 8 shows that operating profit after five years will be €10,850 million under the first approach (A) and €10,771 million under the second approach (B). The results of the calculation are shown in Exhibit 8.

Exhibit 8 Sales and Operating Profit for Unilever, 2018–2023E (€ millions, unless noted)

Approach A	2018	2019E	2020E	2021E	2022E	2023E
Sales	50,982	52,460	53,982	55,547	57,158	58,816
YoY %		3.0%	3.0%	3.0%	3.0%	3.0%
Underlying operating profit	9,359	9,640	9,929	10,227	10,534	10,850
YoY %		3.0%	3.0%	3.0%	3.0%	3.0%
Underlying operating profit margin	18.4%	18.4%	18.4%	18.4%	18.4%	18.4%

Approach B	2018	2019E	2020E	2021E	2022E	2023E
Beauty & Personal Care	20,624	21,263	21,923	22,602	23,303	24,025
Foods & Refreshments	20,227	20,632	21,044	21,465	21,894	22,332
Home Care	10,131	10,557	11,000	11,462	11,943	12,445
Total revenues	50,982	52,451	53,967	55,529	57,140	58,802
Beauty & Personal Care		3.1%	3.1%	3.1%	3.1%	3.1%
Foods & Refreshments		2.0	2.0	2.0	2.0	2.0
Home Care		4.2	4.2	4.2	4.2	4.2
YoY revenue		2.88%	2.89%	2.90%	2.90%	2.91%
Beauty & Personal Care	4,508	4,648	4,792	4,940	5,094	5,251
Foods & Refreshments	3,534	3,605	3,677	3,750	3,825	3,902
Home Care	1,317	1,372	1,430	1,490	1,553	1,618
Total underlying operating profit	9,359	9,625	9,899	10,181	10,471	10,771

Exhibit 8 (Continued)

Approach B	2018	2019E	2020E	2021E	2022E	2023E
Beauty & Personal Care		3.1%	3.1%	3.1%	3.1%	3.1%
Foods & Refreshments		2.0	2.0	2.0	2.0	2.0
Home Care		4.2	4.2	4.2	4.2	4.2
YoY change in underlying profit		2.84%	2.84%	2.85%	2.86%	2.86%
Beauty & Personal Care		21.9%	21.9%	21.9%	21.9%	21.9%
Foods & Refreshments		17.5	17.5	17.5	17.5	17.5
Home Care		13.0	13.0	13.0	13.0	13.0
Total underlying operating margin		18.35%	18.34%	18.33%	18.33%	18.32%

Solution to 2:

Under the first approach (A) a constant 3.0% sales growth rate and a stable 18.4% operating margin are assumed. As a consequence, the operating profit growth rate is in line with the revenue growth rate and constant at 3.0% (see Exhibit 9, Panel A). Under the second approach (see Exhibit 9, Panel B), the high sales growth of 4.2% occurs in the segment with the least amount of sales, Home Care. Because the operating margin in the fastest growing segment is less than the overall average (13.0% vs. 18.4% group average), operating margin for the group falls slightly from 18.4% in 2018 to 18.3% in 2023. This comparison illustrates that the higher sales growth in a lower margin segment puts the company's operating margin under structural pressure. The rate of operating profit growth continues to be slightly less than the rate of sales growth (2.84% vs. 2.88%). Exhibit 9 shows a summary of the sales, operating profit, sales growth, and margin growth.

Exhibit 9 Summary of Results in Exhibit 8

	2018	2019E	2020E	2021E	2022E	2023E
A. Sales and operating profit (€ millions)						
Sales A	50,982	52,511	54,087	55,709	57,381	59,102
Sales B	50,982	52,451	53,967	55,529	57,140	58,802
Sales A–B		60	120	180	240	300
Sales growth A		3.0%	3.0%	3.0%	3.0%	3.0%
Sales growth B		2.88%	2.89%	2.90%	2.90%	2.91%
Operating profit A	9,359	9,640	9,929	10,227	10,534	10,850
Operating profit B	9,359	9,625	9,899	10,181	10,471	10,771
Operating profit A–B		15	30	46	62	79
B. Growth Rates						
Operating profit growth A		3.00%	3.00%	3.00%	3.00%	3.00%

(continued)

Exhibit 9 (Continued)

	2018	2019E	2020E	2021E	2022E	2023E
Operating profit growth B		2.84%	2.84%	2.85%	2.86%	2.86%
Operating profit margin A		18.4%	18.4%	18.4%	18.4%	18.4%
Operating profit margin B		18.35%	18.34%	18.33%	18.33%	18.32%

Solution to 3

As shown in Exhibit 10, the overall underlying operating profit margin improves from 18.36% in 2018 to 18.75% in 2023 because the margin decline in Beauty & Personal Care is more than offset by the margin increase in Foods & Refreshments and the faster growing Home Care segments.

Exhibit 10 Sales and Operating Profit for Unilever 2018–2023E (€ millions, unless noted)

	2018	2019E	2020E	2021E	2022E	2023E
Beauty & Personal Care	20,624	21,263	21,923	22,602	23,303	24,025
Foods & Refreshments	20,227	20,632	21,044	21,465	21,894	22,332
Home Care	10,131	10,557	11,000	11,462	11,943	12,445
Total revenue	50,982	52,451	53,967	55,529	57,140	58,802
Beauty & Personal Care		3.1%	3.1%	3.1%	3.1%	3.1%
Foods & Refreshments		2.0%	2.0%	2.0%	2.0%	2.0%
Home Care		4.2%	4.2%	4.2%	4.2%	4.2%
YoY revenue growth		2.9%	2.9%	2.9%	2.9%	2.9%
Beauty & Personal Care	4,508	4,614	4,713	4,814	4,917	5,021
Foods & Refreshments	3,534	3,641	3,746	3,853	3,963	4,076
Home Care	1,317	1,425	1,540	1,662	1,791	1,929
Total underlying operating profit	9,359	9,681	9,999	10,329	10,671	11,026
Beauty & Personal Care		2.35%	2.15%	2.14%	2.13%	2.12%
Foods & Refreshments		3.04%	2.87%	2.86%	2.86%	2.85%
Home Care		8.212%	8.06%	7.92%	7.79%	7.67%
YoY underlying profit growth		3.44%	3297%	3.3%	3.31%	3.33%
Beauty & Personal Care		21.7%	21.5%	21.3%	21.1%	20.9%
Foods & Refreshments		17.65%	17.8%	17.95%	18.10%	18.25%

Exhibit 10 (Continued)

	2018	2019E	2020E	2021E	2022E	2023E
Home Care		13.5%	14%	14.5%	15%	15.5%
Total underlying operating margin		18.46%	18.53%	18.60%	18.68%	18.75%

MODELING NON-OPERATING COSTS AND OTHER ITEMS

4

d demonstrate methods to forecast the following costs: cost of goods sold, selling general and administrative costs, financing costs, and income taxes;

Line items on the income statement that appear below operating profit also need to be modeled. Some of the most important items included here are interest income, interest expense, taxes, minority interest, and income from affiliates, share count, and unusual charges.

Interest income depends on the amount of cash and investments on the balance sheet as well as the rates of return earned on investments. Interest income is a key component of revenue for banks and insurance companies, but it is relatively less significant to most non-financial companies. Interest expense depends on the level of debt on the balance sheet as well as the interest rate associated with the debt. Analysts should be aware of the effect of changing interest rates on the market value of company's debt and interest expense in the future.

Taxes are primarily determined by jurisdictional regulations but can also be influenced by the nature of a business. Some companies benefit from special tax treatment—for example, from research and development tax credits or accelerated depreciation of fixed assets. Analysts should be aware of any differences between taxes reported on the income statement and cash taxes, which can result in deferred tax assets or liabilities. Analysts should also be aware of any governmental or business changes that can alter tax rates.

The two most significant non-operating expenses in income statement modeling are financing expenses (i.e., interest) and taxes.

4.1 Financing Expenses

When forecasting financing expenses, the capital structure of a company is a key determinant. For practical purposes, the debt level in combination with the interest rate are the main drivers in forecasting debt financing expenses. Usually the notes to the financial statements provide detail about the maturity structure of the company's debt and the corresponding interest rates. This information can be used to estimate future financing expenses.

EXAMPLE 6

Interest Expense Calculations

Dutch grocer Ahold Delhaize, operating in a number of regions, has a debt structure with a relatively high amount of cash on its balance sheet.

Exhibit 11 Ahold's Debt, Interest Income, and Expense

(€ millions)	31 Dec. 2017	30 Dec. 2018	Average
Loans	3,289	3,683	3,486
Other non-current financial liabilities (includes finance leases and cumulative preferred financing shares)	2,098	2,055	2,077
Current financial liabilities	2,210	1,232	1,721
Gross debt	7,597	6,970	7,284
Less: cash, cash equivalents, and short-term financial assets	3,360	5,042	4,201
Net debt	4,237	1,928	3,083
Interest income for 2018			70
Interest expense for 2018			310
Net interest expense			240

Source: Based on information in Ahold Delhaize 2018 annual report.

1 Calculate the interest rate on the average gross debt and interest rate on the average cash position.

2 Calculate the interest rate on the average net debt, assuming the other financial income and expenses are not related to the debt or cash balances.

Solution to 1:

Interest rate on average gross debt is calculated as interest expense divided by average gross debt: (€310 million/€7,284 million) = 4.26% or 4.3%. The interest rate on average cash position is interest income divided by the average cash position (€70 million/€4,201 million) = 1.67%.

Solution to 2:

The interest rate on the average net debt is calculated as net interest expense divided by average net debt (€240 million/€3,083 million) = 7.8%.

4.2 Corporate Income Tax

The final large non-operating item is the tax expense. This is often a large amount that affects profit substantially. Differences in tax rates can be an important driver of value. Generally, there are three types of tax rates:

- The statutory tax rate, which is the tax rate applying to what is considered to be a company's domestic tax base.
- The effective tax rate, which is calculated as the reported tax amount on the income statement divided by the pre-tax income.
- The cash tax rate, which is the tax actually paid (cash tax) divided by pre-tax income.

Differences between cash taxes and reported taxes typically result from timing differences between accounting and tax calculations and are reflected as a deferred tax asset or a deferred tax liability.

In forecasting tax expense and cash taxes, respectively, the effective tax rate and cash tax rate are key. A good understanding of their operational drivers and the financial structure of a company is useful in forecasting these tax rates.

Differences between the statutory tax rate and the effective tax rate can arise for many reasons. Tax credits, withholding tax on dividends, adjustments to previous years, and expenses not deductible for tax purposes are among the reasons for differences. Effective tax rates can differ when companies are active outside the country in which they are domiciled. The effective tax rate becomes a blend of the different tax rates of the countries in which the activities take place in relation to the profit generated in each country. If a company reports a high profit in a country with a high tax rate and a low profit in a country with a low tax rate, the effective tax rate will be the weighted average of the rates and higher than the simple average tax rate of both countries. In some cases, companies have also been able to minimize their taxes by using special purposes entities. For example, some companies create specialized financing and holding companies to minimize the amount of taxable profit reported in high tax rate countries. Although such actions could reduce the effective tax rate substantially, they also create risks if, for example, tax laws change. In general, an effective tax rate that is consistently lower than statutory rates or the effective tax rates reported by competitors may warrant additional attention when forecasting future tax expenses. The notes on the financial statements should disclose other types of items, some of which could contribute to a temporarily high or low effective tax rate. The cash tax rate is used for forecasting cash flows, and the effective tax rate is relevant for projecting earnings on the income statement. In developing an estimated tax rate for forecasts, analysts should adjust for any one-time events. If the income from equity method investees is a substantial part of pre-tax income and also a volatile component of it, the effective tax rate excluding this amount is likely to be a better estimate for the future tax costs for a company. The tax impact from income from participations is disclosed in the notes on the financial statements.

Often, a good starting point for estimating future tax expense is a tax rate based on normalized operating income, before the results from associates and special items. This normalized tax rate should be a good indication of the future tax expense, adjusted for special items, in an analyst's earnings model.

By building a model, the effective tax amount can be found in the profit and loss projections and the cash tax amount on the cash flow statement (or given as supplemental information). The reconciliation between the profit and loss tax amount and the cash flow tax figures should be the change in the deferred tax asset or liability.

EXAMPLE 7

Tax Rate Estimates

ABC, a hypothetical company, operates in Countries A and B. The tax rate in Country A is 40%, and the tax rate in Country B is 10%. In the first year, the company generates an equal amount of profit before tax in each country.

Exhibit 12 Tax Rates That Differ by Jurisdiction

	A	B	Total
Profit before tax	100	100	200
Effective tax rate	40%	10%	25%
Tax	40	10	50
Net profit	60	90	150

1. What will happen to the effective tax rate for the next three years if the profit in Country A is stable but the profit in Country B grows 15% annually?
2. Evaluate the cash tax and effective tax rates for the next three years if the tax authorities in Country A allow some costs (e.g., accelerated depreciation) to be taken sooner for tax purposes. For Country A, the result will be a 50% reduction in taxes paid in the current year but an increase in taxes paid by the same amount in the following year (this happens each year). Assume stable profit before tax in Country A and 15% annual before-tax-profit growth in Country B.
3. Repeat the exercise of Problem 2, but now assume that it is Country B not Country A that allows some costs to be taken sooner for tax purposes and that the tax effect described applies to Country B. Continue to assume stable profit before tax at Country A and 15% annual profit growth in Country B.

Solution to 1:

The effective tax rate will gradually decline because a higher proportion of profit will be generated in the country with the lower tax rate. In Exhibit 13, the effective tax rate declines from 25% in the beginning to 21.9% in the third year.

Exhibit 13 Worksheet for Problem 1

	Year			
	0	1	2	3
Profit before tax, Country A	100.0	100.0	100.0	100.0
Profit before tax, Country B	100.0	115.0	132.3	152.1
Total profit before tax	200.0	215.0	232.3	252.1
Tax, Country A (40%)	40.0	40.0	40.0	40.0
Tax, Country B (10%)	10.0	11.5	13.2	15.2
Total tax	50.0	51.5	53.2	55.2

Exhibit 13 (Continued)

	Year			
	0	1	2	3
Total tax rate %	25.0%	24.0%	22.9%	21.9%
Net profit	150.0	163.5	179.1	196.9

Solution to 2:

The combined cash tax rate (next to last line in Exhibit 14) will be 15% in the first year and then rebound in subsequent years. Only the rate for the first year will benefit from a tax deferral; in subsequent years, the deferral for a given year will be offset by the addition of the amount postponed from the previous year. The combined effective tax rate (last line in Exhibit 14) will be unaffected by the deferral. As shown in Exhibit 14, beginning with the second year, the combined cash tax and effective tax rates decline over time but remain identical to each other.

Exhibit 14 Worksheet for Problem 2

	Year			
	0	1	2	3
Profit before tax	200.0	215.0	232.3	252.1
Tax per income statement	50.0	51.5	53.2	55.2
Tax payment, Country A	20.0	20.0	20.0	20.0
Postponed tax payment, Country A		20.0	20.0	20.0
Tax payment Country B	10.0	11.5	13.2	15.2
Total tax payment	30.0	51.5	53.2	55.2
Cash tax rate	15.0%	24.0%	22.9%	21.9%
Tax rate per income statement	25.0%	24.0%	22.9%	21.9%

Solution to 3:

The combined effective tax rate (last line in Exhibit 15) remains unchanged from Exhibits 13 and 14. Because of the growth assumed for Country B, however, the annual tax postponement will result in a lower cash tax rate in Country B than the effective tax rate in Country B. Consequently, as shown in Exhibit 15, the combined cash tax rate will be less than the effective tax rate.

Exhibit 15 Worksheet for Problem 3

	Year			
	0	1	2	3
Profit before tax	200.0	215.0	232.3	252.1
Tax per income statements	50.0	51.5	53.2	55.2
Tax payment, Country A	40.0	40.0	40.0	40.0
Tax payment, Country B	5.0	5.8	6.6	7.6
Postponed tax payment, Country B		5.0	5.8	6.6
Total tax payment	45.0	50.8	52.4	54.2
Cash tax rate	22.5%	23.6%	22.5%	21.5%
Tax rate per income statement	25.0%	24.0%	22.9%	21.9%

The next section addresses several points to note in modeling dividends, share count, and unusual expenses.

4.3 Income Statement Modeling: Other Items

A company's stated dividend policy helps in modeling future dividend growth. Analysts will often assume that dividends grow each year by a certain dollar amount or as a proportion of net income.

If a company shares an ownership interest in a business unit with a third party, the company may report minority interest expense or income from consolidated affiliates on its income statement. If a company owns more than 50% of an affiliate, it will generally consolidate the affiliate's results with its own and report the portion of income that does not belong to the parent company as minority interest. If a company owns less than 50% of an affiliate, it will not consolidate results but will report its share of income from the affiliate under the equity method. If the affiliate is profitable, minority interest would be a deduction from net income, whereas if a consolidated affiliate generates losses, minority interest would be an addition to net income. In either case, income or expense from these jointly owned businesses can be material.

Share count (shares issued and outstanding) is a key input in the calculation of an intrinsic value estimate and earnings per share. Share count changes for three primary reasons: dilution related to stock options, convertible bonds, and similar securities; issuance of new shares; and share repurchases. The market price of a stock is an important determinant of future share count changes, which can complicate their estimation. Projections for share issuance and repurchases should fit within the analyst's broader analysis of a company's capital structure.

Finally, unusual charges can be almost impossible to predict, particularly past the next couple of years. For this reason, analysts typically exclude unusual charges from their forecasts. But if a company has a habit of frequently classifying certain recurring costs as "unusual," analysts should consider some normalized level of charges in their valuation model.

BALANCE SHEET AND CASH FLOW STATEMENT MODELING

5

e describe approaches to balance sheet modeling;

f describe the relationship between return on invested capital and competitive advantage;

Income statement modeling is the starting point for balance sheet and cash flow statement modeling. Analysts normally have a choice of whether to focus on the balance sheet or cash flow statement; the third financial statement will naturally result from the construction of the other two. Here, we focus on the balance sheet.

Some balance sheet line items—such as retained earnings—flow directly from the income statement, whereas other lines—such as accounts receivable, accounts payable, and inventory—are very closely linked to income statement projections.

A common way to model working capital accounts is through the use of efficiency ratios. For example, analysts may project future accounts receivable by assuming a number of days sales outstanding and combining that assumption with a sales projection. Days sales outstanding is a measure of the number of days, on average, it takes a company to collect revenue from its customers. For example, if annual revenue (assumed to be all credit sales) is \$25 billion and it normally takes 60 days to collect revenue from customers, accounts receivable would be estimated at \$4.1 billion (≈ \$25 billion × 60/365). Analysts can project future inventory by assuming an inventory turnover rate and combining that assumption with a cost of goods sold projection. Inventory turnover is a measure of how much inventory a company keeps on hand, or alternatively, how quickly a company sells through its inventory. In general, if efficiency ratios are held constant, working capital accounts will grow in line with the related income statement accounts.

Working capital projections can be modified by both top-down and bottom-up considerations. In the absence of a specific opinion about working capital, analysts can look at historical efficiency ratios and project recent performance or a historical average to persist in the future, which would be a bottom-up approach. Conversely, analysts may have a specific view of future working capital. For example, if they project economy-wide retail sales to decline unexpectedly, that could result in slower inventory turnover across the retail sector. Because the analysts began with a forecast for a large sector of the economy, this would be considered a top-down approach.

Projections for long-term assets—such as property, plant, and equipment (PP&E)—are less directly tied to the income statement for most companies. Net PP&E primarily changes as a result of capital expenditures and depreciation, both of which are important components of the cash flow statement. Depreciation forecasts are usually based on historical depreciation and disclosure about depreciation schedules, whereas capital expenditure forecasts depend on the analysts' judgment of the future need for new PP&E. Capital expenditures can be thought of as including both **maintenance capital expenditures**, which are necessary to sustain the current business, and **growth capital expenditures**, which are needed to expand the business. All else being equal, maintenance capital expenditure forecasts should normally be higher than depreciation because of inflation.

Finally, analysts must make assumptions about a company's future capital structure. Leverage ratios—such as debt-to-capital, debt-to-equity, and debt-to-EBITDA—can be useful for projecting future debt and equity levels. Analysts should consider historical company practice, management's financial strategy, and the capital requirements implied by other model assumptions when projecting the future capital structure.

Once future income statements and balance sheets are constructed, analysts can use them to determine the rate of **return on invested capital** (ROIC) implied by their assumptions. ROIC measures the profitability of the capital invested by the company's shareholders and debtholders. The numerator for ROIC is usually net operating profit less adjusted taxes (NOPLAT). NOPLAT is basically earnings before interest expense (i.e., earnings available to provide a return to both equityholders and debtholders). The denominator for ROIC is invested capital, which is calculated as operating assets less operating liabilities.(Note that this is just one way of calculating ROIC; there are many variations in the definition of ROIC.) Invested capital can be measured at the beginning of an accounting period or as an average of the beginning and end of the accounting period. ROIC is a better measure of profitability than return on equity because it is not affected by a company's degree of financial leverage. In general, sustainably high ROIC is a sign of a competitive advantage. To increase ROIC, a company must either increase earnings, reduce invested capital, or both. A closely related measure to ROIC, but focusing on pretax operating profit, is **return on capital employed** (ROCE), which is essentially ROIC before tax. This measure is defined as operating profit divided by capital employed (debt and equity capital). As a pretax measure, ROCE can be useful in several contexts, such as peer comparisons of companies in countries with different tax structures, because comparison of underlying profitability would not be biased in favor of companies benefitting from low tax rate regimes. (Note that as in the case of ROIC, the definition of ROCE provided here follows one line of analyst practice; there is no single "authoritative" definition.)

EXAMPLE 8

Balance Sheet Modeling

1 Management at a restaurant chain intends to maintain a 40% debt-to-capital ratio. Management has a track record of meeting its capital structure targets. The restaurant chain is solidly profitable, but earnings are expected to decline 2% annually over the next five years because of increasing competitive pressure. The company does not pay a dividend or repurchase shares, and all earnings are expected to be retained for the next five years. What is *most likely* to happen to the restaurant chain's total debt over this period?

   A Total debt will increase.

   B Total debt will decrease.

   C Total debt will remain the same.

2 Sophie Moreau, a buy-side analyst, is analyzing a French manufacturing company. Working capital and PP&E account for almost all of the company's assets. Moreau believes that the depreciation schedule used by the company is not reflective of economic reality. Rather, she expects PP&E to last twice as long as what is implied by the depreciation schedule, and as such, she projects capital expenditures to be significantly less than depreciation for the next five years. Moreau projects that both earnings and net working capital will grow at a low single-digit rate during this time. What do Moreau's assumptions *most likely* imply for returns on invested capital during the next five years?

   A ROIC will increase.

   B ROIC will decrease.

   C ROIC will stay the same.

Solution to 1:

A is correct. The restaurant chain is profitable and retains all of its earnings. These facts will lead to rising equity on the balance sheet. To maintain a constant debt-to-capital ratio, management will have to increase its debt.

Solution to 2:

A is correct. Earnings are expected to grow over the next five years. Working capital is expected to grow in line with earnings, which would imply a stable ROIC. But net PP&E is expected to decline because depreciation is expected to exceed capital expenditures. Total invested capital will thus grow more slowly than earnings or even shrink, implying improving returns on invested capital.

Once projected income statements and balance sheets have been constructed, future cash flow statements can be projected. Analysts will normally make assumptions about how a company will use its future cash flows—whether for share repurchases, dividends, additional capital expenditures, acquisitions, and so on.

6 SCENARIO ANALYSIS AND SENSITIVITY ANALYSIS

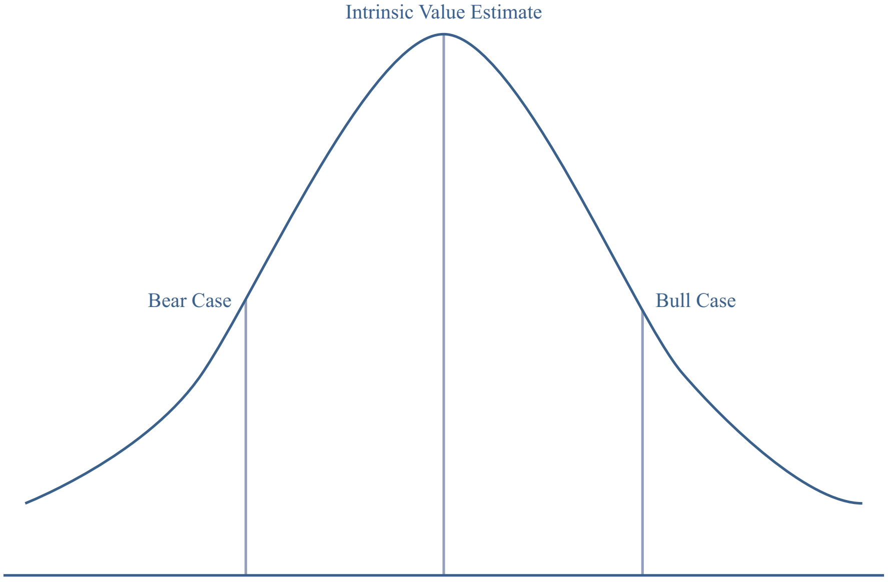
Regardless of the approach used, equity valuation involves a significant degree of uncertainty. Valuing businesses requires making assumptions about the future, which is inherently uncertain. Recognizing this uncertainty, effective analysis should always consider scenarios in addition to the most likely "base case" result.

Sensitivity analysis involves changing one assumption at a time to see the effect on the estimate of intrinsic value. For example, analysts might examine the impact of a different revenue growth rate on the company's valuation. **Scenario analysis** has the same goal but involves changing multiple assumptions at the same time. For example, analysts might simultaneously change assumptions for revenue growth, operating margin, and capital investment. Either sensitivity analysis or scenario analysis can be used to determine a range of potential intrinsic value estimates based on a variety of different assumptions about the future. Analysts can use either tool to estimate the effect on a company's valuation of different assumptions for economic growth, for inflation, for the success of a particular product, and so on.

Value estimates for companies involve varying degrees of uncertainty. Large, mature, slow growing, non-cyclical businesses with well-capitalized balance sheets may be relatively easy to value. In this case, intrinsic value estimates from upside and downside scenarios may be close to the base case. Conversely, for new ventures, companies exposed to technological or regulatory change, companies with significant operating or financial leverage, and so on, the range of potential intrinsic value estimates could be much wider. In this case, analysts may hesitate to make an investment recommendation without substantial confidence that intrinsic value differs markedly from market price.

Analysts should normally think of their valuations as a range of possibilities rather than an estimate at a single point. For most companies, the range will be approximately symmetrical and might be imagined as a bell curve. The base case estimate of intrinsic value would be at the middle of the distribution with, depending on the judgment of the analysts, similar probabilities of upside and downside outcomes, as shown in Exhibit 16. The width of the tails will depend on the level of uncertainty pertaining to forecasts. For example, the large, mature, slow-growing company would have a steep distribution with relatively thin tails, representing a relatively low likelihood of extreme values. The probability distribution of estimates need not be symmetrical. In addition, technically, a distribution of security values would be bounded on the left by zero.

Exhibit 16 Distribution of Intrinsic Value Estimates



For some companies, the range of potential outcomes might be skewed or highly irregular. For example, there might be a relatively high probability that a debt-laden company will have insufficient cash flow to cover interest and principal payments, in which case the value of the equity could be zero. A company with a single untested product may be worth very little or a lot, depending on whether the product turns out to be a success. Scenario analysis can be particularly useful in these cases because there may not be a meaningful "base case." A common approach is for analysts to value such companies by using a probability-weighted average of the various scenarios.

7 THE IMPACT OF COMPETITIVE FACTORS IN PRICES AND COSTS

g explain how competitive factors affect prices and costs;

h evaluate the competitive position of a company based on a Porter's five forces analysis;

Incorporating competition into financial forecasting can be a challenging task. Most of the items that analysts must forecast—including revenues, profit margin, and capital expenditures—are linked to the competitive environment, and competition can affect these items both separately and collectively. Analysts can use various conceptual tools in thinking about how competition will affect forecasts. Such tools provide a way to organize data and ideas. Although there are no "rules" for incorporating competitive analysis in forecasts, it is arguably the analysts' most important job.

Analysts' projections for revenue growth, margin development, capital expenditures, and working capital investment are all based on an estimate of a company's future competitive strength.

One of the tools that analysts can use to think about how competition will affect financial results is Michael Porter's widely used "five forces" framework (see Porter 1980). The framework identifies five forces that affect the intensity of a company's competitive environment and thus cost and price projections. These forces include the following: threat of substitute products, intensity of rivalry among incumbent companies, bargaining power of suppliers, bargaining power of customers, and threat of new entrants.

The first force is the threat of substitute products. If numerous substitutes exist and switching costs are low, companies have limited pricing power. Conversely, if few substitutes exist and/or switching costs are high, companies have greater pricing power.

The second force is the intensity of rivalry among incumbent companies. Pricing power is limited in industries that are fragmented, have limited growth, high exit barriers, high fixed costs, and have basically identical product offerings.

The third force is the bargaining power of suppliers. Companies (and overall industries) whose suppliers have greater ability to increase prices and/or limit the quality and quantity of inputs face downward pressure on profitability. Suppliers' bargaining power is generally a function of relative size, the relative importance the supplier places on a particular product, and the availability of alternatives.

The fourth force is the bargaining power of customers. Companies (and overall industries) whose customers have greater ability to demand lower prices and/or control the quality and quantity of end products face downward pressure on profitability. Buyer power is the reverse of supplier power. Bargaining power of customers is generally lower in markets with a fragmented customer base, a non-standardized product, and high switching costs for the customer.

The fifth force is the threat of new entrants. Companies in industries in which the threat of new entrants is high because of the presence of above-market returns face downward pressure on profitability. In contrast, if there are barriers to entry, it may be costly for new competitors to enter a market. It is easier for incumbents to raise prices and defend their market position when barriers to entry are high.

ANALYSIS OF ANHEUSER-BUSCH INBEV USING PORTER'S FIVE FORCES

The competitive structure a company faces can vary among countries, with implications for modeling revenue growth, profit margins, capital expenditures, and return on investments. For example, Anheuser-Busch (AB) InBev, the largest global brewer, operates in many countries, two of which are Brazil, the world's third largest beer market, and the United Kingdom. AB InBev's competitive position and prospects in the highly consolidated and growing Brazilian market are much more favorable than in the fragmented and declining UK market.

The Brazilian beer market is divided among four players. AmBev (AB InBev's subsidiary in Brazil in which it owns a 61.9% stake) is the dominant brewer with an estimated 65% market share in 2018 versus 20% for Heineken and 12% for Petropolis, Brazil's largest privately owned brewing group. Helped by its dominant market position and strong distribution network, AmBev was able to report an EBITDA margin of nearly 50.4% in 2018 (ri.ambev.com.br), the highest in the global beer industry. The industry participants focus less on price competition and more on expanding distribution and "premiumization" (i.e., selling more expensive beers.) Although the 2015–2018 time period saw challenging trading conditions due to subdued consumer demand, causing years of decline in the market by volume, Brazil is still considered to be a promising market. In this environment, an analyst would likely forecast solid revenue growth for AmBev. Exhibit 17 presents an analysis of the Brazilian beer market using Porter's five forces framework. Most of the competitive forces represent a low threat to profitability (consistent with AmBev's historical profitability), implying that analysts would most likely forecast continued above-average profitability.

Exhibit 17 Analysis of the Brazilian Beer Market Using Porter's Five Forces

Force	Degree	Factors to Consider
Threat of substitutes	Medium	• Beer consumers do not easily shift to other beverages, but such alternatives as wine and spirits are available. • Unlike many other countries, the range of beers is relatively limited.
Rivalry	Low	• AmBev dominates the market with a 65% market share. Its economies of scale in production and distribution yield significant cost advantages relative to competition. • Price competition is limited because of AmBev's cost advantages and because of typically increasing beer volumes.
Bargaining power of suppliers	Low	• The primary inputs (water, hops, barley, and packaging) are basically commodities.
Bargaining power of buyers	Low	• Beer is mostly consumed in bars and restaurants. The owners of these outlets represent a large and highly fragmented group of beer buyers. • The supermarket industry in Brazil is relatively fragmented, and supermarkets are less likely to offer alternatives, such as private labels.
Threat of new entrants	Low	• New entrants face relatively high barriers to entry because of the high costs of building a brewery, establishing a national distribution network, and establishing a nationally known brand name.

The UK beer market is also divided among four players, but the competitive structure is totally different in the United Kingdom than in Brazil. The market is more fragmented with smaller market shares held by the largest players. Heineken, MolsonCoors, AB InBev, and Carlsberg had market shares of 24% (adbrands.net), 18%, 18% (www.ab-inbev.com), and 11% (carlsberggroup.com), respectively, in 2018. Consequently, the British market has no dominant brewer. Given the high fixed costs of a brewery, declining volumes of UK beer consumption, and the highly consolidated customer base, which provides the clients with substantial purchasing power (particularly in the retail channels), price competition is usually intense. A gradual switch from drinking beer in pubs and restaurants ("on-trade") to consumption at home ("off-trade") is making brewers even more exposed to the bargaining power of the dominant retail supermarket (grocers) chains. Increasing taxes on beer and rents faced by pub landlords add to the burden faced by the industry, leading to a steady decline of Britain's pub industry. Profitability has been lower than the beer industry's global average; operating margins are believed to be less than 10%. In this kind of environment, analysts would most likely forecast only very cautious revenue growth, if any. Exhibit 18 presents an analysis of the UK beer market using Porter's five forces framework.

Exhibit 18 Analysis of the UK Beer Market Using Porter's Five Forces

Force	Degree	Factors to Consider
Threat of substitutes	Medium	• Beer consumers do not easily shift to other beverages, but such alternatives as wine, spirits, and cider are available.
Rivalry	High	• The market is relatively fragmented with no dominant market leader and large numbers of small breweries. • Declining beer volumes make price wars more likely.[a] • Brand loyalty is less developed because of the extensive range of alternative beers.
Bargaining power of suppliers	Low	• The primary inputs (water, hops, barley, and packaging) are basically commodities.
Bargaining power of buyers	High	• The large supermarket chains that dominate the grocery sector have significant bargaining power. • Large pub chains in the "on-trade" business (where beer is sold in pubs and restaurants) also have strong bargaining power.
Threat of new entrants	Low	• Barriers to entry are relatively high because of the high costs of building a brewery, establishing a national distribution network (particularly given the history of brewers owning pubs and bars), and establishing a nationally known brand. • Because the United Kingdom consists of islands, companies with breweries in other countries face higher transportation costs than existing participants.

[a] In some declining markets, companies focus on increasing prices to offset declining volumes. But in the case of beer, where the market is very fragmented and thus there is no price leadership, price increases are less viable.

There is a distinction between Porter's five forces and other factors that can affect profitability, such as government regulation and taxes:

> Industry structure, as manifested in the strength of the five competitive forces, determines the industry's long-run profit potential because it determines how the economic value created by the industry is divided… Government is not best understood as a sixth force because government involvement is neither inherently good nor bad for industry profitability. The best way to understand the influence of government on competition is to analyze how specific government policies affect the five competitive forces (Porter 2008, page 10).

EXAMPLE 9

EuroAlco case

In 20X2, EuroAlco was the beer market leader in Eurolandia (a hypothetical country) with a 35% market share. The other large players held 15%, 15%, 10%, and 7% share, respectively. The Eurolandia market is considered a growth market. It historically had high overall alcohol consumption but a relatively low per capita consumption of beer, a product that is attracting interest from the growing, younger population and is further supported by increasing disposable incomes.

Two years earlier, at the start of year 20X1, the Eurolandia government, in its fight to curb alcohol consumption, tripled the excise duty (a special tax) on beer from €0.3 per liter to €0.9 and announced that excise duty will further increase by €0.1 per liter.

In the following year, 20X2, EuroAlco made efforts to strengthen the position of the more expensive brands in its portfolio. These efforts led to a 20% increase in selling costs. Similar to most consumer staple companies, EuroAlco experienced higher production costs. Poor grain harvests put price pressure on buyers of almost all feedstocks, and rising oil prices resulted in higher packaging costs. In 20X2, competing companies were much more cautious with advertising and promotional spending than EuroAlco.

Two analysts research EuroAlco at the start of year 20X3. In making their EuroAlco forecasts, both analysts use market data and the published annual report from EuroAlco (see Exhibit 19). Based on the published data, they consider a number of scenarios and reach different conclusions.

Exhibit 19 EuroAlco Key Financial and Operational Data

(€ millions)	20X0	20X1	20X2	Change (%) 20X1/20X0	Change (%) 20X2/20X1
Retail revenues	9,180	10,248	11,504	11.6%	12.3%
Excise duty	900	2,520	2,900	180%	15.1%
Excise duty of retail price	9.8%	24.6%	25.2%		
VAT (20%)	1,380	1,288	1,434	−6.7	11.3
Net sales consumers	6,900	6,440	7,170	−6.7	11.3
Typical profit for distributors[a]	900	840	935	3	−12
Profit margin–trade	13%	13%	13%		

EuroAlco Key Financial Performance Indicators (€ millions, unless noted)

	20X0	20X1	20X2	Change (%) 20X1/20X0	Change (%) 20X2/20X1
Volume (million hectoliters)	30	28	29	−6.7%	3.6%
Net revenue	6,000	5,600	6,235	−6.7	11.3
Cost of sales	3,150	2,800	3,190	−11.1	13.9
Gross profit	2,850	2,800	3,045	−1.8	8.7
Selling expenses	1,650	1,680	2,088	1.8	24.3

Exhibit 19 (Continued)

EuroAlco Key Financial Performance Indicators (€ millions, unless noted)

	20X0	20X1	20X2	Change (%) 20X1/20X0	Change (%) 20X2/20X1
Administrative expenses	150	140	145	−6.7	3.6
Operating profit	1,050	980	812	−6.7	−17.1
Gross margin	47.5%	50.0%	48.8%		
Selling expenses	27.5%	30.0%	33.5%		
Operating margin	17.5%	17.5%	13.0%		
ROCE	33.9%	32.7%	27.1%		
Capital employed	3,100	3,000	3,000	4	−8

[€ per hectoliter (hl)]	20X0	20X1	20X2	Change (%) 20X1/20X0	Change (%) 20X2/20X1
Retail price/hl	306	366	397	19.6%	8.4%
Excise duty/hl	30	90	100	200	11.1
VAT (value added/ sales tax)	46	46	49.5	0	7.5
Margin for retailer sales	30	30	32.3	0	7.5
Net revenue	200	200	215	0	7.5
Cost of sales	105	100	110	−4.8	10.0
Gross profit	95	100	105	5.3	5
Selling expenses	55	60	72	9.1	20
Administrative expenses	5	5	5	0	0
Operating profit	35	35	28	0	−20

Note: Capital employed includes debt and equity capital.
[a] This is the profit for all companies that buy beer direct from the manufacturers (brewers) for sale to any end user.

1 Both analysts assume that the government will impose a further increase in the excise duty (special tax on beer). They also assume that the excise duty increase will be borne by the consumers, who will face a 10% price increase that will allow the brewers to maintain their net (after-tax) revenues per hectoliter (hl). They assume that half the cost of sales is fixed per hl and half is variable based on volume, that selling expenses will remain unchanged as a percentage of sales, and that administrative expenses are fixed.

A Analyst A expects price elasticity of 0.8, indicating that volume will fall by 8% given the10% retail price increase. Calculate the impact on operating profit and operating profit margin in 20X3 using Exhibit 20.

B Analyst B expects price elasticity of 0.5, indicating that volume will fall by 5% given the 10% retail price increase. Calculate the impact on operating profit and operating profit margin in 20X3 using Exhibit 20.

Exhibit 20 EuroAlco's Costs Structure for 20X2–20X3E (€ millions, unless noted)

		Analyst A		Analyst B	
	20X2	**20X3E**	**YoY%**	**20X3E**	**YoY%**
Volume (million hl)	29	26.7	−8.0%	27.6	−5.0%
Net revenue (€ per hl)	215				
Revenues	6,235				
Cost of sales	3,190				
Gross profit	3,045				
Gross margin	48.8%				
Selling expenses	2,088				
Administrative expenses	145	145		145	
Operating profit	812				
Operating profit margin (%)	13.0				
Cost of sales (fixed)	1,595	1,595		1,595	
Cost of sales (variable)	1,595				
Cost of sales (variable)/hl	55	55		55	
Selling expenses as % of sales	33.5%	33.5%		33.5%	

2 Gross margin improved in 20X1 (50.0%) but fell in 20X2 (48.8%). Cost of sales was relatively high in 20X2 because of high barley costs, an important input for brewing beer. Assume that in 20X2 half of the cost of sales is fixed and half is based on volume. Of the variable part of the cost of sales, assume that half of the amount is related to the barley price in 20X2. Barley prices increased 25% in 20X2. Consider a scenario where no additional taxes are imposed in 20X3, revenues and volumes remain stable and barley prices return to their 20X1 level. Calculate EuroAlco's estimated gross margin for 20X3.

3 EuroAlco's selling expenses increased from 30% of sales in 20X1 to 33.5% of sales in 20X2. Which competitive forces most likely influenced EuroAlco's significant increase in selling expenses?

4 Retailers are the direct customers of brewers. They buy directly from the brewer and sell to the ultimate consumer. Analyst A expects that the increase in mass retailers in Eurolandia will cause brewers' margins to decline. He expects EuroAlco's operating margin will decrease from 13% in 20X2 to 8% in 20X6, with stable sales (€6,235 million) and an unchanged amount of invested capital (€3,000 million). Analyst B also sees the increasing importance of the larger food retailers but expects that EuroAlco can offset potential pricing pressure by offering more attractive trade credit (e.g., allowing the retailers longer payment terms). He thinks operating margin can remain stable at 13% with no sales growth. Capital employed (€3,000 million), however, will double because of the extra investments in inventory and receivables. Describe the analysts' expectations about the impact of large retailers on brewers in terms of Porter's five forces and ROCE. Which of the two scenarios would be better for EuroAlco?

Solution to 1:

Exhibit 21 shows the results for both analysts' projections. Analyst A predicts that operating profit will decrease by 25% to€608 in 20X3, resulting in an operating margin decline from 13.0% in 20X2 to 10.6% in 20X3. Analyst A calculates a revenue decline of 8% to €5,736 based on volume dropping by 8% and a constant price per hl of €215. The decrease in volume reflects the price elasticity of 0.8 and the price increase of 10% as a result of the excise duty increase. Cost of goods sold fell only 4% because part of the costs are fixed. Cost of goods as the sum of fixed and variable costs is €1,595 + [26.68 (hl volume) × €55 (hl cost)] = €1,595 + €1,467 (ignoring rounding error) or €3,062. Analyst A predicts selling expenses will decline in line with sales by 8% and administrative costs will remain unchanged because of their fixed character in the short term.

Analyst B forecasts that operating profit will decline by 16% to €684. Analyst B's calculations follow the same pattern as those of Analyst A, but Analyst B predicts a smaller, 5%, decline in volume. Analyst A's estimates are more pessimistic than those of Analyst B. Note that the net price/hl for the brewer is held constant while the price for the consumer increased 10% as a result of the excise duty increase. Because of Analyst B's more optimistic volume forecast, fixed costs are spread over a higher level of sales than is the case for Analyst A. Consequently, Analyst B will have a higher operating margin estimate than Analyst A. However, both analysts are predicting a decline in operating margin in 20X3.

Exhibit 21 Analysts' Results for EuroAlco's Cost Structure and Projection (€ millions, unless noted)

		Analyst A		Analyst B	
	20X2	20X3E	YoY%	20X3E	YoY%
Volume	29	26.7	–8.0%	27.7	–5.0%
Net revenue (€ per hl)	215	215	0.0	215	0.0
Net revenues	6,235	5,736	–8.0	5,923	–5.0
Cost of sales	3,190	3,062	–4.0	3,110	–2.5
Gross profit	3,045	2,674	–12.2	2,813	–7.6
Gross margin	48.8%	46.6%		47.5%	
Selling expenses	2,088	1,921	–8.0	1,984	–5.0
Administrative expenses	145	145	0.0	145	0.0
Operating profit	812	608	–25%	684	–16%
Operating profit (%)	13%	10.6%		11.55%	
Cost of sales (fixed)	1,595	1,595		1,595	
Cost of sales (variable)	1,595	1,467		1,515	
Cost of sales (per hl)	55	55		55	
Selling expenses as % of sales	33.5%	33.5%		33.5%	

Solution to 2:

If barley prices return to their 20X1 level, they will decline 20% in 20X3. Because volumes are assumed to remain constant, other variable costs will not change. Gross profit in 20X2 was 48.8% of sales, which indicates the cost of sales was 51.2% (100% − 48.8%). Barley is 25% of the cost of sales (because barley represents half of variable costs and variable cost of sales represents half of total cost of sales). Cost of sales is predicted to decline by 25% × 20% = 5%. New cost of sales will be 51.2% − (5% × 51.2%) or 48.6%. Consequently, gross margin is predicted to be 100% − 48.6% = 51.4% in 20X3. Compared with the gross margin of 48.8% in 20X2, gross margin is predicted to increase by 260 bps.

Exhibit 22 Percent of Sales

	20X2	20X3E	YoY%
Revenues	100.00	100.00	0.0
− Barley costs	12.8	10.23	−20.0
− Other costs	12.8	12.8	0.0
Variable cost of sales	25.6	23.02	−10.0
Fixed cost of sales	25.6	25.6	0.0
Total cost of sales	51.2	48.6	−5.0
Gross profit	48.8	51.4	5.2

Solution to 3:

Intra-industry rivalry and threat of substitutes most likely influenced EuroAlco's significant increase in selling costs. By spending more on advertising, EuroAlco wanted to enhance the brand loyalty of its products, thus improving its competitive position versus its brewer rivals and makers of other alcoholic beverages. Furthermore, buyers' bargaining power probably also influenced EuroAlco's increased spending to the extent that advertising creates demand by the ultimate consumer. Strong demand at the ultimate consumer level for EuroAlco's specific brands could enhance the company's bargaining position with its direct customers, the distributors who serve as intermediaries.

Solution to 4:

The increase in mass retailers in EuroAlco is expected to strengthen the bargaining power of buyers relative to brewers. According to Analyst A, this will lead to a lower operating margin of 8%, while Analyst B believes margins can be maintained if the company offers much more favorable credit terms reflected in doubling of capital employed. Analyst A expects operating profit on capital employed to fall from 27.1% (13% × €6,235/€3,000) to 16.6% (8% × €6,235/€3,000). Analyst B's assumptions indicate that the ROCE (operating profit divided by capital employed) in 20X2 of 27% will fall by half to 13.5% as the operating result is earned on double the amount of invested capital (i.e., 13% x €6,235/€6,000). The scenario envisioned by Analyst A is better for EuroAlco.

In summary, Porter's five forces framework and similar analytical tools can help analysts assess the relative profit potential of a company by helping them understand the company's industry and its position within that industry. Understanding the industry and competitive contexts of a company helps analysts estimate whether, for example, sales growth is likely to be relatively high or low (relative to history, relative to the overall growth in the economy or a sector, and/or relative to competing companies) and whether profit margins are likely to be relatively high or low (relative to historical profit margins and relative to competing companies). The process of incorporating an industry and competitive analysis into expectations for future financial performance requires judgment. Suppose analysts observe that a given company is the market leader in a moderately competitive industry with limited buyer and supplier power and relatively high barriers to entry. In broad terms, analysts might project that the company's future revenue growth will be in line with that of the overall industry and that its profit margins and ROIC might be somewhat higher than those of other companies in the industry. But there is no mechanical link between the analysts' observations and projecting the company's future sales growth and profit margin. Instead, the link is more subjective and probabilistic.

INFLATION AND DEFLATION 8

i. explain how to forecast industry and company sales and costs when they are subject to price inflation or deflation;

Inflation and deflation (i.e., the overall increase and decrease in the prices of goods and services) can significantly affect the accuracy of forecasts for a company's future revenue, profit, and cash flow. The impact of inflation or deflation on revenue and expenses differs from company to company. Even within a single company, the impact of inflation or deflation is generally different for revenue and expenses categories.

Some companies are better able to pass on higher input costs by raising the prices at which they sell their output. The ability to pass on price increases can be the result of, for example, strong branding (Coca-Cola) or proprietary technology (Apple). Companies that are well positioned to pass on price increases are, in turn, more likely to have higher and more stable profits and cash flow, relative to competitors.

We first consider the impact of inflation on sales and then on costs.

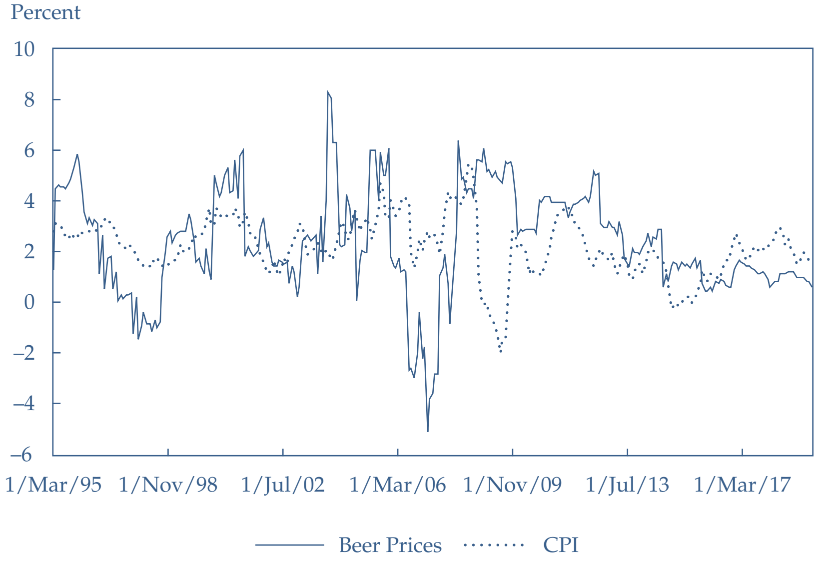
8.1 Sales Projections with Inflation and Deflation

The following analysis addresses the projection of industry sales and company sales in the presence of inflation.

8.1.1 *Industry Sales and Inflation or Deflation*

Most increases in the cost of inputs, such as commodities or labor, will eventually result in higher prices for end products. Industry structure can be an important factor in determining the relationship between increases in input costs and increases in the price of end products. For example, in the United States, the beer market is an oligopoly, with one player, AB InBev, controlling almost half of the market. Moreover, the three-tier structure of the US beer market, in which the producers (the brewers) have to use a third party (the wholesalers) to get their products (beer) to the consumers (café, restaurants, and retailers) results in a fragmented customer base because brewers are not allowed to deliver directly to the end consumer but rather must use wholesale distributors. These wholesalers often differ state by state. Large nationwide retailers, such as Wal-Mart, still have to negotiate with several different wholesalers instead of using their dominant national market position to negotiate directly with the brewers. The industry structure in the United States has likely contributed to increases in beer prices roughly in line with the US Consumer Price Index (CPI). In other words, beer prices have generally risen during years of inflation in input costs and decreased when costs have eased. If necessary, US brewers have been able to increase prices to compensate for costs of inflation. In contrast, European beer companies distribute through a more concentrated customer base—namely, such dominant retail outlets as Carrefour, Tesco, and Ahold—which results in a weaker pricing position for the brewers. Also, the European market lacks an overall dominant brewer. As a result of the industry structure and the lack of underlying volume growth, changes in beer prices in Europe have been on average 100 bps less than customer inflation.

Exhibit 23 US General Inflation and Inflation in Beer Prices

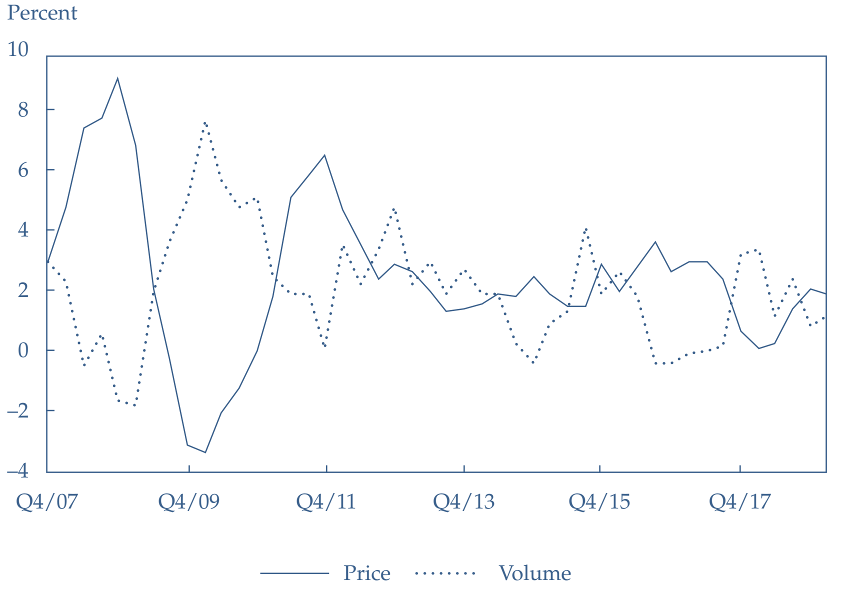


Source: US Federal Reserve Economic Data (FRED).

A company's efforts to pass on inflation through higher prices can have a negative impact on volume if the demand is price elastic, which would be the case if cheaper substitutes are available. If selling prices could be increased 10% while maintaining unit sales volume to offset an increase of 10% in input costs, gross profit margin percentage would be the same but the absolute amount of gross profit would increase. In the short term, however, volumes will usually decline as result of a price increase. The decline would depend not only on the price elasticity of demand but also on the reaction of competitors and the availability of substitutes. Lower input costs also make lower consumer prices possible. The first competitor to lower prices will usually benefit with an uptick in volume. Competitors react quickly, however, resulting in a short-term benefit. The price–volume trade-off can make accurate revenue projections difficult. In an inflationary environment, raising prices too late will result in a profit margin squeeze but acting too soon could result in volume losses. In a deflationary environment, lowering prices too soon will result in a lower gross margin but waiting too long will result in volume losses.

In the highly competitive consumer goods market, pricing is strongly influenced by movements in input prices, which can account for half of the cost of goods sold. In some time periods, customers' price sensitivity has resulted in a strong inverse relationship between volume and pricing. For example, increased input prices for packaging, wheat, and milk forced Anglo-Dutch consumer staple company Unilever to increase prices for its products in 2008 and into 2009. Consequently, volumes deteriorated. But as raw material prices fell in late 2009 and early 2010, the company's prices were lowered and volumes recovered strongly. As the company started to increase prices in 2011, volume growth once again slowed. In 2016, the company faced challenging conditions in several emerging markets as currency devaluation-led cost increases led to weaker volumes. Exhibit 24 illustrates the inverse relationship between Unilever's volume and price.

Exhibit 24 Unilever Overall Revenue Growth by Percentage Change in Volume and Price



Sources: Based on data from Unilever quarterly press releases.

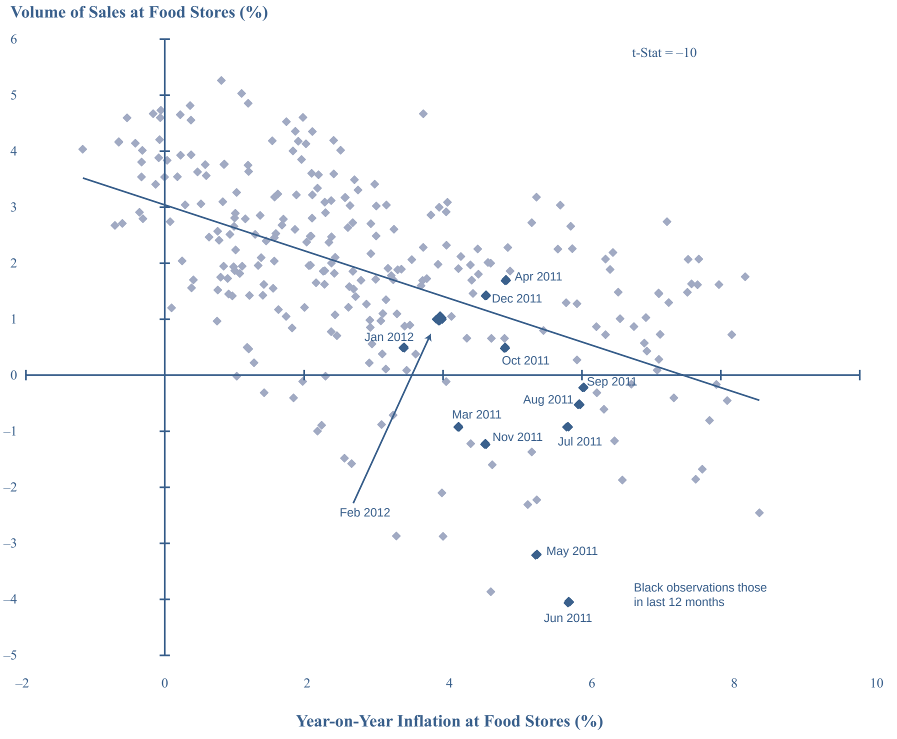
8.1.2 *Company Sales and Inflation or Deflation*

Revenue projections in a model are based on the expected volume and price development. Forecasting revenue for a company faced with inflation in input costs requires some understanding of the price elasticity of the products, the different rates of cost inflation in the countries where the company is active, and, if possible, the likely inflation in costs relevant to a company's individual product categories. Pricing strategy and market position are also important.

The impact of higher prices on volume depends on the price elasticity of demand (i.e., how the quantity demanded varies with price). If demand is relatively price inelastic, revenues will benefit from inflation. If demand is relatively price elastic (i.e., elasticity is greater than unit price elasticity), revenue can decline even if unit prices are raised. For example, a regression of volume on food inflation in UK food stores (shown in Exhibit 25) gives a regression slope coefficient of –0.398. (For every increase by 1 percentage point in year-on-year food prices, year-on-year sales decreased by about 0.4%.)

An analyst covering the UK food retailers can use this information when building forecast profit models. By assuming an expected level of food inflation, volume growth can be estimated and revenue calculated.

Exhibit 25 UK Relationship between Food Inflation and Volume, January 1989–February 2012



Source: Based on data from Datastream. Analysis is the authors'.

The expected pricing component for an international company should take into account the geographic mix of its operations to reflect different rates of the cost of inflation among countries. Of course, strategy and competitive factors, in addition to inflation in input costs, play roles in price setting.

AB InBev's volume growth and pricing have been more robust in emerging markets, for example, thanks to strong demand for its new beer products. The impact of inflation is also an important factor. In its Latin America South division, which then mainly consisted of Argentina, the brewer reported strong 24.7% organic revenue growth in 2011, of which only 2.1% was driven by volume and the remainder by price. As costs increased in line with revenues, operating margin remained more or less stable and organic operating profit growth was high at 27%. With only a limited negative currency impact, reported operating profit increased 24% in US dollars.

High inflation in a company's export market relative to a company's domestic inflation rate generally implies that the export country's currency will come under pressure and any pricing gain may be wiped out by the currency losses. The strong pricing increases AB InBev reported in its Latin America South division were clearly driven by input price inflation. The absence of a negative currency impact should be seen as a positive surprise but not as a typical outcome. A country's currency will usually come under pressure and depreciate if high rates of inflation persist for an extended period of time.

Most analysts adjust for recent high inflation in foreign countries by assuming a normalized growth rate for both revenues and costs after one or two years. This constant currency growth rate is based on an underlying growth rate assumption for the business. This approach can understate revenues in the short term. Other analysts reflect in their forecasts the high impact of inflation on revenues and expense and adjust growth rates for the expected currency (interest rate parity) impact. This approach is also imperfect given the difficulty in projecting currency rates.

Identifying a company's major input costs provides an indication of likely pricing. For a specialist retail bakery chain, for example, the impact of increased grain prices will be more significant than for a diversified standard supermarket chain. Consequently, it seems logical that the bakery is likely to increase its prices by a higher percentage than the grocer in response to increased grain prices.

Company strategy is also an important factor. Faced with rising input prices, a company may decide to preserve its margins by passing on the costs to its customers, or it may decide to accept some margin reduction to increase its market share. In other words, the company may try to gain market share by not fully increasing prices to reflect increased costs. On the one hand, many analysts think Sysco Company (the largest food distributor to restaurants and institutions in North America) has sometimes not passed on food price increases in recessionary conditions from a concern of not financially weakening already recession-affected customers (restaurants, private clubs, schools, nursing homes, etc.). On the other hand, in 2011 and 2012, the large French cognac houses increased the prices for their products substantially in China to reduce strong demand. Because older cognac generates a higher price, it can be more profitable to build an inventory of vintage cognac instead of maximizing short-term volumes.

EXAMPLE 10

Passing on Input Cost Increases or Not

Four food retail analysts are assessing the impact of a potential increase in input costs on the global supermarket chain Carrefour. In this hypothetical scenario, they believe that rising oil prices and packaging prices will impact many of the company's suppliers. They believe that Carrefour is likely to be confronted with 4% inflation in its cost of goods sold (with stable volume). The analysts have their own expectations about how the company will react. Exhibit 26 shows Carrefour's 2018 results, and Exhibit 27 shows the four analysts' estimates of input prices, volume growth, and pricing for the following year.

Exhibit 26 Carrefour Data (€ millions, unless noted)

	2018
Sales	76,000
Cost of goods sold	58,933
Gross profit	17,067
Gross margin	22.5%

Source: Based on data from Carrefour's annual report ("Annual Registration Document") for 2018.

Exhibit 27 Four Analysts' Estimates of Carrefour's Reaction to Inflation

	A	B	C	D
Price increase for revenues	0.0%	2.0%	3.0%	4.0%
Volume growth	5.0%	2.0%	1.0%	–4.0%
Total revenue growth	5.0%	4.0%	3.0%	–0.2%
Input costs increase	4.0%	4.0%	4.0%	4.0%

1 What are each analyst's predictions for gross profit and gross margin?

2 Who has the highest gross margin in his model?

3 Who has the highest absolute gross profit in his model?

Solution to 1:

The results for each analyst are shown in Exhibit 28. For Analyst B, revenues increase 4% [= (1.02 × 1.02) − 1] and cost of goods sold 6.1% [= (1.02 × 1.04) − 1]. The difference between the calculated revenue and cost of goods sold is the new gross profit.

Exhibit 28 Results for Analysts' Predictions (€ millions, unless noted)

	2018	Analyst A	YoY%	Analyst B	YoY%	Analyst C	YoY%	Analyst D	YoY%
Carrefour									
Sales	76,000	79,800	5.0%	79,070	4.0%	79,063	4.0%	75,878	-0.2%
Cost of goods sold	58,933	64,355	9.2%	62,516	6.1%	61,903	5.0%	58,839	–0.2%
Gross profit	17,067	15,445	−9.5%	16,554	−3.0%	17,160	-0.5%	17,040	–0.2%
Gross margin	22.5%	19.4%		20.9%		21.7%		22.5%	

Solution to 2:

The highest gross margin is projected by Analyst D, who assumes that selling prices would increase by 4% to offset rising input costs and keep gross margin stable.

Solution to 3:

The highest gross profit is projected by Analyst C.

8.2 Cost Projections with Inflation and Deflation

The following analysis addresses the forecasting of industry and company costs in the presence of inflation and deflation.

8.2.1 *Industry Costs and Inflation or Deflation*

Familiarity with the specific purchasing characteristics of an industry can also be useful in forecasting costs. For example, long-term price-fixed forward contracts and hedges can delay the impact of price increases. Thus, an analyst forecasting costs for an industry in which companies customarily use such purchasing practices would incorporate any expected input price fluctuations more slowly than for an industry in which the participants do not use long-term contracts or hedges.

Monitoring the underlying drivers of input prices can also be useful in forecasting costs. For example, weather conditions can have a dramatic impact on the price of agricultural products and consequently on the cost base of industries that rely on them. An analyst observing a particular weather pattern may thus be able to incorporate this information into forecasts of costs.

How inflation or deflation affects an industry's cost structure depends on its competitive environment. For example, if the participants within the industry have access to alternative inputs or are vertically integrated, the impact of volatility in input costs can be mitigated. Jacobs Douwe Egberts (JDE) is a coffee company that has been facing high and volatile coffee prices. However, its coffee is a blend of different kinds of beans. By shifting the mix slightly, JDE can keep both taste and costs constant by reducing the amount of the more expensive types of coffee beans in the blend. But if all supplier countries significantly increase the price of coffee simultaneously, JDE cannot play the blending game anymore and will be confronted with overall higher input costs. To sustain its profitability, JDE will have to increase its prices to its clients. But if competition from other companies, such as Nestlé (Nespresso, Dolce Gusto, Nescafe) makes it difficult to increase prices, JDE will have to look for alternatives if it wants to keep its profit margins stable. An easy solution for the short term could be reducing advertising and promotional spending, which usually improves profit. For the longer term, however, it could be harmful for revenues because the company's brand position could be weakened.

For example, in 2010 Russia experienced a heat wave that destroyed large parts of its grain harvest, causing prices for malting barley, a major input for beer, to increase significantly. Carlsberg, as the largest Russian brewer at that time, was particularly hard hit because it had to pay more for its Russian barley and also needed to import grain into the country, incurring additional transportation costs. By increasing imports from Western Europe, Carlsberg also pushed up barley prices in this region, affecting the cost base of other Western European brewers.

8.2.2 *Company Costs and Inflation or Deflation*

In forecasting a company's costs, it is often helpful to segment the cost structure by category and geography. For each item of cost, an assessment should be made about the impact of potential inflation and deflation on input prices. This assessment should take into account the company's ability to substitute cheaper alternatives for expensive inputs or to increase efficiency to offset the impact of increases in input prices. For example, although a jump in raw material prices in 2011 caused Unilever's and Nestlé's gross margins to fall sharply (by 110–170 bps), increases in operational efficiencies, such as reducing advertising spending, enabled both companies to achieve slightly higher overall operating profit margins that year. Example 11 shows the use of common size (percent-of-sales) analysis of inflation in input costs.

EXAMPLE 11

Inflation in Input Costs

Two hypothetical consumer staple companies—chocolate and sweets specialist "Choco A" and a food producer "Sweet B"—have costs that are constantly affected by inflation and deflation. Exhibit 29 presents a common size analysis.

Exhibit 29 Common Size Analysis for Sweet B and Choco A

	Sweet B	Choco A
Net sales	100%	100%
COGS	50%	36%
Gross margin	50%	64%
SG&A	31%	47%
Depreciation	3%	4%
EBIT	16%	13%
Raw materials	22%	22%
Packaging	12%	10%
Other COGS	16%	4%
Total COGS	50%	36%

1 Assume inflation of 10% for all costs (except depreciation) and that the companies are not able to pass on this increase through higher prices (total revenues will remain constant).

A Calculate the gross profit margin for each company. Which company will experience the greatest reduction in gross profit margin?

B Calculate the operating profit margin for each company. Which company will experience the greatest reduction in operating profit (EBIT) margin?

2 Assume inflation of 10% only for the raw material costs (reflected in COGS) and that the companies are not able to pass on this increase through higher prices. Which company will be most affected negatively in terms of gross profit margin and operating profit margin?

Solutions:

Exhibit 30 Effect of Cost Inflation

	All Costs (Except Depreciation) + 10%		Raw Materials + 10%	
	Sweet B	Choco A	Sweet B	Choco A
Net sales	100%	100%	100%	100%
COGS	55%	40%	52%	38%
Gross margin	45%	60%	48%	62%

Exhibit 30 (Continued)

	All Costs (Except Depreciation) + 10%		Raw Materials + 10%	
	Sweet B	Choco A	Sweet B	Choco A
SG&A	34%	52%	31%	47%
Depreciation	3%	4%	3%	4%
EBIT	8%	5%	14%	11%

Solution to 1A:

The company with the highest COGS as a percent of net sales—equivalently, the lowest gross margin—will experience the greatest negative impact. Sweet B has a lower gross margin than Choco A: 50% compared with 64%, as shown in Exhibit 29. After the 10% increase in COGS to 1.10 × 50% = 55%, Sweet B's gross margin will fall to 45%, as shown in Exhibit 30. Sweet B's resulting gross margin of 45% represents a proportional decline of 10% from the initial value of 50%. In contrast, the proportional decline in Choco A's gross margin is approximately 4%/64% = 6%.

Solution to 1B:

Choco A has higher overall costs than Sweet B, primarily as a consequence of its high SG&A expenses. Choco A's operating profit margin will drop to approximately 5%, as shown in Exhibit 30, representing a proportional decline of about 62% compared with a proportional decline of about 8%/16% = 50% for Sweet B.

Solution to 2:

The company with the highest raw material expense component will experience the most negative effect. In this case, raw materials represent 22% of net sales for both Sweet B and Choco A. Gross margin and operating margin will decline by 220 bps for both. This impact is more severe on gross margin on a relative basis for Sweet B (2.2%/50% = 4.4% decline) than for Choco A (2.2%/64% = 3.4% decline). But the relative effect on operating margin will be more severe for Choco A (2.2%/13% = 16.9% decline) than for Sweet B (2.2%/16% = 13.8%).

TECHNOLOGICAL DEVELOPMENTS 9

j evaluate the effects of technological developments on demand, selling prices, costs, and margins;

Technological developments have the potential to change the economics of individual businesses and entire industries. Quantifying the potential impact of such developments on an individual company's earnings involves making certain assumptions about future demand. Such assumptions should be explored through scenario and/or sensitivity analysis so that a range of potential earnings outcomes may be considered. When a technological development results in a new product that threatens to cannibalize demand for an existing product, a unit forecast for the new product combined with an expected cannibalization factor may be used to estimate the impact on future demand for the existing product. When developing an estimate of the cannibalization factor, it may be useful to segment the market if the threat of substitution differs across segments.

Technological developments can affect demand for a product, the quantity supplied of a product, or both. When changes in technology lead to lower manufacturing costs, the supply curve will shift to the right as suppliers produce more of the product at the same price. Conversely, if technology results in the development of attractive substitute products, the demand curve will shift to the left. Consider the following example.

EXAMPLE 12 (HISTORICAL EXAMPLE)

Quantifying the Tablet Market's Potential to Cannibalize Demand for Personal Computers

The worldwide tablet market experienced a major technological development with the introduction of the Apple iPad in April 2010, which was expected to have (and indeed did have) important implications for the manufacturers of desktop and laptop computers. A tablet promised to offer the capabilities of a portable personal computer with a touchscreen interface instead of a keyboard. Another distinguishing feature of tablets is that, unlike the majority of PCs that run on the Microsoft Windows platform, the then-new tablets would run on a non-Microsoft operating system, namely Apple's iOS and Google's Android. Given the tablet's ability to perform many of the most common tasks of a PC—including e-mailing, browsing the web, sharing photos, playing music, watching movies, playing games, keeping a calendar, and managing contacts—an analyst at that time might reasonably have wondered to what extent sales of tablets might cannibalize demand for PCs and the potential impact that might have on Microsoft's sales and earnings. Exhibit 31 presents one approach to answering these questions. It is set at the start of 2012, just over a year after the launch of the iPad. It is presented from the position of an analyst assessing the impact of the tablet on the PC market and Microsoft.

Exhibit 31 Unit and Revenue Projections ($ thousands, unless noted)

PRE-CANNIBALIZATION PC PROJECTIONS	FY2011	FY2012E	FY2013E	FY2014E	3-Year CAGR
Consumer PC shipments	170,022	174,430	184,120	193,811	4.5%
Non-consumer PC shipments	180,881	185,570	195,880	206,189	4.5%
Total global PC shipments	350,903	360,000	380,000	400,000	4.5%
% of which is consumer	48%	48%	48%	48%	
% of which is non-consumer	52%	52%	52%	52%	
Consumer tablet shipments	36,785	82,800	111,250	148,750	59.3%
Non-consumer tablet shipments	1,686	7,200	13,750	26,250	149.7%
Global tablet shipments	38,471	90,000	125,000	175,000	65.7%
% of which is consumer	96%	92%	89%	85%	
% of which is non-consumer	4%	8%	11%	15%	

Exhibit 31 (Continued)

PRE-CANNIBALIZATION PC PROJECTIONS	FY2011	FY2012E	FY2013E	FY2014E	3-Year CAGR
Cannibalization factor, consumer	30%	30%	30%	30%	
Cannibalization factor, non-consumer	10%	10%	10%	10%	
# of consumer PCs cannibalized by tablets	11,036	24,840	33,375	44,625	
# of non-consumer PCs cannibalized by tablets	169	720	1,375	2,625	
Total PCs cannibalized by tablets	11,204	25,560	34,750	47,250	
% of total PCs cannibalized by tablets	3.2%	7.1%	9.1%	11.8%	

POST-CANNIBALIZATION PC PROJECTIONS					
Consumer PC shipments	158,986	149,590	150,745	149,186	−2.1%
Non-consumer PC shipments	180,712	184,850	194,505	203,564	4.0%
Total global PC shipments	339,698	334,440	345,250	352,750	1.3%
Microsoft implied average selling price (ASP)					
Consumer	$85	$85	$85	$85	
Non-consumer	155	155	155	155	
Revenue impact for Microsoft ($ millions)					
Consumer	938	2,111	2,837	3,793	
Non-consumer	26	112	213	407	
Total revenue impact	964	2,223	3,050	4,200	

Notes: CAGR is compound annual growth rate. Non-consumer includes enterprise, education, and government purchasers.
Sources: Based on data from Gartner, JPMorgan, Microsoft, and authors' analysis.

To begin, worldwide market shipments of PCs in FY2011 were 350.9 million units and worldwide shipments of tablets were 38.5 million units (*Source:* Gartner Personal Computer Quarterly Statistics Worldwide Database). Shipments of tablets to consumers represented 96% of total shipments during fiscal year 2011. Next, we estimate the magnitude of the potential substitution effect, or cannibalization factor, that tablets will have on the PC market. Because the cannibalization factor depends on many different variables, including user preferences, end-use application, and whether the purchaser already owns a PC, just to name a few, we use a range of potential estimates. Moreover, we also divide the worldwide PC market into consumer and non-consumer (enterprise, education, and government purchasers) because the degree of substitution is likely to differ between the two. For purposes of illustration, we assume a cannibalization factor of 30% for the consumer market and 10% for the non-consumer market in our base case scenario.

In addition, the base case scenario assumes that non-consumer adoption of tablets increases to 15% of the market from 4% in 2011. Moreover, although the composition of the global PC market is roughly evenly divided between consumers and non-consumers (48% and 52% in fiscal year 2011, respectively), the non-consumer segment is significantly more profitable for Microsoft because approximately 80% of the company's Office products are sold to enterprise, education, and government institutions. The average selling price (ASP) estimates are derived by dividing Microsoft's estimated average revenue for the prior three years by customer type by Microsoft's estimated PC shipments for each type of customer. By multiplying the projected number of PCs cannibalized by tablets by the estimated ASP, we are able to derive an estimate of the revenue impact for Microsoft. For example, in FY2012 it is projected that 24.8 million consumer PCs will be cannibalized by sales of tablets. With an average consumer ASP of \$85, this cannibalization implies a revenue loss for Microsoft of \$2.1 billion (24.8 million units × \$85 ASP per unit = \$2.1 billion).

Once the revenue impact has been projected, the next step is to estimate the impact of lower PC unit volumes on operating costs and margins. We begin by analyzing the cost structure of Microsoft and, more specifically, the breakdown between fixed and variable costs. Most software companies have a cost structure with a relatively high proportion of fixed costs and a low proportion of variable costs because costs related to product development and marketing (mostly fixed) are sunk and unrecoverable, whereas the cost of producing an additional copy of the software (mostly variable) is relatively low. Because very few, if any, companies provide an explicit breakdown of fixed versus variable costs, an estimate almost always needs to be made. One method is to use the formula

$$\%\Delta \text{ (Cost of revenue + Operating expense)}/\%\Delta \text{ revenue},$$

where %Δ is "percent change in," used as a proxy for variable cost percentage. Another approach is to assign an estimate of the percentage of fixed and variable costs to the various components of operating expenses. Both approaches are illustrated in Exhibits 32 and 33.

Exhibit 32 Estimation of Variable Costs for Microsoft, Method 1 (\$ millions)

Selected Operating Segments	FY2009	FY2010	FY2011	FY2011/FY2009 Percent Change
Windows and Windows Live	15,563	18,792	18,778	
Microsoft business division	19,211	19,345	21,986	
Total segment revenue	34,774	38,137	40,764	17%
Windows and Windows Live	6,191	6,539	6,810	
Microsoft business division	8,058	7,703	8,159	
Total operating expense	14,249	14,242	14,969	5%

%Variable cost ≈ %Δ (Cost of revenue + Operating expense)/%Δ revenue ≈ 5%/17% ≈ **29%**.
%Fixed cost ≈ 1 − %Variable cost ≈ 1 − 29% ≈ **71%**.

Exhibit 33 Estimation of Variable Costs for Microsoft, Method 2 ($ millions)

Operating Expenses	FY2009	FY2010	FY2011	FY2009–FY2011 Average	% of Total Op Expense	Estimated % of Cost Fixed	Fixed Cost Contribution
Cost of revenue (excl. depreciation)	10,455	10,595	13,577	11,542	29%	20%	6%
Depreciation expense	1,700	1,800	2,000	1,833	5%	100%	5%
T	12,155	12,395	15,577	13,376	34%		11%
Research and development	9,010	8,714	9,043	8,922	22%	100%	22%
Sales and marketing	12,879	13,214	13,940	13,344	34%	80%	27%
General and admin.	4,030	4,063	4,222	4,105	10%	100%	10%
Total operating expenses	38,074	38,386	42,782	39,747	66%		60%
Estimated percentage of Microsoft's total cost structure that is fixed:							70%

Note: Fiscal year ends in June.
Sources: Microsoft 2011 Form 10-K and authors' analysis.

As can be seen, Microsoft's cost structure appears to consist of approximately 70% fixed costs and 30% variable costs. Note, however, that a growing company like Microsoft will typically re-invest in property, plant, and equipment to support future growth, so even those expenses that appear to be "fixed" will increase over time. To adjust for this expected growth in fixed costs, this example includes an assumption that the change in fixed costs will be half the rate of the change in sales. Variable costs are projected to change at the same rate as sales. As shown in Exhibit 34, after incorporating these assumptions into the projections, an assumed 7.0% compound annual growth rate (CAGR) in revenue through FY2014 would translate into a 10.6% CAGR in operating income $[(36{,}757/27{,}161)^{1/3} - 1 = 0.106$, or 10.6%]. In addition, these assumptions would result in an operating margin expansion of 410 bps over the same period (42.9% – 38.8% = 4.1%, or 410 bps) because of the significant amount of operating leverage that exists as a result of a relatively large fixed cost base. With the further assumptions of no change in other income, a constant effective tax rate, and no change in shares outstanding, the pre-cannibalization model, shown in Exhibit 34, results in projected revenue of $85.7 billion, operating income of $36.8 billion, an operating margin of 42.9%, and earnings per share (EPS) that increases at a CAGR of 10.3% to $3.62 in FY2014.

Exhibit 34 Microsoft Pre-Cannibalization EPS Projections ($ millions)

	FY2011	FY2012E	FY2013E	FY2014E	3-Year CAGR
Revenue	69,943	74,839	80,078	85,683	7.0%
Year-over-year percent change		7.0%	7.0%	7.0%	
Operating Expenses					
Fixed (70%)	29,947	30,996	32,080	33,203	3.5%
Variable (30%)	12,835	13,733	14,694	15,723	7.0%
Total operating expenses	42,782	44,729	46,775	48,926	4.6%
Operating income	27,161	30,110	33,303	36,757	10.6%
Operating margin	38.83%	40.23%	41.59%	42.90%	
Other income (Expense)	910	910	910	910	
Pretax Income	28,071	31,020	34,213	37,667	
Provision for income taxes	4,921	5,438	5,998	6,603	
Effective tax rate	17.53%	17.53%	17.53%	17.53%	
Net income	23,150	25,582	28,215	31,064	
Weighted average shares outstanding, diluted	8,593	8,593	8,593	8,593	
Estimated EPS pre-cannibalization	$2.69	$2.98	$3.28	$3.62	10.3%

In the post-cannibalization scenario, as shown in Exhibit 35, revenue is reduced each year to reflect the expected impact from cannibalization. The expected impact of cannibalization results in a decrease in the CAGR of revenue over the period to 5.2%, down from 7.0% in the pre-cannibalization scenario. Given the reduction in revenue growth and holding the cost structure constant at 70/30 fixed versus variable costs, operating income growth slows to a CAGR of 8.0%, down from 10.6% in the pre-cannibalization scenario. Operating margin at the end of the period is reduced by approximately 100 bps from 42.9% to 41.9% because the company is unable to leverage its fixed cost base to the same degree as a result of slower revenue growth. Overall, in the post-cannibalization scenario, Microsoft is expected to generate revenue of $81.5 billion, operating income of $34.2 billion, an operating margin of 41.9%, and EPS that increase at a CAGR of 7.8% to $3.37 in FY2014. Thus, the cannibalization of PCs as a result of projected growth in the tablet market is expected to reduce the company's annual revenues in FY2014 by $4.2 billion, operating income by $2.6 billion, operating margins by 96 bps, and EPS by $0.25.

Exhibit 35 Microsoft Post-Cannibalization EPS Projections, Base Case Scenario ($ millions, unless noted)

	FY2011	FY2012E	FY2013E	FY2014E	3-Year CAGR
Revenue	69,943	72,616	77,028	81,483	5.2%
Year-over-year percent change		3.8%	6.1%	5.8%	
Operating Expenses					
Fixed (70%)	29,947	30,520	31,447	32,356	2.6%
Variable (30%)	12,835	13,325	14,135	14,952	5.2%
Total operating expenses	42,782	43,845	45,581	47,308	3.4%
Operating income	27,161	28,771	31,446	34,175	8.0%
Operating margin	38.83%	39.62%	40.82%	41.94%	
Other income (Expense)	910	910	910	910	
Pretax income	28,071	29,681	32,356	35,085	
Provision for income taxes	4,921	5,203	5,672	6,151	
Effective tax rate	17.53%	17.53%	17.53%	17.53%	
Net income	23,150	24,478	26,684	28,934	
Weighted average shares outstanding, diluted	8,593	8,593	8,593	8,593	
Estimated EPS post-cannibalization	$2.69	$2.85	$3.11	$3.37	7.8%
Estimated impact on operating margin		−61 bps	−76 bps	−96 bps	
Estimated impact on EPS		($0.13)	($0.18)	($0.25)	−2.6%

Example 13 addresses questions related to the text discussion on cannibalization.

EXAMPLE 13

Estimating the Impact of Cannibalization

Answer the following questions using Exhibits 31 through 35 on Microsoft:

1 Estimate post-cannibalization global PC shipments in FY2012 assuming a cannibalization factor for consumers of 40% and 15% for non-consumers.

2 Using the results derived in Question 1, estimate the post-cannibalization revenue in FY2012 for Microsoft.

3 Using the estimate for post-cannibalization revenue derived in Question 2 and the cost structure provided, estimate post-cannibalization operating income and operating margin in FY2012 for Microsoft. Assume that fixed costs change at half the rate of the change in sales.

4 Using the estimate for operating income derived in Question 3 and the data in the exhibits, calculate the expected post-cannibalization EPS in FY2012 for Microsoft. Assume that other income (expense), the effective tax rate, and the diluted weighted average shares outstanding provided for FY2011 remain constant in FY2012.

Solution to 1:

The number of PCs cannibalized by tablets is equal to the product of the expected number of global tablet shipments, the percentage representation of each category, and the cannibalization factor for the category. Exhibit 31 shows that tablet shipments in FY2012 are projected to be 90 million units. (90 million tablets × 92% consumer representation × 40% consumer cannibalization factor = 33.12 million consumer PCs cannibalized by tablets) + (90 million tablets × 8% non-consumer representation × 15% cannibalization = 1.08 million non-consumer PCs cannibalized by tablets) = 34.2 million total PCs cannibalized by tablets. Post-cannibalization shipments are equal to pre-cannibalization shipments minus expected cannibalization, or 360 million – 34.2 million = 325.8 million.

Solution to 2:

The estimated impact on revenue is equal to the product of the number of PCs cannibalized and the average selling price. Using the results obtained in Question 1 and the ASP data contained in Exhibit 31, the expected revenue impact can be calculated as (33.12 million consumer PCs cannibalized by tablets × $85 ASP = $2.815 billion) + (1.08 million non-consumer PCs cannibalized by tablets × $155 ASP = $167.4 million) = $2.983 billion total impact on revenue for Microsoft. Post-cannibalization revenue is equal to pre-cannibalization revenue minus the estimated impact on revenue from cannibalization, or $74.839 billion – $2.983 billion = $71.856 billion.

Solution to 3:

Exhibit 36 Solution to Problem 3 ($ millions)

	FY2011	FY2012E	Notes:
Revenue	69,943	71,856	Derived from Question 2
YoY %		2.74%	Rate of change in sales used to estimate operating expenses
Operating Expenses			
Fixed (70%)	29,947	30,357	Fixed costs change at half the rate of the change in sales, or 29,947 × (1 + 2.74%/2)
Variable (30%)	12,835	13,186	Variable costs change at the same rate as the change in sales, or 12,835 × (1 + 2.74%)
Total operating expenses	42,782	43,543	Although not shown, operating expenses include cost of goods sold

Exhibit 36 (Continued)

	FY2011	FY2012E	Notes:
Operating income	27,161	28,313	Revenue minus total operating expense, or 71,856 – 43,543 = 28,313
Operating margin	38.8%	39.4%	Operating income divided by revenue, or 28,313/71,856 = 39.4%

Post-cannibalization operating income and operating margin in FY2012 for Microsoft are $28,313 million and 39.4%, respectively.

Solution to 4:

Exhibit 37 Solution to Problem 4 ($ millions, unless noted)

	FY2011	FY2012E	Notes:
Revenue	69,943	71,856	
YoY %		2.74%	
Operating Expenses			
Fixed (70%)	29,947	30,357	
Variable (30%)	12,835	13,186	
Total operating expenses	42,782	43,543	
Operating income	27,161	28,313	
Operating margin	38.8%	39.4%	
Other income (expense)	910	910	
Pretax income	28,071	29,224	Operating income + Other income (expense), or 28,314 + 910 = 29,224
Provision for income taxes	4,921	5,123	Pretax Income × Effective tax rate, or 29,224 × 17.53% = 5,123
Effective tax rate	17.53%	17.53%	
Net income	23,150	24,101	Pretax Income – Provision for income taxes, or 29,224 – 5,123 = 24,101
Weighted average shares outstanding, diluted	8,593	8,593	
EPS post-cannibalization	**$2.69**	**$2.80**	Net income/Wtd Avg Shs Out, or 24,101/8,593 = $2.80

Whenever one is estimating something that depends on many different variables that are difficult to measure, we recommend altering some of the assumptions to generate a range of estimates based on various scenarios. Thus, having developed a forecast under a base case cannibalization scenario, we are able to analyze the sensitivity of the results by altering the cannibalization assumptions. The base case scenario corresponds to the assumptions in the boxed center of the table in Exhibit 38. Exhibit 39 summarizes the results of bull and bear case scenarios, showing the estimated FY2014 EPS under alternative estimated cannibalization factors.

Exhibit 38 Estimated 2014 EPS Sensitivity to Changes in Cannibalization Rates

		Non-Consumer Cannibalization				
		0.0%	5.0%	10.0%	15.0%	20.0%
Consumer Cannibalization	15%	$0.11	$0.12	$0.14	$0.15	$0.16
	20%	$0.15	$0.16	$0.17	$0.19	$0.20
	25%	$0.19	$0.20	$0.21	$0.22	$0.23
	30%	$0.22	$0.24	$\boxed{\$0.25}$	$0.26	$0.27
	35%	$0.26	$0.27	$0.28	$0.30	$0.31
	40%	$0.30	$0.31	$0.32	$0.33	$0.35
	45%	$0.34	$0.35	$0.36	$0.37	$0.38

Exhibit 39 Post-Cannibalization EPS Projections for Bull and Bear Scenarios ($ millions, unless noted)

Bull Case Scenario (Cannibalization Factor: 15% Consumer/5% Non-Consumer)

	FY2011	FY2012E	FY2013E	FY2014E	3-Year CAGR
Revenue	69,943	73,728	78,553	83,583	6.1%
YoY %		5.4%	6.5%	6.4%	
Operating Expenses					
Fixed (70%)	29,947	30,758	31,764	32,781	3.1%
Variable (30%)	12,835	13,529	14,414	15,338	6.1%
Total operating expenses	42,782	44,287	46,179	48,119	4.0%
Operating income	27,161	29,441	32,374	35,464	9.3%
Operating margin	38.83%	39.93%	41.21%	42.43%	
Other income (expense)	910	910	910	910	
Pretax income	28,071	30,351	33,284	36,374	
Provision for income taxes	4,921	5,321	5,835	6,377	
Effective tax rate	17.53%	17.53%	17.53%	17.53%	

Exhibit 39 (Continued)

Bull Case Scenario (Cannibalization Factor: 15% Consumer/5% Non-Consumer)					
	FY2011	FY2012E	FY2013E	FY2014E	3-Year CAGR
Net income	23,150	25,030	27,449	29,998	
Weighted average shares outstanding, diluted	8,593	8,593	8,593	8,593	
Estimated EPS post-cannibalization	**$2.69**	**$2.91**	**$3.19**	**$3.49**	**9.0%**
Estimated impact on operating margin		**−30 bps**	**−38 bps**	**−47 bps**	
Estimated impact on EPS		**($0.06)**	**($0.09)**	**($0.12)**	**−1.3%**

Bear Case Scenario (Cannibalization Factor: 40% Consumer/20% Non-Consumer)					
	FY2011	FY2012E	FY2013E	FY2014E	3-Year CAGR
Revenue	69,943	71,801	75,869	79,812	4.5%
YoY %		2.7%	5.7%	5.2%	
Operating Expenses					
Fixed (70%)	29,947	30,345	31,205	32,016	2.3%
Variable (30%)	12,835	13,175	13,922	14,646	4.5%
Total operating expenses	42,782	43,521	45,127	46,661	2.9%
Operating income	27,161	28,280	30,742	33,151	6.9%
Operating margin	38.83%	39.39%	40.52%	41.54%	
Other income (Expense)	910	910	910	910	
Pretax income	28,071	29,190	31,652	34,061	
Provision for income taxes	4,921	5,117	5,549	5,971	
Effective tax rate	17.53%	17.53%	17.53%	17.53%	
Net income	23,150	24,073	26,103	28,090	
Weighted average shares outstanding, diluted	8,593	8,593	8,593	8,593	
Estimated EPS post-cannibalization	**$2.69**	**$2.80**	**$3.04**	**$3.27**	**6.7%**
Estimated impact on operating margin		**−85 bps**	**−107 bps**	**−136 bps**	
Estimated impact on EPS		**($0.18)**	**($0.25)**	**($0.35)**	**−3.6%**

10 LONG-TERM FORECASTING

k explain considerations in the choice of an explicit forecast horizon;

l explain an analyst's choices in developing projections beyond the short-term forecast horizon;

The choice of the forecast time horizon may be influenced by certain factors, including the investment strategy for which the stock is being considered, cyclicality of the industry, company-specific factors, and the analyst's employer's preferences. Most professionally managed equity investment strategies describe the investment time frame, or average holding period for a stock, in the stated investment objectives of the strategy; the time frame should ideally correspond with average annual turnover of the portfolio. For example, a stated investment time horizon of 3–5 years would imply average annual portfolio turnover between 20–33% (average holding period is calculated as one/portfolio turnover). Cyclicality of the industry may also influence the analyst's choice of time frame because the forecast period should be long enough to allow the business to reach an expected mid-cycle level of sales and profitability. Similar to cyclicality, various company-specific factors, including recent acquisition or restructuring activity, may influence the selection of the forecast period to allow enough time for the realization of the expected benefits from such activity to be reflected in the financial statements. In other cases, there may be no individual analyst choice in the sense that the analyst's employer has specified a particular discounted cash flow (DCF) model with more or less fixed parameters. Much of the discussion so far has focused on various methods of forecasting a company's income statement, balance sheet, and cash flow for an explicit short-term forecast period. Although the underlying principles remain the same if one extends the time horizon, certain considerations and choices are available to the analyst when developing longer-term projections.

Longer-term projections often provide a better representation of the normalized earnings potential of a company than a short-term forecast, especially when certain temporary factors are present. **Normalized earnings** are the expected level of mid-cycle earnings for a company in the absence of any unusual or temporary factors that impact profitability (either positively or negatively). For example, at any given point in time a company's profitability can be impacted by a number of temporary factors, including the stage in the business cycle, recent merger and acquisition activity, and restructuring activity. Similarly, normalized free cash flow can be defined as the expected level of mid-cycle cash flow from operations adjusted for unusual items just described less recurring capital expenditures. By extending the forecast period, an analyst is able to adjust for these unusual or temporary factors and derive an estimate of earnings that the company is likely to earn in a normal year. We will consider various alternatives for two aspects of long-term forecasting: revenue forecasts and terminal value.

As with most income statement projections, a long-term forecast begins with a revenue projection with most of the remaining income statement items subsequently derived from the level or change in revenue. Revenue projection methods were covered earlier.

Case Study: Estimating Normalized Revenue

Exhibit 40 contains 10 years of historical revenue data and four years of estimated normalized data for Continental AG, a global automotive supplier. The accompanying bar chart in Exhibit 41 graphically depicts the data and includes a trendline based on a linear regression of the data. The numerical values for each point along the trend line can be found by using the TREND formula in Microsoft Excel. The TREND formula uses observations on the dependent variable (in this case revenue) and observations on the explanatory (time) variable to perform a linear regression by using least squares criterion to find the best fit. After computing the best fit regression model, the TREND formula returns predicted values associated with new points in time.

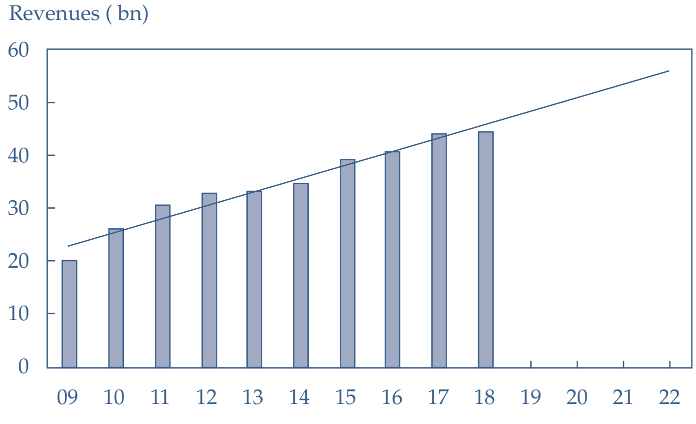
Exhibit 40 Historical and Estimated Revenue Data for Continental AG, 2009–2022E (€ billions)

	2009	2010	2011	2012	2013	2014	2015
Revenue	20.1	26.05	30.5	32.74	33.33	34.51	39.23
Normalized revenue	23.2	25.7	28.2	30.8	33.3	35.8	38.3
Percent above/ below trend	−13.4%	−1.2%	8.0%	6.4%	0.1%	-3.6%	2.4%

	2016	2017	2018	2019E	2020E	2021E	2022E
Revenue	40.55	44.01	44.4				
Normalized revenue	40.8	43.4	45.9	48.4	50.9	53.4	55.9
Percent above/ below trend	−0.7%	1.5%	-3.2%				

Sources: Continental AG annual reports.

Exhibit 41 Historical and Estimated Revenue for Continental AG, 2009–2022E



The "growth relative to GDP growth" and "market growth and market share" methods discussed earlier can also be applied to developing longer-term projections. Once a revenue projection has been established, previously described methods of forecasting costs can be used to complete the income statement, balance sheet, and cash flow statement.

After a financial forecast has been established for the explicit forecast period, an analyst typically estimates a terminal value. Certain considerations should be kept in mind when deriving the terminal value based on long-term projections. For example, when using a historical multiples–based approach to derive the terminal value of a company, an analyst is implicitly assuming that the past is relevant to the future in terms of growth expectations and required rates of return. If a multiple is used to derive the terminal value (TV), the choice of the multiple should be consistent with the long-run expectations for growth and required return. It is common for an analyst to use a historical average multiple as the basis for the target multiple in the terminal value calculation. For example, the analyst could learn that as of the mid-2019 date of the analysis, Continental AG had traded at a median P/E multiple that was 10% lower than that of the FTSEurofirst 300 Index over the previous 10 years. Thus, an analyst could forecast a terminal value (TV) in the year 2022 as the product of 0.9 and the current next 12 months (NTM) P/E of the FTSEurofirst 300, times forecasted EPS for Continental AG in 2022. But in estimating the TV of a company, historical multiples are only relevant to the extent that future growth and profitability are expected to resemble the past. If the future growth or profitability of the company is likely to differ significantly from the historical average, then the target multiple should reflect an expected premium or discount to the historical multiple to reflect this difference in growth and/or profitability.

EXAMPLE 14

Historical Valuation Multiples

Long-term historical average valuation multiples are frequently used in equity analysis as a reference point or as justification of a target multiple at which the shares are expected to trade in the future. Such widespread use is predicated on a belief in mean reversion, the idea that over time the valuation of a stock will revert to its long-term historical average. Of course, this implicitly assumes that the future growth and profitability of the company will resemble the past. If the future outlook differs significantly from the past, the historical average multiple might not be relevant. The multiple may not be computed in the same manner by all analysts. The underlying financial data can be trailing, forward, or current year.

Answer the following assumptions about a hypothetical company with premium, discount, or not applicable on the basis of how you would expect the stock's future multiple to compare with its long-term historical average, keeping all other factors constant. Assumptions:

1. The company is likely to earn higher returns on invested capital in the future.
2. Earnings growth is likely to accelerate in the future.
3. The intensity of competition is likely to increase in the future.
4. The company makes a major acquisition or divestiture.

Solution to 1:

If a company is likely to earn higher returns on invested capital in the future than it has historically, then the company's shares are likely to trade at a premium to their historical average multiple to reflect the expected improvement in profitability.

Solution to 2:

If a company is likely to generate faster EPS growth in the future than it has historically, then the company's shares are likely to trade at a premium to their historical average multiple to reflect the faster expected growth rate.

Solution to 3:

If competition in the industry is expected to intensify in the future, then the company's shares are likely to trade at a discount to their historical average multiple to reflect the likely degradation in profitability.

Solution to 4:

If either the future growth or profitability of a company is expected to change significantly as a result of a major acquisition or divestiture, then the historical valuation multiple of the company should no longer be relied on as either a reference point or as justification of a target multiple at which the shares are expected to trade in the future. The historical valuation multiple would be considered not applicable.

When using a DCF approach to developing terminal value, an analyst should avoid mechanically applying a long-term growth rate to a terminal year free cash flow projection. First, an analyst should consider whether the terminal year free cash flow projection should be normalized before that cash flow is incorporated into a long-term projection. For example, if the explicitly forecasted terminal year free cash flow is "low" (e.g., because of business cycle reasons or capital investment projects), an adjustment to normalize the amount may be warranted. Second, an analyst should consider whether and how the future long-term growth rate will differ from the historical growth rate. For example, even some mature companies may be able to accelerate their long-term growth rate through product innovation and/or market expansion (e.g., Apple), whereas other seemingly well-protected "growers" may experience an unanticipated decline in their business as a result of technological change (e.g., Eastman Kodak Company, a global commercial printing and imaging company).

One of the greatest challenges facing the analyst is anticipating inflection points, when the future will look significantly different from the recent past. Most discounted cash flow models rely on some kind of perpetuity calculation. A perpetuity calculation assumes that the cash flows from the last year of an explicit forecast grow at a constant rate forever. Because the perpetuity can account for a relatively large portion of the overall valuation of the company, it is critical that the cash flow used is representative of a "normalized" or "mid-cycle" result. If the analyst is examining a cyclical company, using a boom year as the starting point for the perpetuity could result in a grossly overstated intrinsic value estimate. Similarly, using a trough year could result in an intrinsic value estimate that is much too low.

Another important consideration is economic disruption, such as occurred in the 2008 global financial crisis. The economy can occasionally experience sudden, unprecedented changes that affect a wide variety of companies. Even a company with a sound strategy and solid operations can be thrown far off course by a sudden economic disruption, particularly if the company has a high degree of financial leverage.

Regulation and technology are also potential drivers of inflection points, and it is important for the analyst to keep a close eye on both. Government actions can have extreme, sudden, and unpredictable impacts on some businesses. Technological advances can turn fast-growing innovators into obsolete dinosaurs in a matter of months. Both regulation and technology affect some industries more than others. Utilities experience intense regulation but may not see a significant technological change for decades. Semiconductor manufacturers must constantly keep up with new technology but experience relatively light regulation. Medical device manufacturers are heavily exposed to both regulation and technological advances.

Finally, long-term growth is a key input in the perpetuity calculation. Some companies and industries can grow faster than the overall economy for long periods of time, causing them to account for an increasing share of the economic pie. Examples include certain smartphone manufacturers, such as Apple and Samsung, and some internet-related companies, such as Amazon and Google. Other companies, such as those in the print media sector, are likely to grow slower than the overall economy or even shrink over time. Using an unrealistic long-term growth rate can put the analyst's estimate of intrinsic value far off the mark.

EXAMPLE 15

Important Considerations When Making Assumptions

1 Turkish Airlines (THYAO.IS) operates in the highly cyclical global airline industry. Operating margins for the last nine years are shown in the following table.

	2010	2011	2012	2013	2014	2015	2016	2017	2018
Operating margin (%)	5.7	1.0	10.8	6.5	5.6	8.6	–2.9	9.0	9.9

On the basis of only the information in the table, which of the following operating margins would *most likely* be appropriate to use in a perpetuity calculation for Turkish Airlines to arrive at a reasonable intrinsic value estimate?

A 6.0%

B 9.0%

C 9.9%

For each of the companies in the following problems, indicate which of the choices is *least likely* to cause a change in the company's outlook.

2 ABC Diesel (hypothetical company), a manufacturer of diesel-power trucks.

A Environmental regulations have been getting tighter in most regions, and consistent with past experience, this need to make the engines less polluting is expected to continue over the next several years.

B Consumers have started switching to trucks with electric engines, threatening ABC's historic strength in diesel engine trucks.

C ABC Diesel has formed a partnership with Electrico (hypothetical), a company involved in research and innovation in electric engines.

3 Abbott Laboratories, a diversified manufacturer of health care products, including pharmaceuticals and medical devices.

A It has become more difficult for medical device manufacturers to receive regulatory approval for new products because of heightened safety concerns.

B A competitor has demonstrated favorable efficacy data on a drug candidate that will compete with an important Abbott product.

C Management reiterates its long-standing approach to capital deployment.

4 Grupo Aeroportuario del Sureste, operator of nine airports in Mexico, especially in the tourist-heavy southeast.

A A technological advance will allow airlines to save 5% on fuel costs, but it is not expected to meaningfully alter passenger volumes. Similar developments in the past have benefited airlines but not airports, whose price per passenger is regulated.

B Global economic disruption has caused a sharp decline in international travel.

C Regulators will allow the construction of a new airport by a competitor in Grupo Aeroportuario del Sureste's service territory.

5 LinkedIn, operator of an online social network for professionals and part of Microsoft Corporation, with limited investment needs and no debt.

A Facebook, another online social network, announces a plan to enhance its offerings in the professional category.

B Regulators announce an investigation of LinkedIn's privacy practices, which could result in significant changes to the service.

C The US Federal Reserve has just increased interest rates. Although this will raise borrowing costs, the rate increase is not expected to have a negative impact on the economy.

Solution to 1:

A is correct. Because the airline industry is cyclical, an estimate of "mid-cycle" or "normalized" operating margin is necessary to estimate a perpetuity value. The nine-year average operating margin was 54%/9 = 6%.

Solution to 2:

A is correct. Although it is important that environmental regulations have been getting stricter, this is consistent with past experience and so does not represent a turning point.

Solution to 3:

C is correct. Management is sticking with its historical approach to capital deployment, so this does not represent a turning point.

Solution to 4:

A is correct. Although the technological advance is good for the airlines, it will not have a meaningful effect on passenger volumes, which will likely prevent the airports from sharing in that benefit. In contrast, both B and C could have a significant impact on the long-run earnings power of Mexican airports.

Solution to 5:

C is correct. Because LinkedIn carries no debt, it is unlikely that higher interest rates will cause a change in the company's outlook.

BUILDING A MODEL: INDUSTRY OVERVIEW AND COMPANY OVERVIEW 11

m demonstrate the development of a sales-based pro forma company model.

This section provides an example of building a company model. The subject company is the Rémy Cointreau Group (Rémy), a French company that sells wines and spirits. After providing a brief overview of the company and industry, we will focus primarily on the mechanics of constructing pro forma income statements, statements of cash flows, and balance sheets. Data sources for this example include the company's fiscal year 2018/19 annual report (year-end 31 March 2019), the company's interim reports, and corresponding investor presentations for additional information on the underlying results of the respective divisions.

11.1 Industry Overview

This industry overview will focus on the cognac industry because it is Rémy's most important business segment, accounting for almost 84% of total operating profit. (In practice, an analyst would also perform a similar industry analysis for the company's other major segments.) An important feature of the cognac market is that supply is limited and demand is growing. Supply is limited because cognac production, similar to champagne, is highly regulated, in this case through the Bureau National Interprofessionnel du Cognac (BNIC). By regulation, cognac can only be produced in a limited geographic area, located around the town of Cognac in southwest France. Furthermore, within the region, production volume is capped each year. About 98% of production is exported. The cognac market is highly concentrated, with the top four players controlling 78% of world volume and 84% of global value. Rémy's market share is about 16% and 18% of global volume and value, respectively (*The Spirits Business,* June 2018). Demand for cognac has been growing because of increasing demand from Asia, particularly China and Singapore, more than offsetting a weakening European market. The global spirits market has grown more than 5% during the 2000–2017 period (*Source:* IWSR drinks market analysis). Simultaneously, Rémy has also seen a product mix improvement because consumers increasingly prefer superior quality and more expensive cognac. Exhibit 42 summarizes Porter's five forces analysis of the cognac industry.

Exhibit 42 Porter's Five Forces Analysis of the Cognac Industry

Force	Degree	Factors to Consider
Threat of substitutes	Low	• Cognac consumers show brand loyalty and do not easily shift to other beverages or high-end spirits.
Rivalry	Low	• Market is consolidated, with four players controlling 78% of the world market in volume and 84% of global value. • Only the European market is fragmented, with less than half of the market controlled by the top four.
Bargaining power of suppliers	Low/medium	• Large number of small independent vineyards supply inputs. • Most of the distillation is carried out by a large body of independent distillers that sell to the big houses.

Exhibit 42 (Continued)

Force	Degree	Factors to Consider
Bargaining power of buyers	Low	• Premium beverages are mostly sold to wine and spirits retail outlets that do not coordinate purchasing. • Premium beverages are mostly consumed in small and fragmented on-premises outlets (restaurants, etc.).
Threat of new entrants	Low	• Producers have long-term contracts with suppliers in the Cognac area. • Barriers to entry are high. ○ Building brands is difficult because they must have heritage/pedigree. ○ Large capital investment is required to build an inventory with "aged" cognac and set up a distribution network.

In summary, the cognac market, Rémy's largest and most profitable operating segment, exhibits a favorable profitability profile. In addition to limited supply and growing demand, the industry faces a generally favorable situation with respect to substitutes, rivalry, suppliers, buyers, and potential new entrants.

11.2 Company Overview

Rémy, whose reporting year ends 31 March, operates three business segments:

1. Cognac. This division, which is named after its main brand Rémy Martin, represented about 69% of FY2019 (year-end 31 March 2019) revenue and 84% of total operating profits.
2. Liqueurs & Spirits. The main brands in this segment are Cointreau, Passoa, Metaxa, Saint Rémy, and Mount Gay. They represented about 23% of FY2019 revenue and 14% of operating profits.
3. Partner Brands. This division includes other companies' brands that are marketed through Rémy's distribution network. They represented about 8% of FY2019 revenue and less than 2% of operating profits. This division's importance has been declining as the company discontinues distribution ("partner brand") contracts.

Segment financial information is summarized in Exhibit 43. As shown, the company's largest business segment is also its most profitable: The Cognac segment earned an operating profit margin of around 30% (= €236million/€774million) in fiscal year 2019.

Exhibit 43 Analysis of Rémy's Turnover and Operating Profit

	FY2017	FY2018	FY2019
Rémy Martin (€ millions)	708	760	774[a]
Liqueurs & Spirits (€ millions)	276	269	264
Partner Brands (€ millions)	111	100	87
Total revenues (€ millions)	1,095	1,127	1,126

(continued)

Exhibit 43 (Continued)

	FY2017	FY2018	FY2019
Operating Profit (€ millions)			
Rémy Martin	185	204	236
Liqueurs & Spirits	58	43	39
Partner Brands	2	5	5
Holding costs[b]	17	16	15
Total operating profit	226	237	264
Operating Profit Margin by Segment			
Rémy Martin	26.2%	26.9%	30.4%
Liqueurs & Spirits	20.8%	16.0%	14.7%
Partner Brands	1.8%	5.3%	5.6%
Holding costs as % of total revenue	−1.7%	−1.4%	−1.3%
Operating margin	20.7%	21.0%	23.5%

[a] 2019 figures reflect adoption of IFRS 15 *Revenue from Contracts with Customers*. Excluding the impact, pro-forma revenues for 2019 would be €1,217 million. There was negligible impact on operating profits.
[b] Holding costs are a deduction.
Source: Based on information in consolidated financial statements of Rémy Cointreau Group on 31 March 2019, Note 17.1.

12 CONSTRUCTION OF PRO FORMA INCOME STATEMENT

m demonstrate the development of a sales-based pro forma company model.

This section will illustrate the construction of pro forma income statements. The forecasts of revenue follow the structure of the company's operating segments.

12.1 Revenue Forecast

The revenue forecasts use primarily a hybrid approach because trends in the individual segments (bottom-up) are combined with the economic development in the relevant regions (top-down). For each segment, the change in revenue is driven by volume, price, and foreign currency estimates that are based on historical trends as adjusted for expected deviations from trend. Price changes refer not only to price changes for a single product but also to changes in price/mix, which is defined as changes in average price that result from selling a different mix of higher and lower priced products. Changes in revenue attributable to volume or price/mix are considered to be organic growth and are shown separately from the impact of foreign exchange (forex impact in the model).

In the Cognac segment, historical volume growth is usually in the 4-6% range. For future years, volume growth is expected to remain robust but be somewhat slower than the 5.9% achieved in 2019 given the global economic slowdown at the time of this projection. The growing number of affluent Asian consumers will likely keep demand high, while developed market consumption is likely to be rather flat. In the model, the assumption is for 5% volume growth in 2020, declining by 1 percentage point to a more modest 4% in 2022.

Price/mix contributed around 5.5%, 7.2%, and 6.0% to the Cognac segment revenue growth in 2017, 2018, and 2019, respectively. Although the impact of price/mix on revenue growth has fluctuated in recent years, it is likely that price/mix will remain a relatively significant contributor to revenue growth in the future given the favorable structure of the industry and the company's efforts to increase the share of revenues accounted for by what it calls "exceptional spirits" (those that cost more than $50 per bottle and are seeing a 10% annual demand growth). A 4% price/mix contribution to revenue growth is assumed in 2020, with the trend maintained into 2021 and 2022. The combined projections for 2020 of 5% volume growth and 4% price/mix impact results in overall organic revenue growth of 9.2%, calculated as $[(1 + 0.05) \times (1 + 0.04) = 1.092] - 1 = 0.092$, or 9.2%.

In addition to the impact of volume and price/mix, Rémy's revenues are affected by movements in exchange rates. Company disclosures indicate that more than 70% of revenues are realized outside the eurozone, whereas most of Rémy's production occurs in the eurozone. The model forecasts no foreign currency impact on revenue in the 2020–2022 forecast period.

Exhibit 44 summarizes historical and projected information for the Cognac segment's revenue.

Exhibit 44 Historical and Projected Information for Rémy Martin (Cognac) Segment Revenue

	2016	2017	2018	2019	2020E	2021E	2022E
Revenues (€ million)	647.8	707.5	760.0	774.4	846	915	989
YoY %	14.7%	9.2%	7.4%	1.9%	9.2%	8.2%	8.2%
Volume growth %	2	4.5	6.0	5.9	5.0	4.0	4.0
Price/mix %	1.2	5.5	7.2	6.0	4.0	4.0	4.0
Organic growth %	3.2	10.0	13.2	11.9	9.2	8.2	8.2
Forex impact & scope change %	11.5	−0.8	−5.8	−4.0	0.0	0.0	0.0
YoY %	14.7	9.2	7.4	7.9	9.2	8.2	8.2

Note: Excluding the impact of IFRS 15 adoption, pre-forma revenue growth in 2019 was 7.9%.
Sources: Based on data from Rémy Cointreau Group and authors' analysis.

A similar analysis can be performed to project revenue for the other segments. Then, the amounts can be summed to derive projected revenue for the company as a whole.

12.2 Cost of Goods Sold

Rémy's gross margin has gradually increased since 2016 from 63% to 68% in 2018 and 2019 (reported at 68.1% on a pro-forma basis). The strong price/mix effect was the underlying driver for the gross margin improvement in the 2016–2019 period, resulting from the strong growth in the high margin region of Asia. With its appetite for expensive cognac, Asia is expected to be the main driver for the volume growth and price/mix will also continue to rise, which should consequently drive the gross margin enhancement in future years. The limited supply of the cognac category and the strong demand for premium spirits, particularly in Asia, justify a high price/mix (see the previous section on "Revenue Forecast") and a corresponding higher gross margin in future periods. Gross margin for the company is expected to widen by 0.6%, 0.8%, and 1.2% in each of the next three year before decelerating as the contribution effect of price/mix on revenue slows over time. Please note that the adoption of IFRS 15 has resulted in a reported gross margin of 63.1% for 2019, a figure used as a basis for the future period forecasts.

12.3 Selling, General, and Administrative (SG&A) Expenses

Distribution costs increased significantly over time, from 26.1% of revenue in 2009 (not shown in the exhibits) to over 38% in 2015. Since then, they have remained at their historically high level because of higher advertising and promotion (A&P) costs and expansion of the company's distribution network. In particular, the setup of Rémy's distribution network in Asia increased the cost base. Rémy is very committed to its brand building and is also diversifying geographically. We estimate a further increase in distribution costs as a percentage of revenue, albeit at a slower rate. The adoption of IFRS 15 has resulted in the distribution costs reported at 30.8% of sales in 2019, and that figure is the basis of future forecasts. Administrative costs as a percentage of revenue have remained within the 8% to 9% range and are assumed to remain constant at 9%, broadly in line with 2019.

Exhibit 45 provides a consolidated income statement for Rémy.

Exhibit 45 Consolidated Historical and Projected Income Statement for Rémy Cointreau Group (€ millions, unless noted)

	2016	2017	2018	2019	2020E	2021E	2022E
Sales	1,051	1,095	1,127	1,126	1,145	1,220	1,300
Cost of sales	385	364	366	415	415	432	448
Gross profit	666	731	761	711	730	788	855
Gross margin	63.4%	66.7%	67.5%	63.1%	63.8%	64.6%	65.5%
Change in gross margin		3.4%	0.8%	–4.4%	0.6%	0.8%	1.2%
Distribution costs	407	417	433	346	350	375	407
Distribution costs as % of sales	38.7%	38.1%	38.4%	30.8%	30.6%	30.7%	31.3%
Administrative expenses	82	89	92	101	103	110	117
Administrative expenses as % of sales	7.8%	8.1%	8.1%	8.9%	9.0%	9.0%	9.0%
Other income from operations	1	–4	–13	2	2	2	2
EBIT	178	221	223	266	279	305	331
EBIT %	17.0%	20.2%	19.8%	23.6%	24.4%	25.0%	25.4%
Depreciation and amortization (add-back)	19	20	22	30	35	46	56
Depreciation and amortization as % of sales	1.8%	1.9%	1.9%	2.7%	3.1%	3.8%	4.3%
EBITDA	198	242	245	296	314	351	387

Exhibit 45 (Continued)

	2016	2017	2018	2019	2020E	2021E	2022E
EBITDA margin	18.8%	22.1%	21.7%	26.3%	27.4%	28.8%	29.7%
Finance costs	24	21	14	14	18	18	18
Other financial expenses	3	11	8	19	0	0	0
Total financial expenses	27	32	22	33	18	18	18
Profit before tax	151	189	201	233	261	287	313
Income tax	44	45	54	68	73	80	88
Tax rate	29.1%	23.5%	26.6%	29.0%	28.0%	28.0%	28.0%
Income from associates	–5	–20	1	–7	0	0	0
Profit from continuing operations	103	125	148	159	188	206	225
Profit/loss from discontinued operations	0	65	0	0	0	0	0
Net profit for the year	103	190	148	159	188	206	225
YoY %							
EPS basic continuing operations in €	2.11	2.55	2.97	3.17	3.69	3.97	4.24
EPS diluted continuing operations in €	2.11	2.42	2.83	3.02	3.52	3.79	4.05
EPS basic total in €	2.11	3.87	2.98	3.18	3.69	3.97	4.24
EPS diluted total in €	2.11	3.68	2.83	3.03	3.52	3.79	4.05
Average number of shares, basic in millions	48.6	49.1	49.8	50.1	51.0	52.0	53.0
Average number of shares, diluted in millions	48.7	51.8	52.4	52.7	53.0	53.0	53.0

Sources: Based on information from Rémy Cointreau Group and authors' analysis.

12.4 Operating Profit by Division

In this section, we estimate operating profit by division for Rémy. If the cost of goods sold, SG&A costs, and other income from operations are subtracted from revenue, the result is EBIT (a proxy for operating profit). This number for consolidated operations should match the cumulative EBIT of the individual segments. For the Cognac segment (Rémy Martin), the forecast of higher revenue growth assumes an improving product mix that will also result in a higher gross margin. But the benefit to gross margin will be somewhat mitigated by higher advertising and distribution costs. For 2020, we estimate an unchanged operating margin, but in the following years, the expectation is that the Cognac segment's operating margin will increase to just over 31.2% by 2022. As a benchmark, this forecast can be compared with the financial results reported by Hennessy (part of LVMH), another cognac brand. That company's operating margin in the wine and spirits segment in 2017–2018 was 30%–32%.

For the other segments, there is not much upside. In the Liqueurs & Spirits division, we assume operating margin to remain at 14%. In total, Rémy Cointreau Group's consolidated operating margin is forecast to improve from 23.5% in 2019 to 25.4% in 2022, largely as a result of growth in the Cognac segment, the most profitable division.

Exhibit 46 Historical and Projected Operating Profit by Segment for Rémy Cointreau Group

	2016	2017	2018	2019	2020E	2021E	2022E
Revenues (€ millions)							
Rémy Martin	648	708	760	774	846	915	989
Liqueurs & Spirits	274	276	267	264	270	275	281
Partner Brands	129	111	100	87	30	30	30
Total revenues	1,051	1,095	1,127	1,126	1,145	1,220	1,300
Operating profit (€ millions)							
Rémy Martin	140	185	204	236	257	283	309
Liqueurs & Spirits	48	58	43	39	38	39	39
Partner Brands	6	2	5	5	1	1	1
Holding costs	15	19	16	15	17	17	18
Total operating profit	178	226	237	264	279	305	331
Operating profit as % of revenue							
Rémy Martin	21.6	26.2	26.9	30.4	30.4	30.9	31.2
Liqueurs & Spirits	17.5	20.8	16.0	14.7	14.0	14.0	14.0
Partner Brands	4.7	1.8	5.3	5.6	4.0	4.0	4.0
Operating margin	17.0	20.7	21.0	23.5	24.9	25.1	25.4
Operating margin change (%)							
Rémy Martin	0.8	4.6	0.7	3.5	0.0	0.5	0.3
Liqueurs & Spirits	−2.2	3.3	−4.8	−1.4	−0.7	0.0	0.0
Partner Brands	−0.6	−2.9	3.5	0.3	−1.6	0.0	0.0
Operating margin change	0.8	3.7	0.4	2.4	0.9	0.6	0.4

Sources: Based on information from Rémy Cointreau Group and authors' analysis.

12.5 Non-Operating Expenses

Two types of non-operating expenses are included in the model: finance expenses (i.e., interest expenses) and income taxes.

Exhibit 47 shows the computation of finance costs. Finance costs require estimating the debt and cash position. Companies usually pay a fixed or variable interest rate on debt. If the interest rate is variable, the rate would be determined from existing market rates. A credit spread is normally applied to an estimate of the benchmark (e.g., reference interest rate) being used. In Exhibit 47, finance costs are fixed and calculated as 3.9% incurred on gross debt at the beginning of the period (€522 billion) minus 1% earned on the cash position (€179 million) at the beginning of the period. Other financial expenses are assumed to be zero. Estimated finance costs for fiscal year 2020 total €18 million.

Exhibit 47 Debt Position and Financial Costs and Income for Rémy (€ millions, unless noted)

	2016	2017	2018	2019	2020E	2021E	2022E
Long-term financial debt	172	393	397	424	424	424	424
Short-term financial debt and accrued interest	333	75	73	98	98	98	98
Gross debt	505	468	470	522	522	522	522
Cash and cash equivalents	–47	–78	–187	–179	–242	–268	–279
Net debt	458	390	283	343	280	254	243
Average gross debt		487	467	496	522	522	522
Average net debt		424	337	314	312	267	249
Net finance costs (cost of financial debt)		21	14	14	18	18	18

Sources: Based on information from Rémy Cointreau Group and authors' analysis.

12.6 Corporate Income Tax Forecast

In line with the average for 2018 and 2019, the tax rate is set at 28% for the future periods. Rémy Cointreau Group has no significant minority interests in any of its subsidiaries.

CONSTRUCTION OF PRO FORMA CASH FLOW STATEMENT AND BALANCE SHEET AND VALUATION INPUTS

13

m demonstrate the development of a sales-based pro forma company model.

To calculate the balance sheet for the end of the year, combine the projections made for the income statement with the expected cash flows during the year. These cash flows could be a direct result of the operational activities, cash proceeds from revenues, and cash outflows from costs. Based on the expected volume trends, the necessary production capacity and corresponding capital investments and cash outlays for the coming years can be budgeted. Dividend payments, share repurchases, and debt redemptions are financial cash flows that will affect the balance sheet and also need to be taken into account.

13.1 Capital Investments and Depreciation Forecasts

Capital investment, or capex, as a percentage of revenue was 3.6% in FY2019. In the previous three years, capex as a percentage of sales was 3.0%, on average. Given the healthy volume growth prospects, we expect capex to increase and assume it will reach 4.5% of sales in 2021. With Rémy's growing fixed asset base, it is logical that depreciation will increase. A gradual growth of depreciation as a percentage of sales with annual increases brings the level to 3.8% of sales in 2021. The breakdowns of capex, depreciation, and amortization are shown in Exhibit 48.

Exhibit 48 Capex, Depreciation, and Amortization Breakdowns

	2016	2017	2018	2019	2020E	2021E	2022E
Depreciation and amortization (€ millions)	19	20	22	30	35	46	56
As % of sales	1.8%	1.9%	1.9%	2.7%	3.1%	3.8%	4.3%
As % of fixed assets	2.6%	2.7%	2.9%	3.9%	4.4%	5.7%	6.9%
Capex (€ millions)	30	31	33	40	40	43	46
Capex as % of sales	2.9%	3.3%	2.9%	3.6%	4.0%	4.5%	5.0%
As % of fixed assets	4.2%	4.8%	4.4%	5.1%	5.8%	6.8%	8.0%
Capex/(depreciation and amortization)	1.6	1.8	1.5	1.3	1.3	1.2	1.2

Sources: Based on information from Rémy Cointreau Group and authors' analysis.

13.2 Working Capital Forecasts

We have assumed that working capital as a percentage of sales will remain similar to what the company experienced in the 2016–2019 period. In Exhibit 49, we include only the relevant balance sheet items related to revenues and costs (i.e., inventories, trade and other receivables, and trade and other payables) and keep the other items constant. Rémy Cointreau Group had positive working capital of more than 80% of its sales in fiscal year 2019. The largest working capital component is inventory, which accounts for almost 111% of annual revenues. Given the strong sales growth, working capital also increases. After a partial return to trend level in 2020, we expect working capital to increase again in absolute terms and to have a negative impact on operational cash flow in 2021 and beyond.

Exhibit 49 Working Capital Development for Rémy

	2016	2017	2018	2019	2020E	2021E	2022E
Inventories (€ millions)	1,108	1,145	1,170	1,246	1,220	1,280	1,350
Trade and other receivables	233	224	209	271	280	290	310
Trade and other payables	−499	−504	−517	−544	−550	−590	−620
Working capital	842	865	863	973	950	980	1,040
Inventories at year end as % of sales	105	105	104	111	107	105	104
Trade and other receivables at year end as % of sales	22	20	19	24	24	24	24
Trade and other payables at year end as % of sales	−48	−46	−46	−48	−48	−48	−48

Exhibit 49 (Continued)

	2016	2017	2018	2019	2020E	2021E	2022E
Working capital at year end as % of sales	80	79	77	86	83	80	80
Inventories as % of sales at year end change		−1	−1	7	−4	−2	−1
Trade and other receivables at year end as % of sales change		−2	−2	5	0	−1	0
Trade and other payables at year end as % of sales change		−2	0	2	0	0	−1
Working capital as % at year end of sales change		−1	−2	10	−3	−3	0
Absolute change in inventories at year end (€ millions)		37	25	75	-26	60	70
Absolute change in trade and other receivables at year end		−9	−14	61	9	10	20
Absolute change in trade and other payables at year end		5	14	27	6	40	30
Absolute change in working capital at year end		24	−3	110	−23	30	60

Sources: Based on information from Rémy Cointreau Group and authors' analysis.

13.3 Forecasted Cash Flow Statement

With the operating profit, capex, and working capital estimates already in place, the cash flow statement is almost automatically generated. The company has raised the dividends from €1.65 per share (from the 2018 earnings) to €2.65 per share (from the 2019 earnings). As shareholders had the option of receiving the dividend in the form of shares, the actual cash dividend payment was relatively modest at €9 million in fiscal year 2019. The company also bought back more than €27 million and €104 million of its shares in 2018 and 2019, respectively. Going forward, the model assumes growing dividends will be paid out in cash during the forecast period.

Exhibit 50 Projected Statement of Cash Flows for Rémy (€ millions)

	2016	2017	2018	2019	2020E	2021E	2022E
Current operating profit	178	226	237	264	278	303	28
Adjustment for depreciation, amortization, and impairments	19	20	22	30	35	46	56
Adjustment for share-based payments	2	2	3	3	0	0	0
Dividends received from associates	1	0	1	1	1	1	1
EBITDA	200	249	262	298	313	350	385
Change in inventories	−6	−26	−33	−70	26	−60	−70

(continued)

Exhibit 50 (Continued)

	2016	2017	2018	2019	2020E	2021E	2022E
Change in trade receivables	21	4	3.5	−46	−9	−10	−20
Change in trade payables	−26	18	16	−46	6	40	30
Change in other receivables and payables	−31	−32	6	0	0	0	0
Change in working capital	−42	−35	−7	−162	23	−30	−60
Net cash flow from operations	157	214	254	137	336	320	325
Other operating income/expense	0	−4	−1	−4	0	0	0
Net financial income	−23	−23	−12	−15	−18	−18	−18
Net income tax	−30	−52	−56	−64	−73	−80	−88
Total other operating cash flow	−53	−79	−70	−83	−91	−98	−105
Net cash flow from operating activities	104	135	185	53	245	221	219
Capital expenditures	−31	−37	−34	−45	−46	−55	−65
Purchase of share in associates	−1	−48					
Disposals	2	2	2	6			
Net other cash flow from investments	1	−1	1	86	0	0	0
Net cash flow from investment activities	29	−84	−32	47	−46	−55	−65
Treasury shares	−1	0	−27	−104	0	0	0
Payment of lease liabilities				−6			
Change in financial debt	−33	−5	0	17	0	0	0
Dividends paid to shareholders	−73	−13	−25	−9	−135	−140	−143
Net cash flow from financing activities	−107	−18	−52	−102	−135	−140	−143
Translation differences cash	5	−2	8	−6	0	0	0
Change in cash and cash equivalents	−27	−31	109	−8	64	26	11
Cash at the beginning	74	47	78	187	179	242	268
Cash at the end	47	78	187	179	242	268	279

Note: Apparent small discrepancies in addition reflect the effects of rounding error.
Sources: Based on information from Rémy Cointreau Group and authors' analysis.

13.4 Forecasted Balance Sheet

The forecasted balance sheet is given in Exhibit 51 and is based on the combination of the projected income statement (Exhibit 45), the projected statement of cash flows (Exhibit 50), and the historical starting balance sheet. The balance sheet items that were not specifically discussed are mostly held constant.

Exhibit 51 Projected Balance Sheet for Rémy (€ millions)

	2016	2017	2018	2019	2020E	2021E	2022E
Brands and other intangible assets	488	526	509	543	543	543	543
Property, plant, and equipment	223	237	243	269	280	289	298
Investments in associates	41	22	20	1	1	1	1
Other financial assets	95	167	166	94	94	94	94
Deferred tax assets	29	30	20	18	18	18	18
Total non-current assets	875	983	959	924	936	945	954
Inventories	1,108	1,145	1,170	1,245	1,220	1,280	1,350
Trade and other receivables	233	224	210	271	280	290	310
Cash and cash equivalents	47	78	187	179	242	268	279
Other current assets	19	12	16.2	5.1	5	5	5
Total current assets	1,407	1,459	1,583	1,700	1,747	1,843	1,944
Total assets	2,282	2,442	2,542	2,625	2,683	2,788	2,899
Share capital	78	80	80	80	80	80	80
Share premium	695	759	805	795	795	795	795
Treasury shares	−9	−8	−20.5	−34	−34	−34	−34
Consolidated reserves	320	446	518	558	611	677	759
Net profit to owners of the company	1,084	1,276	1,383	1,399	1,452	1518	1,600
Translation reserve	27.5	26.8	24	26	25	25	25
Equity attributable to shareholders	1,112	1,302	1,407	1,425	1,477	1,543	1,625
Non-controlling interest	1	2	1	1	1	1	1
Equity	1,113	1,304	1,408	1,426	1,478	1,544	1,626
Long-term financial debt	172	393	397	424	424	424	424
Provision for employee benefits	31	32	33	32	32	32	32
Long-term provisions for liabilities and charges	6	7	7	8	8	8	8
Deferred tax liabilities	101	99	81	62	62	62	62
Total non-current liabilities	309	530	518	526	526	526	526
Short-term financial debt and accrued interest	333	76	73	98	98	98	98
Trade and other payables	499	504	517	5448	550	590	620
Income tax payable	10	11	10	18	18	18	18

(continued)

Exhibit 51 (Continued)

	2016	2017	2018	2019	2020E	2021E	2022E
Short-term provisions for liabilities and charges	13	11	14	2	2	2	2
Derivative financial instruments	1	7	2	10	10	10	10
Liabilities held for sale	2	0			0	0	0
Current liabilities	859	608	616	673	679	679	679
Total equity and liabilities	2,282	2,442	2,542	2,625	2,683	2,788	2,899

Sources: Based on information from Rémy Cointreau Group and authors' analysis.

13.5 Valuation Inputs

In the previous sections, we have built a model that projects the future profit and loss, cash flow statement, and balance sheet for Rémy Cointreau Group. This model is the starting point for valuation. Most company-specific metrics can be found in the model. Valuation estimates can be made based on a variety of metrics, including free cash flow, earnings per share, EBITDA, or EBIT. The company-specific inputs needed to build a discounted cash flow model are shown in Exhibit 52. The first line in Exhibit 52 is from Exhibit 45, excluding other operating expenses. Depreciation and amortization are from Exhibit 48. The remaining data are from Exhibit 50.

Exhibit 52 Calculating Free Cash Flow as Basis for a DCF Valuation (€ millions)

	2019	2020E	2021E	2022E
Normalized operating profit	303	279	305	330
Taxes (28% tax rate)	68	73	80	87
Normalized operating profit after tax	235	206	224	243
Depreciation and amortization	30	35	46	56
	266	241	270	299
Change in working capital	−162	23	−30	−60
Capital expenditures	−40	−46	−55	−65
Free cash flow to the firm	64	218	185	174

Source: Based on the authors' analysis.

14 CONCLUSIONS AND SUMMARY

Industry and company analysis are essential tools of fundamental analysis. The key points made include the following:

- Analysts can use a top-down, bottom-up, or hybrid approach to forecasting income and expenses. Top-down approaches usually begin at the level of the overall economy. Bottom-up approaches begin at the level of the individual company or unit within the company (e.g., business segment). Time-series approaches are considered bottom-up, although time-series analysis can be a tool used in top-down approaches. Hybrid approaches include elements of top-down and bottom-up approaches.

- In a "growth relative to GDP growth" approach to forecasting revenue, the analyst forecasts the growth rate of nominal gross domestic product and industry and company growth relative to GDP growth.
- In a "market growth and market share" approach to forecasting revenue, the analyst combines forecasts of growth in particular markets with forecasts of a company's market share in the individual markets.
- Operating margins that are positively correlated with sales provide evidence of economies of scale in an industry.
- Some balance sheet line items, such as retained earnings, flow directly from the income statement, whereas accounts receivable, accounts payable, and inventory are very closely linked to income statement projections.
- A common way to model working capital accounts is to use efficiency ratios.
- Return on invested capital (ROIC), defined as net operating profit less adjusted taxes divided by the difference between operating assets and operating liabilities, is an after-tax measure of the profitability of investing in a company. High and persistent levels of ROIC are often associated with having a competitive advantage.
- Competitive factors affect a company's ability to negotiate lower input prices with suppliers and to raise prices for products and services. Porter's five forces framework can be used as a basis for identifying such factors.
- Inflation (deflation) affects pricing strategy depending on industry structure, competitive forces, and the nature of consumer demand.
- When a technological development results in a new product that threatens to cannibalize demand for an existing product, a unit forecast for the new product combined with an expected cannibalization factor can be used to estimate the impact on future demand for the existing product.
- Factors influencing the choice of the explicit forecast horizon include the projected holding period, an investor's average portfolio turnover, cyclicality of an industry, company-specific factors, and employer preferences.

REFERENCES

Carruthers, NicolaJune . 2018. "The Spirits Business." Best-Selling Cognac and Brandy Brands." https://www.thespiritsbusiness.com/2018/06/the-top-10-best-selling-cognac-and-brandy-brands/

Hawawini, Gabriel, and Claude Viallet. 2015. *Finance for Executives: Managing for Value Creation*, 5th edition. Mason, Ohio: South-Western Cengage Learning.

Porter, Michael E. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York, NY: The Free Press.

Porter, Michael E. 2008. "The Five Competitive Forces That Shape Strategy." *Harvard Business Review* (January).

PRACTICE PROBLEMS

The following information relates to Questions 1–6

Angela Green, an investment manager at Horizon Investments, intends to hire a new investment analyst. After conducting initial interviews, Green has narrowed the pool to three candidates. She plans to conduct second interviews to further assess the candidates' knowledge of industry and company analysis.

Prior to the second interviews, Green asks the candidates to analyze Chrome Network Systems, a company that manufactures internet networking products. Each candidate is provided Chrome's financial information presented in Exhibit 1.

Exhibit 1 Chrome Network Systems Selected Financial Information ($ millions)

	Year-End:		
	2017	2018	2019
Net sales	46.8	50.5	53.9
Cost of sales	18.2	18.4	18.8
Gross profit	28.6	32.1	35.1
Selling, general, and administrative (SG&A) expenses	19.3	22.5	25.1
Operating income	9.3	9.6	10.0
Interest expense	0.5	0.7	0.6
Income before provision for income tax	8.8	8.9	9.4
Provision for income taxes	2.8	2.8	3.1
Net income	6.0	6.1	6.3

Green asks each candidate to forecast the 2020 income statement for Chrome and to outline the key assumptions used in their analysis. The job candidates are told to include Horizon's economic outlook for 2020 in their analysis, which assumes nominal GDP growth of 3.6%, based on expectations of real GDP growth of 1.6% and inflation of 2.0%.

Green receives the models from each of the candidates and schedules second interviews. To prepare for the interviews, Green compiles a summary of the candidates' key assumptions in Exhibit 2.

© 2015 CFA Institute. All rights reserved.

Exhibit 2 Summary of Key Assumptions Used in Candidates' Models

Metric	Candidate A	Candidate B	Candidate C
Net sales	Net sales will grow at the average annual growth rate in net sales over the 2017–2019 time period.	Industry sales will grow at the same rate as nominal GDP, but Chrome will have a 2 percentage point decline in market share.	Net sales will grow 50 basis points slower than nominal GDP.
Cost of sales	2020 gross margin will be the same as the average annual gross margin over the 2017–2019 time period.	2020 gross margin will decline as costs increase by expected inflation.	2020 gross margin will increase by 20 basis points from 2019.
Selling, general, and administrative (SG&A) expenses	2020 SG&A/net sales ratio will be the same as the average ratio over the 2017–2019 time period.	2020 SG&A will grow at the rate of inflation.	2020 SG&A/net sales ratio will be the same as the 2019 ratio.
Interest expense	2020 interest expense assumes the effective interest rate will be the same as the 2019 rate.	2020 interest expense will be the same as the 2019 interest expense.	2020 interest expense will be the same as the average expense over the 2017–2019 time period.
Income taxes	2020 effective tax rate will be the same as the 2019 rate.	2020 effective tax rate will equal the blended statutory rate of 30%.	2020 effective tax rate will be the same as the average effective tax rate over the 2017–2019 time period.

1 Based on Exhibit 1, which of the following provides the strongest evidence that Chrome displays economies of scale?

A Increasing net sales

B Profit margins that are increasing with net sales

C Gross profit margins that are increasing with net sales

2 Based on Exhibit 2, the job candidate *most likely* using a bottom-up approach to model net sales is:

A Candidate A.

B Candidate B.

C Candidate C.

3 Based on Exhibit 2, the modeling approach used by Candidate B to project future net sales is *most accurately* classified as a:

A hybrid approach.

B top-down approach.

C bottom-up approach.

4 Based on Exhibits 1 and 2, Candidate C's forecast for cost of sales in 2020 is *closest* to:

A $18.3 million.

B $18.9 million.

C $19.3 million.

5 Based on Exhibits 1 and 2, Candidate A's forecast for selling, general, and administrative expenses in 2020 is *closest* to:

A $23.8 million.

B $25.5 million.

C $27.4 million.

6 Based on Exhibit 2, forecasted interest expense will reflect changes in Chrome's debt level under the forecast assumptions used by:

A Candidate A.

B Candidate B.

C Candidate C.

The following information relates to Questions 7–12

Nigel French, an analyst at Taurus Investment Management, is analyzing Archway Technologies, a manufacturer of luxury electronic auto equipment, at the request of his supervisor, Lukas Wright. French is asked to evaluate Archway's profitability over the past five years relative to its two main competitors, which are located in different countries with significantly different tax structures.

French begins by assessing Archway's competitive position within the luxury electronic auto equipment industry using Porter's five forces framework. A summary of French's industry analysis is presented in Exhibit 1.

Exhibit 1 Analysis of Luxury Electronic Auto Equipment Industry Using Porter's Five Forces Framework

Force	Factors to Consider
Threat of substitutes	Customer switching costs are high
Rivalry	Archway holds 60% of world market share; each of its two main competitors holds 15%
Bargaining power of suppliers	Primary inputs are considered basic commodities, and there are a large number of suppliers
Bargaining power of buyers	Luxury electronic auto equipment is very specialized (non-standardized)
Threat of new entrants	High fixed costs to enter industry

French notes that for the year just ended (2019), Archway's cost of goods sold was 30% of sales. To forecast Archway's income statement for 2020, French assumes that all companies in the industry will experience an inflation rate of 8% on the cost of goods sold. Exhibit 2 shows French's forecasts relating to Archway's price and volume changes.

Exhibit 2	Archway's 2020 Forecasted Price and Volume Changes
Average price increase per unit	5.00%
Volume growth	–3.00%

After putting together income statement projections for Archway, French forecasts Archway's balance sheet items. He uses Archway's historical efficiency ratios to forecast the company's working capital accounts.

Based on his financial forecast for Archway, French estimates a terminal value using a valuation multiple based on the company's average price-to-earnings multiple (P/E) over the past five years. Wright discusses with French how the terminal value estimate is sensitive to key assumptions about the company's future prospects. Wright asks French:

> "What change in the calculation of the terminal value would you make if a technological development that would adversely affect Archway was forecast to occur sometime beyond your financial forecast horizon?"

7 Which return metric should French use to assess Archway's five-year historic performance relative to its competitors?

A Return on equity

B Return on invested capital

C Return on capital employed

8 Based on the current competitive landscape presented in Exhibit 1, French should conclude that Archway's ability to:

A pass along price increases is high.

B demand lower input prices from suppliers is low.

C generate above-average returns on invested capital is low.

9 Based on the current competitive landscape presented in Exhibit 1, Archway's operating profit margins over the forecast horizon are *least likely* to:

A decrease.

B remain constant.

C increase.

10 Based on Exhibit 2, Archway's forecasted gross profit margin for 2020 is *closest* to:

A 62.7%.

B 67.0%.

C 69.1%.

11 French's approach to forecasting Archway's working capital accounts would be *most likely* classified as a:

A hybrid approach.

B top-down approach.

C bottom-up approach.

12 The *most appropriate* response to Wright's question about the technological development is to:

A increase the required return.

B decrease the price-to-earnings multiple.

C decrease the perpetual growth rate.

The following information relates to Questions 13–18

Gertrude Fromm is a transportation sector analyst at Tucana Investments. She is conducting an analysis of Omikroon, N.V., a (hypothetical) European engineering company that manufactures and sells scooters and commercial trucks.

Omikroon's petrol scooter division is the market leader in its sector and has two competitors. Omikroon's petrol scooters have a strong brand name and a well-established distribution network. Given the strong branding established by the market leaders, the cost of entering the industry is high. But Fromm anticipates that inexpensive, small imported petrol-fueled motorcycles may become substitutes for Omikroon's petrol scooters.

Fromm uses return on invested capital as the metric to assess Omikroon's performance.

Omikroon has just introduced the first electric scooter to the market at year-end 2019. The company's expectations are as follows:

- Competing electric scooters will reach the market in 2021.
- Electric scooters will not be a substitute for petrol scooters.
- The important research costs in 2020 and 2021 will lead to more efficient electric scooters.

Fromm decides to use a five-year forecast horizon for Omikroon after considering the following factors:

Factor 1 The annual portfolio turnover at Tucana Investments is 30%.

Factor 2 The electronic scooter industry is expected to grow rapidly over the next 10 years.

Factor 3 Omikroon has announced it would acquire a light truck manufacturer that will be fully integrated to its truck division by 2021 and will add 2% to its total revenues.

Fromm uses the base case forecast for 2020 shown in Exhibit 1 to perform the following sensitivity analysis:

- The price of an imported specialty metal used for engine parts increases by 20%.
- This metal constitutes 4% of Omikroon's cost of sales.
- Omikroon will not be able to pass on the higher metal expense to its customers.

Exhibit 1 Omikroon's Selected Financial Forecasts for 2020 Base Case (€ millions)

	Petrol Scooter Division	Commercial Truck Division	Electric Scooter Division	Total
Sales	99.05	45.71	7.62	152.38
Cost of sales				105.38
Gross profit				47.00
Operating profit				9.20

Omikroon will initially outsource its electric scooter parts. But manufacturing these parts in-house beginning in 2021 will imply changes to an existing factory. This factory cost €7 million three years ago and had an estimated useful life of 10 years. Fromm is evaluating two scenarios:

Scenario 1 Sell the existing factory for €5 million. Build a new factory costing €30 million with a useful life of 10 years.

Scenario 2 Refit the existing factory for €27 million.

13 Using Porter's five forces analysis, which of the following competitive factors is likely to have the *greatest* impact on Omikroon's petrol scooter pricing power?

A Rivalry

B Threat of substitutes

C Threat of new entrants

14 The metric used by Fromm to assess Omikroon's performance takes into account:

A degree of financial leverage.

B operating liabilities relative to operating assets.

C competitiveness relative to companies in other tax regimes.

15 Based on Omikroon's expectations, the gross profit margin of Omikroon's electric scooter division in 2021 is *most likely* to be affected by:

A competition.

B research costs.

C cannibalization by petrol scooters.

16 Which factor *best* justifies the five-year forecast horizon for Omikroon selected by Fromm?

A Factor 1

B Factor 2

C Factor 3

17 Fromm's sensitivity analysis will result in a decrease in the 2020 base case gross profit margin *closest to*:

A 0.55 percentage points.

B 0.80 percentage points.

C 3.32 percentage points.

18 Fromm's estimate of growth capital expenditure included in Omikroon's property, plant, and equipment under Scenario 2 should be:

A lower than under Scenario 1.
B the same as under Scenario 1.
C higher than under Scenario 1.

SOLUTIONS

1 C is correct. Economies of scale are a situation in which average costs decrease with increasing sales volume. Chrome's gross margins have been increasing with net sales. Gross margins that increase with sales levels provide evidence of economies of scale, assuming that higher levels of sales reflect increased unit sales. Gross margin more directly reflects the cost of sales than does profit margin.

Metric	2017	2018	2019
Net sales	$46.8	$50.5	$53.9
Gross profit	28.6	32.1	35.1
Gross margin (gross profit/net sales)	61.11%	63.56%	65.12%

2 A is correct. A bottom-up approach for developing inputs to equity valuation models begins at the level of the individual company or a unit within the company. By modeling net sales using the average annual growth rate, Candidate A is using a bottom-up approach. Both Candidate B and Candidate C are using a top-down approach, which begins at the level of the overall economy.

3 B is correct. A top-down approach usually begins at the level of the overall economy. Candidate B assumes industry sales will grow at the same rate as nominal GDP but that Chrome will have a 2 percentage point decline in market share. Candidate B is not using any elements of a bottom-up approach; therefore, a hybrid approach is not being employed.

4 C is correct. Candidate C assumes that the 2020 gross margin will increase by 20 bps from 2019 and that net sales will grow at 50 bps slower than nominal GDP (nominal GDP = Real GDP + Inflation = 1.6% + 2.0% = 3.6%). Accordingly, the 2020 forecasted cost of sales is $19.27 million, rounded to $19.3 million.

Metric	Calculation	Result
2020 gross margin = 2019 gm + 20 bps	$35.1/$53.9 = 65.12% + 0.20% =	65.32%
2020 CoS/net sales = 100% – gross margin	100% – 65.32% =	34.68%
2020 net sales = 2019 net sales × (1 + Nominal GDP – 0.50%)	$53.9 million × (1 + 0.036 – 0.005) = $53.9 million × 1.031 =	$55.57 million
2020 cost of sales = 2020 net sales × CoS/net sales	$55.57 × 34.68% =	$19.27 million

5 B is correct. Candidate A assumes that the 2020 SG&A/net sales will be the same as the average SG&A/net sales over the 2017–2019 time period and that net sales will grow at the annual average growth rate in net sales over the 2017–2019 time period. Accordingly, the 2020 forecasted selling, general, and administrative expenses are $25.5 million.

Metric	Calculation	Result
Average SG&A/net sales, 2017–2019*	(41.24% + 44.55% + 46.57%)/3 =	44.12%
Average annual growth sales in net sales, 2017–2019**	(7.91% + 6.73%)/2 =	7.32%

(continued)

Metric	Calculation	Result
2020 net sales = 2019 net sales × (1 + Average annual growth rate in net sales)	$53.9 million × 1.0732 =	$57.85 million
2020 SG&A = 2020 net sales × Average SG&A/net sales	$57.85 million × 44.12% =	$25.52 million

* SG&A/net sales are calculated as follows:

	2017	2018	2019
Net Sales	$46.8	$50.5	$53.9
SG&A expenses	$19.3	$22.5	$25.1
SG&A-to-sales ratio	41.24%	44.55%	46.57%

** Growth rate in net sales is calculated as follows:

Year	Calculation
2018	($50.5/$46.8) – 1 = 7.91%
2019	($53.9/$50.5) – 1 = 6.73%

6 A is correct. In forecasting financing costs, such as interest expense, the debt/equity structure of a company is a key determinant. Accordingly, a method that recognizes the relationship between the income statement account (interest expense) and the balance sheet account (debt) would be a preferable method for forecasting interest expense when compared with methods that forecast based solely on the income statement account. By using the effective interest rate (interest expense divided by average gross debt), Candidate A is taking the debt/equity structure into account. Candidate B (who forecasts 2020 interest expense to be the same as 2019 interest expense) and Candidate C (who forecasts 2020 interest expense to be the same as the 2017–2019 average interest expense) are not taking the balance sheet into consideration.

7 C is correct. The return on capital employed (ROCE) is a pre-tax return measure that can be useful in the peer comparison of companies in countries with different tax structures. Archway's two main competitors are located in different countries with significantly different tax structures, and therefore, a pre-tax measure of return on capital employed is better than an after-tax measure.

8 A is correct. Porter's five forces framework in Exhibit 1 describes an industry with high barriers to entry, high customer switching costs (suggesting a low threat of substitutes), and a specialized product (suggesting low bargaining power of buyers). Furthermore, the primary production inputs from the large group of suppliers are considered basic commodities (suggesting low bargaining power of suppliers). These favorable industry characteristics will likely enable Archway to pass along price increases and generate above-average returns on invested capital.

9 A is correct. The current favorable characteristics of the industry (high barriers to entry, low bargaining power of suppliers and buyers, low threat of substitutes), coupled with Archway's dominant market share position, will likely lead to Archway's profit margins being at least equal to or greater than current levels over the forecast horizon.

10 C is correct. The calculation of Archway's gross profit margin for 2020, which reflects the industry-wide 8% inflation on cost of goods sold (COGS), is calculated as follows:

Revenue growth	1.85%
Cost of goods sold increase	4.76%
Forecasted revenue (Base revenue = 100)	101.85
Forecasted COGS (Base COGS = 30)	31.43
Forecasted gross profit	70.42
Forecasted gross profit margin	69.14%

$$\begin{aligned}\text{Revenue growth} &= (1 + \text{Price increase for revenue}) \times (1 + \text{Volume growth}) - 1\\ &= (1.05) \times (0.97) - 1\\ &= 1.85\%.\end{aligned}$$

$$\begin{aligned}\text{COGS increase} &= (1 + \text{Price increase for COGS}) \times (1 + \text{Volume growth}) - 1\\ &= (1.08) \times (0.97) - 1\\ &= 4.76\%.\end{aligned}$$

$$\begin{aligned}\text{Forecasted revenue} &= \text{Base revenue} \times \text{Revenue growth increase}\\ &= 100 \times 1.0185\\ &= 101.85.\end{aligned}$$

$$\begin{aligned}\text{Forecasted COGS} &= \text{Base COGS} \times \text{COGS increase}\\ &= 30 \times 1.0476\\ &= 31.43.\end{aligned}$$

$$\begin{aligned}\text{Forecasted gross profit} &= \text{Forecasted revenue} - \text{Forecasted COGS}\\ &= 101.85 - 31.43\\ &= 70.42.\end{aligned}$$

$$\begin{aligned}\text{Forecasted gross profit margin} &= \text{Forecasted gross profit/Forecasted revenue}\\ &= 70.42/101.85\\ &= 69.14\%.\end{aligned}$$

11 C is correct. French is using a bottom-up approach to forecast Archway's working capital accounts by using the company's historical efficiency ratios to project future performance.

12 B is correct. If the future growth or profitability of a company is likely to be lower than the historical average (in this case, because of a potential technological development), then the target multiple should reflect a discount to the historical multiple to reflect this difference in growth and/or profitability. If a multiple is used to derive the terminal value of a company, the choice of the multiple should be consistent with the long-run expectations for growth and required return. French tells Wright he believes that such a technological development may have an adverse impact on Archway beyond the forecast horizon.

13 B is correct. Inexpensive, small imported petro-fueled motorcycles are substitutes for petrol scooters and may increasingly have an impact on Omikroon's petrol scooter pricing power.

14 B is correct. Return on invested capital is net operating profit minus adjusted taxes divided by invested capital, where invested capital is defined as operating assets minus operating liabilities.

15 A is correct. Competition from other electric scooter manufacturers is expected to begin in one year. After this time, competing electric scooters could lead to lower demand for Omikroon's electric scooters and affect Omikroon's gross profit margin.

16 B is correct. The electric scooter market is expected to grow rapidly and so the contribution of Omrikoon's new electric scooter division is forecast to expand significantly over the next 10 years. A is not correct because the investment company's portfolio turnover is not relevant for forecasting Omnrikoon's future results. C is not correct because the light truck division is expected to add only 2% to total revenues in the future.

17 A is correct. The sensitivity analysis consists of an increase of 20% in the price of an input that constitutes 4% of cost of sales. Change in gross profit margin because of that increase is calculated as the change in cost of sales because of price increase divided by sales:

$$= (\text{Cost of sales} \times 0.04 \times 0.2)/\text{Sales}$$
$$= (105.38 \times 0.04 \times 0.2)/152.38$$
$$= 0.0055$$

18 C is correct. In Scenario 2, growth capital expenditure of €27 million for the refit of the existing idle factory is higher than the growth capital expenditure in Scenario 1 of €25 million. The €25 million is the cost of building a new factory for €30 million less the proceeds from the sale of the existing idle factory of €5 million.

READING

23

Discounted Dividend Valuation

by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA

Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA). Thomas R. Robinson, PhD, CFA, CAIA, is at Robinson Global Investment Management LLC, (USA). John D. Stowe, PhD, CFA, is at Ohio University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** compare dividends, free cash flow, and residual income as inputs to discounted cash flow models and identify investment situations for which each measure is suitable;
☐	**b.** calculate and interpret the value of a common stock using the dividend discount model (DDM) for single and multiple holding periods;
☐	**c.** calculate the value of a common stock using the Gordon growth model and explain the model's underlying assumptions;
☐	**d.** calculate the value of non-callable fixed-rate perpetual preferred stock;
☐	**e.** calculate and interpret the implied growth rate of dividends using the Gordon growth model and current stock price;
☐	**f.** calculate and interpret the present value of growth opportunities (PVGO) and the component of the leading price-to-earnings ratio (P/E) related to PVGO;
☐	**g.** calculate and interpret the justified leading and trailing P/Es using the Gordon growth model;
☐	**h.** describe strengths and limitations of the Gordon growth model and justify its selection to value a company's common shares;
☐	**i.** explain the growth phase, transition phase, and maturity phase of a business;
☐	**j.** explain the assumptions and justify the selection of the two-stage DDM, the H-model, the three-stage DDM, or spreadsheet modeling to value a company's common shares;
☐	**k.** describe terminal value and explain alternative approaches to determining the terminal value in a DDM;

(continued)

© 2020 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**l.** calculate and interpret the value of common shares using the two-stage DDM, the H-model, and the three-stage DDM;
☐	**m.** explain the use of spreadsheet modeling to forecast dividends and to value common shares;
☐	**n.** estimate a required return based on any DDM, including the Gordon growth model and the H-model;
☐	**o.** calculate and interpret the sustainable growth rate of a company and demonstrate the use of DuPont analysis to estimate a company's sustainable growth rate;
☐	**p.** evaluate whether a stock is overvalued, fairly valued, or undervalued by the market based on a DDM estimate of value.

1 INTRODUCTION AND PRESENT VALUE METHODS

a compare dividends, free cash flow, and residual income as inputs to discounted cash flow models and identify investment situations for which each measure is suitable;

Common stock represents an ownership interest in a business. A business in its operations generates a stream of cash flows, and as owners of the business, common stockholders have an equity ownership claim on those future cash flows. Beginning with John Burr Williams (1938), analysts have developed this insight into a group of valuation models known as discounted cash flow (DCF) valuation models. DCF models—which view the intrinsic value of common stock as the present value of its expected future cash flows—are a fundamental tool in both investment management and investment research.

Although the principles behind discounted cash flow valuation are simple, applying the theory to equity valuation can be challenging. Four broad steps in applying DCF analysis to equity valuation are:

- choosing the class of DCF model—equivalently, selecting a specific definition of cash flow;
- forecasting the cash flows;
- choosing a discount rate methodology; and
- estimating the discount rate.

In our coverage of this topic, we take the perspective that dividends—distributions to shareholders authorized by a company's board of directors—are an appropriate definition of cash flows. The class of models based on this idea is called dividend discount models, or DDMs. The basic objective of any DDM is to value a stock. The variety of implementations corresponds to different ways to model a company's future stream of dividend payments. The steps of choosing a discount rate methodology and estimating the discount rate involve the same considerations for all DCF models, so they have been presented separately in an earlier discussion.

The sections are organized as follows: We first provide an overview of present value models. We then provide a general statement of the dividend discount model. Forecasting dividends, individually and in detail, into the indefinite future is not generally practicable, so the dividend-forecasting problem is usually simplified. One approach is to assign dividends to a stylized growth pattern. In the subsequent section, we focus on the simplest pattern—dividends growing at a constant rate forever (the constant growth or "Gordon growth" model). We then explain that for some companies, it is more appropriate to view earnings and dividends as having multiple stages of growth. We present multistage dividend discount models along with spreadsheet modeling. We lay out the determinants of dividend growth rates in the last section and conclude with a summary.

1.1 Present Value Models

Present value models as a group constitute a demanding and rigorous approach for valuing assets. In this section, we discuss the economic rationale for valuing an asset as the present value of its expected future cash flows. We also discuss alternative definitions of cash flows and present the major alternative methods for estimating the discount rate.

1.1.1 *Valuation Based on the Present Value of Future Cash Flows*

The value of an asset must be related to the benefits or returns we expect to receive from holding it. Those returns are called the asset's future cash flows (we will define *cash flow* more concretely and technically later). We also need to recognize that a given amount of money received in the future is worth less than the same amount of money received today. Money received today gives us the option of immediately spending and consuming it, so money has a time value. Therefore, when valuing an asset, before adding up the estimated future cash flows, we must **discount** each cash flow back to the present: the cash flow's value is reduced with respect to how far away it is in time. The two elements of discounted cash flow valuation—estimating the cash flows and discounting the cash flows to account for the time value of money—provide the economic rationale for discounted cash flow valuation. In the simplest case, in which the timing and amounts of future cash flows are known with certainty, if we invest an amount equal to the present value of future cash flows at the given discount rate, that investment will replicate all of the asset's cash flows (with no money left over).

For some assets, such as government debt, cash flows may be essentially known with certainty—that is, they are default risk free. The appropriate discount rate for such a risk-free cash flow is a risk-free rate of interest. For example, if an asset has a single, certain cash flow of \$100 to be received in two years, and the risk-free interest rate is 5% a year, the value of the asset is the present value of \$100 discounted at the risk-free rate, $\$100/(1.05)^2 = \$90.70$.

In contrast to risk-free debt, future cash flows for equity investments are not known with certainty—they are risky. Introducing risk makes applying the present value approach much more challenging. The most common approach to dealing with risky cash flows involves two adjustments relative to the risk-free case. First, discount the *expected* value of the cash flows, viewing the cash flows as random variables (note that the expected value of a random quantity is the mean value of its possible outcomes, in which each outcome's weight in the average is its probability of occurrence). Second, adjust the discount rate to reflect the risk of the cash flows.

The following equation expresses the concept that an asset's value is the present value of its (expected) future cash flows:

$$V_0 = \sum_{t=1}^{n} \frac{\text{CF}_t}{(1+r)^t} \tag{1}$$

where

V_0 = the value of the asset at time t = 0 (today)

n = number of cash flows in the life of the asset (n is set equal to ∞ for equities)

CF_t = the cash flow (or the expected cash flow, for risky cash flows) at time t

r = the discount rate or required rate of return

For simplicity, the discount rate in Equation 1 is represented as the same for all periods (i.e., a flat term structure of discount rates is assumed). The analyst has the latitude in this model, however, to apply different discount rates to different cash flows. Such action could reflect different degrees of cash flow riskiness or different risk-free rates at different time horizons. Differences in cash flow riskiness may be caused by differences in business risk, operating risk (use of fixed assets in production), or financial risk or leverage (use of debt in the capital structure). The simple expression given, however, is adequate for this discussion.

Equation 1 gives an asset's value from the perspective of today (t = 0). Likewise, an asset's value at some point in the future equals the value of all subsequent cash flows discounted back to that point in time. Example 1 illustrates these points.

EXAMPLE 1

Value as the Present Value of Future Cash Flows

An asset is expected to generate cash flows of \$100 in one year, \$150 in two years, and \$200 in three years. The value of this asset today, using a 10% discount rate, is

$$\begin{aligned} V_0 &= \frac{100}{(1.10)^1} + \frac{150}{(1.10)^2} + \frac{200}{(1.10)^3} \\ &= 90.909 + 123.967 + 150.263 = \$365.14 \end{aligned}$$

The value at t = 0 is \$365.14. The same logic is used to value an asset at a future date. The value of the asset at t = 1 is the present value, discounted back to t = 1, of all cash flows after this point. This value, V_1, is

$$\begin{aligned} V_1 &= \frac{150}{(1.10)^1} + \frac{200}{(1.10)^2} \\ &= 136.364 + 165.289 = \$301.65 \end{aligned}$$

At any point in time, the asset's value is the value of future cash flows (CF) discounted back to that point. Because V_1 represents the value of CF_2 and CF_3 at t = 1, the value of the asset at t = 0 is also the present value of CF_1 and V_1:

$$\begin{aligned} V_0 &= \frac{100}{(1.10)^1} + \frac{301.653}{(1.10)^1} \\ &= 90.909 + 274.23 = \$365.14 \end{aligned}$$

Finding V_0 as the present value of CF_1, CF_2, and CF_3 is logically equivalent to finding V_0 as the present value of CF_1 and V_1.

In the next section, we present an overview of three alternative definitions of cash flow. The selected cash flow concept defines the type of DCF model we can use: the dividend discount model, the free cash flow model, or the residual income model. We also broadly characterize the types of valuation problems for which analysts often choose a particular model. (Further details are supplied when each model is discussed individually.)

1.1.2 *Streams of Expected Cash Flows*

In present value models of stock valuation, the three most widely used definitions of returns are dividends, free cash flow, and residual income. We discuss each definition in turn.

The dividend discount model defines cash flows as dividends. The basic argument for using this definition of cash flow is that an investor who buys and holds a share of stock generally receives cash returns only in the form of dividends. In practice, analysts usually view investment value as driven by earnings. Does the definition of cash flow as dividends ignore earnings not distributed to shareholders as dividends? Reinvested earnings should provide the basis for increased future dividends. Therefore, the DDM accounts for reinvested earnings when it takes all future dividends into account. Because dividends are less volatile than earnings and other return concepts, the relative stability of dividends may make DDM values less sensitive to short-run fluctuations in underlying value than alternative DCF models. Analysts often view DDM values as reflecting long-run intrinsic value.

A stock either pays dividends or does not pay dividends. A company might not pay dividends on its stock because the company is not profitable and has no cash to distribute. Also, a company might not pay dividends for the opposite reason: because it is very profitable. For example, a company may reinvest all earnings—paying no dividends—to take advantage of profitable growth opportunities. As the company matures and faces fewer attractive investment opportunities, it may initiate dividends. Generally, mature, profitable companies tend to pay dividends and are reluctant to reduce the level of dividends (Grullon, Paye, Underwood, and Weston 2011).

Dividend policy practices have international differences and change through time, even in one market. Typically, research has shown that a lower percentage of companies in US stock markets have paid dividends than have companies in most other markets (He, Ng, Zaiats, and Zhang 2017), although the US sample may have included a disproportionate number of smaller and younger companies, which are less likely to pay dividends (Denis and Osobov 2008). Research has also shown a decline over time in the fraction of companies paying cash dividends in most developed markets such as the United States, Canada, the European Union, the United Kingdom, and Japan (Fama and French 2001; von Eije and Megginson 2008). Although trends and determinants differ across markets, in general, the decline in the proportion of companies paying dividends has been attributed to some or all of the following: a growth in the number of smaller, publicly traded companies with low profitability and high growth potential; an overall reduced propensity to pay dividends (controlling for differences in profitability and growth opportunities); or the increase usage of share repurchases as an alternative way to distribute cash to shareholders (Fama and French 2001; von Eije and Megginson 2008; Julio and Ikenberry 2004).

Analysts will frequently need to value non-dividend-paying shares. Can the DDM be applied to non-dividend-paying shares? In theory it can, as is illustrated later, but in practice it generally is not.

Predicting the timing of dividend initiation and the magnitude of future dividends without any prior dividend data or specifics about dividend policy to guide the analysis is generally not practical. For a non-dividend-paying company, analysts usually prefer a model that defines returns at the company level (as free cash flow or residual income—these concepts are defined shortly) rather than at the stockholder level (as dividends). Another consideration in the choice of models relates to ownership perspective. An investor purchasing a small ownership share lacks the ability to meaningfully influence the timing or magnitude of the distribution of the company's cash to shareholders. That perspective is the one taken in applying a dividend discount model. The only access to the company's value is through the receipt of dividends, and dividend policy is taken as a given. If dividends do not bear an understandable relation to value creation in the company, applying the DDM to value the stock is prone to error.

Generally, the definition of returns as dividends, and the DDM, is most suitable when:

- the company is dividend-paying (i.e., the analyst has a dividend record to analyze);
- the board of directors has established a dividend policy that bears an understandable and consistent relationship to the company's profitability; and
- the investor takes a non-control perspective.

Often, companies with established dividends are seasoned companies, profitable but operating outside the economy's fastest-growing subsectors. Professional analysts often apply a dividend discount model to value the common stock of such companies.

EXAMPLE 2

AB InBev and Diageo plc: Is the DDM an Appropriate Choice?

As director of equity research at a brokerage firm, you have final responsibility in the choice of valuation models. An analyst covering consumer/non-cyclicals has approached you about the use of a dividend discount model for valuing the equity of two companies: Anheuser-Busch InBev SA/NV, referred to as "AB InBev" (Euronext: ABI, NYSE: BUD), and Diageo plc (LSE: DGE, NYSE: DEO). Exhibit 1 gives 15 years of data. (In the table, EPS is earnings per share, DPS is dividends per share, and the payout ratio is DPS divided by EPS.)

Exhibit 1 BUD and DEO: The Earnings and Dividends Record

	BUD			DEO		
Year	EPS ($)	DPS ($)	Payout Ratio (%)	EPS (pence)	DPS (pence)	Payout Ratio (%)
2018	2.17	2.05	94	121.1	65.3	54
2017	3.98	4.33	109	105.5	62.2	59
2016	0.71	3.85	542	89.1	59.2	66
2015	4.96	3.95	80	94.6	56.4	60
2014	5.54	3.52	64	89.3	51.7	58
2013	8.72	2.83	32	97.4	47.4	49
2012	4.40	2.24	51	75.8	43.5	57
2011	3.58	1.55	43	74.1	40.4	55
2010	2.50	1.07	43	64.3	38.1	59
2009	2.90	0.55	19	65.0	36.1	56
2008	1.93	0.35	18	58.9	34.4	58
2007	3.06	3.67	120	55.0	32.7	59
2006	1.81	0.95	52	66.9	31.1	46

Exhibit 1 (Continued)

	BUD			DEO		
Year	**EPS ($)**	**DPS ($)**	**Payout Ratio (%)**	**EPS (pence)**	**DPS (pence)**	**Payout Ratio (%)**
2005	1.17	0.57	49	45.2	29.6	65
2004	NA	NA	–	48.2	27.6	57

Source: Companies' websites and filings on www.sec.gov.

Answer the following questions based on the information in Exhibit 1:

1 State whether a dividend discount model is an appropriate choice for valuing AB InBev. Explain your answer.

2 State whether a dividend discount model is an appropriate choice for valuing Diageo. Explain your answer.

Solution to 1:

Based only on the data given in Exhibit 1, a DDM does not appear to be an appropriate choice for valuing AB InBev. The company's dividends have ranged from $0.35 to $4.33 per share, and the annual payout ratio ranged from 18% to 542%, based on reported information. (The variation of earnings, dividends, and dividend payout reflects the company's history of growth through major mergers and acquisitions. ABInBev was formed when the US company Anheuser-Busch was acquired in 2008 by the Belgian company InBev. InBev itself was originally formed by a merger of the Belgian company Interbrew with the Brazilian company AmBev. Further, in 2016 AB InBev made another major acquisition, purchasing SABMiller.)

Based on the record presented and the company's profile, it is unlikely that there will be a consistent relationship between dividends and earnings. Because dividends are unlikely to adjust to reflect changes in profitability, applying a DDM to ABInBev is probably inappropriate. Valuing ABInBev on another basis, such as a company-level definition of cash flows, appears to be more appropriate.

Valuation is a forward-looking exercise. In practice, an analyst would check for public disclosures concerning changes in dividend policy going forward. In light of the increased debt from the 2016 purchase of SABMiller, ABInBev cut its dividend in 2018 and disclosed in its annual report that paying down its debt is a priority and could "restrict the amount of dividends" it is able to pay.

Solution to 2:

The historical earnings of Diageo show a relatively steady, long-term upward trend, and its dividends have generally followed its growth in earnings. Earnings per share and dividends per share grew at comparable compound annual growth rates of 6.8% and 6.3% during the entire period. In most years, the payout ratio ranged between 50% and 60%. In summary, because Diageo is dividend-paying and because dividends bear an understandable and consistent relationship to earnings, using a DDM to value Diageo is appropriate.

As noted earlier, valuation is a forward-looking exercise, and an analyst would check for public disclosures concerning changes in dividend policy going forward. In its 2018 annual report, Diageo disclosed that it continues to target dividend cover (defined as EPS/DPS) of between 1.8 times and 2.2 times, which implies a payout ratio of between 45% and 56%.

A second definition of returns is free cash flow. The term *cash flow* has been given many meanings in different contexts. Earlier in our coverage the term was used informally, referring to returns to ownership (equity). We now want to give it a more technical meaning, related to accounting usage. Over a given period, a company can add to cash (or use up cash) by selling goods and services. This money is cash flow from operations (for that period). Cash flow from operations is the critical cash flow concept addressing a business's underlying economics. Companies can also generate (or use up) cash in two other ways. First, a company affects cash through buying and selling assets, including investment and disinvestment in plant and equipment. Second, a company can add to or reduce cash through its financing activities. Financing includes debt and equity. For example, issuing bonds increases cash, and buying back stock decreases cash (all else equal).

Internationally, accounting definitions may not be fully consistent with the presented concepts in distinguishing between types of sources and uses of cash. Although the implementation details are not the focus here, an example can be given. US generally accepted accounting principles include a financing item, net interest payments, in cash flow from operating activities. So, careful analysts working with US accounting data often add back after-tax net interest payments to cash flow from operating activities when calculating cash flow from operations. Under International Accounting Standards, companies may or may not include interest expense as an operating cash flow.

Assets supporting current sales may need replacement because of obsolescence or wear and tear, and the company may need new assets to take advantage of profitable growth opportunities. The concept of free cash flow responds to the reality that, for a going concern, some of the cash flow from operations is not "free" but rather needs to be committed to reinvestment and new investment in assets. **Free cash flow to the firm** (FCFF) is cash flow from operations minus capital expenditures. Capital expenditures—reinvestment in new assets, including working capital—are needed to maintain the company as a going concern, so only that part of cash flow from operations remaining after such reinvestment is "free." (This definition is conceptual; free cash flow concepts will be defined in detail later.) FCFF is the part of the cash flow generated by the company's operations that can be withdrawn by bondholders and stockholders without economically impairing the company. Conceptually, the value of common equity is the present value of expected future FCFF—the total value of the company—minus the market value of outstanding debt.

Another approach to valuing equity works with free cash flow to equity. **Free cash flow to equity** (FCFE) is cash flow from operations minus capital expenditures, or FCFF, from which we net all payments to debtholders (interest and principal repayments net of new debt issues). Debt has a claim on the cash of the company that must be satisfied before any money can be paid to stockholders, so money paid on debt is not available to common stockholders. Conceptually, common equity can be valued as the present value of expected FCFE. FCFF is a predebt free cash flow concept; FCFE is a postdebt free cash flow concept. The FCFE model is the baseline free cash flow valuation model for equity, but the FCFF model may be easier to apply in several cases, such as when the company's leverage (debt in its capital structure) is expected to change significantly over time.

Valuation using a free cash flow concept is popular in current investment practice. Free cash flow (FCFF or FCFE) can be calculated for any company. The record of free cash flows can also be examined even for a non-dividend-paying company. FCFE can be viewed as measuring what a company can afford to pay out in dividends. Even for dividend-paying companies, a free cash flow model valuation may be preferred when dividends exceed or fall short of FCFE by significant amounts. FCFE also represents cash flow that can be redeployed outside the company without affecting the company's capital investments. A controlling equity interest can bring about such redeployment. As a result, free cash flow valuation is appropriate for investors who want to take a control perspective. (Even a small shareholder may want to take such a perspective when potential exists for the company to be acquired, because the stock price should reflect the price an acquirer would pay.)

Just as there are cases in which an analyst would find it impractical to apply the DDM, applying the free cash flow approach is a problem in some cases. Some companies have intense capital demands and, as a result, have negative expected free cash flows far into the future. As one example, a retailer may be constantly constructing new outlets and be far from saturating even its domestic market. Even if the retailer is currently very profitable, free cash flow may be negative indefinitely because of the level of capital expenditures. The present value of a series of negative free cash flows is a negative number: The use of a free cash flow model may entail a long forecast horizon to capture the point at which expected free cash flow turns positive. The uncertainty associated with distant forecasts may be considerable. In such cases, the analyst may have more confidence using another approach, such as residual income valuation.

Generally, defining returns as free cash flow and using the FCFE (and FCFF) models are most suitable when:

- the company is not dividend-paying;
- the company is dividend-paying but dividends significantly exceed or fall short of free cash flow to equity;
- the company's free cash flows align with the company's profitability within a forecast horizon with which the analyst is comfortable; and
- the investor takes a control perspective.

The third and final definition of returns that we will discuss in this overview is residual income. Conceptually, **residual income** for a given period is the earnings for that period in excess of the investors' required return on beginning-of-period investment (common stockholders' equity). Suppose shareholders' initial investment is \$200 million, and the required rate of return on the stock is 8%. The required rate of return is investors' **opportunity cost** for investing in the stock: the highest expected return available from other equally risky investments, which is the return that investors forgo when investing in the stock. The company earns \$18 million in the course of a year. How much value has the company added for shareholders?

A return of 0.08 × \$200 million = \$16 million just meets the amount investors could have earned in an equivalent-risk investment (by the definition of opportunity cost). Only the residual or excess amount of \$18 million – \$16 million = \$2 million represents value added, or an economic gain, to shareholders. So, \$2 million is the company's residual income for the period. The residual income approach attempts to match profits to the period in which they are earned (but not necessarily realized as cash). In contrast to accounting net income (which has the same matching objective in principle), however, residual income attempts to measure the value added in excess of opportunity costs.

The residual income model states that a stock's value is book value per share plus the present value of expected future residual earnings. (**Book value per share** is common stockholders' equity divided by the number of common shares outstanding.) In contrast to the dividend and free cash flow models, the residual income model introduces a stock concept, book value per share, into the present value expression. Nevertheless, the residual income model can be viewed as a restatement of the dividend discount model, using a company-level return concept. Dividends are paid out of earnings and are related to earnings and book value (BV) through a simple expression:

BV of equity at t = BV of equity at (t – 1) + Earnings for the period (t – 1) to t – Dividends paid at t

Please note that the foregoing expression is valid assuming that any items that go through the balance sheet (affecting book value) first go through the income statement (reflected in earnings), apart from ownership transactions.

The residual income model is a useful addition to an analyst's toolbox. Because the record of residual income can always be calculated, a residual income model can be used for both dividend-paying and non-dividend-paying stocks. Analysts may choose a residual income approach for companies with negative expected free cash flows within their comfortable forecast horizon. In such cases, a residual income valuation often brings the recognition of value closer to the present as compared with a free cash flow valuation, producing higher value estimates.

The residual income model has an attractive focus on profitability in relation to opportunity costs. Executive compensation schemes are sometimes based on a residual income concept. Knowledgeable application of the residual income model requires a detailed knowledge of accrual accounting; consequently, in cases for which the dividend discount model is suitable, analysts may prefer it as the simpler choice. Management sometimes exercises its discretion within allowable accounting practices to distort the accuracy of its financials as a reflection of economic performance. If the quality of accounting disclosure is good, the analyst may be able to calculate residual income by making appropriate adjustments (to reported net income and book value, in particular). In some cases, the degree of distortion and the quality of accounting disclosure can be such that the application of the residual income model is error-prone.

Generally, the definition of returns as residual income, and the residual income model, is most suitable when:

- the company is not paying dividends, as an alternative to a free cash flow model, or
- the company's expected free cash flows are negative within the analyst's comfortable forecast horizon.

In summary, the three most widely used definitions of returns to investors are dividends, free cash flow, and residual income. Although claims are often made that one cash flow definition is inherently superior to the rest—often following changing fashions in investment practice—a more flexible viewpoint is practical. The analyst may find that one model is more suitable to a particular valuation problem. The analyst may also develop more expertise in applying one type of model. In practice, skill in application—in particular, the quality of forecasts—is frequently decisive for the usefulness of the analyst's work.

In the next section, we present the general form of the dividend discount model as a prelude to discussing the particular implementations of the model that are suitable for different sets of attributes of the company being valued.

2 THE DIVIDEND DISCOUNT MODEL

b calculate and interpret the value of a common stock using the dividend discount model (DDM) for single and multiple holding periods;

Investment analysts use a wide range of models and techniques to estimate the value of common stock, including present value models. In a survey of CFA Institute members with job responsibility for equity analysis, nearly 80% of respondents reported using a discounted cash flow approach (Stowe, Pinto, and Robinson 2018). Earlier we discussed three common definitions of cash flow for use in present value analysis: dividends, free cash flow, and residual income. In this section, we develop the most general form of the dividend discount model.

The DDM is the simplest and oldest present value approach to valuing stock. Recent survey data shows that among the analysts using a discounted cash flow approach to equity valuation, about 35.1% employ a dividend discount model (Stowe, Pinto, and Robinson 2018). Besides its continuing significant position in practice, the DDM has an important place in both academic and practitioner equity research. The DDM is, for these reasons, a basic tool in equity valuation.

2.1 The Expression for a Single Holding Period

From the perspective of a shareholder who buys and holds a share of stock, the cash flows he will obtain are the dividends paid on it and the market price of the share when he sells it. The future selling price should in turn reflect expectations about dividends subsequent to the sale. In this section, we will show how this argument leads to the most general form of the dividend discount model. In addition, the general expression developed for a finite holding period corresponds to one practical approach to DDM valuation. In that approach, the analyst forecasts dividends over a finite horizon, as well as the terminal sales price.

If an investor wishes to buy a share of stock and hold it for one year, the value of that share of stock today is the present value of the expected dividend to be received on the stock plus the present value of the expected selling price in one year:

$$V_0 = \frac{D_1}{(1+r)^1} + \frac{P_1}{(1+r)^1} = \frac{D_1 + P_1}{(1+r)^1} \tag{2}$$

where

V_0 = the value of a share of stock today, at $t = 0$

P_1 = the expected price per share at $t = 1$

D_1 = the expected dividend per share for Year 1, assumed to be paid at the end of the year at $t = 1$

r = the required rate of return on the stock

Equation 2 applies, to a single holding period, the principle that an asset's value is the present value of its future cash flows. In this case, the expected cash flows are the dividend in one year (for simplicity, assumed to be received as one payment at the end of the year) and the price of the stock in one year. Note that throughout the discussion of the DDM, we assume that dividends for a period are paid in one sum at the end of the period.

EXAMPLE 3

DDM Value with a Single Holding Period

Suppose that you expect Carrefour SA (CA: EN Paris) to pay a €0.46 dividend next year. You expect the price of Carrefour stock to be €23.00 in one year. The required rate of return for Carrefour stock is 8%. What is your estimate of the value of Carrefour stock?

Discounting the expected dividend of €0.46 and the expected sales price of €23.00 at the required return on equity of 8%, we obtain

$$V_0 = \frac{D_1 + P_1}{(1+r)^1} = \frac{0.46 + 23.00}{(1+0.08)^1} = \frac{23.46}{1.08} = 21.72.$$

2.2 The Expression for Multiple Holding Periods

If an investor plans to hold a stock for two years, the value of the stock is the present value of the expected dividend in Year 1, plus the present value of the expected dividend in Year 2, plus the present value of the expected selling price at the end of Year 2.

$$V_0 = \frac{D_1}{(1+r)^1} + \frac{D_2}{(1+r)^2} + \frac{P_2}{(1+r)^2} = \frac{D_1}{(1+r)^1} + \frac{D_2 + P_2}{(1+r)^2} \quad (3)$$

The expression for the DDM value of a share of stock for any finite holding period is a straightforward extension of the expressions for one-year and two-year holding periods. For an n-period model, the value of a stock is the present value of the expected dividends for the n periods plus the present value of the expected price in n periods (at $t = n$).

$$V_0 = \frac{D_1}{(1+r)^1} + \cdots + \frac{D_n}{(1+r)^n} + \frac{P_n}{(1+r)^n} \quad (4)$$

If we use summation notation to represent the present value of the first n expected dividends, the general expression for an n-period holding period or investment horizon can be written as

$$V_0 = \sum_{t=1}^{n} \frac{D_t}{(1+r)^t} + \frac{P_n}{(1+r)^n} \quad (5)$$

Equation 5 is significant in DDM application because analysts may make individual forecasts of dividends over some finite horizon (often two to five years) and then estimate the terminal price, P_n, based on one of a number of approaches. (We will discuss valuation using a finite forecasting horizon later.) Example 4 reviews the mechanics of this calculation.

EXAMPLE 4

Finding the Stock Price for a Five-Year Forecast Horizon

For the next five years, the annual dividends of a stock are expected to be \$2.00, \$2.10, \$2.20, \$3.50, and \$3.75. In addition, the stock price is expected to be \$40.00 in five years. If the required return on equity is 10%, what is the value of this stock?

The present values of the expected future cash flows can be written out as

$$V_0 = \frac{2.00}{(1.10)^1} + \frac{2.10}{(1.10)^2} + \frac{2.20}{(1.10)^3} + \frac{3.50}{(1.10)^4} + \frac{3.75}{(1.10)^5} + \frac{40.00}{(1.10)^5}$$

Calculating and summing these present values gives a stock value of V_0 = 1.818 + 1.736 + 1.653 + 2.391 + 2.328 + 24.837 = \$34.76.

The five dividends have a total present value of \$9.926 and the terminal stock value has a present value of \$24.837, for a total stock value of \$34.76.

With a finite holding period, whether one, two, five, or some other number of years, the dividend discount model finds the value of stock as the sum of 1) the present values of the expected dividends during the holding period and 2) the present value of the expected stock price at the end of the holding period. As the holding period is increased by one year, we have an extra expected dividend term. In the limit (i.e., if the holding period extends into the indefinite future), the stock's value is the present value of all expected future dividends.

$$V_0 = \frac{D_1}{(1+r)^1} + \ldots + \frac{D_n}{(1+r)^n} + \ldots \tag{6}$$

This value can be expressed with summation notation as

$$V_0 = \sum_{t=1}^{\infty} \frac{D_t}{(1+r)^t} \tag{7}$$

Equation 7 is the general form of the dividend discount model, first presented by John Burr Williams (1938). Even from the perspective of an investor with a finite investment horizon, the value of stock depends on all future dividends. For that investor, stock value today depends *directly* on the dividends the investor expects to receive before the stock is sold and *indirectly* on the expected dividends after the stock is sold, because those future dividends determine the expected selling price.

Equation 7, by expressing the value of stock as the present value of expected dividends into the indefinite future, presents a daunting forecasting challenge. In practice, of course, analysts cannot make detailed, individual forecasts of an infinite number of dividends. To use the DDM, the forecasting problem must be simplified. Two broad approaches exist, each of which has several variations:

1 Future dividends can be forecast by assigning the stream of future dividends to one of several stylized growth patterns. The most commonly used patterns are:

   - constant growth forever (the Gordon growth model);
   - two distinct stages of growth (the two-stage growth model and the H-model); and
   - three distinct stages of growth (the three-stage growth model).

   The DDM value of the stock is then found by discounting the dividend streams back to the present. We present the Gordon growth model, the two-stage H-model, and three-stage growth models later.

2 A finite number of dividends can be forecast individually up to a terminal point, by using pro forma financial statement analysis, for example. Typically, such forecasts extend from 3 to 10 years into the future. Although some analysts apply the same horizon to all companies under analysis, the horizon selected often depends on the perceived predictability (sometimes called the **visibility**) of the company's earnings. We can then forecast either:

   - the remaining dividends from the terminal point forward by assigning those dividends to a stylized growth pattern, or
   - the share price at the terminal point of our dividend forecasts (**terminal share price**), by using some method (such as taking a multiple of forecasted book value or earnings per share as of that point, based on one of several methods for estimating such multiples).

   The stock's DDM value is then found by discounting the dividends (and forecasted price, if any) back to the present.

Spreadsheets are particularly convenient tools for implementing a DDM with individual dividend forecasts but are useful in all cases. We address spreadsheet modeling at a later stage.

Whether analysts are using dividends or some other definition of cash flow, they generally use one of the foregoing forecasting approaches when valuing stock. The challenge in practice is to choose an appropriate model for a stock's future dividends and to develop quality inputs to that model.

3 THE GORDON GROWTH MODEL: THE GORDON GROWTH MODEL EQUATION AND CHTHE LINKS AMONG DIVIDEND GROWTH, EARNINGS GROWTH, AND VALUE APPRECIATION

c calculate the value of a common stock using the Gordon growth model and explain the model's underlying assumptions;

d calculate the value of non-callable fixed-rate perpetual preferred stock;

h describe strengths and limitations of the Gordon growth model and justify its selection to value a company's common shares;

The Gordon growth model, developed by Gordon and Shapiro (1956) and Gordon (1962), assumes that dividends grow indefinitely at a constant rate. This assumption, applied to the general dividend discount model (Equation 7), leads to a simple and elegant valuation formula that has been influential in investment practice. This section explores the development of the Gordon growth model and illustrates its uses.

3.1 The Gordon Growth Model Equation

The simplest pattern that can be assumed in forecasting future dividends is growth at a constant rate. In mathematical terms, this assumption can be stated as

$$D_t = D_{t-1}(1 + g)$$

where g is the expected constant growth rate in dividends and D_t is the expected dividend payable at time t. Suppose, for example, that the most recent dividend, D_0, was €10. Then, if a 5% dividend growth rate is forecast, the expected dividend at $t = 1$ is $D_1 = D_0(1 + g) = €10 \times 1.05 = €10.5$. For any time t, D_t also equals the $t = 0$ dividend, compounded at g for t periods:

$$D_t = D_0(1 + g)^t \tag{8}$$

To continue the example, at the end of five years the expected dividend is $D_5 = D_0(1 + g)^5 = €10 \times (1.05)^5 = €10 \times 1.276282 = €12.76$. If $D_0(1 + g)^t$ is substituted into Equation 7 for D_t, it gives the Gordon growth model. If all of the terms are written out, they are

$$V_0 = \frac{D_0(1 + g)}{(1 + r)} + \frac{D_0(1 + g)^2}{(1 + r)^2} + \ldots + \frac{D_0(1 + g)^n}{(1 + r)^n} + \ldots \tag{9}$$

Equation 9 is a geometric series; that is, each term in the expression is equal to the previous term times a constant, which in this case is $(1 + g)/(1 + r)$. This equation can be simplified algebraically into a much more compact equation:

$$V_0 = \frac{D_0(1 + g)}{r - g}, \text{ or } V_0 = \frac{D_1}{r - g} \quad (10)$$

The simplification involves the expression for the sum of an infinite geometric progression with the first term equal to a and the growth factor equal to m with $|m| < 1$ [i.e., the sum of $a + am + am^2 + \dots$ is $a/(1 - m)$]. Setting $a = D_1/(1 + r)$ and $m = (1 + g)/(1 + r)$ gives the Gordon growth model.

Both equations are equivalent because $D_1 = D_0(1 + g)$. In Equation 10, it must be specified that the required return on equity must be greater than the expected growth rate: $r > g$. If $r = g$ or $r < g$, Equation 10 as a compact formula for value assuming constant growth is not valid. If $r = g$, dividends grow at the same rate at which they are discounted, so the value of the stock (as the undiscounted sum of all expected future dividends) is infinite. If $r < g$, dividends grow faster than they are discounted, so the value of the stock is infinite. Of course, infinite values do not make economic sense; so constant growth with $r = g$ or $r < g$ does not make sense.

To illustrate the calculation, suppose that an annual dividend of €5 has just been paid (D_0 = €5). The expected long-term growth rate is 5% and the required return on equity is 8%. The Gordon growth model value per share is $D_0(1 + g)/(r - g)$ = (€5 × 1.05)/(0.08 − 0.05) = €5.25/0.03 = €175. When calculating the model value, be careful to use D_1 and not D_0 in the numerator.

The Gordon growth model (Equation 10) is one of the most widely recognized equations in the field of security analysis. Because the model is based on indefinitely extending future dividends, the model's required rate of return and growth rate should reflect long-term expectations. Further, model values are very sensitive to both the required rate of return, r, and the expected dividend growth rate, g. In this model and other valuation models, it is helpful to perform a sensitivity analysis on the inputs, particularly when an analyst is not confident about the proper values.

Earlier we stated that analysts typically apply DDMs to dividend-paying stocks when dividends bear an understandable and consistent relation to the company's profitability. The same qualifications hold for the Gordon growth model. In addition, the Gordon growth model form of the DDM is most appropriate for companies with earnings expected to grow at a rate comparable to or lower than the economy's nominal growth rate. Businesses growing at much higher rates than the economy often grow at lower rates in maturity, and the horizon in using the Gordon growth model is the entire future stream of dividends.

To determine whether the company's growth rate qualifies it as a candidate for the Gordon growth model, an estimate of the economy's nominal growth rate is needed. This growth rate is usually measured by the growth in **gross domestic product**, a money measure of the goods and services produced within a country's borders. National government agencies as well as the World Bank (www.worldbank.org) publish GDP data, which are also available from several secondary sources. Exhibit 2 shows the real GDP growth record for a number of major developed markets.

Exhibit 2 Average Annual Real GDP Growth Rates: 1988–2017

	Period		
Country	**1988–1997**	**1998–2007**	**2008–2017**
Australia	3.2%	3.5%	2.6%
Canada	2.1	3.2	1.6

(continued)

Exhibit 2 (Continued)

Country	Period 1988–1997	1998–2007	2008–2017
Denmark	2.0	2.0	0.8
France	2.2	2.4	0.8
Germany	2.6	1.7	1.3
Italy	1.9	1.5	-0.5
Japan	2.8	1.0	0.5
Netherlands	3.1	2.8	0.9
Sweden	1.4	3.5	1.6
Switzerland	1.5	2.4	1.4
United Kingdom	2.4	2.9	1.1
United States	3.1	3.1	1.5

Source: OECD.

Based on historical and/or forward-looking information, nominal GDP growth can be estimated as the sum of the estimated real growth rate in GDP plus the expected long-run inflation rate. For example, using 10 years of historical data through 2018, one estimate of the underlying real growth rate of the Canadian economy is 1.6%. Adjusting for the Bank of Canada's inflation target of 2% as the expected inflation rate gives an estimate of the Canadian economy's nominal annual growth rate of 1.6% + 2% = 3.6%. Publicly traded companies constitute varying amounts of the total corporate sector but always less than 100%. As a result, the overall growth rate of the public corporate sector can diverge from the nominal GDP growth rate during a long horizon; furthermore, within the public corporate sector, some subsectors may experience persistent growth rate differentials. Nevertheless, an earnings growth rate far above the nominal GDP growth rate is not sustainable in perpetuity.

When forecasting an earnings growth rate far above the economy's nominal growth rate, analysts should use a multistage DDM in which the final-stage growth rate reflects a growth rate that is more plausible relative to the economy's nominal growth rate, rather than using the Gordon growth model.

EXAMPLE 5

Valuation Using the Gordon Growth Model (1)

Joel Williams follows Sonoco Products Company (NYSE: SON), a manufacturer of paper and plastic packaging for both consumer and industrial use. Sonoco appears to have a dividend policy of recognizing sustainable increases in the level of earnings with increases in dividends, typically keeping the dividend payout ratio within a range of 40% to 60%. Williams also notes the following:

- Sonoco's most recent quarterly dividend, declared 13 February 2019, was \$0.41, consistent with a current annual dividend of 4 × \$0.41 = \$1.64 per year.

- His forecasted dividend growth rate is 4.5% per year.
- With a beta (β_i) of 0.95, given an equity risk premium (expected excess return of equities over the risk-free rate, $E(R_M) - R_F$) of 4.5% and a risk-free rate (R_F) of 3%, Sonoco's required return on equity is $r = R_F + \beta_i[E(R_M) - R_F] = 3.0 + 0.95(4.5) = 7.3\%$, using the capital asset pricing model.

Williams believes the Gordon growth model may be an appropriate model for valuing Sonoco.

1 Calculate the Gordon growth model value for Sonoco stock.
2 The current market price of Sonoco stock is \$59.55. Using your answer to Question 1, judge whether Sonoco stock is fairly valued, undervalued, or overvalued.

Solution to 1:

Using Equation 10,

$$V_0 = \frac{D_0(1+g)}{r-g} = \frac{\$1.64 \times 1.045}{0.073 - 0.045} = \frac{\$1.7138}{0.028} = \$61.21$$

Solution to 2:

The market price of \$59.55 is \$1.66, or approximately 2.7% less than the Gordon growth model intrinsic value estimate of \$61.21. Sonoco appears to be slightly undervalued based on the Gordon growth model estimate.

The next example illustrates a Gordon growth model valuation introducing some problems the analyst might face in practice. The example refers to adjusted beta; the most common calculation adjusts raw historical beta toward the overall mean value of one for beta.

EXAMPLE 6

Valuation Using the Gordon Growth Model (2)

As an analyst for a US domestic equity–income mutual fund, Robert Kim is evaluating Middlesex Water Company (NASDAQ: MSEX), a publicly traded water utility, for possible inclusion in the approved list of investments. Kim is conducting the analysis in early 2019.

Not all countries have traded water utility stocks. In the United States, most of the population gets its water from government entities; however, a group of investor-owned water utilities also supplies water to the public. With a market capitalization of about \$880 million as of early 2019, MSEX is among the 10 largest publicly traded US water utilities. MSEX's historical base is the Middlesex System, serving residential, industrial, and commercial customers in a well-developed area of central New Jersey. Through various subsidiaries, MSEX also provides water and wastewater collection and treatment services to areas of southern New Jersey and Delaware.

MSEX's return on equity averaged 8.5% over the past 10 years with relatively little variation, and its profit margins are above industry averages. When MSEX's credit rating was upgraded in 2015, the reasons cited by Standard & Poor's included the company's "improving management of regulatory risk that is expected to result in less volatile profitability measures, moderately improved cash flow measures and the ability to consistently earn closer to its authorized returns" (according to MSEX's Form 8-K filed with the SEC on 24 August 2015). Because MSEX obtains most of its revenue from the regulated business of providing an important staple, water, to a relatively stable population, Kim feels confident in forecasting future earnings and dividend growth. MSEX appears to have a policy of maintaining an average dividend payout ratio between 60% and 70%. Other facts and forecasts include the following:

- MSEX's per-share dividends for 2018 (D_0) were \$0.911.
- Kim forecasts a long-term earnings growth rate of 4.5% per year.
- MSEX's raw beta and adjusted beta are, respectively, 0.70 and 0.80 based on 60 monthly returns. The R^2 associated with beta, however, is under 20%.
- Kim estimates that MSEX's pretax cost of debt is 4.8% based on Standard & Poor's issuer rating of A for MSEX and on the current corporate yield curve.
- Kim's estimate of MSEX's required return on equity is 6.8%.
- MSEX's current market price is \$43.20.

1. Calculate the Gordon growth model estimate of value for MSEX using Kim's required return on equity estimate.
2. State whether MSEX appears to be overvalued, fairly valued, or undervalued based on the Gordon growth model estimate of value.
3. Justify the selection of the Gordon growth model for valuing MSEX.
4. Calculate the CAPM estimate of the required return on equity for MSEX under the assumption that beta reverts to the mean. (Assume an equity risk premium of 4.5% and a risk-free rate of 3% as of the price quotation date.)
5. Calculate the Gordon growth estimate of value using A) the required return on equity from your answer to Question 4, and B) a bond-yield-plus-risk-premium approach with a risk premium of 2.5%.
6. Evaluate the effect of uncertainty in MSEX's required return on equity on the valuation conclusion in Question 2.

Solution to 1:

From Equation 10,

$$V_0 = \frac{D_0(1+g)}{r-g} = \frac{\$0.911(1.045)}{0.068-0.045} = \$41.39$$

Solution to 2:

Because the Gordon growth model estimate of \$41.39 differs from the market price of \$43.20 by a relatively small amount (less than 5%), MSEX appears to be fairly valued.

Solution to 3:

The Gordon growth model, which assumes that dividends grow at a stable rate in perpetuity, is a realistic model for MSEX for the following reasons:

- MSEX profitability is stable as reflected in its return on equity. This stability reflects predictable demand and regulated prices for its product, water.

- Dividends bear an understandable and consistent relationship to earnings, as evidenced by the company's policy of predictable dividend payout ratios.
- Although the company's earnings growth has been higher in recent years, the forecasted earnings growth rate of 4.5% a year seems both attainable and reasonable compared with the historical long-term nominal annual GDP growth for the United States (approximately 4.3% over the 20-year period 1998–2018, based on data from the US Bureau of Economic Analysis).
- The earnings growth forecast for the company does not include a period of forecasted very high or very low growth.

Solution to 4:

The assumption of reversion to the mean is characteristic of adjusted historical beta. The required return on equity as given by the CAPM assuming a risk-free rate of 3% and an equity risk premium of 4.5% is given by the following: 3% + 0.80(4.5%) = 6.6% using adjusted beta, which assumes reversion to the mean of 1.0.

Solution to 5:

A The Gordon growth value of MSEX using a required return on equity of 6.6% is

$$V_0 = \frac{D_0(1+g)}{r-g} = \frac{\$0.911 \times 1.045}{0.066 - 0.045} = \$45.33.$$

$$V_0 = \frac{D_0(1+g)}{r-g} = \frac{\$0.911(1.045)}{0.066 - 0.045} = \$45.33$$

B The bond-yield-plus-risk-premium estimate of the required return on equity is 4.8% + 2.5% = 7.3%. The Gordon growth value of MSEX using a required return on equity of 7.3% is

$$V_0 = \frac{D_0(1+g)}{r-g} = \frac{\$0.911(1.045)}{0.073 - 0.045} = \$34.00$$

Solution to 6:

Using the CAPM estimate of the required return on equity (Question 5A), MSEX appears to be fairly valued; although the estimated value of $45.33 exceeds the current market price, the difference is only around 5%. Further, according to the facts given concerning R^2, beta explains less than 20% of the variation in MSEX's returns. Using a bond-yield-plus-risk-premium approach, MSEX appears to be significantly overvalued ($34.00 is more than 20% lower than the market price of $43.20). No specific evidence, however, supports the particular value of the risk premium selected in the bond-yield-plus-risk-premium approach. In this case, because of the uncertainty in the required return on equity estimate, one has less confidence that MSEX is overvalued. Given the results of the other two approaches, the analyst may view MSEX as relatively fairly valued.

As mentioned earlier, an analyst needs to be aware that Gordon growth model values can be very sensitive to small changes in the values of the required rate of return and expected dividend growth rate. Example 7 illustrates a format for a sensitivity analysis.

EXAMPLE 7

Valuation Using the Gordon Growth Model (3)

In Example 6, the Gordon growth model value for MSEX was estimated as $41.39 based on a current dividend of $0.911, an expected dividend growth rate of 4.5%, and a required return on equity of 6.8%. What if the estimates of r and g each vary by 25 bps? How sensitive is the model value to changes in the estimates of r and g? Exhibit 3 provides information on this sensitivity.

Exhibit 3 Estimated Price Given Uncertain Inputs

	***g* = 4.25%**	***g* = 4.50%**	***g* = 4.75%**
r = 6.55%	$41.29	$46.44	$53.02
***r* = 6.80%**	$37.24	**$41.39**	$46.55
r = 7.05%	$33.92	$37.33	$41.49

A point of interest following from the mathematics of the Gordon growth model is that when the spread between r and g is the widest (r = 7.05% and g = 4.25%), the Gordon growth model value is the smallest ($33.92), and when the spread is the narrowest (r = 6.55% and g = 4.75%), the model value is the largest ($53.02). As the spread goes to zero, in fact, the model value increases without bound. The largest value in Exhibit 3, $53.02, is more than 55% larger than the smallest value, $33.92. Two-thirds of the values in Exhibit 3 are lower than MSEX's current market price of $43.20. All but two of the estimates, however, are within 10% of the current price, which supports the conclusion that MSEX is relatively fairly valued or slightly overvalued. In summary, the best estimate of the value of MSEX given the assumptions is $41.39, bolded in Exhibit 3, but the estimate is quite sensitive to rather small changes in inputs.

Examples 6 and 7 illustrate the application of the Gordon growth model to a utility, a traditional source for such illustrations because of the stability afforded by providing an essential service in a regulated environment. Before applying any valuation model, however, analysts need to know much more about a company than industry membership. For example, if a utility company undertook an aggressive growth-by-acquisition strategy, then its expected growth in income and dividends could potentially diverge significantly from other companies in the industry. Furthermore, many utility holding companies in the United States have major, unregulated business subsidiaries so the traditional picture of steady and slow growth often does not hold.

In addition to individual stocks, analysts have often used the Gordon growth model to value broad equity market indexes, especially in developed markets. Because the value of publicly traded issues typically represents a large fraction of the overall corporate sector in developed markets, such indexes reflect average economic growth rates. Furthermore, in such economies, a sustainable trend value of growth may be identifiable.

The Gordon growth model can also be used to value the non-callable form of a traditional type of preferred stock, **fixed-rate perpetual preferred stock** (stock with a specified dividend rate that has a claim on earnings senior to the claim of common stock, and no maturity date). Perpetual preferred stock has been used particularly by financial institutions such as banks to obtain permanent equity capital while diluting the interests of common equity. Generally, such issues have been callable by the issuer after a certain period, so valuation must take account of the issuer's call option. Valuation of the non-callable form, however, is straightforward.

If the dividend on such preferred stock is D, because payments extend into the indefinite future a **perpetuity** (a stream of level payments extending to infinity) exists in the constant amount of D. With $g = 0$, which is true because dividends are fixed for such preferred stock, the Gordon growth model becomes

$$V_0 = \frac{D}{r} \tag{11}$$

The discount rate, r, capitalizes the amount D, and for that reason is often called a **capitalization rate** in this expression and any other expression for the value of a perpetuity.

EXAMPLE 8

Valuing Noncallable Fixed-Rate Perpetual Preferred Stock

Kansas City Southern Preferred 4% (KSU-P), issued 2 January 1963, has a par value of \$25 per share. Thus, a share pays 0.04(\$25) = \$1.00 in annual dividends. The required return on this security is estimated at 5.5%. Estimate the value of this issue.

Solution:

According to the model in Equation 11, KSU-P preferred stock is worth D/r = 1.00/0.055 = \$18.18.

A perpetual preferred stock has a level dividend, thus a dividend growth rate of zero. Another case is a declining dividend—a negative growth rate. The Gordon growth model also accommodates this possibility, as illustrated in Example 9.

EXAMPLE 9

Gordon Growth Model with Negative Growth

Afton Mines is a profitable company that is expected to pay a \$4.25 dividend next year. Because it is depleting its mining properties, the best estimate is that dividends will decline forever at a rate of 4%. The required rate of return on Afton stock is 9%. What is the value of Afton shares?

Solution:

For Afton, the value of the stock is

$$V_0 = \frac{4.25}{\left[0.09 - (-0.04)\right]}$$
$$= \frac{4.25}{0.13} = \$32.69$$

The negative growth results in a \$32.69 valuation for the stock.

3.2 The Links among Dividend Growth, Earnings Growth, and Value Appreciation in the Gordon Growth Model

The Gordon growth model implies a set of relationships for the growth rates of dividends, earnings, and stock value. With dividends growing at a constant rate g, stock value also grows at g as well. The current stock value is $V_0 = D_1/(r - g)$. Multiplying both sides by $(1 + g)$ gives $V_0(1 + g) = D_1(1 + g)/(r - g)$, which is $V_1 = D_2/(r - g)$. So, both dividends and value have grown at a rate of g (holding r constant). Given a constant payout ratio—a constant, proportional relationship between earnings and dividends—dividends and earnings grow at g.

To summarize, g in the Gordon growth model is the rate of value or capital appreciation (sometimes also called the capital gains yield). Some textbooks state that g is the rate of price appreciation. If prices are efficient (price equals value), price is indeed expected to grow at a rate of g. If there is mispricing (price is different from value), however, the actual rate of capital appreciation depends on the nature of the mispricing and how fast it is corrected, if at all. This topic is discussed in the coverage of return concepts.

Another characteristic of the constant growth model is that the components of total return (dividend yield and capital gains yield) will also stay constant through time, given that price tracks value exactly. The dividend yield, which is D_1/P_0 at $t = 0$, will stay unchanged because both the dividend and the price are expected to grow at the same rate, leaving the dividend yield unchanged through time. For example, consider a stock selling for €50.00 with a **forward dividend yield** (a dividend yield based on the anticipated dividend during the next 12 months) of 2% based on an expected dividend of €1. The estimate of g is 5.50% per year. The dividend yield of 2%, the capital gains yield of 5.50%, and the total return of 7.50% are expected to be the same at $t = 0$ and at any future point in time.

4 SHARE REPURCHASES AND THE IMPLIED DIVIDEND GROWTH RATE

c calculate the value of a common stock using the Gordon growth model and explain the model's underlying assumptions;

e calculate and interpret the implied growth rate of dividends using the Gordon growth model and current stock price;

An issue of increasing importance in many developed markets is share repurchases. Companies can distribute free cash flow to shareholders in the form of share repurchases (also called buybacks) as well as dividends. In the United States, more than half of dividend-paying companies have also been making regular share repurchases (Skinner 2008). Clearly, analysts using DDMs need to understand share repurchases. Share repurchases and cash dividends have several distinctive features:

- Share repurchases involve a reduction in the number of shares outstanding, all else equal. Selling shareholders see their relative ownership position reduced compared with non-selling shareholders.
- Whereas many corporations with established cash dividends are reluctant to reduce or omit cash dividends, corporations generally do not view themselves as committed to maintaining share repurchases at any specified level.

- Cash dividends tend to be more predictable in money terms and more predictable as to timing (Wagner 2007). Although evidence from the United States suggests that, for companies with active repurchase programs, the amount of repurchases during two-year intervals bears a relationship to earnings, companies appear to be opportunistic in timing exactly when to repurchase (Skinner 2008). Thus, share repurchases are generally harder to forecast than the cash dividends of companies with an identifiable dividend policy.
- As a baseline case, share repurchases are neutral in their effect on the wealth of ongoing shareholders if the repurchases are accomplished at market prices.

The analyst could account for share repurchases directly by forecasting the total earnings, total distributions to shareholders (via either cash dividends or share repurchases), and shares outstanding. Experience and familiarity with such models is much less than for DDMs. Focusing on cash dividends, however, DDMs supply accurate valuations consistent with such an approach if the analyst takes account of the effect of expected repurchases on the per-share growth rates of dividends. Correctly applied, the DDM is a valid approach to common stock valuation even when the company being analyzed engages in share repurchases.

4.1 The Implied Dividend Growth Rate

Because the dividend growth rate affects the estimated value of a stock using the Gordon growth model, differences between estimated values of a stock and its actual market value might be explained by different growth rate assumptions. Given price, the expected next-period dividend, and an estimate of the required rate of return, the dividend growth rate reflected in price can be inferred assuming the Gordon growth model. (Actually, it is possible to infer the market-price-implied dividend growth based on other DDMs as well.) An analyst can then judge whether the implied dividend growth rate is reasonable, high, or low, based on what she knows about the company. In effect, the calculation of the implied dividend growth rate provides an alternative perspective on the stock's valuation (fairly valued, overvalued, or undervalued). Example 10 shows how the Gordon growth model can be used to infer the market's implied growth rate for a stock.

EXAMPLE 10

The Growth Rate Implied by the Current Stock Price

Suppose a company has a beta of 1.1. The risk-free rate is 5.6%, and the equity risk premium is 6%. The current dividend of $2.00 is expected to grow at 5% indefinitely. The price of the stock is $40.

1 Estimate the value of the company's stock.

2 Determine the constant dividend growth rate that would be required to justify the market price of $40.

Solution to 1:

The required rate of return is 5.6% + 1.1(6%) = 12.2%. The value of one share, using the Gordon growth model, is

$$\begin{aligned} V_0 &= \frac{D_0(1+g)}{r-g} \\ &= \frac{2.00(1.05)}{0.122-0.05} \\ &= \frac{2.10}{0.072} = \$29.17 \end{aligned}$$

Solution to 2:

The valuation estimate of the model ($29.17) is less than the market value of $40.00, and thus the market price must be forecasting a growth rate above the assumed 5%. Assuming that the model and the required return assumption are appropriate, the growth rate in dividends required to justify the $40 stock price can be calculated by substituting all known values into the Gordon growth model equation except for g:

$$\begin{aligned} 40 &= \frac{2.00(1+g)}{0.122-g} \\ 4.88 - 40g &= 2 + 2g \\ 42g &= 2.88 \\ g &= 0.0686 \end{aligned}$$

An expected dividend growth rate of 6.86% is required for the stock price to be correctly valued at the market price of $40.

5 THE PRESENT VALUE OF GROWTH OPPORTUNITIES, GORDON GROWTH MODEL AND THE PRICE-TO-EARNINGS RATIO, AND ESTIMATING A REQUIRED RETURN USING THE GORDON GROWTH MODEL

f calculate and interpret the present value of growth opportunities (PVGO) and the component of the leading price-to-earnings ratio (P/E) related to PVGO;

g calculate and interpret the justified leading and trailing P/Es using the Gordon growth model;

h describe strengths and limitations of the Gordon growth model and justify its selection to value a company's common shares;

n estimate a required return based on any DDM, including the Gordon growth model and the H-model;

p evaluate whether a stock is overvalued, fairly valued, or undervalued by the market based on a DDM estimate of value.

The value of a stock can be analyzed as the sum of 1) the value of the company without earnings reinvestment and 2) the **present value of growth opportunities** (PVGO). PVGO, also known as the **value of growth**, sums the expected value today of opportunities to profitably reinvest future earnings. More technically, PVGO can be defined as the forecasted total net present value of future projects. In this section, we illustrate this decomposition and discuss how it may be interpreted to gain insight into the market's view of a company's business and prospects.

Earnings growth may increase, leave unchanged, or reduce shareholder wealth depending on whether the growth results from earning returns in excess of, equal to, or less than the opportunity cost of funds. Consider a company with a required return on equity of 10% that has earned €1 per share. The company is deciding whether to pay out current earnings as a dividend or to reinvest them at 10% and distribute the ending value as a dividend in one year. If it reinvests, the present value of investment is €1.10/1.10 = €1.00, equaling its cost, so the decision to reinvest has a net present value (NPV) of zero. If the company were able to earn more than 10% by exploiting a profitable growth opportunity, reinvesting would have a positive NPV, increasing shareholder wealth. Suppose the company could reinvest earnings at 25% for one year: The per-share NPV of the growth opportunity would be €1.25/1.10 – €1 ≈ €0.14. Note that any reinvestment at a positive rate below 10%, although increasing EPS, is not in shareholders' interests. Increases in shareholder wealth occur only when reinvested earnings earn more than the opportunity cost of funds—that is, when investments are in positive NPV projects (condition of profitability as return on equity [ROE] > r, with ROE calculated with the market value of equity rather than the book value of equity in the denominator). Thus, investors actively assess whether and to what degree companies will have opportunities to invest in profitable projects. In principle, companies without prospects for investing in positive NPV projects should distribute most or all earnings to shareholders as dividends so the shareholders can redirect capital to more attractive areas.

A company without positive expected NPV projects is defined as a **no-growth company** (a term for a company without opportunities for *profitable* growth). Such companies should distribute all their earnings in dividends because earnings cannot be reinvested profitably and will be flat in perpetuity, assuming a constant ROE. This flatness occurs because earnings equal ROE × Equity, and equity is constant because retained earnings are not added to it. If assets are in place to support the growth in earnings for the next year (t = 1) compared with the prior year (t = 0), E_1 is the appropriate measure of earnings to use in estimating the no-growth value per share. E_1 is t = 1 earnings, which is the constant level of earnings or the average earnings of a no-growth company if return on equity is viewed as varying about its average level. The **no-growth value per share** is defined as E_1/r, which is the present value of a perpetuity in the amount of E_1 where the capitalization rate, r, is the required rate of return on the company's equity. E_1/r can also be interpreted as the per-share value of assets in place because of the assumption that the company is making no new investments because none are profitable. For any company, the actual value per share is the sum of the no-growth value per share and the present value of growth opportunities:

$$V_0 = \frac{E_1}{r} + \text{PVGO} \tag{12}$$

If prices reflect value ($P_0 = V_0$), P_0 less E_1/r gives the market's estimate of the company's value of growth, PVGO. Referring back to Example 6, suppose that MSEX is expected to have average EPS of \$1.52 if it distributed all earnings as dividends. Its required return of 6.8% and a current price of \$43.20 gives

$$\begin{aligned} \$43.20 &= (\$1.52/0.068) + \text{PVGO} \\ &= \$22.42 + \text{PVGO} \end{aligned}$$

and PVGO = \$43.20 – \$22.42 = \$20.78. So, 48% (\$20.78/\$43.20 = 0.48) of the company's value, as reflected in the market price, is attributable to the value of growth.

Exhibit 4 presents selected data from early 2019 for three companies: Alphabet, Inc. (NASDAQ: GOOGL), McDonald's Corporation (NYSE: MCD), and Macy's, Inc. (NYSE: M). The data indicate that the value of growth represented about 53% of the market value of technology company Alphabet (the parent company of Google) and a much smaller percentage of McDonald's market value and Macy's market value. The negative value for Macy's PVGO could be explained in several ways: It could reflect the expected continued challenges that traditional retailers face from online competition, or it might indicate that the estimated no-growth value per share was too high because the earnings estimate was too high and/or the required return on equity estimate was too low.

Exhibit 4 Estimated PVGO as a Percentage of Price

Company	β	r	E_1	Price	E_1/r	PVGO	PVGO/Price
Alphabet, Inc.	1.16	8.2%	$47.49	$1,236.34	$579.14	$657.20	53.16%
McDonald's Corp	0.52	5.3%	$8.23	$194.12	$155.28	$38.84	20.01%
Macy's Inc.	0.45	5.0%	$3.09	$25.11	$61.80	($36.69)	n.m.

Source: NASDAQ for earnings estimate and S&P equity research for beta.
Note: The required rate of return is estimated using the CAPM with 3.0% for the risk-free rate of return and 4.5% for the equity risk premium.

What determines PVGO? One determinant is the value of a company's options to invest, captured by the word "opportunities." In addition, the flexibility to adapt investments to new circumstances and information is valuable. Thus, a second determinant of PVGO is the value of the company's options to time the start, adjust the scale, or even abandon future projects. This element is the value of the company's **real options** (options to modify projects, in this context). Companies that have good business opportunities and/or a high level of managerial flexibility in responding to changes in the marketplace should tend to have higher values of PVGO than companies that do not have such advantages. This perspective on what contributes to PVGO can provide additional understanding of the results in Exhibit 4.

As an additional aid to an analyst, Equation 12 can be restated in terms of the familiar P/E based on forecasted earnings:

$$\frac{V_0}{E_1} \text{ or } \frac{P_0}{E_1} \text{ or P/E} = \frac{1}{r} + \frac{\text{PVGO}}{E_1} \tag{13}$$

The first term, $1/r$, is the value of the P/E for a no-growth company. The second term is the component of the P/E value that relates to growth opportunities. For MSEX, the P/E is $43.20/$1.52 = 28.4. The no-growth P/E is 1/0.068 = 14.7 and is the multiple at which the company should sell if it has no growth opportunities. The growth component of $20.78/$1.52 = 13.67 reflects anticipated growth opportunities.

As analysts, the distinction between no-growth and growth values is of interest because the value of growth and the value of assets in place generally have different risk characteristics (as the interpretation of PVGO as incorporating the real options suggests).

5.1 Gordon Growth Model and the Price-to-Earnings Ratio

The price-to-earnings ratio is perhaps the most widely recognized valuation indicator, familiar to readers of newspaper financial tables and institutional research reports. Using the Gordon growth model, one can develop an expression for P/E in terms of the fundamentals. This expression has two uses:

- When used with forecasts of the inputs to the model, the analyst obtains a **justified (fundamental) P/E**—the P/E that is fair, warranted, or justified on the basis of fundamentals (given that the valuation model is appropriate). The analyst can then state his view of value in terms not of the Gordon growth model value but of the justified P/E. Because P/E is so widely recognized, this method may be an effective way to communicate the analysis.
- The analyst may also use the expression for P/E to weigh whether the forecasts of earnings growth built into the current stock price are reasonable. What expected earnings growth rate is implied by the actual market P/E? Is that growth rate plausible?

The expression for P/E can be stated in terms of the current (or trailing) P/E (today's market price per share divided by trailing 12 months' earnings per share) or in terms of the leading (or forward) P/E (today's market price per share divided by a forecast of the next 12 months' earnings per share, or sometimes the next fiscal year's earnings per share).

Leading and trailing justified P/E expressions can be developed from the Gordon growth model. Assuming that the model can be applied for a particular stock's valuation, the dividend payout ratio is considered fixed. Define b as the retention rate, the fraction of earnings reinvested in the company rather than paid out in dividends. The dividend payout ratio is then, by definition, $(1 - b)$ = Dividend per share/Earnings per share = D_t/E_t. If $P_0 = D_1/(r - g)$ is divided by next year's earnings per share, E_1, we have

$$\frac{P_0}{E_1} = \frac{D_1/E_1}{r-g} = \frac{1-b}{r-g} \tag{14}$$

This calculation represents a leading P/E, which is current price divided by next year's earnings. Alternatively, if $P_0 = D_0(1 + g)/(r - g)$ is divided by the current-year's earnings per share, E_0, the result is

$$\frac{P_0}{E_0} = \frac{D_0(1+g)/E_0}{r-g} = \frac{(1-b)(1+g)}{r-g} \tag{15}$$

This expression is for trailing P/E, which is current price divided by trailing (current year) earnings.

EXAMPLE 11

The Justified P/E Based on the Gordon Growth Model

Harry Trice wants to use the Gordon growth model to find a justified P/E for the French company L'Oréal SA (EN Paris: OR), a global cosmetics manufacturer. Trice has assembled the following information:

- Current stock price = €242.70.
- Trailing annual earnings per share = €7.08.
- Current level of annual dividends = €3.85.
- Dividend growth rate = 4.25%.
- Risk-free rate = 2.0%.

- Equity risk premium = 5.0%.
- Beta versus the CAC index = 0.72.

1 Calculate the justified trailing and leading P/Es based on the Gordon growth model.
2 Based on the justified trailing P/E and the actual P/E, judge whether L'Oréal is fairly valued, overvalued, or undervalued.

Solution to 1:

For L'Oréal, the required rate of return using the CAPM is

$$\begin{aligned} r_i &= 2.0\% + 0.72(5.0\%) \\ &= 5.6\% \end{aligned}$$

The dividend payout ratio is

$$\begin{aligned} (1 - b) &= D_0/E_0 \\ &= 3.85/7.08 \\ &= 0.54 \end{aligned}$$

The justified leading P/E (based on next year's earnings) is

$$\frac{P_0}{E_1} = \frac{1-b}{r-g} = \frac{0.5438}{0.056-0.0425} = 40.28.$$

$$\frac{P_0}{E_1} = \frac{1-b}{r-g} = \frac{0.5438}{0.056-0.0425} = 40.28$$

The justified trailing P/E (based on trailing earnings) is

$$\frac{P_0}{E_0} = \frac{(1-b)(1+g)}{r-g} = \frac{0.5438(1.0425)}{0.056-0.0425} = 42.00$$

Solution to 2:

Based on a current price of €242.70 and trailing earnings of €7.08, the trailing P/E is €242.70/€7.08 = 34.3. Because the actual P/E of 34.3 is smaller than the justified trailing P/E of 42.0, the conclusion is that L'Oréal appears to be undervalued. The apparent mispricing can also be expressed in terms of price using the Gordon growth model. Using Trice's assumptions, the Gordon growth model assigns a value of 3.85(1.0425)/(0.05 − 0.0425) = €297.31, which is above the current market price of €242.70.

We will later present multistage DDMs. Expressions for the P/E can be developed in terms of the variables of multistage DDMs, but the usefulness of these expressions is not commensurate with their complexity. For multistage models, the simple way to calculate a justified leading P/E is to divide the model value directly by the first year's expected earnings. In all cases, the P/E is explained in terms of the required return on equity, expected dividend growth rate(s), and the dividend payout ratio(s). All else equal, higher prices are associated with higher anticipated dividend growth rates.

5.2 Estimating a Required Return Using the Gordon Growth Model

Under the assumption of efficient prices, the Gordon growth model has been used to estimate a stock's required rate of return, or equivalently, the market-price-implied expected return. The Gordon growth model solved for r is

$$r = \frac{D_0(1+g)}{P_0} + g = \frac{D_1}{P_0} + g \tag{16}$$

As explained in the coverage of return concepts, r in Equation 16 is technically an internal rate of return (IRR). The rate r is composed of two parts: the dividend yield (D_1/P_0) and the capital gains (or appreciation) yield (g).

EXAMPLE 12

Finding the Expected Rate of Return with the Gordon Growth Model

Bob Inguigiatto, CFA, has been given the task of developing mean return estimates for a list of stocks as preparation for a portfolio optimization. On his list is NextEra Energy, Inc. (NYSE: NEE). On analysis, he decides that it is appropriate to model NextEra Energy using the Gordon growth model, and he takes prices as reflecting value. The company paid dividends of \$4.44 in 2018 and in February 2019 announced an increase in quarterly dividends from \$1.11 to \$1.25, implying an annual dividend of \$5.00. The current stock price is \$169.83. The growth rate of dividends per share has averaged around 11.0% per year, based on the past five years. NextEra's recent earnings growth has been affected by non-recurring items, but based on his analysis, Inguigiatto has decided to use 5.50% as his best estimate of the long-term earnings and dividend growth rate. Next year's projected dividend, D_1, is \$5.00(1.055) = \$5.275. Using the Gordon growth model, NextEra Energy's expected rate of return is

$$\begin{aligned} r &= \frac{D_1}{P_0} + g \\ &= \frac{5.275}{169.83} + 0.055 \\ &= 0.0311 + 0.055 \\ &= 0.0860 = 8.60\% \end{aligned}$$

The expected rate of return can be broken into two components: the dividend yield (D_1/P_0 = 3.11%) and the capital gains yield (g = 5.50%).

5.3 The Gordon Growth Model: Concluding Remarks

The Gordon growth model is the simplest practical implementation of discounted dividend valuation. The Gordon growth model is appropriate for valuing the equity of dividend-paying companies when its key assumption of a stable future dividend and earnings growth rate is expected to be satisfied. Broad equity market indexes of developed markets frequently satisfy the conditions of the model fairly well. As a result, analysts have used it to judge whether an equity market is fairly valued or not and for estimating the equity risk premium associated with the current market level. In the multistage models discussed in the next section, the Gordon growth model has often been used to model the last growth stage, when a previously high-growth company matures and the growth rate drops to a long-term sustainable level. In any case in which the model is applied, the analyst must be aware that the model's output is typically sensitive to small changes in the assumed growth rate and required rate of return.

The Gordon growth model is a single-stage DDM because all future periods are grouped into one stage characterized by a single growth rate. For many or even the majority of companies, however, future growth can be expected to consist of multiple stages. Multistage DDMs are the subject of the next section.

6 MULTISTAGE DIVIDEND DISCOUNT MODELS: TWO-STAGE DIVIDEND DISCOUNT MODEL AND VALUING A NON-DIVIDEND PAYING COMPANY

i. explain the growth phase, transition phase, and maturity phase of a business;

j. explain the assumptions and justify the selection of the two-stage DDM, the H-model, the three-stage DDM, or spreadsheet modeling to value a company's common shares;

k. describe terminal value and explain alternative approaches to determining the terminal value in a DDM;

l. calculate and interpret the value of common shares using the two-stage DDM, the H-model, and the three-stage DDM;

p. evaluate whether a stock is overvalued, fairly valued, or undervalued by the market based on a DDM estimate of value.

Earlier we noted that the basic expression for the DDM (Equation 7) is too general for investment analysts to use in practice because one cannot forecast individually more than a relatively small number of dividends. The strongest simplifying assumption—a stable dividend growth rate from now into the indefinite future, leading to the Gordon growth model—is unrealistic for many or even most companies. For many publicly traded companies, practitioners have typically assumed that growth falls into three stages (see Sharpe, Alexander, and Bailey 1999):

- **Growth phase.** A company in its growth phase typically enjoys rapidly expanding markets, high profit margins, and an abnormally high growth rate in earnings per share (**supernormal growth**). Companies in this phase often have negative free cash flow to equity because the company invests heavily in expanding operations. Given high prospective returns on equity, the dividend payout ratios of growth-phase companies are often low or even zero. As the company's markets mature or as unusual growth opportunities attract competitors, earnings growth rates eventually decline.
- **Transition phase.** In this phase, which is a transition to maturity, earnings growth slows as competition puts pressure on prices and profit margins or as sales growth slows because of market saturation. In this phase, earnings growth rates may be above average but declining toward the growth rate for the overall economy. Capital requirements typically decline in this phase, often resulting in positive free cash flow and increasing dividend payout ratios (or the initiation of dividends).
- **Mature phase.** In maturity, the company reaches an equilibrium in which investment opportunities on average just earn their opportunity cost of capital. Return on equity approaches the required return on equity, and earnings growth, the dividend payout ratio, and the return on equity stabilize at levels

that can be sustained long term. The dividend and earnings growth rate of this phase is called the **mature growth rate**. This phase, in fact, reflects the stage in which a company can properly be valued using the Gordon growth model, and that model is one tool for valuing this phase of a current high-growth company's future.

A company may attempt and succeed in restarting the growth phase by changing its strategic focuses and business mix. Technological advances may alter a company's growth prospects for better or worse with surprising rapidity. Nevertheless, this growth-phase picture of a company is a useful approximation. The growth-phase concept provides the intuition for multistage discounted cash flow (DCF) models of all types, including multistage dividend discount models. Multistage models are a staple valuation discipline of investment management firms using DCF valuation models.

A survey of CFA Institute members with job responsibility for equity analysis indicates that, among respondents using a dividend discount model, two-stage and multistage models are used more often than the single-stage model (Stowe, Pinto, and Robinson 2018). Among analysts using a dividend discount model, 55% use a two-stage model, 11% use an H-model (a type of two-stage model), and 50% use a model with more than two stages (Stowe, Pinto, and Robinson 2018). (Because analysts often use more than one model, the response percentages add up to more than 100%).

In the following sections, we present three popular multistage DDMs: the two-stage DDM, the H-model, and the three-stage DDM. Keep in mind that all these models represent stylized patterns of growth; they are attempting to identify the pattern that most accurately approximates an analyst's view of the company's future growth.

6.1 Two-Stage Dividend Discount Model

Two common versions of the two-stage DDM exist. Both versions assume constant growth at a mature growth rate (for example, 7%) in Stage 2. In the first version ("the general two-stage model"), the whole of Stage 1 represents a period of abnormal growth—for example, growth at 15%. The transition to mature growth in Stage 2 is generally abrupt.

In the second version, called the H-model, the dividend growth rate is assumed to decline from an abnormal rate to the mature growth rate during the course of Stage 1. For example, the growth rate could begin at 15% and decline continuously in Stage 1 until it reaches 7%. The second model will be presented after the general two-stage model.

The first two-stage DDM provides for a high growth rate for the initial period, followed by a sustainable and usually lower growth rate thereafter. The two-stage DDM is based on the multiple-period model

$$V_0 = \sum_{t=1}^{n} \frac{D_t}{(1+r)^t} + \frac{V_n}{(1+r)^n} \tag{17}$$

where V_n is used as an estimate of P_n. The two-stage model assumes that the first n dividends grow at an extraordinary short-term rate, g_S:

$$D_t = D_0(1+g_S)^t$$

After time n, the annual dividend growth rate changes to a normal long-term rate, g_L. The dividend at time n + 1 is $D_{n+1} = D_n(1 + g_L) = D_0(1 + g_S)^n(1 + g_L)$, and this dividend continues to grow at g_L. Using D_{n+1}, an analyst can use the Gordon growth model to find V_n:

$$V_n = \frac{D_0(1+g_S)^n(1+g_L)}{r-g_L} \tag{18}$$

To find the value at $t = 0$, V_0, simply find the present value of the first n dividends and the present value of the projected value at time n.

$$V_0 = \sum_{t=1}^{n} \frac{D_0(1+g_S)^t}{(1+r)^t} + \frac{D_0(1+g_S)^n(1+g_L)}{(1+r)^n(r-g_L)} \tag{19}$$

EXAMPLE 13

Valuing a Stock Using the Two-Stage Dividend Discount Model

Carl Zeiss Meditec AG (AFX:GR), 65% owned by the Carl Zeiss Group, provides screening, diagnostic, and therapeutic systems for the treatment of ophthalmologic (vision) problems. Reviewing the issue as of early 2019, when it is trading for €80.55, Hans Mattern, a buy-side analyst covering Meditec, forecasts that the current dividend of €0.55 will grow by 9% per year during the next 10 years. Thereafter, Mattern believes that the growth rate will decline to 5% and remain at that level indefinitely.

Mattern estimates Meditec's required return on equity as 5.88% based on a beta of 0.90 against the equity market benchmark DAX, a 1.2% risk-free rate, and his equity risk premium estimate of 5.2%.

Exhibit 5 shows the calculations of the first 10 dividends and their present values discounted at 5.88%. The terminal stock value at $t = 10$ is

$$\begin{aligned} V_{10} &= \frac{D_0(1+g_S)^n(1+g_L)}{r-g_L} \\ &= \frac{0.55(1.09)^{10}(1.05)}{0.0588-0.05} \\ &= 155.358 \end{aligned}$$

The terminal stock value and its present value are also given.

Exhibit 5 Carl Zeiss Meditec AG

Time	Value	Calculation	D_t or V_t	Present Values $D_t/(1.0588)^t$ or $V_t/(1.0588)^t$
1	D_1	$= 0.55 \times (1 + 0.09)^1$	€0.600	€0.5662
2	D_2	$= 0.55 \times (1 + 0.09)^2$	0.653	0.5829
3	D_3	$= 0.55 \times (1 + 0.09)^3$	0.712	0.6001
4	D_4	$= 0.55 \times (1 + 0.09)^4$	0.776	0.6178
5	D_5	$= 0.55 \times (1 + 0.09)^5$	0.846	0.6360
6	D_6	$= 0.55 \times (1 + 0.09)^6$	0.922	0.6547
7	D_7	$= 0.55 \times (1 + 0.09)^7$	1.005	0.6740
8	D_8	$= 0.55 \times (1 + 0.09)^8$	1.096	0.6938
9	D_9	$= 0.55 \times (1 + 0.09)^9$	1.195	0.7143
10	D_{10}	$= 0.55 \times (1 + 0.09)^{10}$	1.302	0.7353

Exhibit 5 (Continued)

Time	Value	Calculation	D_t or V_t	Present Values $D_t/(1.0588)^t$ or $V_t/(1.0588)^t$
10	V_{10}	$= [0.55 \times (1 + 0.09)^{10} \times 1.05]/(0.0588 - 0.05)$	155.358	87.7395
Total				€94.2145

In this two-stage model, the dividends are forecast during the first stage and then their present values are calculated. The Gordon growth model is used to derive the terminal value (the value of the dividends in the second stage as of the beginning of that stage). As shown in Exhibit 5, the terminal value is $V_{10} = D_{11}/(r - g_L)$. Ignoring rounding errors, the Period 11 dividend is €1.3671 (= D_{10} × 1.05 = €1.302 × 1.05). By using the standard Gordon growth model, V_{10} = €155.36 = €1.3671/(0.0588 − 0.05). The present value of the terminal value is €87.74 = €155.36/1.0588^{10}. The total estimated value of Meditec is €94.21 using this model. Notice that approximately 93% of this value, €87.74, is the present value of V_{10}, and the balance, €94.21 − €87.74 = €6.47, is the present value of the first 10 dividends. If we recall the discussion of the sensitivity of the Gordon growth model to changes in the inputs, we can calculate an interval for the intrinsic value of Meditec by varying the mature growth rate through the range of plausible values.

The two-stage DDM is useful because many scenarios exist in which a company can achieve a supernormal growth rate for a few years, after which time the growth rate falls to a more sustainable level. For example, a company may achieve supernormal growth through possession of a patent, first-mover advantage, or another factor that provides a temporary lead in a specific marketplace. Subsequently, earnings will most likely descend to a level that is more consistent with competition and growth in the overall economy. Accordingly, that is why in the two-stage model, extraordinary growth is often forecast for a few years and normal growth is forecast thereafter. A possible limitation of the two-stage model is that the transition between the initial abnormal growth period and the final steady-state growth period is abrupt.

The accurate estimation of V_n, the **terminal value of the stock** (also known as its **continuing value**) is an important part of the correct use of DDMs. In practice, analysts estimate the terminal value either by applying a multiple to a projected terminal value of a fundamental, such as earnings per share or book value per share, or they estimate V_n using the Gordon growth model. In our coverage of market multiples, we will discuss using price–earnings multiples in this context.

In the examples, a single discount rate, r, is used for all phases, reflecting both a desire for simplicity and lack of a clear objective basis for adjusting the discount rate for different phases. Some analysts, however, use different discount rates for different growth phases.

The following example values P&G (Procter & Gamble Company) by combining the dividend discount model and a P/E valuation model.

EXAMPLE 14

Combining a DDM and P/E Model to Value a Stock

An analyst is reviewing the valuation of Procter & Gamble Company known as "P&G" (NYSE: PG) as of the beginning of 2019 when P&G was selling for \$96.47. In the previous year, P&G paid a \$2.79 dividend that the analyst expects to grow at a rate of 4% annually for the next four years. At the end of Year 4, the analyst expects the dividend to equal 60% of earnings per share and the trailing P/E for P&G to be 22. If the required return on P&G common stock is 6.5%, calculate the per-share value of P&G common stock.

Exhibit 6 summarizes the relevant calculations. When the dividends are growing at 4%, the expected dividends and the present value of each (discounted at 6.5%) are shown. The terminal stock price, V_4, deserves some explanation. As shown in the table, the Year 4 dividend is $\$2.79(1.04)^4 = \$3.2639$. Because dividends at that time are assumed to be 60% of earnings, the EPS projection for Year 4 is $EPS_4 = D_4/0.60 = \$3.2639/0.60 = \$5.4398$. With a trailing P/E of 22.0, the value of P&G at the end of Year 4 would be 22.0(\$5.4398) = \$119.6765. Discounted at 6.5% for four years, the present value of V_4 is \$93.0273.

Exhibit 6 Value of Procter & Gamble Common Stock

Time	Value	Calculation	D_t or V_t	Present Values $D_t/(1.065)^t$ or $V_t/(1.065)^t$
1	D_1	$\$2.79(1.04)^1$	\$2.9016	\$2.7245
2	D_2	$\$2.79\ (1.04)^2$	3.0177	2.6606
3	D_3	$\$2.79\ (1.04)^3$	3.1384	2.5981
4	D_4	$\$2.79\ (1.04)^4$	3.2639	2.5371
4	V_4	$22 \times [2.79\ (1.04)^4/0.60]$ $= 22 \times (3.2639/0.60)$ $= 22 \times 5.4398$	119.6765	93.0273
Total				\$103.5476

The present values of the dividends for Years 1 through 4 sum to \$10.52. The present value of the terminal value of \$119.68 is \$93.03. The estimated total value of P&G's common stock is the sum of these, or \$103.55 per share.

6.2 Valuing a Non-Dividend-Paying Company

The fact that a stock is currently paying no dividends does not mean that the principles of the dividend discount model do not apply. Even though D_0 and/or D_1 may be zero, and the company may not begin paying dividends for some time, the present value of future dividends may still capture the value of the company. Of course, if a company pays no dividends and will never be able to distribute cash to shareholders, the stock is worthless.

To value a non-dividend-paying company using a DDM, generally an analyst can use a multistage DDM model in which the first-stage dividend equals zero. Example 15 illustrates the approach.

EXAMPLE 15

Valuing a Non-Dividend-Paying Stock

Assume that a company is currently paying no dividend and will not pay one for several years. If the company begins paying a dividend of \$1.00 five years from now, and the dividend is expected to grow at 5% thereafter, this future dividend stream can be discounted back to find the value of the company. This company's required rate of return is 11%. Because the expression

$$V_n = \frac{D_{n+1}}{r - g}$$

values a stock at period n using the next period's dividend, the t = 5 dividend is used to find the value at t = 4:

$$V_4 = \frac{D_5}{r - g} = \frac{1.00}{0.11 - 0.05} = \$16.67$$

To find the value of the stock today, simply discount V_4 back for four years:

$$V_0 = \frac{V_4}{(1 + r)^4} = \frac{16.67}{(1.11)^4} = \$10.98$$

The value of this stock, even though it will not pay a dividend until Year 5, is \$10.98.

If a company is not paying a dividend but is very profitable, an analyst might be willing to forecast its future dividends. Of course, for non-dividend-paying, unprofitable companies, such a forecast would be very difficult. Furthermore, as discussed previously, it is usually difficult for the analyst to estimate the timing of the initiation of dividends and the dividend policy that will then be established by the company. Thus, the analyst may prefer a free cash flow or residual income model for valuing such companies.

7 THE H-MODEL AND THREE-STAGE DIVIDEND DISCOUNT MODELS

j explain the assumptions and justify the selection of the two-stage DDM, the H-model, the three-stage DDM, or spreadsheet modeling to value a company's common shares;

k describe terminal value and explain alternative approaches to determining the terminal value in a DDM;

l calculate and interpret the value of common shares using the two-stage DDM, the H-model, and the three-stage DDM;

p evaluate whether a stock is overvalued, fairly valued, or undervalued by the market based on a DDM estimate of value.

The basic two-stage model assumes a constant, extraordinary rate for the supernormal growth period that is followed by a constant, normal growth rate thereafter. The difference in growth rates may be substantial. For instance, in Example 13, the assumed growth rate for Carl Zeiss Meditec was 9% annually for 10 years, followed by a drop to 5% growth in Year 11 and thereafter. In some cases, a smoother transition to the mature phase growth rate would be more realistic. Fuller and Hsia (1984) developed a variant of the two-stage model in which growth begins at a high rate and declines linearly throughout the supernormal growth period until it reaches a normal rate at the end. The value of the dividend stream in the H-model is

$$V_0 = \frac{D_0(1+g_L)}{r-g_L} + \frac{D_0 H(g_S - g_L)}{r-g_L} \quad (20)$$

or

$$V_0 = \frac{D_0(1+g_L) + D_0 H(g_S - g_L)}{r-g_L}$$

where

V_0 = value per share at $t = 0$
D_0 = current dividend
r = required rate of return on equity
H = half-life in years of the high-growth period (i.e., high-growth period = $2H$ years)
g_S = initial short-term dividend growth rate
g_L = normal long-term dividend growth rate after Year $2H$

The first term on the right-hand side of Equation 20 is the present value of the company's dividend stream if it were to grow at g_L forever. The second term is an approximation of the extra value (assuming $g_S > g_L$) accruing to the stock because of its supernormal growth for Years 1 through $2H$ (see Fuller and Hsia 1984 for technical details). Logically, the longer the supernormal growth period (i.e., the larger the value of H, which is one-half the length of the supernormal growth period) and the larger the extra growth rate in the supernormal growth period (measured by g_S minus g_L), the higher the share value, all else equal.

We can provide some intuition on the expression. On average, the expected excess growth rate in the supernormal period will be $(g_S - g_L)/2$. Through $2H$ periods, a total excess amount of dividends (compared with the level given g_L) of $2HD0(g_S - g_L)/2 = D0H(g_S - g_L)$ is expected. This term is the H-model upward adjustment to the first dividend term, reflecting the extra expected dividends as growth declines from g_S to g_L during the first period. Note, however, that the timing of the individual dividends in the first period is not reflected by individually discounting them; the expression is thus an approximation.

To illustrate the expression, if the analyst in Example 13 had forecast a linear decline of the growth rate from 9% to 5% over the next 10 years, his estimate of value of Meditec using the H-model would have been €78.13 (rather than €94.21 as in Example 13):

$$\begin{aligned} V_0 &= \frac{D_0(1+g_L) + D_0 H(g_S - g_L)}{r-g_L} \\ &= \frac{55(1.05) + 55(5)(0.09 - 0.05)}{0.0588 - 0.05} \\ &= 78.13 \end{aligned}$$

Note that an H of 5 corresponds to the 10-year high-growth period of Example 13. Example 16 provides another illustration of the H-model.

EXAMPLE 16

Valuing a Stock with the H-Model

An analyst has decided to use the H-model to estimate the value of a company and has gathered the following facts and forecasts:

- The share price is €41.70.
- The current dividend is €1.77.
- The initial dividend growth rate is 7%, declining linearly during a 10-year period to a final and perpetual growth rate of 4%.
- The analyst estimates the company's required rate of return on equity as 8.0%.

1 Using the H-model and the information given, estimate the company's per-share value.
2 Estimate the value of the company's shares if its normal growth period began immediately.
3 Evaluate whether the company's shares appear to be fairly valued, overvalued, or undervalued.

Solution to 1:

Using the H-model expression gives

$$\begin{aligned} V_0 &= \frac{D_0\left(1 + g_L\right) + D_0 H\left(g_S - g_L\right)}{r - g_L} \\ &= \frac{1.77(1.04) + 1.77(5)(0.07 - 0.04)}{0.08 - 0.04} \\ &= \frac{1.84 + 0.27}{0.04} \\ &= 52.75 \end{aligned}$$

Solution to 2:

If the company experienced normal growth starting now, its estimated value would be the first component of the H-model estimate, €46 (=1.84/0.04). The faster initial growth assumption adds €6.75 (=0.27/0.04) to its value, resulting in an estimated value of €52.75 per share.

Solution to 3:

€52.75 is approximately 26% higher than the company's current market price of €41.70. Thus the company appears to be undervalued.

The H-model is an approximation model that estimates the valuation that would result from discounting all of the future dividends individually. In many circumstances, this approximation is very close. For a long extraordinary growth period (a high H) or for a large difference in growth rates (the difference between g_S and g_L), however, the analyst might abandon the approximation model for the more exact model. Fortunately, the many tedious calculations of the exact model are made fairly easy using a spreadsheet program.

7.1 Three-Stage Dividend Discount Models

There are two popular versions of the three-stage DDM, distinguished by the modeling of the second stage. In the first version ("the general three-stage model"), the company is assumed to have three distinct stages of growth and the growth rate of the second stage is typically constant. For example, Stage 1 could assume 20% growth for three years, Stage 2 could have 10% growth for four years, and Stage 3 could have 5% growth thereafter. In the second version, the growth rate in the middle (second) stage is assumed to decline linearly to the mature growth rate: essentially, the second and third stages are treated as an H-model.

The following example shows how the first type of the three-stage model can be used to value a stock.

EXAMPLE 17

The Three-Stage DDM with Three Distinct Stages

An analyst is analyzing a technology company and makes the following estimates:

- the current required return on equity for the company is 9%; and
- dividends will grow at 14% for the next two years, 12% for the following five years, and 6.75% thereafter.

The company pays a dividend of \$3.30 per year, and its stock currently trades at \$194.98. Based only on the information given, estimate the value of the company's stock using a three-stage DDM approach.

Solution:

Exhibit 7 gives the calculations.

Exhibit 7 Estimated Value Using a Three-Stage DDM

Time	Value	Calculation	D_t or V_t	Present Values $D_t/(1.09)^t$ or $V_t/(1.09)^t$
1	D_1	3.30(1.14)	\$3.7620	\$3.4514
2	D_2	$3.30(1.14)^2$	4.2887	3.6097
3	D_3	$3.30(1.14)^2(1.12)$	4.8033	3.7090
4	D_4	$3.30(1.14)^2(1.12)^2$	5.3797	3.8111
5	D_5	$3.30(1.14)^2(1.12)^3$	6.0253	3.9160
6	D_6	$3.30(1.14)^2(1.12)^4$	6.7483	4.0238
7	D_7	$3.30(1.14)^2(1.12)^5$	7.5581	4.1346
7	V_7	$3.30(1.14)^2(1.12)^5(1.0675)/(0.09 - 0.0675)$	\$358.5908	196.161
Total				\$222.8171

Given these assumptions, the three-stage model indicates that a fair price should be \$222.82, more than 14% above the current market price. Characteristically, the present value of the terminal value of \$196.16 constitutes the overwhelming portion (here, about 88%) of total estimated value.

A second version of the three-stage DDM has a middle stage similar to the first stage in the H-model. In the first stage, dividends grow at a high, constant (supernormal) rate for the whole period. In the second stage, dividends decline linearly as they do in the H-model. Finally, in Stage 3, dividends grow at a sustainable, constant growth rate. The process of using this model involves four steps:

- Gather the required inputs:
  - the current dividend;
  - estimates of the lengths of the first, second, and third stages and the expected growth rate during each stage; and
  - an estimate of the required return on equity.
- Compute the expected dividends in the first stage and find the sum of their present values.
- Apply the H-model expression to the second and third stages to obtain an estimate of their value as of the beginning of the second stage. Then find the present value of this H-value as of today ($t = 0$).
- Sum the values obtained in the second and third steps.

In the first step, analysts often investigate the company more deeply, making explicit, individual earnings and dividend forecasts for the near future (often 3, 5, or 10 years), rather than applying a growth rate to the current level of dividends.

EXAMPLE 18

The Three-Stage DDM with Declining Growth Rates in Stage 2

Elsie Bouvier is evaluating Rhinestone Energy (a hypothetical company) for possible inclusion in a small-cap, growth-oriented portfolio. The company is a diversified energy company involved in oil and gas exploration as well as natural gas distribution. In light of Rhinestone Energy's aggressive program of purchasing oil and gas producing properties, Bouvier expects above-average growth for the next five years. She establishes the following facts and forecasts:

- The current market price is $56.18.
- The current dividend is $0.56.
- Bouvier forecasts an initial five-year period of 11% per year earnings and dividend growth.
- Bouvier anticipates that Rhinestone Energy can grow 6.5% per year as a mature company and allows 10 years for the transition to the mature growth period.
- To estimate the required return on equity using the CAPM, Bouvier uses an adjusted beta of 1.2 based on two years of weekly observations, an estimated equity risk premium of 4.2%, and a risk-free rate based on long bond yields of 3%.
- Bouvier considers any security trading within a band of ±20% of her estimate of intrinsic value to be within a "fair value range."

1 Estimate the required return on Rhinestone Energy's equity using the CAPM. (Use only one decimal place in stating the result.)

2 Estimate the value of Rhinestone Energy's common stock using a three-stage dividend discount model with a linearly declining dividend growth rate in Stage 2.

3 Calculate the percentages of the total value represented by the first stage and by the second and third stages considered as one group.

4 Judge whether Rhinestone Energy's stock is undervalued or overvalued according to Bouvier's perspective.

5 Some analysts are forecasting essentially flat EPS and dividends in the second year. Estimate the value of Rhinestone Energy's stock under the assumptions that EPS is flat in the second year and that 11% growth resumes in the third year.

Solution to 1:

The required return on equity is r = 3% + 1.2(4.2%) = 8%.

Solution to 2:

The first step is to compute the five dividends in Stage 1 and find their present values at 8%. The dividends in Stages 2 and 3 can be valued with the H-model, which estimates their value at the beginning of Stage 2. This value is then discounted back to find the dividends' present value at t = 0.

The calculation of the five dividends in Stage 1 and their present values are given in Exhibit 8. The H-model for calculating the value of the Stage 2 and Stage 3 dividends at the beginning of Stage 2 (t = 5) is

$$V_5 = \frac{D_5(1 + g_L)}{r - g_L} + \frac{D_5 H(g_S - g_L)}{r - g_L}$$

where

$$D_5 = D_0(1 + g_S)^5 = 0.56(1.11)^5 = \$0.9436$$
$$g_S = 11.0\%$$
$$g_L = 6.5\%$$
$$r = 8.0\%$$
$$H = 5 \text{ (the second stage lasts } 2H = 10 \text{ years)}$$

Substituting these values into the equation for the H-model gives V_5 as follows:

$$\begin{aligned} V_5 &= \frac{0.9436(1.065)}{0.08 - 0.065} + \frac{0.9436(5)(0.11 - 0.065)}{0.08 - 0.065} \\ &= 66.9979 + 14.1545 \\ &= \$81.1524 \end{aligned}$$

The present value of V_5 is $\$81.1524/(1.08)^5$ = \$55.2310.

Exhibit 8 Rhinestone Energy

Time	D_t or V_t	Explanation of D_t or V_t	Value of D_t or V_t	PV at 8%
1	D_1	$0.56(1.11)^1$	\$0.6216	\$0.5756
2	D_2	$0.56(1.11)^2$	0.6900	0.5915
3	D_3	$0.56(1.11)^3$	0.7659	0.6080
4	D_4	$0.56(1.11)^4$	0.8501	0.6249
5	D_5	$0.56(1.11)^5$	0.9436	0.6422

Exhibit 8 (Continued)

Time	D_t or V_t	Explanation of D_t or V_t	Value of D_t or V_t	PV at 8%
5	V_5	H-model explained earlier	\$81.1524	55.2310
Total				\$58.2731

According to the three-stage DDM model, the total value of Rhinestone Energy is \$58.27.

Solution to 3:

The sum of the first five present value amounts in the last column of Exhibit 8 is \$3.0422. Thus, the first stage represents \$3.0422/\$58.2731 = 5.2% of total value. The second and third stages together represent 100% − 5.2% = 94.8% of total value (check: \$55.2310/\$58.2731 = 94.8%).

Solution to 4:

The band Bouvier is looking at is \$58.27 ± 0.20(\$58.27), which runs from \$58.27 + \$11.65 = \$69.92 on the upside to \$58.27 − \$11.65 = \$46.62 on the downside. Because the current price of \$56.18 is between \$46.62 and \$69.92, Bouvier would consider Rhinestone Energy to be fairly valued.

Solution to 5:

The estimated value becomes \$52.56 with no growth in Year 2 as shown in Exhibit 9. The value of the second and third stages is given by

$$V_5 = \frac{0.8501(1.065)}{0.08 - 0.065} + \frac{0.8501(5)(0.11 - 0.065)}{0.08 - 0.065} = \$73.1103$$

Exhibit 9 Rhinestone Energy with No Growth in Year 2

Time	D_t or V_t	Explanation of D_t or V_t	Value of D_t or V_t	PV at 8%
1	D_1	$0.56(1.11)^1$	\$0.6216	\$0.5756
2	D_2	No growth in Year 2	0.6216	0.5329
3	D_3	$0.56(1.11)^2$	0.6900	0.5477
4	D_4	$0.56(1.11)^3$	0.7659	0.5629
5	D_5	$0.56(1.11)^4$	0.8501	0.5786
5	V_5	H-model explained earlier	\$73.1103	49.7576
Total				\$52.5553

In Problem 5 of Example 18, the analyst examined the consequences of 11% growth in Year 1 and no growth in Year 2, with 11% growth resuming in Years 3, 4, and 5. In the first stage, analysts may forecast earnings and dividends individually for a certain number of years.

The three-stage DDM with declining growth in Stage 2 has been widely used among companies using a DDM approach to valuation. An example is the DDM adopted by Bloomberg L.P., a financial services company that provides "Bloomberg terminals" to professional investors and analysts. The Bloomberg DDM is a model that provides an estimated value for any stock that the user selects. The DDM is a three-stage model with declining growth in Stage 2. The model uses earnings estimates for assumed Stage 1 and the cost of capital for Stage 3 growth rates, and then it assumes that the Stage 2 rate is a linearly declining rate between the Stage 1 and Stage 3 rates. The model also makes estimates of the required rate of return and the lengths of the three stages, assigning higher-growth companies shorter growth periods (i.e., first stages) and longer transition periods, and slower-growth companies longer growth periods and shorter transition periods. Fixing the total length of the growth and transition phases together at 17 years, the growth stage/transition stage durations for Bloomberg's four growth classifications are 3 years/14 years for "explosive growth" equities, 5 years/12 years for "high growth" equities, 7 years/10 years for "average growth" equities, and 9 years/8 years for "slow/mature growth" equities. Analysts, by tailoring stage specifications to their understanding of the specific company being valued, should be able to improve on the accuracy of valuations compared with a fixed specification.

8 GENERAL MODELING AND ESTIMATING A REQUIRED RETURN USING ANY DDM

m explain the use of spreadsheet modeling to forecast dividends and to value common shares;

n estimate a required return based on any DDM, including the Gordon growth model and the H-model;

DDMs, such as the Gordon growth model and the multistage models presented earlier, assume stylized patterns of dividend growth. An analyst can use *any* assumed dividend pattern, however, to create a spreadsheet to value the stock and to test sensitivity of the value to growth and return assumptions. The following example presents the results of a valuation incorporating dividends that are estimated to change substantially over the forecast period.

EXAMPLE 19

Finding the Value of a Stock with Varying Dividend Assumptions

Yang Co. is expected to pay a $21.00 dividend next year. An analyst estimates that the dividend will decline by 10% annually for the following three years (i.e., the "growth rate" will equal –10%). In Year 5, Yang is expected to sell off assets worth $100 per share. The Year 5 dividend, which includes a distribution of some of the proceeds of the asset sale, is expected to be $60. In Year 6, the dividend is expected to decrease to $40 and to be maintained at $40 for one additional year. The dividend is then expected to grow by 5% annually thereafter. If the required rate of return is 12%, what is the value of one share of Yang?

Solution:

The value is shown in Exhibit 10. Each dividend, its present value discounted at 12%, and an explanation are included in the table. The final row treats the dividends from $t = 8$ forward as a Gordon growth model because after Year 7, the dividend grows at a constant 5% annually. V_7 is the value of these dividends at $t = 7$.

Exhibit 10 Value of Yang Co. Stock

Year	D_t or V_t	Value of D_t or V_t	Present Value at 12%	Explanation of D_t or V_t
1	D_1	$21.00	$18.75	Dividend set at $21
2	D_2	18.90	15.07	Previous dividend × 0.90
3	D_3	17.01	12.11	Previous dividend × 0.90
4	D_4	15.31	9.73	Previous dividend × 0.90
5	D_5	60.00	34.05	Set at $60
6	D_6	40.00	20.27	Set at $40
7	D_7	40.00	18.09	Set at $40
7	V_7	600.00	271.41	$V_7 = D_8/(r - g)$ $V_7 = (40.00 \times 1.05)/(0.12 - 0.05)$
Total			$399.48	

As the table in Example 19 shows, the total present value of Yang Co.'s dividends is $399.48. In this example, the terminal value of the company (V_n) at the end of the first stage is found using the Gordon growth model and a mature growth rate of 5%.

Several alternative approaches to estimating g are available in this context:

- Use the formula g = (b in the mature phase) × (ROE in the mature phase). We will discuss the expression $g = b \times \text{ROE}$ later. Analysts estimate mature-phase ROE in several ways, such as the following:
  - The DuPont decomposition of ROE based on forecasts for the components of the DuPont expression.
  - Setting ROE = r, the required rate of return on equity, based on the assumption that in the mature phase companies can do no more than earn investors' opportunity cost of capital.
  - Setting ROE in the mature phase equal to the median industry ROE.
- The analyst may estimate the growth rate, g, with other models by relating the mature growth rate to macroeconomic, including industry, growth projections.

When the analyst uses the sustainable growth expression, the earnings retention ratio, b, may be empirically based. For example, Bloomberg L.P.'s model has been assuming that $b = 0.55$ in the mature phase, equivalent to a dividend payout ratio of 45%, a long-run average payout ratio for mature dividend-paying companies in the United States. In addition, sometimes analysts project the dividend payout ratio for the company individually.

EXAMPLE 20

A Sustainable Growth Rate Calculation

An analyst is estimating the dividend growth rate of a company to incorporate in the final stage of a multistage dividend discount model. Assume the company's payout ratio is 25% and its ROE is equal to its estimated required return on equity of 9%. An estimate of the sustainable growth rate can be derived using the expression

$$g = (b \text{ in the mature phase}) \times (\text{ROE in the mature phase})$$
$$= 0.75(9\%) = 6.75\%.$$

The analyst's estimate of the company's sustainable dividend growth rate is 6.75%.

8.1 Estimating a Required Return Using Any DDM

We have focused on finding the value of a security using assumptions for dividends, required rates of return, and expected growth rates. Given current price and all inputs to a DDM except for the required return, an IRR can be calculated. Such an IRR has been used as a required return estimate (although reusing it in a DDM is not appropriate because it risks circularity). This IRR can also be interpreted as the expected return on the issue implied by the market price—essentially, an efficient market expected return. In the following discussion, keep in mind that if price does not equal intrinsic value, the expected return will need to be adjusted to reflect the additional component of return that accrues when the mispricing is corrected, as discussed earlier.

In some cases, finding the IRR is very easy. In the Gordon growth model, $r = D_1/P_0 + g$. The required return estimate is the dividend yield plus the expected dividend growth rate. For a security with a current price of \$10, an expected dividend of \$0.50, and expected growth of 8%, the required return estimate is 13%.

For the H-model, the expected rate of return can be derived as

$$r = \left(\frac{D_0}{P_0}\right)\left[(1 + g_L) + H(g_S - g_L)\right] + g_L \tag{21}$$

When the short- and long-term growth rates are the same, this model reduces to the Gordon growth model. For a security with a current dividend of \$1, a current price of \$20, and an expected short-term growth rate of 10% declining over 10 years (H = 5) to 6%, the expected rate of return would be

$$r = \left(\frac{\$1}{\$20}\right)\left[(1 + 0.06) + 5(0.10 - 0.06)\right] + 0.06 = 12.3\%$$

For multistage models and spreadsheet models, finding a single equation for the rate of return can be more difficult. The process generally used is similar to that of finding the IRR for a series of varying cash flows. Using a computer or trial and error, the analyst must find the rate of return such that the present value of future expected dividends equals the current stock price.

EXAMPLE 21

Finding the Expected Rate of Return for Varying Expected Dividends

An analyst expects Johnson & Johnson's (NYSE: JNJ)) dividend of $3.60 for 2019 to grow by 7.0% for six years and then grow by 5% into perpetuity. A recent price for JNJ as of early 2019 is $136.61. What is the IRR on an investment in JNJ's stock?

In estimating the expected rate of return with a two-stage model, using trial and error is one approach. Having a good initial approximation is helpful. In this case, the expected rate of return formula from the Gordon growth model and JNJ's long-term growth rate can be used to find a first approximation: r = ($3.60 × 1.07)/$136.61 + 0.05 = 7.8%. Because the estimated growth rate for the first six years is higher than the long-term growth rate of 5%, the implied estimated rate of return must be above 7.8%. Exhibit 11 shows the value estimate of JNJ's shares for two discount rates, 8% and 8.5%.

Exhibit 11 Estimation of Required Return: Johnson & Johnson

Time	D_t	Present Value of D_t and V_6 at r = 8%	Present Value of D_t and V_6 at r = 8.5%
1	$3.8520	$3.5667	$3.5502
2	$4.1216	$3.5336	$3.5011
3	$4.4101	$3.5009	$3.4527
4	$4.7188	$3.4685	$3.4050
5	$5.0491	$3.4363	$3.3579
6	$5.4025	$3.4045	$3.3114
7	$5.6726		
Subtotal 1	(t = 1 to 6)	$20.91	$20.58
Subtotal 2	(t = 7 to ∞)	$119.16	$99.34
Total		$140.07	$119.92
Market Price		$136.61	$136.61

In the exhibit, the amount labeled "Subtotal 1" is the present value of the expected dividends for Years 1 through 6. The amount labeled "Subtotal 2" is the present value of the terminal value, $V_6/(1 + r)^6 = [D_7/(r - g)]/(1 + r)^6$. For r = 8%, that present value is $[5.6726/(0.08 - 0.05)]/(1.08)^6$ = $119.16. The present value for other values of r is found similarly.

Using 8.0% as the discount rate, the value estimate for JNJ is $140.07, which is about 2.5% larger than JNJ's market price of $136.61. This fact indicates that the IRR is greater than 8%. With an 8.5% discount rate, the present value of $119.92 is significantly less than the market price. Thus, the IRR is slightly more than 8%. The IRR can be determined to be 8.08% using a spreadsheet. For example, using the Goal Seek function of Excel: In the "set cell" parameter, enter the reference for the cell that contains the Total present value; in the "by changing" parameter, enter the current price as an amount; and in the "by changing cell" parameter, enter the reference for the cell that contains the discount rate.

8.2 Multistage DDM: Concluding Remarks

Multistage dividend discount models can accommodate a variety of patterns of future streams of expected dividends.

In general, multistage DDMs make stylized assumptions about growth based on a lifecycle view of business. The first stage of a multistage DDM frequently incorporates analysts' individual earnings and dividend forecasts for the next two to five years (sometimes longer). The final stage is often modeled using the Gordon growth model based on an assumption of the company's long-run sustainable growth rate. In the case of the H-model, the transition to the mature growth phase happens smoothly during the first stage. In the case of the standard two-stage model, the growth rate typically transitions immediately to mature growth rate in the second period. In three-stage models, the middle stage is a stage of transition. Using a spreadsheet, an analyst can model an almost limitless variety of cash flow patterns.

Multistage DDMs have several limitations. Often, the present value of the terminal stage represents more than three-quarters of the total value of shares. Terminal value can be very sensitive to the growth and required return assumptions. Furthermore, technological innovation can make the lifecycle model a crude representation.

9 THE FINANCIAL DETERMINANTS OF GROWTH RATES: SUSTAINABLE GROWTH RATE AND DIVIDEND GROWTH RATE, RETENTION RATE, AND ROE ANALYSIS

o calculate and interpret the sustainable growth rate of a company and demonstrate the use of DuPont analysis to estimate a company's sustainable growth rate;

In a number of examples earlier, we have implicitly used the relationship that the dividend growth rate (g) equals the earning retention ratio (b) multiplied by the return on equity (ROE). In this section, we explain this relationship and show how it can be combined with a method of analyzing return on equity, called DuPont analysis, as a simple tool for forecasting dividend growth rates.

9.1 Sustainable Growth Rate

We define the **sustainable growth rate** as the rate of dividend (and earnings) growth that can be sustained for a given level of return on equity, assuming that the capital structure is constant through time and that additional common stock is not issued. The reason for studying this concept is that it can help in estimating either 1) the stable growth rate in a Gordon growth model valuation or 2) the mature growth rate in a multistage DDM in which the Gordon growth formula is used to find the terminal value of the stock.

The expression to calculate the sustainable growth rate is

$$g = b \times \text{ROE}, \tag{22}$$

where

g = dividend growth rate

b = earnings retention rate (1 – Dividend payout ratio)

ROE = return on equity

More precisely, in Equation 22 the retention rate should be multiplied by the rate of return expected to be earned on new investment. Analysts commonly assume that the rate of return is well approximated by the return on equity, as shown in Equation 22; however, whether that is actually the case should be investigated by the analyst on a case-by-case basis.

Example 22 illustrates the fact that growth in shareholders' equity is driven by reinvested earnings alone (no new issues of equity and debt growing at the rate g). Note that in scenarios in which debt is growing at g, the capital structure is constant. If the capital structure is not constant, ROE will not be constant in general because ROE depends on leverage.

EXAMPLE 22

Example Showing $g = b \times$ ROE

Suppose that a company's ROE is 25% and its retention rate is 60%. According to the expression for the sustainable growth rate, the dividends should grow at $g = b \times \text{ROE} = 0.60 \times 25\% = 15\%$.

To demonstrate the working of the expression, suppose that, in the year just ended, a company began with shareholders' equity of \$1,000,000, earned \$250,000 net income, and paid dividends of \$100,000. The company begins the next year with \$1,000,000 + 0.60(\$250,000) = \$1,000,000 + \$150,000 = \$1,150,000 of shareholders' equity. No additions to equity are made from the sale of additional shares.

If the company again earns 25% on equity, net income will be 0.25 × \$1,150,000 = \$287,500, which is a \$287,500 – \$250,000 = \$37,500 or a \$37,500/\$250,000 = 0.15% increase from the prior year level. The company retains 60% of earnings, 60% × \$287,500 = \$172,500, and pays out the other 40%, 40% × \$287,500 = \$115,000 as dividends. Dividends for the company grew from \$100,000 to \$115,000, which is exactly a 15% growth rate. With the company continuing to earn 25% each year on the 60% of earnings that is reinvested in the company, dividends would continue to grow at 15%.

Equation 22 implies that the higher the return on equity, the higher the dividend growth rate, all else constant. That relation appears to be reliable. Another implication of the expression is that the lower (higher) the earnings retention ratio, the lower (higher) the growth rate in dividends, holding all else constant; this relationship has been called *the dividend displacement of earnings*. Of course, all else may not be equal—the return on reinvested earnings may not be constant at different levels of investment, or companies with changing future growth prospects may change their dividend policy. Furthermore, research has shown that dividend-paying companies had higher future growth rates during the period studied, indicating that caution is appropriate in assuming that dividends displace earnings (Arnott and Asness 2003; ap Gwilym, Seaton, Suddason, and Thomas 2006; Zhou and Ruland 2006).

A practical logic for defining *sustainable* in terms of growth through internally generated funds (retained earnings) is that external equity (secondary issues of stock) is considerably more costly than internal equity (reinvested earnings), for several reasons including the investment banker fees associated with secondary equity issues.

In general, continuous issuance of new stock is not a practical funding alternative for companies. Growth of capital through issuance of new debt, however, can sometimes be sustained for considerable periods. Further, if a company manages its capital structure to a target percentage of debt to total capital (debt and common stock), it will need to issue debt to maintain that percentage as equity grows through reinvested earnings. (This approach is one of a variety of observed capital structure policies.) In addition, the earnings retention ratio nearly always shows year-to-year variation in actual companies. For example, earnings may have transitory components that management does not want to reflect in dividends. The analyst may thus observe actual dividend growth rates straying from the growth rates predicted by Equation 22 because of these effects, even when her input estimates are unbiased. Nevertheless, the equation can be useful as a simple expression for approximating the average rate at which dividends can grow over a long horizon.

9.2 Dividend Growth Rate, Retention Rate, and ROE Analysis

Thus far we have seen that a company's sustainable growth, as defined earlier, is a function of its ability to generate return on equity (which depends on investment opportunities) and its retention rate. We now expand this model by examining what drives ROE. Remember that ROE is the return (net income) generated on the equity invested in the company:

$$\text{ROE} = \frac{\text{Net income}}{\text{Shareholders' equity}} \tag{23}$$

If a company has a ROE of 15%, it generates \$15 of net income for every \$100 invested in stockholders' equity. For purposes of analyzing ROE, it can be related to several other financial ratios. For example, ROE can be related to return on assets (ROA) and the extent of financial leverage (equity multiplier):

$$\text{ROE} = \frac{\text{Net income}}{\text{Total assets}} \times \frac{\text{Total assets}}{\text{Shareholders' equity}} \tag{24}$$

Therefore, a company can increase its ROE either by increasing ROA or through the use of leverage (assuming the company can borrow at a rate lower than it earns on its assets).

This model can be expanded further by breaking ROA into two components, profit margin and turnover (efficiency):

$$\text{ROE} = \frac{\text{Net income}}{\text{Sales}} \times \frac{\text{Sales}}{\text{Total assets}} \times \frac{\text{Total assets}}{\text{Shareholders' equity}} \tag{25}$$

The first term is the company's profit margin. A higher profit margin will result in a higher ROE. The second term measures total asset turnover, which is the company's efficiency. A turnover of one indicates that a company generates \$1 in sales for every \$1 invested in assets. A higher turnover will result in higher ROE. The last term is the equity multiplier, which measures the extent of leverage, as noted earlier. This relationship is widely known as the DuPont model or analysis of ROE. Although ROE can be analyzed further using a five-way analysis, the three-way analysis will provide insight into the determinants of ROE that are pertinent to our understanding of the growth rate. By combining Equations 22 and 25, we can see that the dividend growth rate is equal to the retention rate multiplied by ROE:

$$\begin{aligned} g = {} & \frac{\text{Net income} - \text{Dividends}}{\text{Net income}} \times \frac{\text{Net income}}{\text{Sales}} \\ & \times \frac{\text{Sales}}{\text{Total assets}} \times \frac{\text{Total assets}}{\text{Shareholders' equity}} \end{aligned} \tag{26}$$

This expansion of the sustainable growth expression has been called the PRAT model. Growth is a function of profit margin (P), retention rate (R), asset turnover (A), and financial leverage (T). The profit margin and asset turnover determine ROA. The other two factors, the retention rate and financial leverage, reflect the company's financial policies. So, the growth rate in dividends can be viewed as determined by the company's ROA and financial policies. Analysts may use Equation 26 to forecast a company's dividend growth rate in the mature growth phase.

Theoretically, the sustainable growth rate expression and this expansion of it based on the DuPont decomposition of ROE hold exactly only when ROE is calculated using beginning-of-period shareholders' equity, as illustrated in Example 22. Such calculation assumes that retained earnings are not available for reinvestment until the end of the period. Analysts and financial databases more frequently prefer to use average total assets in calculating ROE and, practically, DuPont analysis is frequently performed using that definition. The following example illustrates the logic behind this equation.

EXAMPLE 23

ROA, Financial Policies, and the Dividend Growth Rate

Baggai Enterprises (a fictional company) has an ROA of 10%, retains 30% of earnings, and has an equity multiplier of 1.25. Mondale Enterprises also has an ROA of 10%, but it retains two-thirds of earnings and has an equity multiplier of 2.00.

1. What are the sustainable dividend growth rates for (A) Baggai Enterprises and (B) Mondale Enterprises?
2. Identify the drivers of the difference in the sustainable growth rates of Baggai Enterprises and Mondale Enterprises.

Solution to 1:

A. Baggai's dividend growth rate should be g = 0.30 × 10% × 1.25 = 3.75%.
B. Mondale's dividend growth rate should be g = (2/3) × 10% × 2.00 = 13.33%.

Solution to 2:

Because Mondale has the higher retention rate and higher financial leverage, its dividend growth rate is much higher.

If growth is being forecast for the next five years, an analyst should use the expectations of the four factors driving growth during this five-year period. If growth is being forecast into perpetuity, an analyst should use very long-term forecasts for these variables.

To illustrate the calculation and implications of the sustainable growth rate using the expression for ROE given by the DuPont formula, assume the growth rate is g = b × ROE = 0.60 (15%) = 9%. The ROE of 15% was based on a profit margin of 5%, an asset turnover of 2.0, and an equity multiplier of 1.5. Given fixed ratios of sales-to-assets and assets-to-equity, sales, assets, and debt will also be growing at 9%. Because dividends are fixed at 40% of income, dividends will grow at the same rate as income, or 9%. If the company increased dividends faster than 9%, this growth rate would not be sustainable using internally generated funds. Earning retentions would be reduced, and the company would be unable to finance the assets required for sales growth without external financing.

An analyst should be careful in projecting historical financial ratios into the future when using this analysis. Although a company may have grown at 25% a year for the last five years, this rate of growth is probably not sustainable indefinitely. Abnormally high ROEs, which may have driven that growth, are unlikely to persist indefinitely because of competitive forces and possibly other reasons, such as adverse changes in technology or demand. In the following example, an above-average terminal growth rate is plausibly forecasted because the company has positioned itself in businesses that may have relatively high margins on an ongoing basis.

EXAMPLE 24

Forecasting Growth with the PRAT Formula

An analyst is estimating a mature-phase growth rate for International Business Machines (NYSE: IBM) to use in her multistage dividend discount model. The company's ROE for 2018 was around 52%, and over the past 10 years, IBM's retention rate has averaged around 62%. Applying the formula for sustainable growth rate that was described previously [namely, g = (b in the mature phase) × (ROE in the mature phase] would yield an unrealistic long-term growth rate, particularly given the decline in the company's sales and earnings over the past several years. Further, IBM's annual investment in property, plant, and equipment has also declined—from \$3.7 billion in 2014 to \$3.4 billion in 2018.

The analyst therefore decides to estimate the company's growth rate using the DuPont decomposition and PRAT formula. A decomposition of IBM's ROE for the past 10 years is shown in Exhibit 12. In addition, the exhibit shows a benchmark based on the median values of ROE components for a group of firms with the same two-digit SIC code as IBM.

Exhibit 12 ROE Decomposition for IBM

Year	ROE	Profit Margin	Asset Turnover	Financial Leverage
2018	52.0%	11.0%	0.65	7.35
2017	32.7%	7.3%	0.63	7.12
2016	65.1%	14.9%	0.68	6.44
2015	92.5%	16.1%	0.74	7.75
2014	101.3%	13.0%	0.79	9.90
2013	72.3%	16.5%	0.79	5.54
2012	88.0%	15.9%	0.88	6.32
2011	78.7%	14.8%	0.92	5.78
2010	64.4%	14.9%	0.88	4.92
2009	59.3%	14.0%	0.88	4.82

Benchmark Average			
ROE	Profit Margin	Asset Turnover	Financial Leverage
13.5%	10.5%	0.62	2.07

IBM's ROE is much higher than the benchmark average, primarily because of much higher financial leverage. Its profit margin and asset turnover do not differ significantly from the benchmark average.

Suppose the analyst believes that IBM's profit margin and asset turnover will be roughly the same as the benchmark average. The analyst also believes that capital investment will continue to decline in IBM's maturity stage, and cash flow that was previously used for investment will be used to retire debt and pay dividends. The analyst forecasts a financial leverage ratio of 2.0, similar to the industry benchmark. The analyst also sees the dividend payout ratio continuing its recent rise and ultimately reaching a level of 50%.

Based on a profit margin of 10.5%, an asset turnover ratio of 0.62, and financial leverage of 2.0, a forecast of ROE in the maturity phase is (10.5%)(0.62)(2.0) = 13.0%. Therefore, based on this analysis, the estimate of the sustainable growth rate for IBM would be g = (0.50)(13.0%) = 6.5%.

FINANCIAL MODELS AND DIVIDENDS 10

m explain the use of spreadsheet modeling to forecast dividends and to value common shares

Analysts can also forecast dividends by building more-complex models of the company's total operating and financial environment. The company's ability to pay dividends in the future can be predicted using one of these models. The following example shows the dividends that a highly profitable and rapidly growing company can pay when its growth rates and profit margins decline because of increasing competition over time.

EXAMPLE 25

A Model for Forecasting Dividends Using More-Detailed Assumptions

An analyst is preparing a forecast of dividends for Hoshino Distributors (a fictional company) for the next five years. He uses a model with the following assumptions:

- Sales are $100 million in Year 1. They grow by 20% in Year 2, 15% in Year 3, and 10% in Years 4 and 5.
- Operating profits (earnings before interest and taxes, or EBIT) are 20% of sales in Years 1 and 2, 18% of sales in Year 3, and 16% of sales in Years 4 and 5.
- Interest expenses are 10% of total debt for the current year.
- The income tax rate is 40%.

- Hoshino pays out 20% of earnings in dividends in Years 1 and 2, 30% in Year 3, 40% in Year 4, and 50% in Year 5.
- Retained earnings are added to equity in the next year.
- Total assets are 80% of the current year's sales in all years.
- In Year 1, debt is $40 million and shareholders' equity is $40 million. Debt equals total assets minus shareholders' equity. Shareholders' equity will equal the previous year's shareholders' equity plus the addition to retained earnings from the previous year.
- Hoshino has 4 million shares outstanding.
- The required return on equity is 15%.
- The value of the company at the end of Year 5 is expected to be 10.0 times earnings.

The analyst wants to estimate the current value per share of Hoshino. Exhibit 13 adheres to the foregoing modeling assumptions. Total dividends and earnings are found at the bottom of the income statement.

Exhibit 13 Hoshino Distributors Pro Forma Financial Statements (in millions)

	Year 1	Year 2	Year 3	Year 4	Year 5
Income statement					
Sales	$100.00	$120.00	$138.00	$151.80	$166.98
EBIT	20.00	24.00	24.84	24.29	26.72
Interest	4.00	4.83	5.35	5.64	6.18
EBT	16.00	19.17	19.49	18.65	20.54
Taxes	6.40	7.67	7.80	7.46	8.22
Net income	9.60	11.50	11.69	11.19	12.32
Dividends	1.92	2.30	3.51	4.48	6.16
Balance sheet					
Total assets	$80.00	$96.00	$110.40	$121.44	$133.58
Total debt	40.00	48.32	53.52	56.38	61.81
Equity	40.00	47.68	56.88	65.06	71.77

Dividing the total dividends by the number of outstanding shares gives the dividend per share for each year shown in the following table. The present value of each dividend, discounted at 15%, is also shown.

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
DPS	$0.480	$0.575	$0.877	$1.120	$1.540	$4.59
PV	0.417	0.435	0.577	0.640	0.766	2.84

The earnings per share in Year 5 are $12.32 million divided by 4 million shares, or $3.08 per share. Given a P/E of 10, the market price in Year 5 is predicted to be $30.80. Discounted at 15%, the required return on equity by assumption, the present value of this price is $15.31. Adding the present values of the five dividends, which sum to $2.84, gives a total stock value today of $18.15 per share.

SUMMARY

We have provided an overview of DCF models of valuation, discussed the estimation of a stock's required rate of return, and presented in detail the dividend discount model.

- In DCF models, the value of any asset is the present value of its (expected) future cash flows

$$V_0 = \sum_{t=1}^{n} \frac{\mathrm{CF}_t}{(1+r)^t}$$

  where V_0 is the value of the asset as of t = 0 (today), CF_t is the (expected) cash flow at time t, and r is the discount rate or required rate of return. For infinitely lived assets such as common stocks, n runs to infinity.

- Several alternative streams of expected cash flows can be used to value equities, including dividends, free cash flow, and residual income. A discounted dividend approach is most suitable for dividend-paying stocks in which the company has a discernible dividend policy that has an understandable relationship to the company's profitability and the investor has a non-control (minority ownership) perspective.
- The free cash flow approach (FCFF or FCFE) might be appropriate when the company does not pay dividends, dividends differ substantially from FCFE, free cash flows align with profitability, or the investor takes a control (majority ownership) perspective.
- The residual income approach can be useful when the company does not pay dividends (as an alternative to a FCF approach) or free cash flow is negative.
- The DDM with a single holding period gives stock value as

$$V_0 = \frac{D_1}{(1+r)^1} + \frac{P_1}{(1+r)^1} = \frac{D_1 + P_1}{(1+r)^1}$$

  where D_1 is the expected dividend at Time 1 and V_0 is the stock's (expected) value at Time 0. Assuming that V_0 is equal to today's market price, P_0, the expected holding-period return is

$$r = \frac{D_1 + P_1}{P_0} - 1 = \frac{D_1}{P_0} + \frac{P_1 - P_0}{P_0}$$

- The expression for the DDM for any given finite holding period n and the general expression for the DDM are, respectively,

$$V_0 = \sum_{t=1}^{n} \frac{D_t}{(1+r)^t} + \frac{P_n}{(1+r)^n} \text{ and } V_0 = \sum_{t=1}^{\infty} \frac{D_t}{(1+r)^t}$$

- There are two main approaches to the problem of forecasting dividends. First, an analyst can assign the entire stream of expected future dividends to one of several stylized growth patterns. Second, an analyst can forecast a finite number of dividends individually up to a terminal point and value the remaining dividends either by assigning them to a stylized growth pattern or by forecasting share price as of the terminal point of the dividend forecasts.

- The Gordon growth model assumes that dividends grow at a constant rate g forever, so that $D_t = D_{t-1}(1 + g)$. The dividend stream in the Gordon growth model has a value of

$$V_0 = \frac{D_0(1+g)}{r-g}, \text{ or } V_0 = \frac{D_1}{r-g} \text{ where } r > g$$

- The value of non-callable fixed-rate perpetual preferred stock is $V_0 = D/r$, where D is the stock's (constant) annual dividend.
- Assuming that price equals value, the Gordon growth model estimate of a stock's expected rate of return is

$$r = \frac{D_0(1+g)}{P_0} + g = \frac{D_1}{P_0} + g$$

- Given an estimate of the next-period dividend and the stock's required rate of return, the Gordon growth model can be used to estimate the dividend growth rate implied by the current market price (making a constant growth rate assumption).
- The present value of growth opportunities is the part of a stock's total value, V_0, that comes from profitable future growth opportunities in contrast to the value associated with assets already in place. The relationship is $V_0 = E_1/r + \text{PVGO}$, where E_1/r is defined as the no-growth value per share.
- The leading price-to-earnings ratio (P_0/E_1) and the trailing price-to-earnings ratio (P_0/E_0) can be expressed in terms of the Gordon growth model as, respectively,

$$\frac{P_0}{E_1} = \frac{D_1/E_1}{r-g} = \frac{1-b}{r-g} \text{ and } \frac{P_0}{E_0} = \frac{D_0(1+g)/E_0}{r-g} = \frac{(1-b)(1+g)}{r-g}$$

  The foregoing expressions give a stock's justified price-to-earnings ratio based on forecasts of fundamentals (given that the Gordon growth model is appropriate).
- The Gordon growth model may be useful for valuing broad-based equity indexes and the stock of businesses with earnings that are expected to grow at a stable rate comparable to or lower than the economy's nominal growth rate.
- Gordon growth model values are very sensitive to the assumed growth rate and required rate of return.
- For many companies, growth falls into phases. In the growth phase, a company enjoys an abnormally high growth rate in earnings per share, called supernormal growth. In the transition phase, earnings growth slows. In the mature phase, the company reaches an equilibrium in which such factors as earnings growth and the return on equity stabilize at levels that can be sustained long term. Analysts often apply multistage DCF models to value the stock of a company with multistage growth prospects.
- The two-stage dividend discount model assumes different growth rates in Stage 1 and Stage 2:

$$V_0 = \sum_{t=1}^{n} \frac{D_0(1+g_S)^t}{(1+r)^t} + \frac{D_0(1+g_S)^n(1+g_L)}{(1+r)^n(r-g_L)}$$

  where g_S is the expected dividend growth rate in the first period and g_L is the expected growth rate in the second period.

- The terminal stock value, V_n, is sometimes found with the Gordon growth model or with some other method, such as applying a P/E multiplier to forecasted EPS as of the terminal date.
- The H-model assumes that the dividend growth rate declines linearly from a high supernormal rate to the normal growth rate during Stage 1 and then grows at a constant normal growth rate thereafter:

$$V_0 = \frac{D_0(1 + g_L)}{r - g_L} + \frac{D_0 H(g_S - g_L)}{r - g_L} = \frac{D_0(1 + g_L) + D_0 H(g_S - g_L)}{r - g_L}$$

- There are two basic three-stage models. In one version, the growth rate in the middle stage is constant. In the second version, the growth rate declines linearly in Stage 2 and becomes constant and normal in Stage 3.
- In addition to valuing equities, the IRR of a DDM, assuming assets are correctly priced in the marketplace, has been used to estimate required returns. For simpler models (such as the one-period model, the Gordon growth model, and the H-model), well-known formulas may be used to calculate these rates of return. For many dividend streams, however, the rate of return must be found by trial and error, producing a discount rate that equates the present value of the forecasted dividend stream to the current market price.
- Multistage DDM models can accommodate a wide variety of patterns of expected dividends. Even though such models may use stylized assumptions about growth, they can provide useful approximations.
- Dividend growth rates can be obtained from analyst forecasts, statistical forecasting models, or company fundamentals. The sustainable growth rate depends on the ROE and the earnings retention rate, b: $g = b \times \text{ROE}$. This expression can be expanded further, using the DuPont formula, as

$$g = \frac{\text{Net income} - \text{Dividends}}{\text{Net income}} \times \frac{\text{Net income}}{\text{Sales}} \times \frac{\text{Sales}}{\text{Total assets}} \times \frac{\text{Total assets}}{\text{Shareholders' equity}}$$

REFERENCES

ap Gwilym, Owain, James Seaton, Karina Suddason, and Stephen Thomas. 2006. "International Evidence on the Payout Ratio, Earnings, Dividends, and Returns." *Financial Analysts Journal* 62 (1): 36–53.

Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal* 59 (1): 70–87.

Brockman, Paul, and Emre Unlu. 2009. "Dividend Policy, Creditor Rights, and the Agency Costs of Debt." *Journal of Financial Economics* 92 (2): 276–99.

Denis, David J., and Igor Osobov. 2008. "Why Do Firms Pay Dividends? International Evidence on the Determinants of Dividend Policy." *Journal of Financial Economics* 89 (1): 62–82.

Fama, Eugene F., and Kenneth R. French. 2001. "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?" *Journal of Financial Economics* 60 (1): 3–43.

Fuller, Russell J., and Chi-Cheng Hsia. 1984. "A Simplified Common Stock Valuation Model." *Financial Analysts Journal* 40 (5): 49–56.

Gordon, Myron J. 1962. *The Investment, Financing, and Valuation of the Corporation*. Homeward, IL: Richard D. Irwin.

Gordon, Myron J., and Eli Shapiro. 1956. "Capital Equipment Analysis: The Required Rate of Profit." *Management Science* 3 (1): 102–10.

Grullon, Gustavo, Bradley Paye, Shane Underwood, and James P. Weston. 2011. "Has the Propensity to Pay Out Declined?" *Journal of Financial and Quantitative Analysis* 46 (1): 1–24.

He, Wen, Lilian Ng, Nataliya Zaiats, and Bohui Zhang. 2017. "Dividend Policy and Earnings Management across Countries." *Journal of Corporate Finance* 42:267–86.

Julio, Brandon, and David L. Ikenberry. 2004. "Reappearing Dividends." *Journal of Applied Corporate Finance* 16 (4): 89–100.

Skinner, Douglas J. 2008. "The Evolving Relation between Earnings, Dividends, and Stock Repurchases." *Journal of Financial Economics* 87 (3): 582–609.

Stowe, J., J Pinto, and T. Robinson.Equity Valuation: A Survey of Professional Practice", October 2018, Review of Financial Economics.

von Eije, J. Henk, and William L. Megginson. 2008. "Dividends and Share Repurchases in the European Union." *Journal of Financial Economics* 89:347–74.

Wanger, Ralph. 2007. "More Dividends, Please." *CFA Magazine* 18 (2): 8–11.

Williams, John Burr. 1938. *The Theory of Investment Value*. Cambridge, MA: Harvard University Press.

Zhou, Ping, and William Ruland. 2006. "Dividend Payout and Future Earnings Growth." *Financial Analysts Journal* 62 (3): 58–69.

PRACTICE PROBLEMS

1 Amy Tanner is an analyst for a US pension fund. Her supervisor has asked her to value the stocks of two companies: Bright Lights Electronics ("BL Electronics") and Fast Clean Autos ("FC Autos"). Tanner wants to evaluate the DDM's appropriateness for valuing the two companies and has compiled the following data for 2012 through 2019.

	BL Electronics			FC Autos		
Year	EPS ($)	DPS ($)	Payout Ratio	EPS ($)	DPS ($)	Payout Ratio
2019	2.17	1.15	0.53	−68.45	1.00	−0.01
2018	1.99	1.03	0.52	−3.50	1.00	−0.29
2017	1.76	0.91	0.52	−18.50	2.00	−0.11
2016	1.61	0.82	0.51	4.94	2.00	0.40
2015	1.55	0.77	0.50	5.03	2.00	0.40
2014	1.51	0.73	0.48	3.35	2.00	0.60
2013	1.41	0.66	0.47	1.77	2.00	1.13
2012	1.27	0.57	0.45	6.68	2.00	0.30

For each of the stocks, explain whether the DDM is appropriate for valuing the stock.

2 Vincent Nguyen, an analyst, is examining the stock of Green Mountain Airways ("GM Airways") as of the beginning of 2018. He notices that the consensus forecast by analysts is that the stock will pay a £4 dividend per share in 2019 (based on 21 analysts) and a £5 dividend in 2020 (based on 10 analysts). Nguyen expects the stock price at the end of 2020 to be £250. He has estimated that the required rate of return on the stock is 11%. Assume all dividends are paid at the end of the year.

A Using the DDM, estimate the value of GM Airways stock at the end of 2019.

B Using the DDM, estimate the value of GM Airways stock at the end of 2018.

3 Justin Owens is an analyst for an equity mutual fund that invests in British stocks. At the beginning of 2018, Owens is examining domestic stocks for possible inclusion in the fund. One of the stocks that he is analyzing is Sage Broadcasting Group ("Sage Broadcasting"). The stock has paid dividends per share of £9, £12.20, and £15.50 at the end of 2015, 2016, and 2017, respectively. The consensus forecast by analysts is that the stock will pay a dividend per share of £18.66 at the end of 2018 (based on 19 analysts) and £20.20 at the end of 2019 (based on 17 analysts). Owens has estimated that the required rate of return on the stock is 11%.

A Compare the compound annual growth rate in dividends from 2015 to 2017, inclusive (i.e., from a beginning level of £9 to an ending level of £15.50), with the consensus predicted compound annual growth rate in dividends from 2017 to 2019, inclusive.

B Owens believes that Sage Broadcasting has matured such that the dividend growth rate will be constant going forward at half the consensus compound annual growth rate from 2017 to 2019, inclusive, computed in Part A. Using

© 2017 CFA Institute. All rights reserved.

the growth rate forecast of Owens as the constant growth rate from 2017 onwards, estimate the value of the stock as of the end of 2017 given an 11% required rate of return on equity.

C State the relationship between estimated value and *r* and estimated value and *g*.

4 During the period 1989–2018, the average annual growth rate in earnings of S&P 500 Index companies was 14.6%, and the average annual growth rate in dividends paid was 6.2%.

Assume the following:

- Dividends will continue to grow at the 1989–2018 rate.
- The required return on the index is 8%.
- Companies in the S&P 500 Index collectively paid $456.3 billion in dividends in 2018.

Estimate the aggregate value of the S&P 500 Index component companies at the beginning of 2019 using the Gordon growth model.

5 Great Plains Energy is a public utility holding company that listed its 4.5% cumulative perpetual preferred stock series E on the NYSE Euronext in March 1952. The par value of the preferred stock is $100. If the required rate of return on this stock is 5.6%, estimate the value of the stock.

6 German Resources is involved in coal mining. The company is currently profitable and is expected to pay a dividend of €4 per share next year. The company has suspended exploration, however, and because its current mature operations exhaust the existing mines, you expect that the dividends paid by the company will decline forever at an 8% rate. The required return on German Resource's stock is 11%. Using the DDM, estimate the value of the stock.

7 Maspeth Robotics shares are currently selling for €24 and have paid a dividend of €1 per share for the most recent year. The following additional information is given:

- The risk-free rate is 4%;
- The shares have an estimated beta of 1.2; and
- The equity risk premium is estimated at 5%.

Based on the foregoing information, determine the constant dividend growth rate that would be required to justify the market price of €24.

8 You believe the Gordon (constant) growth model is appropriate to value the stock of Reliable Electric Corp. The company had an EPS of $2 in 2019. In the absence of additional investment, EPS will remain at $2. The retention ratio is 0.60. The company is expected to earn an ROE of 14% on its investments, and the required rate of return is 11%. Assume that all dividends are paid at the end of the year.

A Calculate the company's sustainable growth rate.

B Estimate the value of the company's stock at the beginning of 2020.

C Calculate the present value of growth opportunities.

D Determine the fraction of the company's value that comes from its growth opportunities.

9 Stellar Baking Company in Australia has a trailing P/E of 14. Analysts predict that Stellar's dividends will continue to grow at its recent rate of 4.5% per year into the indefinite future. Given a current dividend and EPS of A$0.7 per share and A$2.00 per share, respectively, and a required rate of return on equity of 8%, determine whether Stellar Baking Company is undervalued, fairly valued, or overvalued. Justify your answer.

10 Mohan Gupta is the portfolio manager of an India-based equity fund. He is analyzing the value of FFR&A Ltd. Gupta has concluded that the DDM is appropriate to value this company.

During the last five years (fiscal year ending 31 March 2014 to fiscal year ending 31 March 2018), the company has paid dividends per share of Rs. 5.50, 6.50, 7.00, 8.00, and 9.00, respectively. These dividends suggest an average annual growth rate in DPS of just above 13%. Gupta has decided to use a three-stage DDM with a linearly declining growth rate in Stage 2. He considers FFR&A to be an average growth company and estimates Stage 1 (the growth stage) to be 6 years and Stage 2 (the transition stage) to be 10 years. Gupta estimates the growth rate to be 14% in Stage 1 and 10% in Stage 3. He has estimated the required return on equity for the company to be 16%. Estimate the current value of the stock.

11 You are analyzing the stock of Ansell Limited (ASX: ANN), a healthcare company, as of early 2019. The stock price is A$24.54. The company's dividend per share for the fiscal year ending 30 June 2008 was A$0.58. You expect the dividend to increase by 5% for the next three years and then increase by 4% per year forever. You estimate the required return on equity of Ansell Limited to be 7%.

A Estimate the value of ANN using a two-stage dividend discount model.

B Judge whether ANN is undervalued, fairly valued, or overvalued.

12 Sime Natural Cosmetics Ltd. has a dividend yield of 2% based on the current dividend and a mature-phase dividend growth rate of 5% a year. The current dividend growth rate is 10% a year, but the growth rate is expected to decline linearly to its mature-phase value during the next six years.

A If Sime Natural Cosmetics is fairly priced in the marketplace, what is the expected rate of return on its shares?

B If Sime were in its mature growth phase right now, would its expected return be higher or lower, holding all other facts constant?

13 Kazuo Uto is analyzing the stock of a Japanese company and has concluded that a multistage DDM is appropriate to value the stock. The company's ROE has declined from 16.7% in the fiscal year ending in 2015 to 12.7% in the fiscal year ending in 2019. The dividend payout ratio has increased from 11.5% in 2015 to 22.3% in 2019. Uto estimates the company will reach a mature stage in four years and that in the mature phase, the company's ROE will be 11%, approximately equal to estimated required return on equity. He has also estimated that the payout ratio in the mature phase will be 40%, significantly greater than its payout ratio in 2018 but less than the average payout of about 50% for a benchmark group of Japanese companies.

A Calculate the sustainable growth rate for the company in the mature phase.

B With reference to the formula for the sustainable growth rate, a colleague of Uto asserts that the greater the earnings retention ratio, the greater the sustainable growth rate because g is a positive function of b. The colleague argues that Brother should decrease payout ratio. Explain the flaw in that argument.

14 An analyst wants to estimate a company's sustainable growth rate using the PRAT model. For this purpose, she has compiled the data in the following table. Assets and equity values are for the end of the year; the analyst uses averages of beginning and ending balance sheet values in computing ratios based on total assets and shareholders' equity. For example, average total assets for 2017 would be computed as (148,786 + 132,628)/2 = $140,707. *Note*: All numbers except for EPS and DPS are in $ millions.

Item	2017	2016	2015	2014
Net income	$18,688	$17,138	$14,099	$13,328
Sales	214,091	204,892	193,641	150,865
Total assets	148,786	132,628	125,833	93,208
Shareholders' equity	77,088	68,935	62,676	45,230
EPS	8.77	7.80	6.54	6.28
DPS	2.26	2.01	1.75	1.53

A Compute the average value of each PRAT component during 2015–2017.

B Using the overall mean value of the average component values calculated in Part A, estimate the company's sustainable growth rate.

C Judge whether the company has reached a mature growth stage.

15 Casey Hyunh is trying to value the stock of Resources Limited. To easily see how a change in one or more of her assumptions affects the estimated value of the stock, she is using a spreadsheet model. The model has projections for the next four years based on the following assumptions.

- Sales will be $300 million in Year 1.
- Sales will grow at 15% in Years 2 and 3 and 10% in Year 4.
- Operating profits (EBIT) will be 17% of sales in each year.
- Interest expense will be $10 million per year.
- Income tax rate is 30%.
- Earnings retention ratio will stay at 0.60.
- The per-share dividend growth rate will be constant from Year 4 forward, and this final growth rate will be 200 bps less than the growth rate from Year 3 to Year 4.

The company has 10 million shares outstanding. Hyunh has estimated the required return on Resources' stock to be 13%.

A Estimate the value of the stock at the end of Year 4 based on the foregoing assumptions.

B Estimate the current value of the stock using the foregoing assumptions.

C Hyunh is wondering how a change in the projected sales growth rate would affect the estimated value. Estimate the current value of the stock if the sales growth rate in Year 3 is 10% instead of 15%.

The following information relates to Questions 16–21

Jacob Daniel is the chief investment officer at a US pension fund sponsor, and Steven Rae is an analyst for the pension fund who follows consumer/non-cyclical stocks. At the beginning of 20X9, Daniel asks Rae to value the equity of Tasty Foods Company for its possible inclusion in the list of approved investments. Tasty Foods Company is involved in the production of frozen foods that are sold under its own brand name to retailers.

Rae is considering whether a dividend discount model would be appropriate for valuing Tasty Foods. He has compiled the information in the following table for the company's EPS and DPS during the last five years. The quarterly dividends paid by the company have been added to arrive at the annual dividends. Rae has also computed the dividend payout ratio for each year as DPS/EPS and the growth rates in EPS and DPS.

Year	EPS ($)	DPS ($)	Payout Ratio	Growth in EPS (%)	Growth in DPS (%)
20X8	2.12	0.59	0.278	2.9	3.5
20X7	2.06	0.57	0.277	2.5	5.6
20X6	2.01	0.54	0.269	6.3	5.9
20X5	1.89	0.51	0.270	6.2	6.3
20X4	1.78	0.48	0.270		

Rae notes that the company's EPS has been increasing at an average rate of 4.48% per year. The dividend payout ratio has remained fairly stable, and dividends have increased at an average rate of 5.30%. In view of a history of dividend payments by the company and the understandable relationship dividend policy bears to the company's earnings, Rae concludes that the DDM is appropriate to value the equity of Tasty Foods. Further, he expects the company's moderate growth rate to persist and decides to use the Gordon growth model.

Rae uses the CAPM to compute the return on equity. He uses the annual yield of 4% on the 10-year Treasury bond as the risk-free return. He estimates the expected US equity risk premium, with the S&P 500 Index used as a proxy for the market, to be 6.5% per year. The estimated beta of Tasty Foods against the S&P 500 Index is 1.10. Accordingly, Rae's estimate for the required return on equity for Tasty Foods is 0.04 + 1.10(0.065) = 0.1115, or 11.15%.

Using the past growth rate in dividends of 5.30% as his estimate of the future growth rate in dividends, Rae computes the value of Tasty Foods stock. He shows his analysis to Alex Renteria, his colleague at the pension fund who specializes in the frozen foods industry. Renteria concurs with the valuation approach used by Rae but disagrees with the future growth rate he used. Renteria believes that the stock's current price of $8.42 is the fair value of the stock.

16 Which of the following is *closest* to Rae's estimate of the stock's value?

- **A** $10.08.
- **B** $10.54.
- **C** $10.62.

17 What is the stock's justified trailing P/E based on the stock's value estimated by Rae?

- **A** 5.01.
- **B** 5.24.
- **C** 5.27.

18 Rae considers a security trading within a band of ±10% of his estimate of intrinsic value to be within a "fair value range." By that criterion, the stock of Tasty Foods is:

A undervalued.

B fairly valued.

C overvalued.

19 The beta of Tasty Foods stock of 1.10 that Rae used in computing the required return on equity was based on monthly returns for the last 10 years. If Rae uses daily returns for the last five years, the beta estimate is 1.25. If a beta of 1.25 is used, what would be Rae's estimate of the value of Tasty Foods stock?

A \$8.64.

B \$9.10.

C \$20.13.

20 Renteria has suggested that the market price of Tasty Foods stock is its fair value. What is the implied growth rate of dividends given the stock's market price? Use the required return on equity based on a beta of 1.10.

A 3.87%.

B 5.30%.

C 12.1%.

21 If Renteria is correct that the current price of Tasty Foods stock is its fair value, what is the expected capital gains yield on the stock?

A 3.87%.

B 4.25%.

C 5.30%.

The following information relates to Questions 22–27

Brian Dobson, an analyst at UK-based globally diversified equity mutual fund, has been assigned the task of estimating a fair value of the common stock of Charmed Energy. Dobson is aware of several approaches that could be used for this purpose. After carefully considering the characteristics of the company and its competitors, he believes Charmed will have extraordinary growth for the next few years and normal growth thereafter. So, he has concluded that a two-stage DDM is the most appropriate for valuing the stock.

Charmed pays semi-annual dividends. The total dividends during 2016, 2017, and 2018 have been C\$0.114, C\$0.15, and C\$0.175, respectively. These imply a growth rate of 32% in 2017 and 17% in 2018. Dobson believes that the growth rate will be 14% in the next year. He has estimated that the first stage will include the next eight years.

Dobson is using the CAPM to estimate the required return on equity for Charmed. He has estimated that the company's beta, as measured against the S&P/TSX Composite Index (formerly TSE 300 Composite Index), is 0.84. The Canadian risk-free rate, as measured by the annual yield on the 10-year government bond, is 4.1%. The equity risk premium for the Canadian market is estimated at 5.5%. Based on these data, Dobson has estimated that the required return on Charmed Energy's stock is 0.041 + 0.84(0.055) = 0.0872, or 8.72%. Dobson is doing the analysis in January 2019, and the stock price at that time is C$17.

Dobson realizes that even within the two-stage DDM, there could be some variations in the approach. He would like to explore how these variations affect the stock's valuation. Specifically, he wants to estimate the value of the stock for each of the following approaches separately.

I. The dividend growth rate will be 14% throughout the first stage of eight years. The dividend growth rate thereafter will be 7%.

II. Instead of using the estimated stable growth rate of 7% in the second stage, Dobson wants to use his estimate that eight years later, Charmed Energy's stock will be worth 17 times its earnings per share (trailing P/E of 17). He expects that the earnings retention ratio at that time will be 0.70.

III. In contrast to the first approach, in which the growth rate declines abruptly from 14% in the eighth year to 7% in the ninth, the growth rate would decline linearly from 14% in the first year to 7% in the ninth.

22 What is the terminal value of the stock based on the first approach?

- **A** C$17.65.
- **B** C$31.06.
- **C** C$33.09.

23 In the first approach, what proportion of the stock's total value is represented by the value of second stage?

- **A** 0.10.
- **B** 0.52.
- **C** 0.90.

24 What is the stock's terminal value based on the second approach (earnings multiple)?

- **A** C$12.12.
- **B** C$28.29.
- **C** C$33.09.

25 What is the stock's current value based on the second approach?

- **A** C$16.24.
- **B** C$17.65.
- **C** C$28.29.

26 Based on the third approach (the H-model), the stock is:

- **A** undervalued.
- **B** fairly valued.
- **C** overvalued.

27 Dobson is wondering what the consequences would be if the duration of the first stage was assumed to be 11 years instead of 8, with all the other assumptions and estimates remaining the same. Considering this change, which of the following is true?

- **A** In the second approach, the proportion of the total value of the stock represented by the second stage would not change.

B The total value estimated using the third approach would increase.

C Using this new assumption and the first approach will lead Dobson to conclude that the stock is overvalued.

The following information relates to Questions 28–36

Gianna Peters is an investment analyst who focuses on dividend-paying stocks. Peters uses a DCF approach to stock selection. She is meeting with her staff to evaluate portfolio holdings based on a bottom-up screening of stocks listed in the United States and Canada. Peters and her staff begin by reviewing the characteristics of the following portfolio candidates.

Company ABC

A Canadian company in the consumer staples sector with a required rate of return of 7.35%. Recent media reports suggest that ABC might be a takeover candidate. Peters and her team estimate that if the incumbent Canadian prime minister's party retains its power, the company's current annual dividend of C$0.65 per share will grow 12% a year for the next four years and then stabilize at a 3.5% growth rate a year indefinitely. If a new government takes office in Canada, however, then the team estimates that ABC will likely not experience the elevated 12% short-run growth because of new regulatory and tax changes, and instead it will grow by 3.5% indefinitely.

Company XYZ

A mid-sized US company in the utilities sector with a required rate of return of 10%. Peters and her team believe that because of a recent restructuring, the company is unlikely to pay dividends for the next three years. The team expects XYZ to pay an annual dividend of US$1.72 per share beginning four years from now, however. Thereafter, the dividend is expected to grow indefinitely at 4% even though the current price implies a growth rate of 6% during this same period.

Company JZY

A large US company in the telecom sector with a required rate of return of 8%. The stock is currently trading at US$32.76 per share with an implied earnings growth rate of 5.3%. Peters believes that because JZY is mature and has a stable capital structure, the company will grow at its sustainable growth rate. Over the past 10 years, the company's return on equity (ROE) has averaged 8.17% and its payout ratio has averaged 40%. Recently, the company paid an annual dividend of US$0.84 per share.

Peters asks a newly hired analyst, Kurt Thomas, to comment on the evaluation approach for these three stocks. Thomas makes the following statements:

1 A free cash flow valuation model would not be appropriate to evaluate Company ABC if the firm becomes a takeover candidate.

2 A dividend discount model cannot be applied to Company XYZ if dividends are suspended for a few years.

3 A dividend discount model is suitable for evaluating the stock of Company JZY because of the historically consistent payout ratio.

Peters then asks the team to examine the growth opportunities of three Canadian stocks currently held in the portfolio. These stocks are listed in Exhibit 1. Peters believes that the stocks are fairly valued.

Exhibit 1 Selected Stock Characteristics

Stock	Required Rate of Return	Next Year's Forecasted EPS (C$)	Current Price per Share (C$)
ABTD	10.5%	7.30	80.00
BKKQ	8.0%	2.12	39.00
CPMN	12.0%	1.90	27.39

28 Which of the following statements made by Thomas is *correct*?

A Statement 1

B Statement 2

C Statement 3

29 Assuming the incumbent government retains office in Canada, Peters and her team estimate that the current value of Company ABC stock would be *closest* to:

A C$22.18.

B C$23.60.

C C$25.30.

30 Assuming a new government takes office in Canada, Peters and her team estimate that the current intrinsic value of Company ABC would be *closest* to:

A C$9.15.

B C$16.88.

C C$17.47.

31 Assume that a new government takes office in Canada. If Peters and her team use the Gordon growth model and assume that Company ABC stock is fairly valued, then which of the following would *most likely* be true?

A The total return of ABC stock will be 10.85%.

B The dividend yield of ABC stock will be 3.85%.

C The stock price of ABC will grow at 7.35% annually.

32 If the team uses the dividend discount model, the current intrinsic value of Company XYZ stock would be *closest to*:

A US$19.58.

B US$20.36.

C US$21.54.

33 The dividend growth rate implied in the stock price of Company XYZ suggests that XYZ's stock price is *most likely*:

A undervalued.

B fairly valued.

C overvalued.

34 Based on the relationship between the implied growth rate and the sustainable growth rate, Peters' team should conclude that Company JZY's stock price is *most likely*:

A undervalued.

B fairly valued.

C overvalued.

35 Based on Exhibit 1, the stock with the largest present value of growth opportunities (PVGO) is:

A ABTD.

B BKKQ.

C CPMN.

36 Based on Exhibit 1, the growth component of the leading P/E is largest for:

A ABTD.

B BKKQ.

C CPMN.

The following information relates to Questions 37–46

June Withers is analyzing four stocks in the processed food industry as of 31 December 2019. All stocks pay a dividend at the end of each year.

Ukon Corporation

Withers estimates a required rate of return for Ukon Corporation of 8% and notes that the dividend for 2019 was EUR 2.315 per share. Her first valuation approach is a basic two-stage DDM, with dividends growing at a rate of 5% from 2020 through 2023, after which time dividends will grow at a sustainable rate of 3%. Her second valuation approach is the H-model, assuming that dividend growth of 5% in 2020 declines linearly during the years 2021 through 2023 to the 3% growth rate after 2023. Exhibit 1 summarizes Withers's dividend growth assumptions.

Exhibit 1 Ukon Corporation Dividend Growth Assumptions, by Model

Model	Period	Rate
Two-stage DDM	2020 through 2023	5%
	Beginning 2024	3%
H-model	2020	5%
	2021 through 2023	Declining linearly to 3.5%
	Beginning 2024	3%

Venus Company

Withers has assembled the data on Venus Company in Exhibit 2. After analyzing competitive pressures and financial conditions in the industry, she predicts that Venus Company will lose market share because of new entrants but will stabilize within a few years. The required rate of return for Venus Company is 8%. Beginning with a per-share dividend of USD3.15 in 2019, she develops two scenarios regarding Venus Company's dividend growth. The scenarios, shown in Exhibit 2, are summarized as follows:

- In Scenario 1, the growth rate will fall in a linear manner over the years 2020 through 2023 from 8% to 4%. Using the H-model, Withers calculates a value of USD58.79 per share of Venus Company stock.
- In Scenario 2, the growth rate falls from 8% in 2019 to 6% in 2020 and 2021, to 5% in 2022 and 2023, and then to a sustainable rate of 3% for 2024 and beyond.

Exhibit 2 Venus Company Dividend Growth Scenarios

Scenario	Period	Rate
Scenario 1	2020 through 2023	Declining linearly to 4%
	Beginning 2024	Remaining stable at 4%
Scenario 2	2020 and 2021	6%
	2022 and 2023	5%
	Beginning 2024	Remaining stable at 3%

Wakuni Corporation

Withers evaluates Wakuni Corporation and uses recent financial data from Exhibit 3 to calculate a sustainable growth based on the DuPont model. In addition to this estimate, she performs a sensitivity analysis on the sustainable growth rate whereby the dividend payout ranges from 0% to 10% and the return on equity ranges from 8% to 12%.

Exhibit 3 Selected Data for Wakuni Corporation (JPY billions)

Net income	43,923
Sales	423,474
Total assets, average during year	486,203
Shareholders' equity, beginning of year	397,925
Dividends paid	1,518

Xavier Corporation

In her analysis of the stock of Xavier Corporation, Withers observes that it has a dividend of USD2 per share and a stock price of USD52. Two analyst interns have offered estimates of the company's required rate of return and dividend growth rate, as shown in Exhibit 4.

Exhibit 4 Xavier Corporation Required Rate of Return and Dividend Growth Rates (Estimates)

	Intern 1	Intern 2
Required rate of return	8.3%	7.8%
Growth rate, first four years	5.0%	4.8%
Growth rate, beyond first four years	3.6%	4.0%

37 Based on Exhibit 1, when Withers applies the first valuation approach to Ukon Corporation, the estimated value of the stock at the end of the first stage represents the:

A present value of the dividends beyond year 2023.

B present value of the dividends for years 2020 through 2023.

C sum of the present value of the dividends for 2020 through 2023 and the present value of dividends beyond year 2023.

38 Using her first valuation approach and Exhibit 1, Withers's forecast of the per share stock value of Ukon Corporation at the end of 2019 should be *closest to*:

A EUR48.

B EUR50.

C EUR51.

39 Using Withers's assumptions for the H-model and the basic two-stage dividend discount model, the forecasted Ukon stock price at the end of the year 2023 for the H-model should be:

A lower than the basic two-stage model.

B the same as the basic two-stage model.

C higher than the basic two-stage model.

40 Under her Scenario 1 and based on Exhibit 2, the required rate of return that Withers used for Venus Company stock valuation is *closest* to:

A 8.0%.

B 9.6%.

C 10.0%.

41 Under Scenario 2 and based on Exhibit 2, Withers estimates that the value of the Venus Company stock to be *closest* to:

A USD69.73.

B USD71.03.

C USD72.98.

42 Using the data in Exhibit 3, Withers can estimate the sustainable growth of the Wakuni Corporation as being *closest* to:

A 10.66%.

B 11.04%.

C 14.05%.

43 Withers's sensitivity analysis of Wakuni Corporation should produce a range of sustainable growth estimates between:

A 0.0% and 1.2%.

B 7.2% and 12.0%.

C 8.0% and 13.3%.

44 Based on Exhibit 4 and Intern 1's analysis, Xavier Corporation's sustainable dividend payout ratio is *closest* to:

A 43.4%.

B 44.6%.

C 56.6%.

45 Based on Exhibit 4, Intern 2 should conclude that the Xavier stock is:

A underpriced.

B fairly priced.

C overpriced.

46 Based on Exhibit 4 and Intern 1's estimate of the required rate of return and the dividend growth rate for the first four years, the growth rate beyond the first four years consistent with the current price of USD52 is *closest* to:

A 3.80%.

B 4.17%.

C 4.23%.

The following information relates to Questions 47–53

BJL Financial provides clients with professional investment management services that are tailored to the specific needs of each client. The firm's portfolio manager, Angelique Kwaza, has called a meeting with the senior analyst, Samira Khan, to discuss the quarterly rebalancing of three client portfolios. The valuation model used in the analyses is the discounted dividend model.

- Client 1 has a portfolio with significant exposure to dividend-paying stocks.
- Client 2 is interested in including preferred stock in the portfolio.
- Client 3 has a growth-oriented equity-only portfolio.

Khan has identified two utilities (ABC and XYZ) for possible inclusion in Client 1's portfolio, as shown in Exhibit 1. She uses a discount rate of 7% for both common stocks.

Exhibit 1 Candidate Stocks for Client 1

Stock	Company Description
ABC	■ ABC is a publicly traded utility with an expected constant growth rate for earnings and dividends of 3.5%. ■ The most recent year's dividend payout is 70%. The expected dividend payout in future years is 60%. ■ The common stock price is $14.49 per share.
XYZ	■ XYZ is a publicly traded utility with several unregulated business subsidiaries. ■ The company generates 3% growth in dividends and has an annual dividend payout of 80%. No changes in dividend growth or payout are expected. ■ The common stock price is $10 per share. ■ The current year earnings are $0.45 per share, and next year's earnings are expected to be $0.50 per share.

Kwaza asks Khan to investigate the most appropriate models for valuing utility companies. She tells Khan about the following points mentioned in various research reports on the utilities sector.

Report 1: A resurgence in domestic manufacturing activity will generate long-term growth in earnings and dividends that exceeds the cost of equity.

Report 2: Share repurchases are expected to increase. The report expresses confidence in the forecasts regarding the magnitude and timing of these repurchases.

Report 3: The report forecasts earnings growth of 4.5%. The key growth drivers are increases in population and business creation associated with stable GDP growth of 2.75%.

For Client 2's portfolio, Khan has identified the non-callable perpetual preferred stocks of Standard Company and Main Company.

- The Standard Company's preferred stock pays 2.75% on a par value of $100. Khan believes it to be fairly valued at a market price of $49.60.
- The perpetual preferred stock of Main Company has a par value of $50 per share and pays an annual dividend of 5.5%. Khan estimates a capitalization rate at 6%. The current market price of Main Company preferred stock is $42.

Finally, Khan has identified three stocks, shown in Exhibit 2, as likely candidates for Client 3's portfolio.

Exhibit 2	Candidate Stocks for Client 3
Stock	**Company Description**
BIOK	■ BIOK is a profitable biotech firm that currently pays an annual dividend of $1.20 per share. ■ The current annual dividend growth rate is 15%. ■ Patent protection runs out in eight years, after which dividend growth will likely decline at a steady rate over three years before stabilizing at a mature growth rate.
CCAX	■ CCAX builds communication software for state and federal prisons and detention facilities. ■ The company is expected to hold its cash dividends steady at $0.56 per share for six years as it builds out facilities and acquires properties. ■ Dividends are expected to grow at the nominal GDP growth rate after the next six years.
HLTV	■ HLTV is a health care equipment and services firm that is expected to maintain a stable dividend payout ratio. ■ Earnings are forecast to grow over the next two years by 27% annually. ■ After that, earnings will likely grow by 12% annually for another 10 years before stabilizing at a mature growth rate.

47 Based on the Gordon growth model, the justified leading P/E for ABC stock is *closest* to:

- **A** 17.1.
- **B** 17.7.
- **C** 20.0.

48 Based on its justified leading P/E and the Gordon growth model, XYZ stock is:

- **A** undervalued.
- **B** fairly valued.
- **C** overvalued.

49 Which sector report *best* describes a situation in which the Gordon growth model could be used to value utility stocks?

- **A** Report 1
- **B** Report 2
- **C** Report 3

50 Based on Khan's estimate of the capitalization rate, Main Company's preferred stock is:

- **A** undervalued.
- **B** fairly valued.
- **C** overvalued.

51 The capitalization rate of the preferred stock of Standard Company is *closest* to:

- **A** 2.75%.
- **B** 4.96%.
- **C** 5.54%.

52 Based on Exhibit 2, which stock can most appropriately be valued using a three-stage DDM with the second and third stages being treated as an H-model?

A BIOK

B CCAX

C HLTV

53 Which of the following models is *most* appropriate for valuing HLTV?

A H-model

B Three-stage DDM

C Gordon growth model

SOLUTIONS

1 Both companies are dividend-paying and have an established history of dividend payments that can provide some help in forecasting future dividends. In the case of BL Electronics, EPS has been increasing steadily from 2010 to 2017 and DPS has shown increases consistent with this trend. For example, EPS increased by \$0.23 from 2015 to 2016 and DPS increased by \$0.12. Then EPS increased by \$0.18 from 2016 to 2017 and DPS increased by \$0.12. The payout ratios have also been increasing gradually during the period examined. Dividends appear to be at least somewhat predictable given earnings forecasts. Overall, the DDM seems to be an appropriate model for valuing BL Electronics. In the case of FC Autos, however, dividends do not have a discernable relationship to the company's profitability. For example, DPS was \$2 in 2010 when the company was doing well and had an EPS of \$6.68, but DPS continued to be \$2 in 2015 when EPS was –\$18.50. The company continued to pay dividends in 2017, which was the third consecutive year of a negative EPS: In 2017, EPS had fallen to –\$68.45. The lack of a clear relationship of dividends to operating results suggests that the DDM is not appropriate for valuing FC Autos.

2 **A** Discounting the expected dividend of £5 in 2020 and the expected stock price of £250 at the end of 2020,

$$V_1 = \frac{D_2 + P_2}{(1+r)^1} = \frac{5 + 250}{(1 + 0.11)^1} = \frac{255}{1.11} = 229.73$$

B One way to answer this question is to use a DDM for two holding periods. Accordingly, discounting the expected dividend of £5 in 2020 and the expected stock price of £250 at the end of 2020 for two periods, and discounting the expected dividend of £4 in 2019 for one period,

$$V_0 = \frac{D_1}{(1+r)^1} + \frac{D_2 + P_2}{(1+r)^2} = \frac{4}{(1+0.11)^1} + \frac{5+250}{(1+0.11)^2}$$

$$= \frac{4}{1.11} + \frac{255}{1.11^2} = 3.60 + 206.96 = 210.57$$

based on full precision, or £210.56 with intermediate rounding. Another way to answer this question is to use the answer to Part A and a DDM for one holding period. Accordingly, discounting the expected dividend of £4 in 2019 and the expected stock price of £229.73 at the end of 2019 for one period,

$$V_0 = \frac{D_1 + V_1}{(1+r)^1} = \frac{4 + 229.73}{(1+0.11)^1} = \frac{233.73}{1.11} = 210.57$$

3 **A** The growth rate from 2015 to 2017 is $(15.50/9)^{1/2} - 1 = 0.312$, or 31.2%. The consensus predicted growth rate from 2017 to 2019 is $(20.20/15.50)^{1/2} - 1 = 0.142$, or 14.2%. Thus, the consensus forecast is for a sharp decline in the dividend growth rate for 2018 to 2019.

B Half of the growth rate computed in Part A = 14.2/2 = 7.1%. Based on this growth rate, D_1 = £15.50(1.071) = £16.60, rounded. Using the Gordon growth model,

$$V_0 = \frac{D_1}{r - g} = \frac{16.60}{0.11 - 0.071} = 425.64$$

or 425.65 based on not rounding the numerator.

C The estimated value of Sage Broadcasting would decrease as r increases and increase as g increases, all else equal.

4 Applying the Gordon growth model with the assumed 6.2% dividend growth rate results in an estimated value of $26,921.7 billion for the S&P 500 Index.

$$V_0 = \frac{D_1}{r - g} = \frac{456.3(1.062)}{(0.08 - 0.062)} = \$26{,}921.7 \text{ billion}$$

5 The preferred stock pays 4.5% of $100, or $4.50 in annual dividends. The dividend is fixed, so $g = 0$. Therefore, using the Gordon growth model with zero growth,

$$V_0 = \frac{D_1}{r} = \frac{4.50}{0.056} = \$80.36$$

6 This problem can be addressed using the Gordon growth model with constant expected negative growth. The estimated value of the stock is

$$V_0 = \frac{D_1}{r - g} = \frac{4}{0.11 - (-0.08)} = 21.05$$

7 Using the CAPM, the required rate of return on Maspeth Robotics shares is 4% + 1.2(5%) = 10%. Therefore, the constant dividend growth rate implied by a market price of €24 is 5.6%, as shown:

$$\begin{aligned} V_0 &= \frac{D_0(1 + g)}{r - g} \\ 24 &= \frac{1.00(1 + g)}{0.10 - g} \\ 2.4 - 24g &= 1.00 + g \\ 25g &= 1.4 \\ g &= 0.056 \text{ or } 5.6 \text{ percent} \end{aligned}$$

8 **A** With $b = 0.60$, the dividend payout ratio = $1 - b = 1 - 0.60 = 0.40$. Sustainable growth rate $g = b(\text{ROE}) = 0.60(0.14) = 0.084$, or 8.4%.

B The company paid a dividend per share of $1 - b(\text{EPS}) = 0.40(\$2) = \$0.80$ in 2019. The estimated value at the beginning of 2020 is

$$V_0 = \frac{D_1}{r - g} = \frac{0.80(1 + 0.0840)}{0.1100 - 0.0840} = \$33.35$$

C If the company were a no-growth company—that is, it paid out all its earnings and did not reinvest any—its earnings would stay the same. The value of such a company would be the value of a perpetuity, which is $D/r = E/r$ = $2/0.11 = $18.18. This amount is the no-growth value per share. So, PVGO = $33.35 − $18.18 = $15.17.

D The fraction of the company's value that comes from its growth opportunities is 15.17/33.35 = 0.4549, or 45.49%.

9 The payout ratio is A\$0.70/A\$2.00 = 0.35 = 1 − b, where b is the earnings retention ratio. Therefore, the justified trailing P/E based on fundamentals is 10.45, as shown here:

$$\frac{P_0}{E_0} = \frac{(1-b)(1+g)}{r-g}$$
$$= \frac{0.35(1+0.045)}{0.08-0.045}$$
$$= 10.45$$

Because the market trailing P/E of 14 is greater than 10.45, Stellar Baking Company shares appear to be overvalued (i.e., selling at a higher-than-warranted P/E).

10 The dividends in Stages 2 and 3 can be valued with the H-model, which estimates their value at the beginning of Stage 2. In this case, V_6 would capture the value of Stages 2 and 3 dividends. V_6 would then be discounted to the present. Also, the present values of dividends D_1 through D_6 need to be added to the present value of V_6.

$$V_6 = \frac{D_6(1 + g_L) + D_6 H(g_S - g_L)}{r - g_L}$$

where

$D_6 = D_0(1 + g_S)^6 = 9(1.14)^6 = 19.7548$
$r = 0.16$
$H = 10/2 = 5$
$g_S = 0.14$
$g_L = 0.10$

$$V_6 = \frac{19.7548(1.10) + 19.7548(5)(0.14 - 0.10)}{0.16 - 0.10} = 428.02$$

PV of $V_6 = 428.02/1.16^6 = 175.68$

PV of $D_1 = 9(1.14) / 1.16 = 8.8448$

PV of $D_2 = 9(1.14)^2 / 1.16^2 = 8.6923$

PV of $D_3 = 9(1.14)^3 / 1.16^3 = 8.5425$

PV of $D_4 = 9(1.14)^4 / 1.16^4 = 8.3952$

PV of $D_5 = 9(1.14)^5 / 1.16^5 = 8.2504$

PV of $D_6 = 9(1.14)^6 / 1.16^6 = 8.1082$

Value of stock = 8.8448 + 8.6923 + 8.5425 + 8.3952 + 8.2504 + 8.1082 + 175.68 = Rs. 226.51

11 A Let r be the required rate of return. Also let t = 0 indicate the valuation date in early 2019. Because the dividend growth rate becomes constant from the middle of 2022 (t = 3), this can be solved using the two-stage dividend discount model as follows:

$D_1 = 0.58(1.05) = 0.6090$

$D_2 = 0.58(1.05)^2 = 0.6395$

$D_3 = 0.58(1.05)^3 = 0.6714$

$D_4 = D_3\,(1.04) = 0.6714(1.04) = 0.6983$

And $V_3 = 0.6983/(0.07 - 0.04) = 23.2767$

V_0 can be expressed as

$$\begin{aligned} V_0 &= \frac{D_1}{1+r} + \frac{D_2}{(1+r)^2} + \frac{D_3}{(1+r)^3} + \frac{V_3}{(1+r)^3} \\ &= \frac{0.6090}{(1.07)} + \frac{0.6395}{(1.07)^2} + \frac{0.6714}{(1.07)^3} + \frac{23.2767}{(1.07)^3} \\ &= 0.5692 + 0.5585 + 0.5481 + 19.0001 \\ &= \text{A\$}20.68 \end{aligned}$$

B Because ANN's estimated value of A$20.68 is less than the market price of A$24.54, ANN appears to be overvalued at the market price.

12 A Use the H-model expression, with $H = 6/2 = 3$ and long-term and short-term dividend growth rates of 0.05 and 0.10, respectively, produces an expected return of 7.4% as shown:

$$\begin{aligned} r &= \left(\frac{D_0}{P_0}\right)\left[\left(1+g_L\right) + H\left(g_S - g_L\right)\right] + g_L \\ &= 0.02\left[\left(1+\ 0.05\right) + 3\left(0.10 - 0.05\right)\right] + 0.05 \\ &= 0.024 + 0.05 \\ &= 0.074 \end{aligned}$$

B In this case, the long- and short-term dividend growth rates are identical and the expected return is lower:

$$\begin{aligned} r &= \left(\frac{D_0}{P_0}\right)\left[\left(1+g_L\right) + H\left(g_S - g_L\right)\right] + g_L \\ &= 0.02\left[\left(1+0.05\right) + 3\left(0.05 - 0.05\right)\right] + 0.05 \\ &= 0.021 + 0.05 \\ &= 0.071 \end{aligned}$$

It is intuitive that a higher dividend growth rate is associated with a higher expected return if all the other inputs (such as the assumed required rate of return) are held constant.

13 A The formula for sustainable growth rate is

g = (b in the mature phase) × (ROE in the mature phase)

Because the dividend payout ratio in the mature phase is estimated to be 40%, or 0.40, the retention ratio b is expected to $1 - 0.40 = 0.60$. Therefore, given the 11% per year forecasted ROE,

$g = 0.60(11\%) = 6.6\%$.

B Based on the formula for sustainable growth rate, as b increases, growth rate increases, holding all else constant. However, all else may not be constant. In particular, the return accruing to additional investments may be lower, leading to a lower overall ROE. If that is the case and the company lowers the payout ratio to below 0.40 (thus increasing b to above 0.60), ROE would be expected to decline, which may lead to a lower growth rate.

14 A The four components of PRAT are computed for 2017 as follows:

P (Profit margin) = NI/Sales = 18,688/214,091 = 0.0873

R (Retention) = b = (EPS – DPS)/EPS = (8.77 – 2.26)/8.77 = 0.7423

A (Asset turnover) = Sales/Average total assets = 214,091/0.5(148,786 + 132,628) = 1.5215

T (Leverage) = Average total assets/Average shareholders' equity = (148,786 + 132,628)/(77,088 + 68,935) = 1.9272

The components are similarly computed for the other years and summarized in the following table. Their average values are also included.

	Needed for Solution to B	Solution to A		
Item	**Average**	**2017**	**2016**	**2015**
P (Profit margin)	0.0812	0.0873	0.0836	0.0728
R (Retention)	0.7390	0.7423	0.7423	0.7324
A (Asset turnover)	1.6250	1.5215	1.5855	1.7681
T (Leverage)	1.9736	1.9272	1.9638	2.0299

B Using the average values for each component,

$$g = \text{PRAT} = (0.0812)(0.7390)(1.6250)(1.9736) = 0.1924 \text{ or } 19.2\%$$

The sustainable growth rate for the company based on the PRAT expression is 19.2%.

C Given that the high value of g does not seem sustainable indefinitely, it appears that the company has not yet reached the mature phase.

15 A The following table provides the details from the spreadsheet model. The constant growth rate after Year 4 is 2% less than that in Year 4. So,

$$g = 0.1180 - 0.0200 = 0.098, \text{ or } 9.8\%.$$

$$V_4 = D_4(1 + g)/(r - g) = 1.80(1.098)/(0.13 - 0.098) = \$61.76$$

Year	1	2	3	4
Sales ($ millions)	300.00	345.00	396.75	436.43
EBIT	51.00	58.65	67.45	74.19
Interest ($ millions)	10.00	10.00	10.00	10.00
EBT	41.00	48.65	57.45	64.19
Taxes (30%)	12.30	14.60	17.23	19.26
Net income	28.70	34.06	40.21	44.93
Dividends	11.48	13.62	16.09	17.97
DPS	1.15	1.36	1.61	1.80
Growth rate of DPS		18.26%	18.38%	11.80%
PV of DPS	1.02	1.07	1.12	1.10
$V_4 = D_4(1 + g)/(r - g)$				61.76
PV of V_4				$37.87

B V_0 = Sum of PV of DPS and PV of V_4 = 1.02 + 1.07 + 1.12 + 1.10 + 61.76/$(1 + 0.13)^4$ = $42.18

C The following table provides the details if the sales growth rate in Year 3 is 10%:

Year	**1**	**2**	**3**	**4**
Sales ($ millions)	300.00	345.00	379.50	417.45
EBIT	51.00	58.65	64.52	70.97
Interest ($ millions)	10.00	10.00	10.00	10.00
EBT	41.00	48.65	54.52	60.97
Taxes (%)	12.30	14.60	16.35	18.29
Net income	28.70	34.06	38.16	42.68
Dividends	11.48	13.62	15.26	17.07
DPS	1.15	1.36	1.53	1.71
Growth rate of DPS		18.26%	12.50%	11.76%
PV of DPS	1.02	1.07	1.06	1.05
$V_4 = D_4(1 + g)/(r - g)$				57.93
PV of V_4				$35.53

V_0 = Sum of PV of DPS and PV of V_4

= 1.02 + 1.07 + 1.06 + 1.05 + 35.53

= $39.73

16 C is correct. Using the Gordon growth model,

$$V_0 = \frac{D_1}{r - g} = \frac{0.59(1 + 0.0530)}{0.1115 - 0.0530} = \$10.62$$

17 A is correct. The justified trailing P/E or P_0/E_0 is V_0/E_0, where V_0 is the fair value based on the stock's fundamentals. The fair value V_0 computed earlier is $10.62 and E_0 is $2.12. So, the justified trailing P/E is 10.62/2.12 = 5.01.

18 A is correct. Rae's estimate of the intrinsic value is $10.62. So, the band Rae is looking at is $10.62 ± 0.10($10.62), which runs from $10.62 + $1.06 = $11.68 on the upside to $10.62 – $1.06 = $9.56 on the downside. Because $8.42 is less than $9.56, Rae would consider Tasty Foods to be undervalued.

19 B is correct. Using a beta of 1.25, Rae's estimate for the required return on equity for Tasty Foods is 0.04 + 1.25(0.065) = 0.1213, or 12.13%. The estimated value of the stock is

$$V_0 = \frac{D_1}{r - g} = \frac{0.59 \times (1 + 0.0530)}{0.1213 - 0.0530} = \$9.10$$

20 A is correct. The price of the stock is $8.42. If this price is also the fair value of the stock,

$$V_0 = 8.42 = \frac{D_1}{r - g} = \frac{0.59 \times (1 + g)}{0.1115 - g}$$

$$0.9388 - 8.42g = 0.59 + 0.59g$$

$$9.01g = 0.3488$$

$$g = 0.0387 \text{ or } 3.87 \text{ percent}$$

21 A is correct. If the stock is fairly priced in the market as per the Gordon growth model, the stock price is expected to increase at g, the expected growth rate in dividends. The implied growth rate in dividends, if price is the fair value, is 3.87%. Therefore, the expected capital gains yield is 3.87%.

22 B is correct. The following table provides the calculations needed to compute the value of the stock using the first approach, including the calculations for the terminal value V_8. As the table shows, the terminal value V_8 = C$31.0550.

Time	Value	Calculation	D_t or V_t	Present Values $D_t/(1.0872)^t$ or $V_t/(1.0872)^t$
1	D_1	C$0.175(1.14)	C$0.1995	C$0.1835
2	D_2	$0.175(1.14)^2$	0.2274	0.1924
3	D_3	$0.175(1.14)^3$	0.2593	0.2018
4	D_4	$0.175(1.14)^4$	0.2956	0.2116
5	D_5	$0.175(1.14)^5$	0.3369	0.2218
6	D_6	$0.175(1.14)^6$	0.3841	0.2326
7	D_7	$0.175(1.14)^7$	0.4379	0.2439
8	D_8	$0.175(1.14)^8$	0.4992	0.2557
8	V_8	$0.175(1.14)^8(1.07)/(0.0872 - 0.07)$	31.0550	15.9095
Total				C$17.6528

23 C is correct. As shown in the foregoing table, the value of the second stage = PV of V_8 = C$15.9095. The total value is C$17.6528. As a proportion, the second stage represents 15.9095/17.6528 = 0.90 of the total value.

24 B is correct.

$V_8/E_8 = 17$

$D_8/E_8 = 1 - 0.70 = 0.30$

From the table with the calculation details for the solution to Problem 22, D_8 = C$0.4992. So, $0.4992/E_8 = 0.30$, which means that $E_8 = 0.4992/0.30 = 1.6640$.

$V_8/E_8 = 17$ implies that $V_8/1.6640 = 17$, which gives $V_8 = 17(1.6640)$ = C$28.2880.

25 A is correct. As computed earlier, $V_8 = 17(1.6640)$ = C$28.2880.

PV of $V_8 = 28.2880/1.0872^8 = 14.4919$

From the table with the calculation details for the solution to Problem 22,

Sum of PV of D_1 through D_8 = 1.7433

So, the value of stock $V_0 = 14.4919 + 1.7433$ = C$16.2352.

26 C is correct. Using the H-model,

$$V_0 = \frac{D_0(1 + g_L) + D_0 H(g_S - g_L)}{r - g_L}$$

where

$D_0 = 0.175$
$r = 0.0872$
$H = 4$

$g_S = 0.14$
$g_L = 0.07$
$$V_0 = \frac{0.175(1.07) + 0.175(4)(0.14 - 0.07)}{0.0872 - 0.07} = 13.7355.$$

The market price is C$17, which is greater than C$13.7355. So, the stock is overvalued in the market.

27 B is correct. If the extraordinary growth rate of 14% is expected to continue for a longer duration, the stock's value would increase. Choice A is false because given that the first stage is longer (11 years instead of 8), the terminal value is being calculated at a later point in time. So, its present value would be smaller. Moreover, the first stage has more years and contributes more to the total value. Overall, the proportion contributed by the second stage would be smaller. Choice C is false because the intrinsic value of the stock would be higher and the appropriate conclusion would be that the stock would be undervalued to a greater extent based on the first approach.

28 C is correct. A dividend discount model is especially useful when dividend policy bears an understandable and consistent relationship to the company's profitability. The relatively consistent dividend payout ratio suggests Company JZY would be a suitable candidate for a dividend discount model.

29 B is correct. The value of ABC stock can be computed as follows:

Given: Dividend (D_0) = C$0.65, Return ($r$) = 7.35%, Short-term growth (g_S) = 12% for 4 years, Long-term growth (g_L) = 3.5% thereafter.

Then:

$D_1 = D_0(1 + g_S)^1 = 0.65(1.12) = \text{C\$}0.7280$

$D_2 = D_0(1 + g_S)^2 = 0.65(1.12)^2 = \text{C\$}0.8154$

$D_3 = D_0(1 + g_S)^3 = 0.65(1.12)^3 = \text{C\$}0.9132$

$D_4 = D_0(1 + g_S)^4 = 0.65(1.12)^4 = \text{C\$}1.0228$

$P_4 = [D_4\ (1 + g_L)]/(r - g_L) = [D_4\ (1.035)]/(0.0735 - 0.035) = \text{C\$}27.4960.$

$V_0 = D_1/(1 + r)^1 + \dots + D_4/(1 + r)^4 + P_4/(1 + r)^4.$
$V_0 = [0.7280/(1.0735)^1] + [0.8154/(1.0735)^2] +[\ 0.9132/(1.0735)^3] + [1.0228/(1.0735)^4] + [27.4960/(1.0735)^4]$
$= \text{C\$}23.5984$ (rounded to C$23.60).

30 C is correct. The value of ABC would be calculated using the Gordon growth model as follows:

$V_0 = [D_0(1 + g)]/(r - g) = [0.65\ (1.035)]/(0.0735 - 0.035) = \text{C\$}17.47.$

31 B is correct. In the Gordon growth model, Total return = Dividend yield + Capital gains yield (i.e., constant growth rate). When a stock is fairly valued, the expected total return will equal the required return or discount rate (i.e., 7.35%). In the case of ABC, the total return is 7.35% and the capital gains yield is 3.5%. Therefore, the dividend yield is 7.35% – 3.5% = 3.85%.

32 C is correct. The current value of XYZ stock would be calculated as follows:

$V_0 = [P_3/(1 + r)^3]$, where $P_3 = D_4/(r - g)$.

Given $D_4 = 1.72$, $r = 10\%$, and $g = 4\%$,

$V_0 = [1.72/(0.10 - 0.04)]/(1.10)^3 = \text{US\$}21.54.$

33 C is correct. The dividend growth rate implied in the stock price of XYZ (i.e., 6%) is greater than the growth rate assumed by the analyst (i.e., 4%), suggesting that XYZ is overvalued.

34 C is correct. The sustainable growth rate of JZY stock = g = Retention ratio × ROE = 0.60 × 0.0817 = 4.9%. JZY stock's implied growth rate of 5.3% is higher than the sustainable growth rate of 4.9%. Consequently, the stock is overvalued—that is, the intrinsic value of the stock will be less than its current market price.

The current intrinsic value of JZY stock is as follows:

$$\begin{aligned} V_0 &= [D_0(1+g)]/(r-g) \\ &= [0.84\ (1.0490)]/(0.08-0.0490) \\ &= \text{US\$}28.42 < \text{US\$}32.76 \end{aligned}$$

35 B is correct. BKKQ has the largest PVGO, calculated as follows:

PVGO (ABTD) = $P_0 - E_1/r$ = 80.00 – [7.30/0.105] = C$10.48

PVGO (BKKQ) = $P_0 - E_1/r$ = 39.00 – [2.12/0.08] = C$12.50

PVGO (CPMN) = $P_0 - E_1/r$ = 27.39 – [1.90/0.12] = C$11.56

where P_0 is the current price per share, E_1 is the forecasted earnings per share, and r is the required rate of return.

36 C is correct. The leading P/E is calculated as follows:

$$P_0/E_1 = [1/r] + [\text{PVGO}/E_1],$$

where $1/r$ captures the no-growth component of P/E and PVGO/E_1 captures the growth component of the P/E.

PVGO is computed as follows:

PVGO (ABTD) = $P_0 - E_1/r$ = 80.00 – [7.30/0.105] = C$10.48

PVGO (BKKQ) = $P_0 - E_1/r$ = 39.00 – [2.12/0.08] = C$12.50

PVGO (CPMN) = $P_0 - E_1/r$ = 27.39 – [1.90/0.12] = C$11.56

where P_0 is the current price per share, E_1 is the forecasted earnings per share, and r is the required rate of return.

The growth component of the P/E for each stock [PVGO/E_1] is as follows:

ABTD: 10.48/7.30 = 1.44×

BKKQ: 12.50/2.12 = 5.90×

CPMN: 11.56/1.90 = 6.08×

37 A is correct because the estimated value of the stock at the end of the first stage of a basic two-stage DDM (terminal value) is the present value of all dividends beyond the first stage. The first stage is 2020 through 2023, and the second stage begins in 2024, so the terminal value (that is, the value of the stock at the end of 2023) is the present value of future dividends beyond 2023.

38 C is correct based on Withers's assumptions applied to the dividend valuation model.

The stock value as of the end of 2019 equals the present value of all future dividends in 2020 through 2022 plus the present value of the terminal value at the end of 2022. The forecasted stock value equals EUR51.254:

Year	Dividend	Terminal Value	D_t or V_t	Present Value of D_t or V_t
2020	2.315(1.05) = 2.431		2.431	2.251
2021	2.431(1.05) = 2.553		2.553	2.189
2022	2.553(1.05) = 2.681		2.681	2.128
2023	2.681(1.05) = 2.815	57.980	60.795	44.686
2024	2.815(1.03) = 2.899			
Total				51.254

The terminal value at the end of 2023 is calculated using the dividend in the first year beyond the first stage, divided by the difference between the required rate of return and the growth rate in the second stage.

$$\text{Terminal value at end of 2023} = \frac{2.815(1.03)}{(0.08 - 0.03)} = 57.980$$

39 A is correct. During the first stage, the basic two-stage model has higher (i.e., 5%) growth than the H-model, in which growth is declining linearly from 5.0% to 3.5%. Higher growth rates result in higher forecasted dividends and stock prices at the beginning of the sustained growth phase. Because the long-term dividend growth rates are the same for both models, the difference in forecasted stock price arises from growth rate differences in the first stage.

Therefore, the dividend at the end of the first stage will be lower for the H-model than for the basic two-stage DDM, and the terminal value will be lower in the H-model than in the two-stage model. Specifically, the 2023 dividends will be 2.734 (i.e., 2.315 × 1.05 × 1.045 × 1.04 × 1.035) for the H-model versus 2.815 [i.e., 2.315 × $(1.05)^4$] for the basic two-stage DDM.

40 C is correct, based on Exhibit 2 and the H-model.

Estimate the required rate of return using Equation 21:

$$r = \frac{D_0}{P_0}\left[(1 + g_L) + H(g_S - g_L)\right] + g_L$$

Substitute the following:

$D_0 = 3.15$
$g_S = 8\%$
$g_L = 4\%$
$H = 4 \div 2 = 2$

The model thus produces

$$r = \frac{3.15}{58.79}\left[(1 + 0.04) + 2(0.08 - 0.04)\right] + 0.04$$
$$= (0.053581 \times 1.12) + 0.04$$
$$= 0.060010 + 0.04 = 0.10001 \approx 10\%.$$

41 B is correct based on the present value of forecasted dividends. The dividend at the end of 2019, based on case material, is USD3.15 per share.

Year	Dividend per Share, Prior Year	Growth Rate during Year	Dividend per Share, Current Year	Terminal Value	D_t or V_t	Present Value of D_t or V_t
2020	3.150	6%	3.339		3.339	3.092
2021	3.339	6%	3.539		3.539	3.034
2022	3.539	5%	3.716		3.716	2.950
2023	3.716	5%	3.902	80.381	84.283	61.951
					Total	71.027

$$\text{Terminal value at the end of 2023} = \frac{3.902(1.03)}{(0.08 - 0.03)} = 80.381$$

42 A is correct, based on the use of average total assets and beginning-of-year shareholders' equity.

$$g = \frac{\text{Net income} - \text{Dividends}}{\text{Net income}} \times \frac{\text{Net income}}{\text{Sales}} \times \frac{\text{Sales}}{\text{Total assets}} \times \frac{\text{Total assets}}{\text{Shareholders' equity}}$$

To calculate sustainable growth,

$$g = \frac{43{,}923 - 1{,}518}{43{,}923} \times \frac{43{,}923}{423{,}474} \times \frac{423{,}474}{486{,}203} \times \frac{486{,}203}{397{,}925}$$

$$= 96.544\% \times 10.372\% \times 87.100\% \times 122.200\%$$
$$= 10.658\%$$

43 B is correct because the sustainable growth is the product of the return on equity and the retention ratio. If the payout ratio ranges from 0% to 10%, the percentage of earnings retained by the firm ranges from 100% to 90%.

Sensitivity: Sustainable Growth Rates

Return on Equity	Retention Ratio	
	90%	100%
8%	7.2%	8.0%
12%	10.8%	12.0%

44 C is correct because it is based on the sustainable growth rate and the required rate of return:

$$\begin{aligned}\text{Sustainable growth rate} &= (b \text{ in mature phase}) \times (\text{Return on equity})\\ &= (1 - \text{Dividend payout}) \times (\text{Return on equity})\\ 0.036 &= (1 - \text{Dividend payout}) \times 0.083\end{aligned}$$

Solving for the dividend payout ratio, the dividend payout = 56.627% ≈ 56.6%.

45 A is correct. Intern 2 values Xavier stock at USD56.372 per share, which is higher than the current price of USD52.

$$D_1 = 2.000 \times (1.048)^1 = 2.096$$

$$D_2 = 2.000 \times (1.048)^2 = 2.197$$

$$D_3 = 2.000 \times (1.048)^3 = 2.302$$

$$D_4 = 2.000 \times (1.048)^4 = 2.413$$

$$D_5 = 2.000 \times (1.048)^4 \times 1.04 = 2.510$$

$$\text{Value per share} = \frac{2.096}{(1+0.078)^1} + \frac{2.197}{(1+0.078)^2} + \frac{2.302}{(1+0.078)^3} + \frac{2.413 + \dfrac{2.510}{(0.078-0.04)}}{(1+0.078)^4}$$

$$= \text{USD}56.372$$

46 B is correct. The candidate can arrive at the answer one of two ways. The first way is to use Equation 19 and solve for g_L:

$$P_0 = \left[\sum_{t=1}^{n} \frac{D_0(1+g_S)^t}{(1+r)^t}\right] + \left[\frac{D_0(1+g_S)^n(1+g_L)}{(1+r)^n(r-g_L)}\right]$$

Insert the known values:

$$\text{USD}52 = \sum_{t=1}^{4} \frac{2(1+0.05)^t}{(1+0.083)^t} + \frac{2(1+0.05)^4(1+g_L)}{(1+0.083)^4(0.083-g_L)}$$

$$= 7.4089 + \frac{2.431(1+g_L)}{1.37567(0.083-g_L)}$$

Solve for g_L:

$$g_L = 4.172\%.$$

Check:

$$7.4089 + \frac{2.431(1+0.04127)}{1.3757(0.083-0.04172)} = 7.4089 + 44.5830 \approx 52.00$$

The second way is to use Equation 19 and substitute the different choices to determine the value that produces a value of USD52 per share:

$$\text{USD}52 = \sum_{t=1}^{4} \frac{2(1+0.05)^t}{(1+0.083)^t} + \frac{2(1+0.05)^4(1+0.0417)}{(1+0.083)^4(0.083-0.0417)}$$

47 A is correct. The justified leading P/E is calculated as

$$\frac{P_0}{E_1} = \frac{(1-b)}{(r-g)}$$

where b is the retention ratio, $1 - b$ is the dividend payout ratio, r is the discount rate, and g is the long-term growth rate.

ABC's dividend payout rate, $1 - b$, is given as 0.60. For Company ABC, the justified leading P/E is

$$\frac{P_0}{E_1} = \frac{(1-b)}{(r-g)} = \frac{(0.60)}{(0.07-0.035)} \approx 17.1$$

48 B is correct. The justified leading P/E is calculated as

$$\frac{P_0}{E_1} = \frac{(1-b)}{(r-g)}$$

where b is the retention ratio, $1 - b$ is the dividend payout ratio, r is the discount rate, and g is the long-term growth rate.

The justified leading P/E is

$$\frac{P_0}{E_1} = \frac{0.8}{(0.07 - 0.03)} = 20$$

XYZ's actual leading P/E is

$$\frac{P_0}{E_1} = \frac{\$10}{\$0.50} = 20$$

Because the justified leading P/E equals the actual leading P/E, the stock is fairly valued.

49 B is correct because the Gordon growth model can accurately value companies that are repurchasing shares when the analyst can appropriately adjust the dividend growth rate for the impact of share repurchases.

50 A is correct. The value of a share of Main Company's preferred stock is

$$V_0 = \frac{D}{r} = \frac{\$50 \times 0.055}{0.06} = \frac{\$2.75}{0.06} = \$45.83$$

The current price of a share of Main Company's preferred stock is \$42, so the stock is currently undervalued.

51 C is correct. The value of non-callable fixed-rate perpetual preferred stock is calculated as

$$V_0 = \frac{D}{r} \rightarrow r = \frac{D}{V_0}$$

where D is the constant dividend per share and r is the discount rate. The discount rate of a perpetuity is often called the capitalization rate.

For Standard Company, the dividend is $D = 2.75\% \times \$100 = \$2.75$.

Therefore,

$$r = \frac{\$2.75}{\$49.60} = 5.54\%$$

52 A is correct because the dividend growth is declining linearly during the second stage of a three-stage DDM used to value BIOK. As noted in the text, a three-stage valuation clearly has an H-model process in the second and third stages. In contrast, abrupt—rather than linearly declining—dividend growth rates are implied for CCAX and HLTV.

53 B is correct because HLTV is forecast to have three growth stages: the growth phase (2 years at 27%), the transition phase (10 years at 12%), and the mature phase. Because the earnings growth has three stages and the dividend payout ratio is stable, a three-stage DDM is appropriate.

Glossary

Abandonment option The ability to terminate a project at some future time if the financial results are disappointing.
Abnormal earnings See *residual income.*
Abnormal return The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.
Absolute convergence The idea that developing countries, regardless of their particular characteristics, will eventually catch up with the developed countries and match them in per capita output.
Absolute valuation model A model that specifies an asset's intrinsic value.
Absolute version of PPP An extension of the law of one price whereby the prices of goods and services will not differ internationally once exchange rates are considered.
Accounting estimates Estimates used in calculating the value of assets or liabilities and in the amount of revenue and expense to allocate to a period. Examples of accounting estimates include, among others, the useful lives of depreciable assets, the salvage value of depreciable assets, product returns, warranty costs, and the amount of uncollectible receivables.
Accumulated benefit obligation The actuarial present value of benefits (whether vested or non-vested) attributed, generally by the pension benefit formula, to employee service rendered before a specified date and based on employee service and compensation (if applicable) before that date. The accumulated benefit obligation differs from the projected benefit obligation in that it includes no assumption about future compensation levels.
Accuracy The percentage of correctly predicted classes out of total predictions. It is an overall performance metric in classification problems.
Acquirer The company in a merger or acquisition that is acquiring the target.
Acquiring company See *acquirer*.
Acquisition The purchase of some portion of one company by another; the purchase may be for assets, a definable segment of another entity, or the entire company.
Activation function A functional part of a neural network's node that transforms the total net input received into the final output of the node. The activation function operates like a light dimmer switch that decreases or increases the strength of the input.
Active factor risk The contribution to active risk squared resulting from the portfolio's different-than-benchmark exposures relative to factors specified in the risk model.
Active return The return on a portfolio minus the return on the portfolio's benchmark.
Active risk The standard deviation of active returns.
Active risk squared The variance of active returns; active risk raised to the second power.
Active share A measure of how similar a portfolio is to its benchmark. A manager who precisely replicates the benchmark will have an active share of zero; a manager with no holdings in common with the benchmark will have an active share of one.
Active specific risk The contribution to active risk squared resulting from the portfolio's active weights on individual assets as those weights interact with assets' residual risk.
Adjusted funds from operations (AFFO) Funds from operations adjusted to remove any non-cash rent reported under straight-line rent accounting and to subtract maintenance-type capital expenditures and leasing costs, including leasing agents' commissions and tenants' improvement allowances.
Adjusted present value As an approach to valuing a company, the sum of the value of the company, assuming no use of debt, and the net present value of any effects of debt on company value.
Adjusted R^2 A measure of goodness-of-fit of a regression that is adjusted for degrees of freedom and hence does not automatically increase when another independent variable is added to a regression.
Administrative regulations or administrative law Rules issued by government agencies or other regulators.
Advanced set An arrangement in which the reference interest rate is set at the time the money is deposited.
Advanced settled An arrangement in which a forward rate agreement (FRA) expires and settles at the same time, at the FRA expiration date.
Agency costs Costs associated with the conflict of interest present when a company is managed by non-owners. Agency costs result from the inherent conflicts of interest between managers and equity owners.
Agency costs of equity The smaller the stake managers have in the company, the less their share in bearing the cost of excessive perquisite consumption—consequently, the less their desire to give their best efforts in running the company.
Agency issues Conflicts of interest that arise when the agent in an agency relationship has goals and incentives that differ from the principal to whom the agent owes a fiduciary duty. Also called *agency problems* or *principal–agent problems*.
Agglomerative clustering A bottom-up hierarchical clustering method that begins with each observation being treated as its own cluster. The algorithm finds the two closest clusters, based on some measure of distance (similarity), and combines them into one new larger cluster. This process is repeated iteratively until all observations are clumped into a single large cluster.
Allowance for loan losses A balance sheet account; it is a contra asset account to loans.
Alpha The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.
American Depositary Receipt A negotiable certificate issued by a depositary bank that represents ownership in a non-US company's deposited equity (i.e., equity held in custody by the depositary bank in the company's home market).
Analysis of variance (ANOVA) The analysis that breaks the total variability of a dataset (such as observations on the dependent variable in a regression) into components representing different sources of variation. With reference to regression, ANOVA provides the inputs for an *F*-test of

the significance of the regression as a whole, as well as the inputs for the coefficient of determination and the standard error of the estimate.

Application programming interface (API) A set of well-defined methods of communication between various software components and typically used for accessing external data.

Arbitrage (1) The simultaneous purchase of an undervalued asset or portfolio and sale of an overvalued but equivalent asset or portfolio in order to obtain a riskless profit on the price differential. Taking advantage of a market inefficiency in a risk-free manner. (2) The condition in a financial market in which equivalent assets or combinations of assets sell for two different prices, creating an opportunity to profit at no risk with no commitment of money. In a well-functioning financial market, few arbitrage opportunities are possible. (3) A risk-free operation that earns an expected positive net profit but requires no net investment of money.

Arbitrage-free models Term structure models that project future interest rate paths that emanate from the existing term structure. Resulting prices are based on a no-arbitrage condition.

Arbitrage-free valuation An approach to valuation that determines security values consistent with the absence of any opportunity to earn riskless profits without any net investment of money.

Arbitrage opportunity An opportunity to conduct an arbitrage; an opportunity to earn an expected positive net profit without risk and with no net investment of money.

Arbitrage portfolio The portfolio that exploits an arbitrage opportunity.

Ask price The price at which a trader will sell a specified quantity of a security. Also called *ask*, *offer price*, or *offer*.

Asset-based approach Approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities.

Asset-based valuation An approach to valuing natural resource companies that estimates company value on the basis of the market value of the natural resources the company controls.

Asset beta The unlevered beta; reflects the business risk of the assets; the asset's systematic risk.

Asset purchase An acquisition in which the acquirer purchases the target company's assets and payment is made directly to the target company.

Asymmetric information The differential of information between corporate insiders and outsiders regarding the company's performance and prospects. Managers typically have more information about the company's performance and prospects than owners and creditors.

At market contract When a forward contract is established, the forward price is negotiated so that the market value of the forward contract on the initiation date is zero.

Authorized participants (APs) A special group of institutional investors who are authorized by the ETF issuer to participate in the creation/redemption process. APs are large broker/dealers, often market makers.

Autocorrelations The correlations of a time series with its own past values.

Autoregressive model (AR) A time series regressed on its own past values in which the independent variable is a lagged value of the dependent variable.

Backtesting The process that approximates the real-life investment process, using historical data, to assess whether an investment strategy would have produced desirable results.

Backward integration A merger involving the purchase of a target ahead of the acquirer in the value or production chain; for example, to acquire a supplier.

Backward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving backward through the network's layers.

Backwardation A condition in futures markets in which the spot price exceeds the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is higher than the longer-term futures contract price.

Bag-of-words (BOW) A collection of a distinct set of tokens from all the texts in a sample dataset. BOW does not capture the position or sequence of words present in the text.

Bankruptcy A declaration provided for by a country's laws that typically involves the establishment of a legal procedure that forces creditors to defer their claims.

Barbell portfolio Fixed-income portfolio that combines short and long maturities.

Base error Model error due to randomness in the data.

Basic earnings per share (EPS) Net earnings available to common shareholders (i.e., net income minus preferred dividends) divided by the weighted average number of common shares outstanding during the period.

Basis The difference between the spot price and the futures price. As the maturity date of the futures contract nears, the basis converges toward zero.

Basis trade A trade based on the pricing of credit in the bond market versus the price of the same credit in the CDS market. To execute a basis trade, go long the "underpriced" credit and short the "overpriced" credit. A profit is realized as the implied credit prices converge.

Bear hug A tactic used by acquirers to circumvent target management's objections to a proposed merger by submitting the proposal directly to the target company's board of directors.

Bearish flattening Term structure shift in which short-term bond yields rise more than long-term bond yields, resulting in a flatter yield curve.

Benchmark value of the multiple In using the method of comparables, the value of a price multiple for the comparison asset; when we have comparison assets (a group), the mean or median value of the multiple for the group of assets.

Best ask The offer to sell with the lowest ask price. Also called *best offer* or *inside ask*.

Best bid The offer to buy with the highest bid price. Also called the *inside bid*.

Best offer See *best ask*.

Bias error Describes the degree to which a model fits the training data. Algorithms with erroneous assumptions produce high bias error with poor approximation, causing underfitting and high in-sample error.

Bid–ask spread The ask price minus the bid price.

Bid price The price at which a trader will buy a specified quantity of a security. Also called *bid.*

Bill-and-hold basis Sales on a bill-and-hold basis involve selling products but not delivering those products until a later date.

Blockage factor An illiquidity discount that occurs when an investor sells a large amount of stock relative to its trading volume (assuming it is not large enough to constitute a controlling ownership).

Bond indenture A legal contract specifying the terms of a bond issue.

Bond risk premium The expected excess return of a default-free long-term bond less that of an equivalent short-term bond.

Bond yield plus risk premium method An estimate of the cost of common equity that is produced by summing the before-tax cost of debt and a risk premium that captures the additional yield on a company's stock relative to its bonds. The additional yield is often estimated using historical spreads between bond yields and stock yields.

Bonding costs Costs borne by management to assure owners that they are working in the owners' best interest (e.g., implicit cost of non-compete agreements).

Bonus issue of shares *See stock dividend.*

Book value Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value of equity Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value per share The amount of book value (also called carrying value) of common equity per share of common stock, calculated by dividing the book value of shareholders' equity by the number of shares of common stock outstanding.

Bootstrap aggregating (or bagging) A technique whereby the original training dataset is used to generate *n* new training datasets or bags of data. Each new bag of data is generated by random sampling with replacement from the initial training set.

Bootstrapping The use of a forward substitution process to determine zero-coupon rates by using the par yields and solving for the zero-coupon rates one by one, from the shortest to longest maturities.

Bottom-up approach With respect to forecasting, an approach that usually begins at the level of the individual company or a unit within the company.

Breakup value The value derived using a sum-of-the-parts valuation.

Breusch–Pagan test A test for conditional heteroskedasticity in the error term of a regression.

Bullet portfolio A fixed-income portfolio concentrated in a single maturity.

Bullish flattening Term structure change in which the yield curve flattens in response to a greater decline in long-term rates than short-term rates.

Bullish steepening Term structure change in which short-term rates fall by more than long-term yields, resulting in a steeper term structure.

Buy-side analysts Analysts who work for investment management firms, trusts, bank trust departments, and similar institutions.

Buyback *See share repurchase.*

Callable bond Bond that includes an embedded call option that gives the issuer the right to redeem the bond issue prior to maturity, typically when interest rates have fallen or when the issuer's credit quality has improved.

Canceled shares Shares that were issued, subsequently repurchased by the company, and then retired (cannot be reissued).

Cannibalization Cannibalization occurs when an investment takes customers and sales away from another part of the company.

Capital charge The company's total cost of capital in money terms.

Capital deepening An increase in the capital-to-labor ratio.

Capital rationing A capital rationing environment assumes that the company has a fixed amount of funds to invest.

Capital structure The mix of debt and equity that a company uses to finance its business; a company's specific mixture of long-term financing.

Capitalization of earnings method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capitalization rate The divisor in the expression for the value of perpetuity. In the context of real estate, it is the divisor in the direct capitalization method of estimating value. The cap rate equals net operating income divided by value.

Capitalized cash flow method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity. Also called *capitalized cash flow model.*

Capitalized income method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capped floater Floating-rate bond with a cap provision that prevents the coupon rate from increasing above a specified maximum rate. It protects the issuer against rising interest rates.

Carry arbitrage model A no-arbitrage approach in which the underlying instrument is either bought or sold along with an opposite position in a forward contract.

Carry benefits Benefits that arise from owning certain underlyings; for example, dividends, foreign interest, and bond coupon payments.

Carry costs Costs that arise from owning certain underlyings. They are generally a function of the physical characteristics of the underlying asset and also the interest forgone on the funds tied up in the asset.

Cash available for distribution See *adjusted funds from operations.*

Cash-generating unit The smallest identifiable group of assets that generates cash inflows that are largely independent of the cash inflows of other assets or groups of assets.

Cash offering A merger or acquisition that is to be paid for with cash; the cash for the merger might come from the acquiring company's existing assets or from a debt issue.

Cash settlement A procedure used in certain derivative transactions that specifies that the long and short parties settle the derivative's difference in value between them by making a cash payment.

Catalyst An event or piece of information that causes the marketplace to re-evaluate the prospects of a company.

Categorical dependent variables An alternative term for qualitative dependent variables.

CDS spread A periodic premium paid by the buyer to the seller that serves as a return over a market reference rate required to protect against credit risk.

Ceiling analysis A systematic process of evaluating different components in the pipeline of model building. It helps to understand what part of the pipeline can potentially improve in performance by further tuning.

Centroid The center of a cluster formed using the *k*-means clustering algorithm.

Chain rule of forecasting A forecasting process in which the next period's value as predicted by the forecasting equation is substituted into the right-hand side of the equation to give a predicted value two periods ahead.

Cheapest-to-deliver The debt instrument that can be purchased and delivered at the lowest cost yet has the same seniority as the reference obligation.

Classification and regression tree A supervised machine learning technique that can be applied to predict either a categorical target variable, producing a classification tree, or a continuous target variable, producing a regression tree. CART is commonly applied to binary classification or regression.

Clean surplus relation The relationship between earnings, dividends, and book value in which ending book value is equal to the beginning book value plus earnings less dividends, apart from ownership transactions.

Club convergence The idea that only rich and middle-income countries sharing a set of favorable attributes (i.e., are members of the "club") will converge to the income level of the richest countries.

Cluster A subset of observations from a dataset such that all the observations within the same cluster are deemed "similar."

Clustering The sorting of observations into groups (clusters) such that observations in the same cluster are more similar to each other than they are to observations in other clusters.

Cobb–Douglas production function A function of the form $Y = K^{\alpha} L^{1-\alpha}$ relating output (Y) to labor (L) and capital (K) inputs.

Coefficient of determination The percentage of the variation of the dependent variable that is explained by the independent variable. Also referred to as the "*R*-squared" or "R^2."

Cointegrated Describes two time series that have a long-term financial or economic relationship such that they do not diverge from each other without bound in the long run.

Collateral return The component of the total return on a commodity futures position attributable to the yield for the bonds or cash used to maintain the futures position. Also called *collateral yield*.

Collection frequency (CF) The number of times a given word appears in the whole corpus (i.e., collection of sentences) divided by the total number of words in the corpus.

Commercial real estate properties Income-producing real estate properties; properties purchased with the intent to let, lease, or rent (in other words, produce income).

Commodity swap A type of swap involving the exchange of payments over multiple dates as determined by specified reference prices or indexes relating to commodities.

Common size statements Financial statements in which all elements (accounts) are stated as a percentage of a key figure, such as revenue for an income statement or total assets for a balance sheet.

Company fundamental factors Factors related to the company's internal performance, such as factors relating to earnings growth, earnings variability, earnings momentum, and financial leverage.

Company share-related factors Valuation measures and other factors related to share price or the trading characteristics of the shares, such as earnings yield, dividend yield, and book-to-market value.

Comparables Assets used as benchmarks when applying the method of comparables to value an asset. Also called *comps*, *guideline assets*, or *guideline companies*.

Competition laws A law that promotes or maintains market competition by regulating anti-competitive conduct. Known as "antitrust law" in the United States, "anti-monopoly law" in China and Russia, and often referred to as "trade practices law" in the United Kingdom and Australia.

Compiled financial statements Financial statements that are not accompanied by an auditor's opinion letter.

Complexity A term referring to the number of features, parameters, or branches in a model and to whether the model is linear or non-linear (non-linear is more complex).

Composite variable A variable that combines two or more variables that are statistically strongly related to each other.

Comprehensive income All changes in equity other than contributions by, and distributions to, owners; income under clean surplus accounting; includes all changes in equity during a period except those resulting from investments by owners and distributions to owners. Comprehensive income equals net income plus other comprehensive income.

Comps Assets used as benchmarks when applying the method of comparables to value an asset.

Concentrated ownership Ownership structure consisting of an individual shareholder or a group (controlling shareholders) with the ability to exercise control over the corporation.

Conditional convergence The idea that convergence of per capita income is conditional on the countries having the same savings rate, population growth rate, and production function.

Conditional heteroskedasticity Heteroskedasticity in the error variance that is correlated with the values of the independent variable(s) in the regression.

Conditional VaR (CVaR) The weighted average of all loss outcomes in the statistical (i.e., return) distribution that exceed the VaR loss. Thus, CVaR is a more comprehensive measure of tail loss than VaR is. Sometimes referred to as the *expected tail loss* or *expected shortfall*.

Confusion matrix A grid used for error analysis in classification problems, it presents values for four evaluation metrics including true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Conglomerate discount The discount possibly applied by the market to the stock of a company operating in multiple, unrelated businesses.

Conglomerate merger A merger involving companies that are in unrelated businesses.

Consolidation The combining of the results of operations of subsidiaries with the parent company to present financial statements as if they were a single economic unit. The assets, liabilities, revenues, and expenses of the subsidiaries are combined with those of the parent company, eliminating intercompany transactions.

Constant dividend payout ratio policy A policy in which a constant percentage of net income is paid out in dividends.

Constant returns to scale The condition that if all inputs into the production process are increased by a given percentage, then output rises by that same percentage.

Contango A condition in futures markets in which the spot price is lower than the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is lower than the longer-term futures contract price.

Contingent consideration Potential future payments to the seller that are contingent on the achievement of certain agreed-on occurrences.

Continuing earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *persistent earnings*, or *underlying earnings*.

Continuing residual income Residual income after the forecast horizon.

Continuing value The analyst's estimate of a stock's value at a particular point in the future.

Control premium An increment or premium to value associated with a controlling ownership interest in a company.

Convergence The property by which as expiration approaches, the price of a newly created forward or futures contract will approach the price of a spot transaction. At expiration, a forward or futures contract is equivalent to a spot transaction in the underlying.

Conversion period For a convertible bond, the period during which bondholders have the right to convert their bonds into shares.

Conversion price For a convertible bond, the price per share at which the bond can be converted into shares.

Conversion rate (or ratio) For a convertible bond, the number of shares of common stock that a bondholder receives from converting the bond into shares.

Conversion value For a convertible bond, the value of the bond if it is converted at the market price of the shares. Also called *parity value*.

Convertible bond Bond with an embedded conversion option that gives bondholders the right to convert their bonds into the issuer's common stock during a pre-determined period at a pre-determined price.

Convexity A measure of how interest rate sensitivity changes with a change in interest rates.

Core earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *persistent earnings*, or *underlying earnings*.

Core real estate investment style Investing in high-quality, well-leased, core property types with low leverage (no more than 30% of asset value) in the largest markets with strong, diversified economies. It is a conservative strategy designed to avoid real estate–specific risks, including leasing, development, and speculation in favor of steady returns. Hotel properties are excluded from the core categories because of the higher cash flow volatility resulting from single-night leases and the greater importance of property operations, brand, and marketing.

Corpus A collection of text data in any form, including list, matrix, or data table forms.

Cost approach An approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities. In the context of real estate, this approach estimates the value of a property based on what it would cost to buy the land and construct a new property on the site that has the same utility or functionality as the property being appraised.

Cost of carry model A model that relates the forward price of an asset to the spot price by considering the cost of carry (also referred to as future-spot parity model).

Cost of debt The cost of debt financing to a company, such as when it issues a bond or takes out a bank loan.

Cost of equity The required rate of return on common stock.

Covariance stationary Describes a time series when its expected value and variance are constant and finite in all periods and when its covariance with itself for a fixed number of periods in the past or future is constant and finite in all periods.

Covered bonds A senior debt obligation of a financial institution that gives recourse to the originator/issuer and a predetermined underlying collateral pool.

Covered interest rate parity The relationship among the spot exchange rate, the forward exchange rate, and the interest rates in two currencies that ensures that the return on a hedged (i.e., covered) foreign risk-free investment is the same as the return on a domestic risk-free investment. Also called *interest rate parity*.

Cox-Ingersoll-Ross model A general equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is directly related to the level of interest rates.

Creation basket The list of securities (and share amounts) the authorized participant (AP) must deliver to the ETF manager in exchange for ETF shares. The creation basket is published each business day.

Creation/redemption The process in which ETF shares are created or redeemed by authorized participants transacting with the ETF issuer.

Creation units Large blocks of ETF shares transacted between the authorized participant (AP) and the ETF manager that are usually but not always equal to 50,000 shares of the ETF.

Credit correlation The correlation of credit (or default) risks of the underlying single-name CDS contained in an index CDS.

Credit curve The credit spreads for a range of maturities of a company's debt.

Credit default swap A derivative contract between two parties in which the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit derivative A derivative instrument in which the underlying is a measure of the credit quality of a borrower.

Credit event The event that triggers a payment from the credit protection seller to the credit protection buyer.

Credit protection buyer One party to a credit default swap; the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit protection seller One party to a credit default swap; the seller makes a promise to pay compensation for credit losses resulting from the default.

Credit risk The risk that the borrower will not repay principal and interest. Also called *default risk*.

Credit valuation adjustment The value of the credit risk of a bond in present value terms.

Cross-validation A technique for estimating out-of-sample error directly by determining the error in validation samples.

Current exchange rate For accounting purposes, the spot exchange rate on the balance sheet date.

Current rate method Approach to translating foreign currency financial statements for consolidation in which all assets and liabilities are translated at the current exchange rate. The current rate method is the prevalent method of translation.

Curvature One of the three factors (the other two are level and steepness) that empirically explain most of the changes in the shape of the yield curve. A shock to the curvature factor affects mid-maturity interest rates, resulting in the term structure becoming either more or less hump-shaped.

Curve trade Buying a CDS of one maturity and selling a CDS on the same reference entity with a different maturity.

Cyclical businesses Businesses with high sensitivity to business- or industry-cycle influences.

Data mining The practice of determining a model by extensive searching through a dataset for statistically significant patterns.

Data preparation (cleansing) The process of examining, identifying, and mitigating (i.e., cleansing) errors in raw data.

Data snooping The subconscious or conscious manipulation of data in a way that produces a statistically significant result (i.e., the p-value is sufficiently small or the t-statistic sufficiently large to indicate statistical significance), such as by running multiple simulations and naively accepting the best result. Also known as p-hacking.

Data wrangling (preprocessing) This task performs transformations and critical processing steps on cleansed data to make the data ready for ML model training (i.e., preprocessing), and includes dealing with outliers, extracting useful variables from existing data points, and scaling the data.

"Dead-hand" provision A poison pill provision that allows for the redemption or cancellation of a poison pill provision only by a vote of continuing directors (generally directors who were on the target company's board prior to the takeover attempt).

Debt rating An objective measure of the quality and safety of a company's debt based upon an analysis of the company's ability to pay the promised cash flows. It includes an analysis of any indentures.

Deep learning Algorithms based on deep neural networks, ones with many hidden layers (more than two), that address highly complex tasks, such as image classification, face recognition, speech recognition, and natural language processing.

Deep neural networks Neural networks with many hidden layers—at least 2 but potentially more than 20—that have proven successful across a wide range of artificial intelligence applications.

Default risk See *credit risk*.

Defined benefit pension plans Plan in which the company promises to pay a certain annual amount (defined benefit) to the employee after retirement. The company bears the investment risk of the plan assets.

Defined contribution pension plans Individual accounts to which an employee and typically the employer makes contributions, generally on a tax-advantaged basis. The amounts of contributions are defined at the outset, but the future value of the benefit is unknown. The employee bears the investment risk of the plan assets.

Definitive merger agreement A contract signed by both parties to a merger that clarifies the details of the transaction, including the terms, warranties, conditions, termination details, and the rights of all parties.

Delay costs Implicit trading costs that arise from the inability to complete desired trades immediately. Also called *slippage.*

Delta The relationship between the option price and the underlying price, which reflects the sensitivity of the price of the option to changes in the price of the underlying. Delta is a good approximation of how an option price will change for a small change in the stock.

Dendrogram A type of tree diagram used for visualizing a hierarchical cluster analysis; it highlights the hierarchical relationships among the clusters.

Dependent variable The variable whose variation about its mean is to be explained by the regression; the left-side variable in a regression equation. Also referred to as the *explained variable*.

Depository Trust and Clearinghouse Corporation A US-headquartered entity providing post-trade clearing, settlement, and information services.

Descriptive statistics The study of how data can be summarized effectively.

Diluted earnings per share (Diluted EPS)Net income, minus preferred dividends, divided by the weighted average number of common shares outstanding considering all dilutive securities (e.g., convertible debt and options); the EPS that would result if all dilutive securities were converted into common shares.

Dilution A reduction in proportional ownership interest as a result of the issuance of new shares.

Dimension reduction A set of techniques for reducing the number of features in a dataset while retaining variation across observations to preserve the information contained in that variation.

Diminishing marginal productivity When each additional unit of an input, keeping the other inputs unchanged, increases output by a smaller increment.

Direct capitalization method In the context of real estate, this method estimates the value of an income-producing property based on the level and quality of its net operating income.

Discount To reduce the value of a future payment in allowance for how far away it is in time; to calculate the present value of some future amount. Also, the amount by which an instrument is priced below its face value.

Discount factor The present value or price of a risk-free single-unit payment when discounted using the appropriate spot rate.

Discount for lack of control An amount or percentage deducted from the pro rata share of 100% of the value of an equity interest in a business to reflect the absence of some or all of the powers of control.

Discount for lack of marketability An amount of percentage deducted from the value of an ownership interest to reflect the relative absence of marketability.

Discount function Discount factors for the range of all possible maturities. The spot curve can be derived from the discount function and vice versa.

Discount rate Any rate used in finding the present value of a future cash flow.

Discounted abnormal earnings model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Discounted cash flow (DCF) analysis In the context of merger analysis, an estimate of a target company's value found by discounting the company's expected future free cash flows to the present.

Discounted cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows. In the context of real estate, this method estimates the value of an income-producing property based on discounting future projected cash flows.

Discounted cash flow model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Discriminant analysis A multivariate classification technique used to discriminate between groups, such as companies that either will or will not become bankrupt during some time frame.

Dispersed ownership Ownership structure consisting of many shareholders, none of which has the ability to individually exercise control over the corporation.

Divestiture The sale, liquidation, or spin-off of a division or subsidiary.

Dividend A distribution paid to shareholders based on the number of shares owned.

Dividend coverage ratio The ratio of net income to dividends.

Dividend discount model (DDM) A present value model of stock value that views the intrinsic value of a stock as present value of the stock's expected future dividends.

Dividend displacement of earnings The concept that dividends paid now displace earnings in all future periods.

Dividend imputation tax system A taxation system that effectively assures corporate profits distributed as dividends are taxed just once and at the shareholder's tax rate.

Dividend index point A measure of the quantity of dividends attributable to a particular index.

Dividend payout ratio The ratio of cash dividends paid to earnings for a period.

Dividend policy The strategy a company follows with regard to the amount and timing of dividend payments.

Dividend rate The annualized amount of the most recent dividend.

Dividend yield Annual dividends per share divided by share price.

Divisive clustering A top-down hierarchical clustering method that starts with all observations belonging to a single large cluster. The observations are then divided into two clusters based on some measure of distance (similarity). The algorithm then progressively partitions the intermediate clusters into smaller ones until each cluster contains only one observation.

Document frequency (DF) The number of documents (texts) that contain a particular token divided by the total number of documents. It is the simplest feature selection method and often performs well when many thousands of tokens are present.

Document term matrix (DTM) A matrix where each row belongs to a document (or text file), and each column represents a token (or term). The number of rows is equal to the number of documents (or text files) in a sample text dataset. The number of columns is equal to the number of tokens from the BOW built using all the documents in the sample dataset. The cells typically contain the counts of the number of times a token is present in each document.

Dominance An arbitrage opportunity when a financial asset with a risk-free payoff in the future must have a positive price today.

Double taxation system Corporate earnings are taxed twice when paid out as dividends. First, corporate pretax earnings are taxed regardless of whether they will be distributed as dividends or retained at the corporate level. Second, dividends are taxed again at the individual shareholder level.

Downstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary) such that the investor company records a profit on its income statement. An example is a sale of inventory by the investor company to the associate or by a parent to a subsidiary company.

Dual-class shares Shares that grant one share class superior or even sole voting rights, whereas the other share class has inferior or no voting rights.

Due diligence Investigation and analysis in support of a recommendation; the failure to exercise due diligence may sometimes result in liability according to various securities laws.

Dummy variable A type of qualitative variable that takes on a value of 1 if a particular condition is true and 0 if that condition is false.

Duration A measure of the approximate sensitivity of a security to a change in interest rates (i.e., a measure of interest rate risk).

Dutch disease A situation in which currency appreciation driven by strong export demand for resources makes other segments of the economy (particularly manufacturing) globally uncompetitive.

Earnings surprise The difference between reported EPS and expected EPS. Also referred to as *unexpected earnings.*

Earnings yield EPS divided by price; the reciprocal of the P/E.

Economic profit See *residual income.*

Economic sectors Large industry groupings.

Economic value added (EVA®) A commercial implementation of the residual income concept; the computation of EVA® is the net operating profit after taxes minus the cost of capital, where these inputs are adjusted for a number of items.

Economies of scale A situation in which average costs per unit of good or service produced fall as volume rises. In reference to mergers, the savings achieved through the consolidation of operations and elimination of duplicate resources.

Edwards–Bell–Ohlson model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Effective convexity Sensitivity of duration to changes in interest rates.

Effective duration Sensitivity of the bond's price to a 100 bps parallel shift of the benchmark yield curve, assuming no change in the bond's credit spread.

Effective spread Two times the difference between the execution price and the midpoint of the market quote at the time an order is entered.

Eigenvalue A measure that gives the proportion of total variance in the initial dataset that is explained by each eigenvector.

Eigenvector A vector that defines new mutually uncorrelated composite variables that are linear combinations of the original features.

Embedded options Contingency provisions found in a bond's indenture or offering circular representing rights that enable their holders to take advantage of interest rate movements. They can be exercised by the issuer, by the bondholder, or automatically depending on the course of interest rates.

Ensemble learning A technique of combining the predictions from a collection of models to achieve a more accurate prediction.

Ensemble method The method of combining multiple learning algorithms, as in ensemble learning.

Enterprise value Total company value (the market value of debt, common equity, and preferred equity) minus the value of cash and investments.

Enterprise value multiple A valuation multiple that relates the total market value of all sources of a company's capital (net of cash) to a measure of fundamental value for the entire company (such as a pre-interest earnings measure).

Equilibrium The condition in which supply equals demand.

Equity carve-out A form of restructuring that involves the creation of a new legal entity and the sale of equity in it to outsiders.

Equity charge The estimated cost of equity capital in money terms.

Equity REITs REITs that own, operate, and/or selectively develop income-producing real estate.

Equity swap A swap transaction in which at least one cash flow is tied to the return on an equity portfolio position, often an equity index.

Error autocorrelations The autocorrelations of the error term.

Error term The difference between an observation and its expected value, where the expected value is based on the true underlying population relation between the dependent and independent variables. Also known simply as the *error*.

ESG integration An ESG investment approach that focuses on systematic consideration of material ESG factors in asset allocation, security selection, and portfolio construction decisions for the purpose of achieving the product's stated investment objectives.

Estimated parameters With reference to a regression analysis, the estimated values of the population intercept and population slope coefficients in a regression.

***Ex ante* tracking error** A measure of the degree to which the performance of a given investment portfolio might be expected to deviate from its benchmark; also known as *relative VaR*.

***Ex ante* version of PPP** The hypothesis that expected changes in the spot exchange rate are equal to expected differences in national inflation rates. An extension of relative purchasing power parity to expected future changes in the exchange rate.

Ex-dividend Trading ex-dividend refers to shares that no longer carry the right to the next dividend payment.

Ex-dividend date The first date that a share trades without (i.e., "ex") the right to receive the declared dividend for the period.

Excess earnings method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Exchange ratio The number of shares that target stockholders are to receive in exchange for each of their shares in the target company.

Exercise date The date when employees actually exercise stock options and convert them to stock.

Exercise value The value of an option if it were exercised. Also sometimes called *intrinsic value*.

Expanded CAPM An adaptation of the CAPM that adds to the CAPM a premium for small size and company-specific risk.

Expectations approach A procedure for obtaining the value of an option derived from discounting at the risk-free rate its expected future payoff based on risk neutral probabilities.

Expected exposure The projected amount of money an investor could lose if an event of default occurs, before factoring in possible recovery.

Expected holding-period return The expected total return on an asset over a stated holding period; for stocks, the sum of the expected dividend yield and the expected price appreciation over the holding period.

Expected shortfall See *conditional VaR*.

Expected tail loss See *conditional VaR*.

Exploratory data analysis (EDA) The preliminary step in data exploration, where graphs, charts, and other visualizations (heat maps and word clouds) as well as quantitative methods (descriptive statistics and central tendency measures) are used to observe and summarize data.

Exposure to foreign exchange risk The risk of a change in value of an asset or liability denominated in a foreign currency due to a change in exchange rates.

Extendible bond Bond with an embedded option that gives the bondholder the right to keep the bond for a number of years after maturity, possibly with a different coupon.

External growth Company growth in output or sales that is achieved by buying the necessary resources externally (i.e., achieved through mergers and acquisitions).

Extra dividend *See special dividend*.

F1 score The harmonic mean of precision and recall. F1 score is a more appropriate overall performance metric (than accuracy) when there is unequal class distribution in the dataset and it is necessary to measure the equilibrium of precision and recall.

Factor A common or underlying element with which several variables are correlated.

Factor betas An asset's sensitivity to a particular factor; a measure of the response of return to each unit of increase in a factor, holding all other factors constant.

Factor portfolio See *pure factor portfolio*.

Factor price The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors.

Factor risk premium The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors. Also called *factor price*.

Factor sensitivity See *factor betas*.

Failure to pay When a borrower does not make a scheduled payment of principal or interest on any outstanding obligations after a grace period.

Fair market value The market price of an asset or liability that trades regularly.

Fair value The amount at which an asset (or liability) could be bought (or incurred) or sold (or settled) in a current transaction between willing parties, that is, other than in a forced or liquidation sale. As defined in IFRS and US GAAP, it is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date.

Feature engineering A process of creating new features by changing or transforming existing features.

Feature selection A process whereby only pertinent features from the dataset are selected for model training. Selecting fewer features decreases model complexity and training time.

Features The independent variables (X's) in a labeled dataset.

Financial contagion A situation in which financial shocks spread from their place of origin to other locales. In essence, a faltering economy infects other, healthier economies.

Financial distress Heightened uncertainty regarding a company's ability to meet its various obligations because of lower or negative earnings.

Financial transaction A purchase involving a buyer having essentially no material synergies with the target (e.g., the purchase of a private company by a company in an unrelated industry or by a private equity firm would typically be a financial transaction).

First-differencing A transformation that subtracts the value of the time series in period $t - 1$ from its value in period t.

First-order serial correlation Correlation between adjacent observations in a time series.

Fitting curve A curve which shows in- and out-of-sample error rates (E_{in} and E_{out}) on the y-axis plotted against model complexity on the x-axis.

Fixed price tender offer Offer made by a company to repurchase a specific number of shares at a fixed price that is typically at a premium to the current market price.

Fixed-rate perpetual preferred stock Non-convertible, non-callable preferred stock with a specified dividend rate that has a claim on earnings senior to the claim of common stock, and no maturity date.

Flight to quality During times of market stress, investors sell higher-risk asset classes such as stocks and commodities in favor of default-risk-free government bonds.

Flip-in pill A poison pill takeover defense that dilutes an acquirer's ownership in a target by giving other existing target company shareholders the right to buy additional target company shares at a discount.

Flip-over pill A poison pill takeover defense that gives target company shareholders the right to purchase shares of the acquirer at a significant discount to the market price, which has the effect of causing dilution to all existing acquiring company shareholders.

Float Amounts collected as premium and not yet paid out as benefits.

Floored floater Floating-rate bond with a floor provision that prevents the coupon rate from decreasing below a specified minimum rate. It protects the investor against declining interest rates.

Flotation cost Fees charged to companies by investment bankers and other costs associated with raising new capital.

Forced conversion For a convertible bond, when the issuer calls the bond and forces bondholders to convert their bonds into shares, which typically happens when the underlying share price increases above the conversion price.

Foreign currency transactions Transactions that are denominated in a currency other than a company's functional currency.

Forward curve The term structure of forward rates for loans made on a specific initiation date.

Forward dividend yield A dividend yield based on the anticipated dividend during the next 12 months.

Forward integration A merger involving the purchase of a target that is farther along the value or production chain; for example, to acquire a distributor.

Forward P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Forward price The fixed price or rate at which the transaction, scheduled to occur at the expiration of a forward contract, will take place. This price is agreed to at the initiation date of the forward contract.

Forward pricing model The model that describes the valuation of forward contracts.

Forward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving forward through the network's layers.

Forward rate An interest rate determined today for a loan that will be initiated in a future period.

Forward rate agreement An over-the-counter forward contract in which the underlying is an interest rate on a deposit. A forward rate agreement (FRA) calls for one party to make a fixed interest payment and the other to make an interest payment at a rate to be determined at contract expiration.

Forward rate model The forward pricing model expressed in terms of spot and forward interest rates.

Forward rate parity The proposition that the forward exchange rate is an unbiased predictor of the future spot exchange rate.

Forward value The monetary value of an existing forward contract.

Franking credit A tax credit received by shareholders for the taxes that a corporation paid on its distributed earnings.

Free cash flow The actual cash that would be available to the company's investors after making all investments necessary to maintain the company as an ongoing enterprise (also referred to as free cash flow to the firm); the internally generated funds that can be distributed to the company's investors (e.g., shareholders and bondholders) without impairing the value of the company.

Free cash flow hypothesis The hypothesis that higher debt levels discipline managers by forcing them to make fixed debt service payments and by reducing the company's free cash flow.

Free cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows.

Free cash flow to equity The cash flow available to a company's common shareholders after all operating expenses, interest, and principal payments have been made and necessary investments in working and fixed capital have been made.

Free cash flow to equity model A model of stock valuation that views a stock's intrinsic value as the present value of expected future free cash flows to equity.

Free cash flow to the firm The cash flow available to the company's suppliers of capital after all operating expenses (including taxes) have been paid and necessary investments in working and fixed capital have been made.

Free cash flow to the firm model A model of stock valuation that views the value of a firm as the present value of expected future free cash flows to the firm.

Frequency analysis The process of quantifying how important tokens are in a sentence and in the corpus as a whole. It helps in filtering unnecessary tokens (or features).

Friendly transaction A potential business combination that is endorsed by the managers of both companies.
Functional currency The currency of the primary economic environment in which an entity operates.
Fundamental factor models A multifactor model in which the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices.
Fundamentals Economic characteristics of a business, such as profitability, financial strength, and risk.
Funds available for distribution (FAD) See *adjusted funds from operations*.
Funds from operations (FFO) Net income (computed in accordance with generally accepted accounting principles) *plus* (1) gains and losses from sales of properties and (2) depreciation and amortization.
Futures price The price at which the parties to a futures contract agree to exchange the underlying (or cash). In commodity markets, the price agreed on to deliver or receive a defined quantity (and often quality) of a commodity at a future date.
Futures value The monetary value of an existing futures contract.
FX carry trade An investment strategy that involves taking long positions in high-yield currencies and short positions in low-yield currencies.
Gamma A measure of how sensitive an option's delta is to a change in the underlying. The change in a given instrument's delta for a given small change in the underlying's value, holding everything else constant.
Generalize When a model retains its explanatory power when predicting out-of-sample (i.e., using new data).
Generalized least squares A regression estimation technique that addresses heteroskedasticity of the error term.
Going-concern assumption The assumption that the business will maintain its business activities into the foreseeable future.
Going-concern value A business's value under a going-concern assumption.
Goodwill An intangible asset that represents the excess of the purchase price of an acquired company over the value of the net identifiable assets acquired.
Grant date The day that stock options are granted to employees.
Green bond Bonds in which the proceeds are designated by issuers to fund a specific project or portfolio of projects that have environmental or climate benefits.
Greenmail The purchase of the accumulated shares of a hostile investor by a company that is targeted for takeover by that investor, usually at a substantial premium over market price.
Greenwashing The risk that a green bond's proceeds are not actually used for a beneficial environmental or climate-related project.
Grid search A method of systematically training a model by using various combinations of hyperparameter values, cross validating each model, and determining which combination of hyperparameter values ensures the best model performance.
Gross domestic product A money measure of the goods and services produced within a country's borders over a stated period.
Gross lease A lease under which the tenant pays a gross rent to the landlord, who is responsible for all operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.
Ground truth The known outcome (i.e., target variable) of each observation in a labelled dataset.
Growth accounting equation The production function written in the form of growth rates. For the basic Cobb–Douglas production function, it states that the growth rate of output equals the rate of technological change plus α multiplied by the growth rate of capital plus $(1 - \alpha)$ multiplied by the growth rate of labor.
Growth capital expenditures Capital expenditures needed for expansion.
Growth option The ability to make additional investments in a project at some future time if the financial results are strong. Also called *expansion option*.
Guideline assets Assets used as benchmarks when applying the method of comparables to value an asset.
Guideline companies Assets used as benchmarks when applying the method of comparables to value an asset.
Guideline public companies Public-company comparables for the company being valued.
Guideline public company method A variation of the market approach; establishes a value estimate based on the observed multiples from trading activity in the shares of public companies viewed as reasonably comparable to the subject private company.
Guideline transactions method A variation of the market approach; establishes a value estimate based on pricing multiples derived from the acquisition of control of entire public or private companies that were acquired.
Harmonic mean A type of weighted mean computed by averaging the reciprocals of the observations and then taking the reciprocal of that average.
Hazard rate The probability that an event will occur, given that it has not already occurred.
Hedonic index Unlike a repeat-sales index, a hedonic index does not require repeat sales of the same property. It requires only one sale. The way it controls for the fact that different properties are selling each quarter is to include variables in the regression that control for differences in the characteristics of the property, such as size, age, quality of construction, and location.
Heteroskedastic With reference to the error term of regression, having a variance that differs across observations.
Heteroskedasticity The property of having a nonconstant variance; refers to an error term with the property that its variance differs across observations.
Heteroskedasticity-consistent standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.
Hierarchical clustering An iterative unsupervised learning procedure used for building a hierarchy of clusters.
Highest and best use The concept that the best use of a vacant site is the use that would result in the highest value for the land. Presumably, the developer that could earn the highest risk-adjusted profit based on time, effort, construction and development cost, leasing, and exit value would be the one to pay the highest price for the land.
Historical exchange rates For accounting purposes, the exchange rates that existed when the assets and liabilities were initially recorded.

Historical scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Historical simulation A simulation method that uses past return data and a random number generator that picks observations from the historical series to simulate an asset's future returns.

Historical simulation method The application of historical price changes to the current portfolio.

Historical stress testing The process that tests how investment strategies would perform under some of the most negative (i.e., adverse) combinations of events and scenarios.

Ho–Lee model The first arbitrage-free term structure model. The model is calibrated to market data and uses a binomial lattice approach to generate a distribution of possible future interest rates.

Holding period return The return that an investor earns during a specified holding period; a synonym for total return.

Holdout samples Data samples that are not used to train a model.

Homoskedasticity The property of having a constant variance; refers to an error term that is constant across observations.

Horizontal merger A merger involving companies in the same line of business, usually as competitors.

Horizontal ownership Companies with mutual business interests (e.g., key customers or suppliers) that have cross-holding share arrangements with each other.

Hostile transaction An attempt to acquire a company against the wishes of the target's managers.

Human capital The accumulated knowledge and skill that workers acquire from education, training, or life experience.

Hybrid approach With respect to forecasting, an approach that combines elements of both top-down and bottom-up analyses.

Hyperparameter A parameter whose value must be set by the researcher before learning begins.

I-spreads Shortened form of "interpolated spreads" and a reference to a linearly interpolated yield.

Illiquidity discount A reduction or discount to value that reflects the lack of depth of trading or liquidity in that asset's market.

Impairment Diminishment in value as a result of carrying (book) value exceeding fair value and/or recoverable value.

Impairment of capital rule A legal restriction that dividends cannot exceed retained earnings.

Implementation shortfall The difference between the money return (or value) on a notional or paper portfolio and the actual portfolio return (or value).

Implied volatility The standard deviation that causes an option pricing model to give the current option price.

In-sample forecast errors The residuals from a fitted time-series model within the sample period used to fit the model.

iNAVs "Indicated" net asset values are intraday "fair value" estimates of an ETF share based on its creation basket.

Income approach A valuation approach that values an asset as the present discounted value of the income expected from it. In the context of real estate, this approach estimates the value of a property based on an expected rate of return. The estimated value is the present value of the expected future income from the property, including proceeds from resale at the end of a typical investment holding period.

Incremental VaR (IVaR) A measure of the incremental effect of an asset on the VaR of a portfolio by measuring the difference between the portfolio's VaR while including a specified asset and the portfolio's VaR with that asset eliminated.

Indenture A written contract between a lender and borrower that specifies the terms of the loan, such as interest rate, interest payment schedule, or maturity.

Independent board directors Directors with no material relationship with the company with regard to employment, ownership, or remuneration.

Independent regulators Regulators recognized and granted authority by a government body or agency. They are not government agencies per se and typically do not rely on government funding.

Independent variable A variable used to explain the dependent variable in a regression; a right-side variable in a regression equation. Also referred to as the *explanatory variable*.

Index CDS A type of credit default swap that involves a combination of borrowers.

Indicator variable A variable that takes on only one of two values, 0 or 1, based on a condition. In simple linear regression, the slope is the difference in the dependent variable for the two conditions. Also referred to as a *dummy variable*.

Industry structure An industry's underlying economic and technical characteristics.

Information gain A metric which quantifies the amount of information that the feature holds about the response. Information gain can be regarded as a form of non-linear correlation between Y and X.

Information ratio (IR) Mean active return divided by active risk; or alpha divided by the standard deviation of diversifiable risk.

Informational frictions Forces that restrict availability, quality, and/or flow of information and its use.

Inside ask See *best ask*.

Inside bid See *best bid*.

Inside spread The spread between the best bid price and the best ask price. Also called the *market bid-ask spread*, *inside bid-ask spread*, or *market spread*.

Insiders Corporate managers and board directors who are also shareholders of a company.

Inter-temporal rate of substitution The ratio of the marginal utility of consumption *s* periods in the future (the numerator) to the marginal utility of consumption today (the denominator).

Intercept The expected value of the dependent variable when the independent variable in a simple linear regression is equal to zero.

Interest rate risk The risk that interest rates will rise and therefore the market value of current portfolio holdings will fall so that their current yields to maturity then match comparable instruments in the marketplace.

Interlocking directorates Corporate structure in which individuals serve on the board of directors of multiple corporations.

Internal rate of return Abbreviated as IRR. Rate of return that discounts future cash flows from an investment to the exact amount of the investment; the discount rate that makes the present value of an investment's costs (outflows) equal to the present value of the investment's benefits (inflows).

International Fisher effect The proposition that nominal interest rate differentials across currencies are determined by expected inflation differentials.

Intrinsic value The value of an asset given a hypothetically complete understanding of the asset's investment characteristics; the value obtained if an option is exercised based on current conditions. The difference between the spot exchange rate and the strike price of a currency.

Inverse price ratio The reciprocal of a price multiple—for example, in the case of a P/E, the "earnings yield" E/P (where P is share price and E is earnings per share).

Investment value The value to a specific buyer, taking account of potential synergies based on the investor's requirements and expectations.

ISDA Master Agreement A standard or "master" agreement published by the International Swaps and Derivatives Association. The master agreement establishes the terms for each party involved in the transaction.

Judicial law Interpretations of courts.

Justified (fundamental) P/E The price-to-earnings ratio that is fair, warranted, or justified on the basis of forecasted fundamentals.

Justified price multiple The estimated fair value of the price multiple, usually based on forecasted fundamentals or comparables.

***K*-fold cross-validation** A technique in which data (excluding test sample and fresh data) are shuffled randomly and then are divided into k equal sub-samples, with $k – 1$ samples used as training samples and one sample, the kth, used as a validation sample.

***K*-means** A clustering algorithm that repeatedly partitions observations into a fixed number, k, of non-overlapping clusters.

***K*-nearest neighbor** A supervised learning technique that classifies a new observation by finding similarities ("nearness") between this new observation and the existing data.

Kalotay–Williams–Fabozzi (KWF) model An arbitrage-free term structure model that describes the dynamics of the log of the short rate and assumes constant drift, no mean reversion, and constant volatility.

Key rate durations Sensitivity of a bond's price to changes in specific maturities on the benchmark yield curve. Also called *partial durations*.

***k*th-order autocorrelation** The correlation between observations in a time series separated by k periods.

Labeled dataset A dataset that contains matched sets of observed inputs or features (X's) and the associated output or target (Y).

Labor force Everyone of working age (ages 16 to 64) who either is employed or is available for work but not working.

Labor force participation rate The percentage of the working age population that is in the labor force.

Labor productivity The quantity of real GDP produced by an hour of labor. More generally, output per unit of labor input.

Labor productivity growth accounting equation States that potential GDP growth equals the growth rate of the labor input plus the growth rate of labor productivity.

Lack of marketability discount An extra return to investors to compensate for lack of a public market or lack of marketability.

LASSO Least absolute shrinkage and selection operator is a type of penalized regression which involves minimizing the sum of the absolute values of the regression coefficients. LASSO can also be used for regularization in neural networks.

Latency The elapsed time between the occurrence of an event and a subsequent action that depends on that event.

Law of one price A principle that states that if two investments have the same or equivalent future cash flows regardless of what will happen in the future, then these two investments should have the same current price.

Leading dividend yield Forecasted dividends per share over the next year divided by current stock price.

Leading P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Learning curve A curve that plots the accuracy rate (= 1 – error rate) in the validation or test samples (i.e., out-of-sample) against the amount of data in the training sample, which is thus useful for describing under- and overfitting as a function of bias and variance errors.

Learning rate A parameter that affects the magnitude of adjustments in the weights in a neural network.

Level One of the three factors (the other two are steepness and curvature) that empirically explain most yield curve shape changes. A shock to the level factor changes the yield for all maturities by an almost identical amount.

Leveraged buyout A transaction whereby the target company management team converts the target to a privately held company by using heavy borrowing to finance the purchase of the target company's outstanding shares.

Leveraged recapitalization A post-offer takeover defense mechanism that involves the assumption of a large amount of debt that is then used to finance share repurchases. The effect is to dramatically change the company's capital structure while attempting to deliver a value to target shareholders in excess of a hostile bid.

Libor–OIS spread The difference between Libor and the overnight indexed swap rate.

Limit order book The book or list of limit orders to buy and sell that pertains to a security.

Lin-log model A regression model in which the independent variable is in logarithmic form.

Linear classifier A binary classifier that makes its classification decision based on a linear combination of the features of each data point.

Linear regression Regression that models the straight-line relationship between the dependent and independent variables. Also known as *least squares regression* and *ordinary least squares regression*.

Linear trend A trend in which the dependent variable changes at a constant rate with time.

Liquidating dividend A dividend that is a return of capital rather than a distribution from earnings or retained earnings.

Liquidation To sell the assets of a company, division, or subsidiary piecemeal, typically because of bankruptcy; the form of bankruptcy that allows for the orderly satisfaction of creditors' claims after which the company ceases to exist.

Liquidation value The value of a company if the company were dissolved and its assets sold individually.

Liquidity preference theory A term structure theory that asserts liquidity premiums exist to compensate investors for the added interest rate risk they face when lending long term.

Liquidity premium The premium or incrementally higher yield that investors demand for lending long term.

Local currency The currency of the country where a company is located.

Local expectations theory A term structure theory that contends the return for all bonds over short periods is the risk-free rate.

Log-lin model A regression model in which the dependent variable is in logarithmic form.

Log-linear model With reference to time-series models, a model in which the growth rate of the time series as a function of time is constant.

Log-log model A regression model in which both the dependent and independent variables are in logarithmic form. Also known as the *double-log model*.

Log-log regression model A regression that expresses the dependent and independent variables as natural logarithms.

Logistic regression (logit model) A qualitative-dependent-variable multiple regression model based on the logistic probability distribution.

Long/short credit trade A credit protection seller with respect to one entity combined with a credit protection buyer with respect to another entity.

Look-ahead bias The bias created by using information that was unknown or unavailable in the time periods over which backtesting is conducted, such as company earnings and macroeconomic indicator values.

Lookback period The time period used to gather a historical data set.

Loss given default The amount that will be lost if a default occurs.

Macroeconomic factor model A multifactor model in which the factors are surprises in macroeconomic variables that significantly explain equity returns.

Macroeconomic factors Factors related to the economy, such as the inflation rate, industrial production, or economic sector membership.

Maintenance capital expenditures Capital expenditures needed to maintain operations at the current level.

Majority shareholders Shareholders that own more than 50% of a corporation's shares.

Majority-vote classifier A classifier that assigns to a new data point the predicted label with the most votes (i.e., occurrences).

Managerialism theories Theories that posit that corporate executives are motivated to engage in mergers to maximize the size of their company rather than shareholder value (a form of agency cost).

Marginal VaR (MVaR) A measure of the effect of a small change in a position size on portfolio VaR.

Market approach Valuation approach that values an asset based on pricing multiples from sales of assets viewed as similar to the subject asset.

Market conversion premium per share For a convertible bond, the difference between the market conversion price and the underlying share price, which allows investors to identify the premium or discount payable when buying a convertible bond rather than the underlying common stock.

Market conversion premium ratio For a convertible bond, the market conversion premium per share expressed as a percentage of the current market price of the shares.

Market efficiency A finance perspective on capital markets that deals with the relationship of price to intrinsic value. The **traditional efficient markets formulation** asserts that an asset's price is the best available estimate of its intrinsic value. The **rational efficient markets formulation** asserts that investors should expect to be rewarded for the costs of information gathering and analysis by higher gross returns.

Market fragmentation Trading the same instrument in multiple venues.

Market impact The effect of the trade on transaction prices. Also called *price impact.*

Market timing Asset allocation in which the investment in the market is increased if one forecasts that the market will outperform T-bills.

Market value of invested capital The market value of debt and equity.

Mature growth rate The earnings growth rate in a company's mature phase; an earnings growth rate that can be sustained long term.

Maximum drawdown The worst cumulative loss ever sustained by an asset or portfolio. More specifically, maximum drawdown is the difference between an asset's or a portfolio's maximum cumulative return and its subsequent lowest cumulative return.

Mean reversion The tendency of a time series to fall when its level is above its mean and rise when its level is below its mean; a mean-reverting time series tends to return to its long-term mean.

Mean square error (MSE) The sum of squares error divided by the degrees of freedom, $n - k - 1$; in a simple linear regression, $n - k - 1 = n - 2$.

Mean square regression (MSR) The sum of squares regression divided by the number of independent variables k; in a simple linear regression, $k = 1$.

Merger The absorption of one company by another; two companies become one entity and one or both of the pre-merger companies ceases to exist as a separate entity.

Metadata Data that describes and gives information about other data.

Method based on forecasted fundamentals An approach to using price multiples that relates a price multiple to forecasts of fundamentals through a discounted cash flow model.

Method of comparables An approach to valuation that involves using a price multiple to evaluate whether an asset is relatively fairly valued, relatively undervalued, or relatively overvalued when compared to a benchmark value of the multiple.

Midquote price The average, or midpoint, of the prevailing bid and ask prices.

Minority interest The proportion of the ownership of a subsidiary not held by the parent (controlling) company.

Minority shareholders Shareholders that own less than 50% of a corporation's shares.

Mispricing Any departure of the market price of an asset from the asset's estimated intrinsic value.

Mixed offering A merger or acquisition that is to be paid for with cash, securities, or some combination of the two.

Model specification With reference to regression, the set of variables included in the regression and the regression equation's functional form.

Molodovsky effect The observation that P/Es tend to be high on depressed EPS at the bottom of a business cycle and tend to be low on unusually high EPS at the top of a business cycle.

Momentum indicators Valuation indicators that relate either price or a fundamental (such as earnings) to the time series of their own past values (or in some cases to their expected value).

Monetary assets and liabilities Assets and liabilities with value equal to the amount of currency contracted for, a fixed amount of currency. Examples are cash, accounts receivable, accounts payable, bonds payable, and mortgages payable. Inventory is not a monetary asset. Most liabilities are monetary.

Monetary/non-monetary method Approach to translating foreign currency financial statements for consolidation in which monetary assets and liabilities are translated at the current exchange rate. Non-monetary assets and liabilities are translated at historical exchange rates (the exchange rates that existed when the assets and liabilities were acquired).

Monetizing Unwinding a position to either capture a gain or realize a loss.

Monitoring costs Costs borne by owners to monitor the management of the company (e.g., board of director expenses).

Monte Carlo simulation A technique that uses the inverse transformation method for converting a randomly generated uniformly distributed number into a simulated value of a random variable of a desired distribution. Each key decision variable in a Monte Carlo simulation requires an assumed statistical distribution; this assumption facilitates incorporating non-normality, fat tails, and tail dependence as well as solving high-dimensionality problems.

Mortgages Loans with real estate serving as collateral for the loans.

Multicollinearity A regression assumption violation that occurs when two or more independent variables (or combinations of independent variables) are highly but not perfectly correlated with each other.

Multiple linear regression Linear regression involving two or more independent variables.

Multiple linear regression model A linear regression model with two or more independent variables.

Mutual information Measures how much information is contributed by a token to a class of texts. MI will be 0 if the token's distribution in all text classes is the same. MI approaches 1 as the token in any one class tends to occur more often in only that particular class of text.

Mutually exclusive projects Mutually exclusive projects compete directly with each other. For example, if Projects A and B are mutually exclusive, you can choose A or B, but you cannot choose both.

N-grams A representation of word sequences. The length of a sequence varies from 1 to n. When one word is used, it is a unigram; a two-word sequence is a bigram; and a 3-word sequence is a trigram; and so on.

***n*-Period moving average** The average of the current and immediately prior $n - 1$ values of a time series.

Naked credit default swap A position where the owner of the CDS does not have a position in the underlying credit.

Name entity recognition An algorithm that analyzes individual tokens and their surrounding semantics while referring to its dictionary to tag an object class to the token.

Negative serial correlation Serial correlation in which a positive error for one observation increases the chance of a negative error for another observation, and vice versa.

Net asset balance sheet exposure When assets translated at the current exchange rate are greater in amount than liabilities translated at the current exchange rate. Assets exposed to translation gains or losses exceed the exposed liabilities.

Net asset value The difference between assets and liabilities, all taken at current market values instead of accounting book values.

Net asset value per share Net asset value divided by the number of shares outstanding.

Net lease A lease under which the tenant pays a net rent to the landlord and an additional amount based on the tenant's pro rata share of the operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Net liability balance sheet exposure When liabilities translated at the current exchange rate are greater assets translated at the current exchange rate. Liabilities exposed to translation gains or losses exceed the exposed assets.

Net operating income Gross rental revenue minus operating costs but before deducting depreciation, corporate overhead, and interest expense. In the context of real estate, a measure of the income from the property after deducting operating expenses for such items as property taxes, insurance, maintenance, utilities, repairs, and insurance but before deducting any costs associated with financing and before deducting federal income taxes. It is similar to EBITDA in a financial reporting context.

Net regulatory burden The private costs of regulation less the private benefits of regulation.

Network externalities The impact that users of a good, a service, or a technology have on other users of that product; it can be positive (e.g., a critical mass of users makes a product more useful) or negative (e.g., congestion makes the product less useful).

Neural networks Highly flexible machine learning algorithms that have been successfully applied to a variety of supervised and unsupervised tasks characterized by nonlinearities and interactions among features.

No-arbitrage approach A procedure for obtaining the value of an option based on the creation of a portfolio that replicates the payoffs of the option and deriving the option value from the value of the replicating portfolio.

No-growth company A company without positive expected net present value projects.

No-growth value per share The value per share of a no-growth company, equal to the expected level amount of earnings divided by the stock's required rate of return.

Non-cash rent An amount equal to the difference between the average contractual rent over a lease term (the straight-line rent) and the cash rent actually paid during a period. This figure is one of the deductions made from FFO to calculate AFFO.

Non-convergence trap A situation in which a country remains relatively poor, or even falls further behind, because it fails to implement necessary institutional reforms and/or adopt leading technologies.

Non-monetary assets and liabilities Assets and liabilities that are not monetary assets and liabilities. Non-monetary assets include inventory, fixed assets, and intangibles, and non-monetary liabilities include deferred revenue.

Non-renewable resources Finite resources that are depleted once they are consumed; oil and coal are examples.

Non-residential properties Commercial real estate properties other than multi-family properties, farmland, and timberland.

Nonearning assets Cash and investments (specifically cash, cash equivalents, and short-term investments).

Nonstationarity With reference to a random variable, the property of having characteristics, such as mean and variance, that are not constant through time.

Normal EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normalized EPS*.

Normalized earnings The expected level of mid-cycle earnings for a company in the absence of any unusual or temporary factors that affect profitability (either positively or negatively).

Normalized EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normal EPS*.

Normalized P/E P/E based on normalized EPS data.

Notional amount The amount of protection being purchased in a CDS.

NTM P/E Next 12-month P/E: current market price divided by an estimated next 12-month EPS.

Off-the-run A series of securities or indexes that were issued/created prior to the most recently issued/created series.

On-the-run The most recently issued/created series of securities or indexes.

One hot encoding The process by which categorical variables are converted into binary form (0 or 1) for machine reading. It is one of the most common methods for handling categorical features in text data.

One-sided durations Effective durations when interest rates go up or down, which are better at capturing the interest rate sensitivity of bonds with embedded options that do not react symmetrically to positive and negative changes in interest rates of the same magnitude.

One-tier board Board structure consisting of a single board of directors, composed of executive (internal) and non-executive (external) directors.

Opportunity cost The value that investors forgo by choosing a particular course of action; the value of something in its best alternative use.

Optimal capital structure The capital structure at which the value of the company is maximized.

Option-adjusted spread (OAS) Constant spread that, when added to all the one-period forward rates on the interest rate tree, makes the arbitrage-free value of the bond equal to its market price.

Orderly liquidation value The estimated gross amount of money that could be realized from the liquidation sale of an asset or assets, given a reasonable amount of time to find a purchaser or purchasers.

Organic growth Company growth in output or sales that is achieved by making investments internally (i.e., excludes growth achieved through mergers and acquisitions).

Other comprehensive income Changes to equity that bypass (are not reported in) the income statement; the difference between comprehensive income and net income.

Other post-employment benefits Promises by the company to pay benefits in the future, such as life insurance premiums and all or part of health care insurance for its retirees.

Out-of-sample forecast errors The differences between actual and predicted values of time series outside the sample period used to fit the model.

Overfitting When a model fits the training data too well and so does not generalize well to new data.

Overnight indexed swap (OIS) rate An interest rate swap in which the periodic floating rate of the swap equals the geometric average of a daily unsecured overnight rate (or overnight index rate).

Pairs trading An approach to trading that uses pairs of closely related stocks, buying the relatively undervalued stock and selling short the relatively overvalued stock.

Par curve A hypothetical yield curve for coupon-paying Treasury securities that assumes all securities are priced at par.

Par swap A swap in which the fixed rate is set so that no money is exchanged at contract initiation.

Parametric method A method of estimating VaR that uses the historical mean, standard deviation, and correlation of security price movements to estimate the portfolio VaR. Generally assumes a normal distribution but can be adapted to non-normal distributions with the addition of skewness and kurtosis. Sometimes called the *variance–covariance method* or the *analytical method*.

Partial regression coefficients The slope coefficients in a multiple regression. Also called *partial slope coefficients*.

Partial slope coefficients The slope coefficients in a multiple regression. Also called *partial regression coefficients*.

Parts of speech An algorithm that uses language structure and dictionaries to tag every token in the text with a corresponding part of speech (i.e., noun, verb, adjective, proper noun, etc.).

Payout amount The loss given default times the notional.

Payout policy The principles by which a company distributes cash to common shareholders by means of cash dividends and/or share repurchases.

Payouts Cash dividends and the value of shares repurchased in any given year.

Pecking order theory The theory that managers consider how their actions might be interpreted by outsiders and thus order their preferences for various forms of corporate financing. Forms of financing that are least visible to outsiders (e.g., internally generated funds) are most preferable to managers and those that are most visible (e.g., equity) are least preferable.

PEG ratio The P/E-to-growth ratio, calculated as the stock's P/E divided by the expected earnings growth rate.

Penalized regression A regression that includes a constraint such that the regression coefficients are chosen to minimize the sum of squared residuals *plus* a penalty term that increases in size with the number of included features.

Pension obligation The present value of future benefits earned by employees for service provided to date.

Perfect capital markets Markets in which, by assumption, there are no taxes, transaction costs, or bankruptcy costs and in which all investors have equal ("symmetric") information.

Perpetuity A perpetual annuity, or a set of never-ending level sequential cash flows, with the first cash flow occurring one period from now.

Persistent earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *continuing earnings*, or *underlying earnings*.

Pet projects Projects in which influential managers want the corporation to invest. Often, unfortunately, pet projects are selected without undergoing normal capital budgeting analysis.

Physical settlement Involves actual delivery of the debt instrument in exchange for a payment by the credit protection seller of the notional amount of the contract.

Point-in-time data Data consisting of the exact information available to market participants as of a given point in time. Point-in-time data is used to address look-ahead bias.

Poison pill A pre-offer takeover defense mechanism that makes it prohibitively costly for an acquirer to take control of a target without the prior approval of the target's board of directors.

Poison puts A pre-offer takeover defense mechanism that gives target company bondholders the right to sell their bonds back to the target at a pre-specified redemption price, typically at or above par value; this defense increases the need for cash and raises the cost of the acquisition.

Portfolio balance approach A theory of exchange rate determination that emphasizes the portfolio investment decisions of global investors and the requirement that global investors willingly hold all outstanding securities denominated in each currency at prevailing prices and exchange rates.

Positive serial correlation Serial correlation in which a positive error for one observation increases the chance of a positive error for another observation; a negative error for one observation increases the chance of a negative error for another observation.

Potential GDP The maximum amount of output an economy can sustainably produce without inducing an increase in the inflation rate. The output level that corresponds to full employment with consistent wage and price expectations.

Precision In error analysis for classification problems it is ratio of correctly predicted positive classes to all predicted positive classes. Precision is useful in situations where the cost of false positives (FP), or Type I error, is high.

Preferred habitat theory A term structure theory that contends that investors have maturity preferences and require yield incentives before they will buy bonds outside of their preferred maturities.

Premise of value The status of a company in the sense of whether it is assumed to be a going concern or not.

Premium leg The series of payments the credit protection buyer promises to make to the credit protection seller.

Premiums Amounts paid by the purchaser of insurance products.

Present value model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Present value of growth opportunities The difference between the actual value per share and the no-growth value per share. Also called *value of growth*.

Presentation currency The currency in which financial statement amounts are presented.

Price improvement When trade execution prices are better than quoted prices.

Price momentum A valuation indicator based on past price movement.

Price multiples The ratio of a stock's market price to some measure of value per share.

Price-setting option The operational flexibility to adjust prices when demand varies from what is forecast. For example, when demand exceeds capacity, the company could benefit from the excess demand by increasing prices.

Price-to-earnings ratio (P/E) The ratio of share price to earnings per share.

Priced risk Risk for which investors demand compensation for bearing (e.g., equity risk, company-specific factors, macroeconomic factors).

Principal components analysis (PCA) An unsupervised ML technique used to transform highly correlated features of data into a few main, uncorrelated composite variables.

Principle of no arbitrage In well-functioning markets, prices will adjust until there are no arbitrage opportunities.

Prior transaction method A variation of the market approach; considers actual transactions in the stock of the subject private company.

Private market value The value derived using a sum-of-the-parts valuation.

Probability of default The probability that a bond issuer will not meet its contractual obligations on schedule.

Probability of survival The probability that a bond issuer will meet its contractual obligations on schedule.

Procedural law The body of law that focuses on the protection and enforcement of the substantive laws.

Production-flexibility option The operational flexibility to alter production when demand varies from forecast. For example, if demand is strong, a company may profit from employees working overtime or from adding additional shifts.

Project sequencing To defer the decision to invest in a future project until the outcome of some or all of a current project is known. Projects are sequenced through time, so that investing in a project creates the option to invest in future projects.

Projection error The vertical (perpendicular) distance between a data point and a given principal component.

Prospective P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Protection leg The contingent payment that the credit protection seller may have to make to the credit protection buyer.

Protection period Period during which a bond's issuer cannot call the bond.

Provision for loan losses An income statement expense account that increases the amount of the allowance for loan losses.

Proxy fight An attempt to take control of a company through a shareholder vote.

Proxy statement A public document that provides the material facts concerning matters on which shareholders will vote.

Prudential supervision Regulation and monitoring of the safety and soundness of financial institutions to promote financial stability, reduce system-wide risks, and protect customers of financial institutions.

Pruning A regularization technique used in CART to reduce the size of the classification or regression tree—by pruning, or removing, sections of the tree that provide little classifying power.

Purchasing power gain A gain in value caused by changes in price levels. Monetary liabilities experience purchasing power gains during periods of inflation.

Purchasing power loss A loss in value caused by changes in price levels. Monetary assets experience purchasing power loss during periods of inflation.

Purchasing power parity (PPP) The idea that exchange rates move to equalize the purchasing power of different currencies.

Pure expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *unbiased expectations theory*.

Pure factor portfolio A portfolio with sensitivity of 1 to the factor in question and a sensitivity of 0 to all other factors.

Putable bond Bond that includes an embedded put option, which gives the bondholder the right to put back the bonds to the issuer prior to maturity, typically when interest rates have risen and higher-yielding bonds are available.

Qualitative dependent variables Dummy variables used as dependent variables rather than as independent variables.

Quality of earnings analysis The investigation of issues relating to the accuracy of reported accounting results as reflections of economic performance. Quality of earnings analysis is broadly understood to include not only earnings management but also balance sheet management.

Random forest classifier A collection of a large number of decision trees trained via a bagging method.

Random walk A time series in which the value of the series in one period is the value of the series in the previous period plus an unpredictable random error.

Rational efficient markets formulation See *market efficiency.*

Readme files Text files provided with raw data that contain information related to a data file. They are useful for understanding the data and how they can be interpreted correctly.

Real estate investment trusts Tax-advantaged entities (companies or trusts) that own, operate, and—to a limited extent—develop income-producing real estate property.

Real estate operating companies Regular taxable real estate ownership companies that operate in the real estate industry in countries that do not have a tax-advantaged REIT regime in place or that are engage in real estate activities of a kind and to an extent that do not fit in their country's REIT framework.

Real interest rate parity The proposition that real interest rates will converge to the same level across different markets.

Real options Options that relate to investment decisions such as the option to time the start of a project, the option to adjust its scale, or the option to abandon a project that has begun.

Rebalance return A return from rebalancing the component weights of an index.

Recall Also known as *sensitivity*, in error analysis for classification problems it is the ratio of correctly predicted positive classes to all actual positive classes. Recall is useful in situations where the cost of false negatives (FN), or Type II error, is high.

Reconstitution When dealers recombine appropriate individual zero-coupon securities and reproduce an underlying coupon Treasury.

Recovery rate The percentage of the loss recovered.

Redemption basket The list of securities (and share amounts) the authorized participant (AP) receives when it redeems ETF shares back to the ETF manager. The redemption basket is published each business day.

Reference entity The borrower (debt issuer) covered by a single-name CDS.

Reference obligation A particular debt instrument issued by the borrower that is the designated instrument being covered.

Regime With reference to a time series, the underlying model generating the times series.

Regression analysis A tool for examining whether a variable is useful for explaining another variable.

Regression coefficients The intercept and slope coefficient(s) of a regression.

Regular expression (regex) A series of texts that contains characters in a particular order. Regex is used to search for patterns of interest in a given text.

Regularization A term that describes methods for reducing statistical variability in high-dimensional data estimation problems.

Regulatory arbitrage Entities identify and use some aspect of regulations that allows them to exploit differences in economic substance and regulatory interpretation or in foreign and domestic regulatory regimes to their (the entities') advantage.

Regulatory burden The costs of regulation for the regulated entity.

Regulatory capture Theory that regulation often arises to enhance the interests of the regulated.

Regulatory competition Regulators may compete to provide a regulatory environment designed to attract certain entities.

Reinforcement learning Machine learning in which a computer learns from interacting with itself or data generated by the same algorithm.

Relative-strength indicators Valuation indicators that compare a stock's performance during a period either to its own past performance or to the performance of some group of stocks.

Relative valuation models A model that specifies an asset's value relative to the value of another asset.

Relative VaR See *ex ante tracking error*.

Relative version of PPP The hypothesis that changes in (nominal) exchange rates over time are equal to national inflation rate differentials.

Renewable resources Resources that can be replenished, such as a forest.

Rental price of capital The cost per unit of time to rent a unit of capital.

Repeat sales index As the name implies, this type of index relies on repeat sales of the same property. In general, the idea supporting this type of index is that because it is the same property that sold twice, the change in value between the two sale dates indicates how market conditions have changed over time.

Replacement cost In the context of real estate, the value of a building assuming it was built today using current construction costs and standards.

Reporting unit For financial reporting under US GAAP, an operating segment or one level below an operating segment (referred to as a component).

Required rate of return The minimum rate of return required by an investor to invest in an asset, given the asset's riskiness.

Residential properties Properties that provide housing for individuals or families. Single-family properties may be owner-occupied or rental properties, whereas multi-family properties are rental properties even if the owner or manager occupies one of the units.

Residual The difference between an observation and its predicted value, where the predicted value is based on the estimated linear relation between the dependent and independent variables using sample data.

Residual autocorrelations The sample autocorrelations of the residuals.

Residual income Earnings for a given period, minus a deduction for common shareholders' opportunity cost in generating the earnings. Also called *economic profit* or *abnormal earnings*.

Residual income method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Residual income model (RIM) A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share. Also called *discounted abnormal earnings model* or *Edwards–Bell–Ohlson model*.

Residual loss Agency costs that are incurred despite adequate monitoring and bonding of management.

Restructuring Reorganizing the capital structure of a firm.

Return on capital employed Operating profit divided by capital employed (debt and equity capital).

Return on invested capital A measure of the after-tax profitability of the capital invested by the company's shareholders and debtholders.

Reverse carry arbitrage A strategy involving the short sale of the underlying and an offsetting opposite position in the derivative.

Reverse stock split A reduction in the number of shares outstanding with a corresponding increase in share price but no change to the company's underlying fundamentals.

Reverse stress testing A risk management approach in which the user identifies key risk exposures in the portfolio and subjects those exposures to extreme market movements.

Reviewed financial statements A type of non-audited financial statements; typically provide an opinion letter with representations and assurances by the reviewing accountant that are less than those in audited financial statements.

Rho The change in a given derivative instrument for a given small change in the risk-free interest rate, holding everything else constant. Rho measures the sensitivity of the option to the risk-free interest rate.

Risk budgeting The allocation of an asset owner's total risk appetite among groups or divisions (in the case of a trading organization) or among strategies and managers (in the case of an institutional or individual investor).

Risk decomposition The process of converting a set of holdings in a portfolio into a set of exposures to risk factors.

Risk factors Variables or characteristics with which individual asset returns are correlated. Sometimes referred to simply as *factors*.

Risk parity A portfolio allocation scheme that weights stocks or factors based on an equal risk contribution.

Robust standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Roll When an investor moves its investment position from an older series to the most current series.

Roll return The component of the return on a commodity futures contract attributable to rolling long futures positions forward through time. Also called *roll yield*.

Rolling down the yield curve A maturity trading strategy that involves buying bonds with a maturity longer than the intended investment horizon. Also called *riding the yield curve*.

Rolling windows A backtesting method that uses a rolling-window (or walk-forward) framework, rebalances the portfolio after each period, and then tracks performance over time. As new information arrives each period, the investment manager optimizes (revises and tunes) the model and readjusts stock positions.

Root mean squared error (RMSE) The square root of the average squared forecast error; used to compare the out-of-sample forecasting performance of forecasting models.

Sale-leaseback A situation in which a company sells the building it owns and occupies to a real estate investor and the company then signs a long-term lease with the buyer to continue to occupy the building. At the end of the lease, use of the property reverts to the landlord.

Sales comparison approach In the context of real estate, this approach estimates value based on what similar or comparable properties (comparables) transacted for in the current market.

Scaled earnings surprise Unexpected earnings divided by the standard deviation of analysts' earnings forecasts.

Scaling The process of adjusting the range of a feature by shifting and changing the scale of the data. Two of the most common ways of scaling are normalization and standardization.

Scatter plot A chart in which two variables are plotted along the axis and points on the chart represent pairs of the two variables. In regression, the dependent variable is plotted on the vertical axis and the independent variable is plotted along the horizontal axis. Also known as a scattergram and a *scatter diagram*.

Scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Scree plots A plot that shows the proportion of total variance in the data explained by each principal component.

Screening The application of a set of criteria to reduce a set of potential investments to a smaller set having certain desired characteristics.

Seasonality A characteristic of a time series in which the data experience regular and predictable periodic changes; for example, fan sales are highest during the summer months.

Secured overnight financing rate (SOFR) A daily volume-weighted index of rates on qualified cash borrowings collateralized by US Treasuries that is expected to replace Libor as a floating reference rate for swaps.

Securities offering A merger or acquisition in which target shareholders are to receive shares of the acquirer's common stock as compensation.

Security selection risk See *active specific risk*.

Segmented markets theory A term structure theory that contends yields are solely a function of the supply and demand for funds of a particular maturity.

Self-regulating organizations (SROs) Self-regulating bodies that are given recognition and authority, including enforcement power, by a government body or agency.

Self-regulatory bodies Private, non-governmental organizations that both represent and regulate their members. Some self-regulating organizations are also independent regulators.

Sell-side analysts Analysts who work at brokerages.

Sensitivity analysis A technique for exploring how a target variable (e.g., portfolio returns) and risk profiles are affected by changes in input variables (e.g., the distribution of asset or factor returns).

Sentence length The number of characters, including spaces, in a sentence.

Serially correlated With reference to regression errors, errors that are correlated across observations.

Service period For employee stock options, usually the period between the grant date and the vesting date.

Settled in arrears An arrangement in which the interest payment is made (i.e., settlement occurs) at the maturity of the underlying instrument.

Settlement In the case of a credit event, the process by which the two parties to a CDS contract satisfy their respective obligations.

Shaping risk The sensitivity of a bond's price to the changing shape of the yield curve.

Share repurchase A transaction in which a company buys back its own shares. Unlike stock dividends and stock splits, share repurchases use corporate cash.

Shareholder activism Strategies used by shareholders to attempt to compel a company to act in a desired manner.

Shareholders' equity Total assets minus total liabilities.

Shark repellents A pre-offer takeover defense mechanism involving the corporate charter (e.g., staggered boards of directors and supermajority provisions).

Simple linear regression (SLR) A regression that summarizes the relation between the dependent variable and a single independent variable.

Simulation A technique for exploring how a target variable (e.g. portfolio returns) would perform in a hypothetical environment specified by the user, rather than a historical setting.

Single-name CDS Credit default swap on one specific borrower.

Sinking fund bond A bond that requires the issuer to set aside funds over time to retire the bond issue, thus reducing credit risk.

Slope coefficient The coefficient of an independent variable that represents the average change in the dependent variable for a one-unit change in the independent variable.

Soft margin classification An adaptation in the support vector machine algorithm which adds a penalty to the objective function for observations in the training set that are misclassified.

Special dividend A dividend paid by a company that does not pay dividends on a regular schedule or a dividend that supplements regular cash dividends with an extra payment.

Spin-off A form of restructuring in which shareholders of a parent company receive a proportional number of shares in a new, separate entity; shareholders end up owning stock in two different companies where there used to be one.

Split-off A form of restructuring in which shareholders of the parent company are given shares in a newly created entity in exchange for their shares of the parent company.

Split-rate tax system In reference to corporate taxes, a split-rate system taxes earnings to be distributed as dividends at a different rate than earnings to be retained. Corporate profits distributed as dividends are taxed at a lower rate than those retained in the business.

Spot curve The term structure of spot rates for loans made today.

Spot price The current price of an asset or security. For commodities, the current price to deliver a physical commodity to a specific location or purchase and transport it away from a designated location.

Spot rate The interest rate that is determined today for a risk-free, single-unit payment at a specified future date.

Spot yield curve The term structure of spot rates for loans made today.

Stabilized NOI In the context of real estate, the expected NOI when a renovation is complete.

Stable dividend policy A policy in which regular dividends are paid that reflect long-run expected earnings. In contrast to a constant dividend payout ratio policy, a stable dividend policy does not reflect short-term volatility in earnings.

Standard error of the estimate A measure of the fit of a regression line, calculated as the square root of the mean square error. Also known as the *standard error of the regression* and the *root mean square error*.

Standard error of the forecast A measure of the uncertainty associated with a forecasted value of the dependent variable that depends on the standard error of the estimate, the variability of the independent variable, the deviation of the forecasted independent variable from the mean in the regression, and the number of observations.

Standard error of the slope coefficient The standard error of the slope, which in a simple linear regression is the ratio of the model's standard error of the estimate (s_e) to the square root of the variation of the independent variable.

Standardized beta With reference to fundamental factor models, the value of the attribute for an asset minus the average value of the attribute across all stocks, divided by the standard deviation of the attribute across all stocks.

Standardized unexpected earnings Unexpected earnings per share divided by the standard deviation of unexpected earnings per share over a specified prior time period.

Static trade-off theory of capital structure A theory pertaining to a company's optimal capital structure. The optimal level of debt is found at the point where additional debt would cause the costs of financial distress to increase by a greater amount than the benefit of the additional tax shield.

Statistical factor model A multifactor model in which statistical methods are applied to a set of historical returns to determine portfolios that best explain either historical return covariances or variances.

Statutes Laws enacted by legislative bodies.

Statutory merger A merger in which one company ceases to exist as an identifiable entity and all its assets and liabilities become part of a purchasing company.

Steady-state rate of growth The constant growth rate of output (or output per capita) that can or will be sustained indefinitely once it is reached. Key ratios, such as the capital–output ratio, are constant on the steady-state growth path.

Steepness The difference between long-term and short-term yields that constitutes one of the three factors (the other two are level and curvature) that empirically explain most of the changes in the shape of the yield curve.

Stock dividend A type of dividend in which a company distributes additional shares of its common stock to shareholders instead of cash.

Stock purchase An acquisition in which the acquirer gives the target company's shareholders some combination of cash and securities in exchange for shares of the target company's stock.

Stop-loss limit Constraint used in risk management that requires a reduction in the size of a portfolio, or its complete liquidation, when a loss of a particular size occurs in a specified period.

Straight bond An underlying option-free bond with a specified issuer, issue date, maturity date, principal amount and repayment structure, coupon rate and payment structure, and currency denomination.

Straight-line rent The average annual rent under a multi-year lease agreement that contains contractual increases in rent during the life of the lease.

Straight-line rent adjustment See *non-cash rent.*

Straight voting Voting structure in which shareholders are granted the right of one vote for each share owned.

Stranded assets Assets that are obsolete or not economically viable.

Strategic transaction A purchase involving a buyer that would benefit from certain synergies associated with owning the target firm.

Stress tests A risk management technique that assesses the portfolio's response to extreme market movements.

Stripping A dealer's ability to separate a bond's individual cash flows and trade them as zero-coupon securities.

Subsidiary merger A merger in which the company being purchased becomes a subsidiary of the purchaser.

Substantive law The body of law that focuses on the rights and responsibilities of entities and relationships among entities.

Succession event A change of corporate structure of the reference entity, such as through a merger, a divestiture, a spinoff, or any similar action, in which ultimate responsibility for the debt in question is unclear.

Sum of squares error (SSE) The sum of the squared deviations of (1) the value of the dependent variable and (2) the value of the dependent variable based on the estimated regression line. Also referred to as the *residual sum of squares.*

Sum of squares regression (SSR) The sum of the squared deviations of (1) the value of the dependent variable based on the estimated regression line and (2) the mean of the dependent variable.

Sum of squares total (SST) The sum of the squared deviations of the dependent variable from its mean; the variation of the dependent variable. Also referred to as the *total sum of squares.*

Sum-of-the-parts valuation A valuation that sums the estimated values of each of a company's businesses as if each business were an independent going concern.

Summation operator A functional part of a neural network's node that multiplies each input value received by a weight and sums the weighted values to form the total net input, which is then passed to the activation function.

Supernormal growth Above-average or abnormally high growth rate in earnings per share.

Supervised learning Machine learning where algorithms infer patterns between a set of inputs (the *X*'s) and the desired output (*Y*). The inferred pattern is then used to map a given input set into a predicted output.

Support vector machine A linear classifier that determines the hyperplane that optimally separates the observations into two sets of data points.

Survivorship bias The bias that results when data as of a given date reflects only those entities that have survived to that date. Entities can include any element of an index or list that is constituted through time: stocks, investment funds, etc. Survivorship bias is a form of look-ahead bias.

Sustainable growth rate The rate of dividend (and earnings) growth that can be sustained over time for a given level of return on equity, keeping the capital structure constant and without issuing additional common stock.

Swap curve The term structure of swap rates.

Swap rate The "price" that swap traders quote among one another. It is the rate at which the present value of all the expected floating-rate payments received over the life of the floating-rate bond equal the present value of all the expected fixed-rate payments made over the life of the fixed-rate bond.

Swap rate curve The term structure of swap rates.

Swap spread The difference between the fixed rate on an interest rate swap and the rate on a Treasury note with equivalent maturity; it reflects the general level of credit risk in the market.

Systematic risk Risk that affects the entire market or economy; it cannot be avoided and is inherent in the overall market. Systematic risk is also known as non-diversifiable or market risk.

Systemic risk The risk of failure of the financial system.

Tail risk The risk that losses in extreme events could be greater than would be expected for a portfolio of assets with a normal distribution.

Takeover A merger; the term may be applied to any transaction but is often used in reference to hostile transactions.

Takeover premium The amount by which the takeover price for each share of stock must exceed the current stock price in order to entice shareholders to relinquish control of the company to an acquirer.

Tangible book value per share Common shareholders' equity minus intangible assets reported on the balance sheet, divided by the number of shares outstanding.

Target In machine learning, the dependent variable (*Y*) in a labeled dataset; the company in a merger or acquisition that is being acquired.

Target capital structure A company's chosen proportions of debt and equity.

Target company See *target.*

Target payout ratio A strategic corporate goal representing the long-term proportion of earnings that the company intends to distribute to shareholders as dividends.

Taxable REIT subsidiaries Subsidiaries that pay income taxes on earnings from non-REIT-qualifying activities like merchant development or third-party property management.

Technical indicators Momentum indicators based on price.

TED spread A measure of perceived credit risk determined as the difference between Libor and the T-bill yield of matching maturity.

Temporal method A variation of the monetary/non-monetary translation method that requires not only monetary assets and liabilities, but also non-monetary assets and liabilities that are measured at their current value on the balance sheet date to be translated at the current exchange rate. Assets and liabilities are translated at rates consistent with the timing of their measurement value. This method is typically used when the functional currency is other than the local currency.

Tender offer A public offer whereby the acquirer invites target shareholders to submit ("tender") their shares in return for the proposed payment.

Term frequency (TF) Ratio of the number of times a given token occurs in all the texts in the dataset to the total number of tokens in the dataset.

Term premium The additional return required by lenders to invest in a bond to maturity net of the expected return from continually reinvesting at the short-term rate over that same time horizon.

Terminal price multiples The price multiple for a stock assumed to hold at a stated future time.

Terminal share price The share price at a particular point in the future.

Terminal value of the stock The analyst's estimate of a stock's value at a particular point in the future. Also called *continuing value of the stock*.

Test sample A data sample that is used to test a model's ability to predict well on new data.

Theta The change in a derivative instrument for a given small change in calendar time, holding everything else constant. Specifically, the theta calculation assumes nothing changes except calendar time. Theta also reflects the rate at which an option's time value decays.

Time series A set of observations on a variable's outcomes in different time periods.

Tobin's *q* The ratio of the market value of debt and equity to the replacement cost of total assets.

Token The equivalent of a word (or sometimes a character).

Tokenization The process of splitting a given text into separate tokens. Tokenization can be performed at the word or character level but is most commonly performed at word level.

Top-down approach With respect to forecasting, an approach that usually begins at the level of the overall economy. Forecasts are then made at more narrowly defined levels, such as sector, industry, and market for a specific product.

Total factor productivity (TFP) A multiplicative scale factor that reflects the general level of productivity or technology in the economy. Changes in total factor productivity generate proportional changes in output for any input combination.

Total invested capital The sum of market value of common equity, book value of preferred equity, and face value of debt.

Tracking error The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking risk*.

Tracking risk The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking error*.

Trailing dividend yield The reciprocal of current market price divided by the most recent annualized dividend.

Trailing P/E A stock's current market price divided by the most recent four quarters of EPS (or the most recent two semi-annual periods for companies that report interim data semi-annually). Also called *current P/E*.

Training sample A data sample that is used to train a model.

Tranche CDS A type of credit default swap that covers a combination of borrowers but only up to pre-specified levels of losses.

Transaction exposure The risk of a change in value between the transaction date and the settlement date of an asset of liability denominated in a foreign currency.

Treasury shares/stock Shares that were issued and subsequently repurchased by the company.

Trend A long-term pattern of movement in a particular direction.

Triangular arbitrage An arbitrage transaction involving three currencies that attempts to exploit inconsistencies among pairwise exchange rates.

Trimming Also called truncation, it is the process of removing extreme values and outliers from a dataset.

Triple-net leases Common leases in the United States and Canada that require each tenant to pay its share of the following three operating expenses: common area maintenance and repair expenses; property taxes; and building insurance costs. Also known as *NNN leases*.

Two-tier board Board structure consisting of a supervisory board that oversees a management board.

Unbiased expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *pure expectations theory*.

Unconditional heteroskedasticity Heteroskedasticity of the error term that is not correlated with the values of the independent variable(s) in the regression.

Uncovered interest rate parity The proposition that the expected return on an uncovered (i.e., unhedged) foreign currency (risk-free) investment should equal the return on a comparable domestic currency investment.

Underlying earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *core earnings*, or *persistent earnings*.

Unexpected earnings The difference between reported EPS and expected EPS. Also referred to as an *earnings surprise*.

Unit root A time series that is not covariance stationary is said to have a unit root.

Unsupervised learning Machine learning that does not make use of labeled data.

Upfront payment The difference between the credit spread and the standard rate paid by the protection buyer if the standard rate is insufficient to compensate the protection seller. Also called *upfront premium*.

Upfront premium See *upfront payment*.

Upstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary company) such that the associate company records a profit on its income statement. An example is a sale of inventory by the associate to the investor company or by a subsidiary to a parent company.

Validation sample A data sample that is used to validate and tune a model.

Valuation The process of determining the value of an asset or service either on the basis of variables perceived to be related to future investment returns or on the basis of comparisons with closely similar assets.

Value additivity An arbitrage opportunity when the value of the whole equals the sum of the values of the parts.

Value at risk (VaR) The minimum loss that would be expected a certain percentage of the time over a certain period of time given the assumed market conditions.

Value of growth The difference between the actual value per share and the no-growth value per share.

Variance error Describes how much a model's results change in response to new data from validation and test samples. Unstable models pick up noise and produce high variance error, causing overfitting and high out-of-sample error.

Vasicek model A partial equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is constant.

Vega The change in a given derivative instrument for a given small change in volatility, holding everything else constant. A sensitivity measure for options that reflects the effect of volatility.

Venture capital investors Private equity investors in development-stage companies.

Vertical merger A merger involving companies at different positions of the same production chain; for example, a supplier or a distributor.

Vertical ownership Ownership structure in which a company or group that has a controlling interest in two or more holding companies, which in turn have controlling interests in various operating companies.

Vested benefit obligation The actuarial present value of vested benefits.

Vesting date The date that employees can first exercise stock options.

Visibility The extent to which a company's operations are predictable with substantial confidence.

Voting caps Legal restrictions on the voting rights of large share positions.

Web spidering (scraping or crawling) programs Programs that extract raw content from a source, typically web pages.

Weighted average cost of capital (WACC) A weighted average of the after-tax required rates of return on a company's common stock, preferred stock, and long-term debt, where the weights are the fraction of each source of financing in the company's target capital structure.

Weighted harmonic mean See *harmonic mean*.

White-corrected standard errors A synonym for robust standard errors.

White knight A third party that is sought out by the target company's board to purchase the target in lieu of a hostile bidder.

White squire A third party that is sought out by the target company's board to purchase a substantial minority stake in the target—enough to block a hostile takeover without selling the entire company.

Winner's curse The tendency for the winner in certain competitive bidding situations to overpay, whether because of overestimation of intrinsic value, emotion, or information asymmetries.

Winsorization The process of replacing extreme values and outliers in a dataset with the maximum (for large value outliers) and minimum (for small value outliers) values of data points that are not outliers.

Write-down A reduction in the value of an asset as stated in the balance sheet.

Yield curve factor model A model or a description of yield curve movements that can be considered realistic when compared with historical data.

Zero A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

Zero-coupon bond A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

CFA Institute®
CFA Program

EQUITY AND FIXED INCOME

CFA® Program Curriculum
2022 • LEVEL II • VOLUME 4

WILEY

CONTENTS

How to Use the CFA Program Curriculum

Congratulations on your decision to enter the Chartered Financial Analyst (CFA®) Program. This exciting and rewarding program of study reflects your desire to become a serious investment professional. You are embarking on a program noted for its high ethical standards and the breadth of knowledge, skills, and abilities (competencies) it develops. Your commitment should be educationally and professionally rewarding.

The credential you seek is respected around the world as a mark of accomplishment and dedication. Each level of the program represents a distinct achievement in professional development. Successful completion of the program is rewarded with membership in a prestigious global community of investment professionals. CFA charterholders are dedicated to life-long learning and maintaining currency with the ever-changing dynamics of a challenging profession. CFA Program enrollment represents the first step toward a career-long commitment to professional education.

The CFA exam measures your mastery of the core knowledge, skills, and abilities required to succeed as an investment professional. These core competencies are the basis for the Candidate Body of Knowledge (CBOK™). The CBOK consists of four components:

- A broad outline that lists the major CFA Program topic areas (www.cfainstitute.org/programs/cfa/curriculum/cbok);
- Topic area weights that indicate the relative exam weightings of the top-level topic areas (www.cfainstitute.org/programs/cfa/curriculum);
- Learning outcome statements (LOS) that advise candidates about the specific knowledge, skills, and abilities they should acquire from readings covering a topic area (LOS are provided in candidate study sessions and at the beginning of each reading); and
- CFA Program curriculum that candidates receive upon exam registration.

Therefore, the key to your success on the CFA exams is studying and understanding the CBOK. The following sections provide background on the CBOK, the organization of the curriculum, features of the curriculum, and tips for designing an effective personal study program.

BACKGROUND ON THE CBOK

CFA Program is grounded in the practice of the investment profession. CFA Institute performs a continuous practice analysis with investment professionals around the world to determine the competencies that are relevant to the profession, beginning with the Global Body of Investment Knowledge (GBIK®). Regional expert panels and targeted surveys are conducted annually to verify and reinforce the continuous feedback about the GBIK. The practice analysis process ultimately defines the CBOK. The CBOK reflects the competencies that are generally accepted and applied by investment professionals. These competencies are used in practice in a generalist context and are expected to be demonstrated by a recently qualified CFA charterholder.

© 2021 CFA Institute. All rights reserved.

The CFA Institute staff—in conjunction with the Education Advisory Committee and Curriculum Level Advisors, who consist of practicing CFA charterholders—designs the CFA Program curriculum in order to deliver the CBOK to candidates. The exams, also written by CFA charterholders, are designed to allow you to demonstrate your mastery of the CBOK as set forth in the CFA Program curriculum. As you structure your personal study program, you should emphasize mastery of the CBOK and the practical application of that knowledge. For more information on the practice analysis, CBOK, and development of the CFA Program curriculum, please visit www.cfainstitute.org.

ORGANIZATION OF THE CURRICULUM

The Level II CFA Program curriculum is organized into 10 topic areas. Each topic area begins with a brief statement of the material and the depth of knowledge expected. It is then divided into one or more study sessions. These study sessions should form the basic structure of your reading and preparation. Each study session includes a statement of its structure and objective and is further divided into assigned readings. An outline illustrating the organization of these study sessions can be found at the front of each volume of the curriculum.

The readings are commissioned by CFA Institute and written by content experts, including investment professionals and university professors. Each reading includes LOS and the core material to be studied, often a combination of text, exhibits, and in-text examples and questions. End of Reading Questions (EORQs) followed by solutions help you understand and master the material. The LOS indicate what you should be able to accomplish after studying the material. The LOS, the core material, and the EORQs are dependent on each other, with the core material and EORQs providing context for understanding the scope of the LOS and enabling you to apply a principle or concept in a variety of scenarios.

The entire readings, including the EORQs, are the basis for all exam questions and are selected or developed specifically to teach the knowledge, skills, and abilities reflected in the CBOK.

You should use the LOS to guide and focus your study because each exam question is based on one or more LOS and the core material and practice problems associated with the LOS. As a candidate, you are responsible for the entirety of the required material in a study session.

We encourage you to review the information about the LOS on our website (www.cfainstitute.org/programs/cfa/curriculum/study-sessions), including the descriptions of LOS "command words" on the candidate resources page at www.cfainstitute.org.

FEATURES OF THE CURRICULUM

End of Reading Questions/Solutions *All End of Reading Questions (EORQs) as well as their solutions are part of the curriculum and are required material for the exam.* In addition to the in-text examples and questions, these EORQs help demonstrate practical applications and reinforce your understanding of the concepts presented. Some of these EORQs are adapted from past CFA exams and/or may serve as a basis for exam questions.

Glossary For your convenience, each volume includes a comprehensive Glossary. Throughout the curriculum, a **bolded** word in a reading denotes a term defined in the Glossary.

Note that the digital curriculum that is included in your exam registration fee is searchable for key words, including Glossary terms.

LOS Self-Check We have inserted checkboxes next to each LOS that you can use to track your progress in mastering the concepts in each reading.

Source Material The CFA Institute curriculum cites textbooks, journal articles, and other publications that provide additional context or information about topics covered in the readings. As a candidate, you are not responsible for familiarity with the original source materials cited in the curriculum.

Note that some readings may contain a web address or URL. The referenced sites were live at the time the reading was written or updated but may have been deactivated since then.

Some readings in the curriculum cite articles published in the *Financial Analysts Journal®*, which is the flagship publication of CFA Institute. Since its launch in 1945, the *Financial Analysts Journal* has established itself as the leading practitioner-oriented journal in the investment management community. Over the years, it has advanced the knowledge and understanding of the practice of investment management through the publication of peer-reviewed practitioner-relevant research from leading academics and practitioners. It has also featured thought-provoking opinion pieces that advance the common level of discourse within the investment management profession. Some of the most influential research in the area of investment management has appeared in the pages of the *Financial Analysts Journal*, and several Nobel laureates have contributed articles.

Candidates are not responsible for familiarity with *Financial Analysts Journal* articles that are cited in the curriculum. But, as your time and studies allow, we strongly encourage you to begin supplementing your understanding of key investment management issues by reading this, and other, CFA Institute practice-oriented publications through the Research & Analysis webpage (www.cfainstitute.org/en/research).

Errata The curriculum development process is rigorous and includes multiple rounds of reviews by content experts. Despite our efforts to produce a curriculum that is free of errors, there are times when we must make corrections. Curriculum errata are periodically updated and posted by exam level and test date online (www.cfainstitute.org/en/programs/submit-errata). If you believe you have found an error in the curriculum, you can submit your concerns through our curriculum errata reporting process found at the bottom of the Curriculum Errata webpage.

DESIGNING YOUR PERSONAL STUDY PROGRAM

Create a Schedule An orderly, systematic approach to exam preparation is critical. You should dedicate a consistent block of time every week to reading and studying. Complete all assigned readings and the associated problems and solutions in each study session. Review the LOS both before and after you study each reading to ensure that

you have mastered the applicable content and can demonstrate the knowledge, skills, and abilities described by the LOS and the assigned reading. Use the LOS self-check to track your progress and highlight areas of weakness for later review.

Successful candidates report an average of more than 300 hours preparing for each exam. Your preparation time will vary based on your prior education and experience, and you will probably spend more time on some study sessions than on others.

You should allow ample time for both in-depth study of all topic areas and additional concentration on those topic areas for which you feel the least prepared.

CFA INSTITUTE LEARNING ECOSYSTEM (LES)

As you prepare for your exam, we will email you important exam updates, testing policies, and study tips. Be sure to read these carefully.

Your exam registration fee includes access to the CFA Program Learning Ecosystem (LES). This digital learning platform provides access, even offline, to all of the readings and End of Reading Questions found in the print curriculum organized as a series of shorter online lessons with associated EORQs. This tool is your one-stop location for all study materials, including practice questions and mock exams.

The LES provides the following supplemental study tools:

Structured and Adaptive Study Plans The LES offers two ways to plan your study through the curriculum. The first is a structured plan that allows you to move through the material in the way that you feel best suits your learning. The second is an adaptive study plan based on the results of an assessment test that uses actual practice questions.

Regardless of your chosen study path, the LES tracks your level of proficiency in each topic area and presents you with a dashboard of where you stand in terms of proficiency so that you can allocate your study time efficiently.

Flashcards and Game Center The LES offers all the Glossary terms as Flashcards and tracks correct and incorrect answers. Flashcards can be filtered both by curriculum topic area and by action taken—for example, answered correctly, unanswered, and so on. These Flashcards provide a flexible way to study Glossary item definitions.

The Game Center provides several engaging ways to interact with the Flashcards in a game context. Each game tests your knowledge of the Glossary terms a in different way. Your results are scored and presented, along with a summary of candidates with high scores on the game, on your Dashboard.

Discussion Board The Discussion Board within the LES provides a way for you to interact with other candidates as you pursue your study plan. Discussions can happen at the level of individual lessons to raise questions about material in those lessons that you or other candidates can clarify or comment on. Discussions can also be posted at the level of topics or in the initial Welcome section to connect with other candidates in your area.

Practice Question Bank The LES offers access to a question bank of hundreds of practice questions that are in addition to the End of Reading Questions. These practice questions, only available on the LES, are intended to help you assess your mastery of individual topic areas as you progress through your studies. After each practice question, you will receive immediate feedback noting the correct response and indicating the relevant assigned reading so you can identify areas of weakness for further study.

Mock Exams The LES also includes access to three-hour Mock Exams that simulate the morning and afternoon sessions of the actual CFA exam. These Mock Exams are intended to be taken after you complete your study of the full curriculum and take practice questions so you can test your understanding of the curriculum and your readiness for the exam. If you take these Mock Exams within the LES, you will receive feedback afterward that notes the correct responses and indicates the relevant assigned readings so you can assess areas of weakness for further study. We recommend that you take Mock Exams during the final stages of your preparation for the actual CFA exam. For more information on the Mock Exams, please visit www.cfainstitute.org.

PREP PROVIDERS

You may choose to seek study support outside CFA Institute in the form of exam prep providers. After your CFA Program enrollment, you may receive numerous solicitations for exam prep courses and review materials. When considering a prep course, make sure the provider is committed to following the CFA Institute guidelines and high standards in its offerings.

Remember, however, that there are no shortcuts to success on the CFA exams; reading and studying the CFA Program curriculum *is* the key to success on the exam. The CFA Program exams reference only the CFA Institute assigned curriculum; no prep course or review course materials are consulted or referenced.

SUMMARY

Every question on the CFA exam is based on the content contained in the required readings and on one or more LOS. Frequently, an exam question is based on a specific example highlighted within a reading or on a specific practice problem and its solution. To make effective use of the CFA Program curriculum, please remember these key points:

1. All pages of the curriculum are required reading for the exam.
2. All questions, problems, and their solutions are part of the curriculum and are required study material for the exam. These questions are found at the end of the readings in the print versions of the curriculum. In the LES, these questions appear directly after the lesson with which they are associated. The LES provides immediate feedback on your answers and tracks your performance on these questions throughout your study.
3. We strongly encourage you to use the CFA Program Learning Ecosystem. In addition to providing access to all the curriculum material, including EORQs, in the form of shorter, focused lessons, the LES offers structured and adaptive study planning, a Discussion Board to communicate with other candidates, Flashcards, a Game Center for study activities, a test bank of practice questions, and online Mock Exams. Other supplemental study tools, such as eBook and PDF versions of the print curriculum, and additional candidate resources are available at www.cfainstitute.org.
4. Using the study planner, create a schedule and commit sufficient study time to cover the study sessions. You should also plan to review the materials, answer practice questions, and take Mock Exams.
5. Some of the concepts in the study sessions may be superseded by updated rulings and/or pronouncements issued after a reading was published. Candidates are expected to be familiar with the overall analytical framework contained in the assigned readings. Candidates are not responsible for changes that occur after the material was written.

FEEDBACK

At CFA Institute, we are committed to delivering a comprehensive and rigorous curriculum for the development of competent, ethically grounded investment professionals. We rely on candidate and investment professional comments and feedback as we work to improve the curriculum, supplemental study tools, and candidate resources.

Please send any comments or feedback to info@cfainstitute.org. You can be assured that we will review your suggestions carefully. Ongoing improvements in the curriculum will help you prepare for success on the upcoming exams and for a lifetime of learning as a serious investment professional.

Equity Valuation

STUDY SESSIONS

Study Session 10	Equity Valuation (3)

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to analyze and evaluate equity securities using appropriate valuation concepts and techniques. The candidate should also be able to estimate risk and expected return of equities in global contexts.

Companies across the world differ widely in their operating and reporting models and risk–return considerations. A privately held, early stage financial technology startup with few physical assets or cash flows will look and operate differently than a mature auto manufacturer with complex operations across the globe. Fortunately, equity valuation methods exist that, based on the fundamental inputs available, can be applied to value the business, investment, or transaction in question. In each case, determining the most appropriate method to apply requires a sound understanding of the company and its industry.

© 2021 CFA Institute. All rights reserved.

EQUITY VALUATION

STUDY SESSION 10

Equity Valuation (3)

This study session presents additional valuation methods for estimating a company's intrinsic value. The free cash flow model, which takes available cash flows for distribution as the basis for valuation, is presented as an alternative to the dividend discount model, which uses actual dividends distributed. Relative valuation, using price and enterprise value multiples and which includes the comparables and forecasted fundamentals methods, comes next. Residual income valuation, useful when dividends or cash flows are minimal or volatile, or when difficulties exist in forecasting long-term terminal values, follows. The main approaches for valuing private company equity (income, market, asset based) conclude the session.

READING ASSIGNMENTS

Reading 24	Free Cash Flow Valuation by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, and John D. Stowe, PhD, CFA
Reading 25	Market-Based Valuation: Price and Enterprise Value Multiples by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA
Reading 26	Residual Income Valuation by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA
Reading 27	Private Company Valuation by Raymond D. Rath, ASA, CFA, CEIV

© 2021 CFA Institute. All rights reserved.

READING

24

Free Cash Flow Valuation

by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA

Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA). Thomas R. Robinson, PhD, CFA, CAIA, Robinson Global Investment Management LLC, (USA). John D. Stowe, PhD, CFA, is at Ohio University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** compare the free cash flow to the firm (FCFF) and free cash flow to equity (FCFE) approaches to valuation;
☐	**b.** explain the ownership perspective implicit in the FCFE approach;
☐	**c.** explain the appropriate adjustments to net income, earnings before interest and taxes (EBIT), earnings before interest, taxes, depreciation, and amortization (EBITDA), and cash flow from operations (CFO) to calculate FCFF and FCFE;
☐	**d.** calculate FCFF and FCFE;
☐	**e.** describe approaches for forecasting FCFF and FCFE;
☐	**f.** compare the FCFE model and dividend discount models;
☐	**g.** explain how dividends, share repurchases, share issues, and changes in leverage may affect future FCFF and FCFE;
☐	**h.** evaluate the use of net income and EBITDA as proxies for cash flow in valuation;
☐	**i.** explain the single-stage (stable-growth), two-stage, and three-stage FCFF and FCFE models and justify the selection of the appropriate model given a company's characteristics;
☐	**j.** estimate a company's value using the appropriate free cash flow model(s);
☐	**k.** explain the use of sensitivity analysis in FCFF and FCFE valuations;
☐	**l.** describe approaches for calculating the terminal value in a multistage valuation model; and
☐	**m.** evaluate whether a stock is overvalued, fairly valued, or undervalued based on a free cash flow valuation model.

© 2020 CFA Institute. All rights reserved.

1 THE INTRODUCTION TO FREE CASH FLOWS AND FCFF AND FCFE VALUATION APPROACHES

a compare the free cash flow to the firm (FCFF) and free cash flow to equity (FCFE) approaches to valuation;

b explain the ownership perspective implicit in the FCFE approach;

Discounted cash flow (DCF) valuation views the intrinsic value of a security as the present value of its expected future cash flows. When applied to dividends, the DCF model is the discounted dividend approach or dividend discount model (DDM). Our coverage extends DCF analysis to value a company and its equity securities by valuing free cash flow to the firm (FCFF) and free cash flow to equity (FCFE). Whereas dividends are the cash flows actually paid to stockholders, free cash flows are the cash flows *available* for distribution to shareholders.

Unlike dividends, FCFF and FCFE are not readily available data. Analysts need to compute these quantities from available financial information, which requires a clear understanding of free cash flows and the ability to interpret and use the information correctly. Forecasting future free cash flows is a rich and demanding exercise. The analyst's understanding of a company's financial statements, its operations, its financing, and its industry can pay real "dividends" as he or she addresses that task. Many analysts consider free cash flow models to be more useful than DDMs in practice. Free cash flows provide an economically sound basis for valuation.

A study of professional analysts substantiates the importance of free cash flow valuation (Pinto, Robinson, Stowe 2019). When valuing individual equities, 92.8% of analysts use market multiples and 78.8% use a discounted cash flow approach. When using discounted cash flow analysis, 20.5% of analysts use a residual income approach, 35.1% use a dividend discount model, and 86.9% use a discounted free cash flow model. Of those using discounted free cash flow models, FCFF models are used roughly twice as frequently as FCFE models. Analysts often use more than one method to value equities, and it is clear that free cash flow analysis is in near universal use.

Analysts like to use free cash flow as the return (either FCFF or FCFE) whenever one or more of the following conditions is present:

- The company does not pay dividends.
- The company pays dividends, but the dividends paid differ significantly from the company's capacity to pay dividends.
- Free cash flows align with profitability within a reasonable forecast period with which the analyst is comfortable.
- The investor takes a "control" perspective. With control comes discretion over the uses of free cash flow. If an investor can take control of the company (or expects another investor to do so), dividends may be changed substantially; for example, they may be set at a level approximating the company's capacity to pay dividends. Such an investor can also apply free cash flows to uses such as servicing the debt incurred in an acquisition.

Common equity can be valued directly by finding the present value of FCFE or indirectly by first using an FCFF model to estimate the value of the firm and then subtracting the value of non-common-stock capital (usually debt) to arrive at an estimate of the value of equity. The purpose of the coverage in the subsequent sections is to develop the background required to use the FCFF or FCFE approaches to value a company's equity.

In the next section, we define the concepts of free cash flow to the firm and free cash flow to equity and then present the two valuation models based on discounting of FCFF and FCFE. We also explore the constant-growth models for valuing FCFF and FCFE, which are special cases of the general models. The subsequent sections turn to the vital task of calculating and forecasting FCFF and FCFE. They also explain multistage free cash flow valuation models and present some of the issues associated with their application. Analysts usually value operating assets and non-operating assets separately and then combine them to find the total value of the firm, an approach described in the last section on this topic.

1.1 FCFF and FCFE Valuation Approaches

The purpose of this section is to provide a conceptual understanding of free cash flows and the valuation models based on them. A detailed accounting treatment of free cash flows and more-complicated valuation models follow in subsequent sections.

1.1.1 *Defining Free Cash Flow*

Free cash flow to the firm is the cash flow available to the company's suppliers of capital after all operating expenses (including taxes) have been paid and necessary investments in working capital (e.g., inventory) and fixed capital (e.g., equipment) have been made. FCFF is the cash flow from operations minus capital expenditures. A company's suppliers of capital include common stockholders, bondholders, and, sometimes, preferred stockholders. The equations analysts use to calculate FCFF depend on the accounting information available.

Free cash flow to equity is the cash flow available to the company's holders of common equity after all operating expenses, interest, and principal payments have been paid and necessary investments in working and fixed capital have been made. FCFE is the cash flow from operations minus capital expenditures minus payments to (plus receipts from) debtholders.

The way in which free cash flow is related to a company's net income, cash flow from operations, and measures such as EBITDA (earnings before interest, taxes, depreciation, and amortization) is important: The analyst must understand the relationship between a company's reported accounting data and free cash flow in order to forecast free cash flow and its expected growth. Although a company reports cash flow from operations (CFO) on the statement of cash flows, CFO is *not* free cash flow. Net income and CFO data can be used, however, in determining a company's free cash flow.

The advantage of FCFF and FCFE over other cash-flow concepts is that they can be used directly in a DCF framework to value the firm or to value equity. Other cash flow– or earnings-related measures, such as CFO, net income, EBIT, and EBITDA, do not have this property because they either double-count or omit cash flows in some way. For example, EBIT and EBITDA are before-tax measures, and the cash flows available to investors (in the firm or in the equity of the firm) must be after tax. From the stockholders' perspective, EBITDA and similar measures do not account for differing capital structures (the after-tax interest expenses or preferred dividends) or for the funds that bondholders supply to finance investments in operating assets. Moreover, these measures do not account for the reinvestment of cash flows that the company makes in capital assets and working capital to maintain or maximize the long-run value of the firm.

Using free cash flow in valuation is more challenging than using dividends because in forecasting free cash flow, the analyst must integrate the cash flows from the company's operations with those from its investing and financing activities. Because FCFF is the after-tax cash flow going to all suppliers of capital to the firm, the value of the firm is estimated by discounting FCFF at the weighted average cost of capital

(WACC). An estimate of the value of equity is then found by subtracting the value of debt from the estimated value of the firm. The value of equity can also be estimated directly by discounting FCFE at the required rate of return for equity (because FCFE is the cash flow going to common stockholders, the required rate of return on equity is the appropriate risk-adjusted rate for discounting FCFE).

The two free cash flow approaches for valuing equity, FCFF and FCFE, theoretically should yield the same estimates if all inputs reflect identical assumptions. An analyst may prefer to use one approach rather than the other, however, because of the characteristics of the company being valued. For example, if the company's capital structure is relatively stable, using FCFE to value equity is more direct and simpler than using FCFF. The FCFF model is often chosen, however, in two other cases:

- *A levered company with negative FCFE.* In this case, working with FCFF to value the company's equity might be easiest. The analyst would discount FCFF to find the present value of operating assets, adding the value of excess cash ("excess" in relation to operating needs) and marketable securities and of any other significant non-operating assets to get total firm value. He or she would then subtract the market value of debt to obtain an estimate of the intrinsic value of equity.
- *A levered company with a changing capital structure.* First, if historical data are used to forecast free cash flow growth rates, FCFF growth might reflect fundamentals more clearly than does FCFE growth, which reflects fluctuating amounts of net borrowing. Second, in a forward-looking context, the required return on equity might be expected to be more sensitive to changes in financial leverage than changes in the WACC, making the use of a constant discount rate difficult to justify.

Specialized DCF approaches are also available to facilitate the equity valuation when the capital structure is expected to change. The **adjusted present value** (APV) approach is one example of such models. In the APV approach, firm value is calculated as the sum of (1) the value of the company under the assumption that debt is not used (i.e., unlevered firm value) and (2) the net present value of any effects of debt on firm value (such as any tax benefits of using debt and any costs of financial distress). In this approach, the analyst estimates unlevered company value by discounting FCFF (under the assumption of no debt) at the unlevered cost of equity (the cost of equity given that the firm does not use debt). For more info, see Luehrman (1997), who explained APV in a capital budgeting context.

In the following section, we present the general form of the FCFF valuation model and the FCFE valuation model.

1.1.2 *Present Value of Free Cash Flow*

The two distinct approaches to using free cash flow for valuation are the FCFF valuation approach and the FCFE valuation approach. The general expressions for these valuation models are similar to the expression for the general dividend discount model. In the DDM, the value of a share of stock equals the present value of forecasted dividends from Time 1 through infinity discounted at the required rate of return for equity.

1.1.2.1 Present Value of FCFF The FCFF valuation approach estimates the value of the firm as the present value of future FCFF discounted at the weighted average cost of capital:

$$\text{Firm value} = \sum_{t=1}^{\infty} \frac{\text{FCFF}_t}{(1 + \text{WACC})^t}. \quad (1)$$

Because FCFF is the cash flow available to all suppliers of capital, using WACC to discount FCFF gives the total value of all of the firm's capital. The value of equity is the value of the firm minus the market value of its debt:

$$\text{Equity value} = \text{Firm value} - \text{Market value of debt}. \tag{2}$$

Dividing the total value of equity by the number of outstanding shares gives the value per share.

The cost of capital is the required rate of return that investors should demand for a cash flow stream like that generated by the company being analyzed. WACC depends on the riskiness of these cash flows. The calculation and interpretation of WACC were discussed earlier under the topic of return concepts; that is, WACC is the weighted average of the after (corporate) tax required rates of return for debt and equity, where the weights are the proportions of the firm's total market value from each source, debt and equity. As an alternative, analysts may use the weights of debt and equity in the firm's target capital structure when those weights are known and differ from market value weights. The formula for WACC is

$$\begin{aligned}\text{WACC} &= \frac{\text{MV}(\text{Debt})}{\text{MV}(\text{Debt}) + \text{MV}(\text{Equity})} r_d (1 - \text{Tax rate}) \\ &\quad + \frac{\text{MV}(\text{Equity})}{\text{MV}(\text{Debt}) + \text{MV}(\text{Equity})} r.\end{aligned} \tag{3}$$

MV(Debt) and MV(Equity) are the current market values of debt and equity, not their book or accounting values, and the ratios of MV(Debt) and MV(Equity) to the total market value of debt plus equity define the weights in the WACC formula. The quantities $r_d(1 - \text{Tax rate})$ and r are, respectively, the after-tax cost of debt and the after-tax cost of equity (in the case of equity, one could just write "cost of equity" because net income, the income belonging to equity, is after tax). In Equation 3, the tax rate is in principle the marginal corporate income tax rate.

1.1.2.2 Present Value of FCFE The value of equity can also be found by discounting FCFE at the required rate of return on equity, r:

$$\text{Equity value} = \sum_{t=1}^{\infty} \frac{\text{FCFE}_t}{(1 + r)^t}. \tag{4}$$

Because FCFE is the cash flow remaining for equity holders after all other claims have been satisfied, discounting FCFE by r (the required rate of return on equity) gives the value of the firm's equity. Dividing the total value of equity by the number of outstanding shares gives the value per share.

1.1.3 *Single-Stage (Constant-Growth) FCFF and FCFE Models*

In the DDM approach, the Gordon (constant- or stable-growth) model makes the assumption that dividends grow at a constant rate. The assumption that free cash flows grow at a constant rate leads to a single-stage (stable-growth) FCFF or FCFE model.

1.1.3.1 Constant-Growth FCFF Valuation Model Assume that FCFF grows at a constant rate, g, such that FCFF in any period is equal to FCFF in the previous period multiplied by $(1 + g)$:

$$\text{FCFF}_t = \text{FCFF}_{t-1}(1 + g).$$

If FCFF grows at a constant rate,

$$\text{Firm value} = \frac{\text{FCFF}_1}{\text{WACC} - g} = \frac{\text{FCFF}_0(1 + g)}{\text{WACC} - g}. \tag{5}$$

Subtracting the market value of debt from the firm value gives the value of equity.

EXAMPLE 1

Using the Constant-Growth FCFF Valuation Model

Cagiati Enterprises has FCFF of 700 million Swiss francs (CHF) and FCFE of CHF620 million. Cagiati's before-tax cost of debt is 5.7%, and its required rate of return for equity is 11.8%. The company expects a target capital structure consisting of 20% debt financing and 80% equity financing. The tax rate is 33.33%, and FCFF is expected to grow forever at 5.0%. Cagiati Enterprises has debt outstanding with a market value of CHF2.2 billion and has 200 million outstanding common shares.

1 What is Cagiati's weighted average cost of capital?
2 What is the value of Cagiati's equity using the FCFF valuation approach?
3 What is the value per share using this FCFF approach?

Solution to 1:

From Equation 3, WACC is calculated as follows:

$$\text{WACC} = 0.20(5.7\%)(1 - 0.3333) + 0.80(11.8\%) = 10.2\%.$$

Solution to 2:

The firm value of Cagiati Enterprises is the present value of FCFF discounted by using WACC. For FCFF growing at a constant 5% rate, the result is

$$\text{Firm value} = \frac{\text{FCFF}_1}{\text{WACC} - g} = \frac{\text{FCFF}_0(1+g)}{\text{WACC} - g} = \frac{700(1.05)}{0.102 - 0.05}$$
$$= \frac{735}{0.052} = \text{CHF}14{,}134.6 \text{ million.}$$

The value of equity is the value of the firm minus the value of debt:

$$\text{Equity value} = 14{,}134.6 - 2{,}200 = \text{CHF}11{,}934.6 \text{ million.}$$

Solution to 3:

Dividing CHF11,934.6 million by the number of outstanding shares gives the estimated value per share, V_0:

$$V_0 = \text{CHF}11{,}934.6 \text{ million}/200 \text{ million shares}$$
$$= \text{CHF}59.67 \text{ per share.}$$

1.1.3.2 Constant-Growth FCFE Valuation Model The constant-growth FCFE valuation model assumes that FCFE grows at constant rate g. FCFE in any period is equal to FCFE in the preceding period multiplied by $(1 + g)$:

$$\text{FCFE}_t = \text{FCFE}_{t-1}(1 + g).$$

The value of equity if FCFE is growing at a constant rate is

$$\text{Equity value} = \frac{\text{FCFE}_1}{r - g} = \frac{\text{FCFE}_0(1+g)}{r - g}. \qquad (6)$$

The discount rate is r, the required rate of return on equity. Note that the growth rate of FCFF and the growth rate of FCFE need not be and frequently are not the same.

In this section, we presented the basic ideas underlying free cash flow valuation and the simplest implementation, single-stage free cash flow models. The next section examines the precise definition of free cash flow and introduces the issues involved in forecasting free cash flow.

FORECASTING FREE CASH FLOW AND COMPUTING FCFF FROM NET INCOME 2

c explain the appropriate adjustments to net income, earnings before interest and taxes (EBIT), earnings before interest, taxes, depreciation, and amortization (EBITDA), and cash flow from operations (CFO) to calculate FCFF and FCFE;

d calculate FCFF and FCFE;

e describe approaches for forecasting FCFF and FCFE;

Estimating FCFF or FCFE requires a complete understanding of the company and its financial statements. To provide a context for the estimation of FCFF and FCFE, we first discuss the calculation of free cash flows, including the relationship between free cash flow and accounting measures of income. We then describe approaches to forecasting free cash flow. For most of this section, we assume that the company has two sources of capital: debt and common stock. We then incorporate preferred stock as a third source of capital.

2.1 Computing FCFF from Net Income

FCFF is the cash flow available to the company's suppliers of capital after all operating expenses (including taxes) have been paid and operating investments have been made. The company's suppliers of capital include bondholders and common shareholders (plus, occasionally, holders of preferred stock, which we ignore until later). Keeping in mind that a noncash charge is a charge or expense that does not involve the outlay of cash, we can write the expression for FCFF as follows:

$$
\begin{aligned}
\text{FCFF} = {} & \text{Net income available to common shareholders (NI)} \\
& \text{Plus: Net noncash charges (NCC)} \\
& \text{Plus: Interest expense} \times (1 - \text{Tax rate}) \\
& \text{Less: Investment in fixed capital (FCInv)} \\
& \text{Less: Investment in working capital (WCInv).}
\end{aligned}
$$

This equation can be written more compactly as

$$\text{FCFF} = \text{NI} + \text{NCC} + \text{Int}(1 - \text{Tax rate}) - \text{FCInv} - \text{WCInv}. \quad (7)$$

Consider each component of FCFF. The starting point in Equation 7 is net income available to common shareholders—usually, but not always, the bottom line in an income statement. It represents income after depreciation, amortization, interest expense, income taxes, and the payment of dividends to preferred shareholders (but not payment of dividends to common shareholders).

To derive cash flow from net income, it is necessary to make adjustments for any items that involved decreases and increases in net income but did not involve cash inflows or outflows. These items are referred to as noncash charges (NCC). If noncash decreases in net income exceed the increases, as is usually the case, the total adjustment is positive. If noncash increases exceed noncash decreases, the total adjustment is negative. The most common noncash charge is depreciation expense.

The depreciation expense reduces net income but is not a cash outflow. Depreciation expense is thus one (the most common) noncash charge that must be added back in computing FCFF. In the case of intangible assets, there is a similar noncash charge, amortization expense, which must be added back. Other noncash charges vary from company to company and are discussed later.

After-tax interest expense must be added back to net income to arrive at FCFF. This step is required because interest expense net of the related tax savings was deducted in arriving at net income, but interest is a cash flow available to one of the company's capital providers (i.e., the company's creditors). In many countries, interest is tax deductible (reduces taxes) for the company (borrower) and taxable for the recipient (lender). As we explain later, when we discount FCFF, we use an after-tax cost of capital. For consistency, we thus compute FCFF by using the after-tax interest paid. Note that we could compute WACC on a pretax basis and compute FCFF by adding back interest paid with no tax adjustment. Whichever approach is adopted, the analyst must use mutually consistent definitions of FCFF and WACC.

Similar to the treatment of after-tax interest expense, dividends on preferred stock that are deducted in arriving at net income available to common shareholders must be added back to derive FCFF. The reason for the add-back is that preferred stock dividends are also a cash flow available to one of the company's capital providers and thus constitute part of overall FCFF.

Investments in fixed capital represent the outflows of cash to purchase the fixed capital necessary to support the company's current and future operations. These investments are capital expenditures for long-term assets, such as the property, plant, and equipment (PP&E) necessary to support the company's operations. Necessary capital expenditures may also include intangible assets, such as trademarks. In the case of a cash acquisition of another company instead of a direct acquisition of PP&E, the cash purchase amount can also be treated as a capital expenditure that reduces the company's free cash flow (note that this treatment is conservative because it reduces FCFF). In the case of large acquisitions (and all noncash acquisitions), analysts must take care in evaluating the impact on future free cash flow. If a company receives cash in disposing of any of its fixed capital, the analyst must deduct this cash in calculating investment in fixed capital. For example, suppose a company sells equipment for $100,000. This cash inflow would reduce the company's cash outflows for investments in fixed capital.

The company's statement of cash flows is an excellent source of information on capital expenditures as well as on sales of fixed capital. Analysts should be aware that some companies acquire fixed capital without using cash—for example, through an exchange for stock or debt. Such acquisitions do not appear in a company's statement of cash flows but, if material, must be disclosed in the footnotes. Although noncash exchanges do not affect historical FCFF, if the capital expenditures are necessary and may be made in cash in the future, the analyst should use this information in forecasting future FCFF.

Finally, the adjustment for net increases in working capital represents the net investment in current assets (such as accounts receivable) less current liabilities (such as accounts payable). Analysts can find this information by examining either the company's balance sheet or its statement of cash flows.

Although working capital is often defined as current assets minus current liabilities, working capital for cash flow and valuation purposes is defined to exclude cash and short-term debt (which includes notes payable and the current portion of long-term debt). When finding the net increase in working capital for the purpose of calculating free cash flow, we define working capital to exclude cash and cash equivalents as well as notes payable and the current portion of long-term debt. Cash and cash equivalents are excluded because a change in cash is what we are trying to explain. Notes payable and the current portion of long-term debt are excluded because they are liabilities with explicit interest costs that make them financing items rather than operating items.

Example 2 shows the adjustments to net income required to find FCFF.

EXAMPLE 2

Calculating FCFF from Net Income

Cane Distribution, Inc., incorporated on 31 December 2017 with initial capital infusions of $224,000 of debt and $336,000 of common stock, acts as a distributor of industrial goods. The company managers immediately invested the initial capital in fixed capital of $500,000 and working capital of $60,000. Working capital initially consisted solely of inventory. The fixed capital consisted of nondepreciable property of $50,000 and depreciable property of $450,000. The depreciable property has a 10-year useful life with no salvage value. Exhibits 1, 2, and 3 provide Cane's financial statements for the three years following incorporation. Starting with net income, calculate Cane's FCFF for each year.

Exhibit 1 Cane Distribution, Inc., Income Statement (in Thousands)

	Years Ending 31 December		
	2018	2019	2020
Earnings before interest, taxes, depreciation, and amortization (EBITDA)	$200.00	$220.00	$242.00
Depreciation expense	45.00	49.50	54.45
Operating income	155.00	170.50	187.55
Interest expense (at 7%)	15.68	17.25	18.97
Income before taxes	139.32	153.25	168.58
Income taxes (at 30%)	41.80	45.97	50.58
Net income	$97.52	$107.28	$118.00

Exhibit 2 Cane Distribution, Inc., Balance Sheet (in Thousands)

	Years Ending 31 December			
	2017	2018	2019	2020
Cash	$0.00	$108.92	$228.74	$360.54
Accounts receivable	0.00	100.00	110.00	121.00
Inventory	60.00	66.00	72.60	79.86
Current assets	60.00	274.92	411.34	561.40
Fixed assets	500.00	500.00	550.00	605.00
Less: Accumulated depreciation	0.00	45.00	94.50	148.95
Total assets	$560.00	$729.92	$866.84	$1,017.45
Accounts payable	$0.00	$50.00	$55.00	$60.50
Current portion of long-term debt	0.00	0.00	0.00	0.00
Current liabilities	0.00	50.00	55.00	60.50

(continued)

Exhibit 2 (Continued)

	Years Ending 31 December			
	2017	2018	2019	2020
Long-term debt	224.00	246.40	271.04	298.14
Common stock	336.00	336.00	336.00	336.00
Retained earnings	0.00	97.52	204.80	322.80
Total liabilities and equity	$560.00	$729.92	$866.84	$1,017.45

Exhibit 3 Cane Distribution, Inc., Working Capital (in Thousands)

	Years Ending 31 December			
	2017	2018	2019	2020
Current assets excluding cash				
Accounts receivable	$0.00	$100.00	$110.00	$121.00
Inventory	60.00	66.00	72.60	79.86
Total current assets excluding cash	60.00	166.00	182.60	200.86
Current liabilities excluding short-term debt				
Accounts payable	0.00	50.00	55.00	60.50
Working capital	$60.00	$116.00	$127.60	$140.36
Increase in working capital		$56.00	$11.60	$12.76

Solution:

Following the logic in Equation 7, we calculate FCFF from net income as follows: We add noncash charges (here, depreciation) and after-tax interest expense to net income and then subtract the investment in fixed capital and the investment in working capital. The format for presenting the solution follows the convention that parentheses around a number indicate subtraction. The calculation follows (in thousands):

	Years Ending 31 December		
	2018	2019	2020
Net income	$97.52	$107.28	$118.00
Noncash charges – Depreciation	45.00	49.50	54.45
Interest expense × (1 – Tax rate)	10.98	12.08	13.28
Investment in fixed capital	(0.00)	(50.00)	(55.00)
Investment in working capital	(56.00)	(11.60)	(12.76)
Free cash flow to the firm	$97.50	$107.26	$117.97

COMPUTING FCFF FROM THE STATEMENT OF CASH FLOWS 3

c explain the appropriate adjustments to net income, earnings before interest and taxes (EBIT), earnings before interest, taxes, depreciation, and amortization (EBITDA), and cash flow from operations (CFO) to calculate FCFF and FCFE;

d calculate FCFF and FCFE;

e describe approaches for forecasting FCFF and FCFE;

FCFF is the cash flow that is available to all providers of capital (debt and equity). Analysts frequently use cash flow from operations, taken from the statement of cash flows, as a starting point to compute free cash flow because CFO incorporates adjustments for noncash expenses (such as depreciation and amortization) as well as for net investments in working capital.

In most cases, companies include interest paid as part of operating cash flow. Under US generally accepted accounting principles (GAAP), companies must include interest paid in operating cash flow. Under International Financial Reporting Standards (IFRS), companies may include interest paid in either financing or operating. According to Gordon, Henry, Jorgensen, and Linthicum (2017), most IFRS-reporting European firms choose to classify interest paid within the operating cash flow section of the statement of cash flows. This will be discussed later. Assuming that interest paid is included in operating cash flow, FCFF can be estimated as follows:

$$\begin{aligned}\text{Free cash flow to the firm} &= \text{Cash flow from operations}\\ &\quad \text{Plus: Interest expense} \times (1 - \text{Tax rate})\\ &\quad \text{Less: Investment in fixed capital,}\end{aligned}$$

or

$$\text{FCFF} = \text{CFO} + \text{Int}(1 - \text{Tax rate}) - \text{FCInv}. \quad (8)$$

To reiterate, as with the calculation shown as Equation 7, the after-tax interest expense is added back because it was previously taken out of net income but must be included in FCFF because it is a component of the total cash flows available to all suppliers of the firm's capital. In comparison with Equation 7, neither depreciation nor the investment in working capital appears in Equation 8 because both are already included in CFO. Example 3 illustrates the use of CFO to calculate FCFF. In this example, the operating section of the statement of cash flows begins with net income and presents each adjustment required to derive operating cash flow. This presentation, known as the "indirect" method because it derives operating cash flows indirectly from net income via adjustments, is the most common presentation of the statement of cash flows.

EXAMPLE 3

Calculating FCFF from CFO

Use the information from the statement of cash flows given in Exhibit 4 to calculate FCFF for the three years 2018–2020. The tax rate (as given in Exhibit 1) is 30%.

Exhibit 4 Cane Distribution, Inc., Statement of Cash Flows: Indirect Method (in Thousands)

	Years Ending 31 December		
	2018	**2019**	**2020**
Cash flow from operations			
Net income	$97.52	$107.28	$118.00
Plus: Depreciation	45.00	49.50	54.45
Increase in accounts receivable	(100.00)	(10.00)	(11.00)
Increase in inventory	(6.00)	(6.60)	(7.26)
Increase in accounts payable	50.00	5.00	5.50
Cash flow from operations	86.52	145.18	159.69
Cash flow from investing activities			
Purchases of PP&E	0.00	(50.00)	(55.00)
Cash flow from financing activities			
Borrowing (repayment)	22.40	24.64	27.10
Total cash flow	108.92	119.82	131.80
Beginning cash	0.00	108.92	228.74
Ending cash	$108.92	$228.74	$360.54
Notes:			
Cash paid for interest	($15.68)	($17.25)	($18.97)
Cash paid for taxes	($41.80)	($45.98)	($50.57)

Solution:

As shown in Equation 8, FCFF equals CFO plus after-tax interest expense minus the investment in fixed capital:

	Years Ending 31 December		
	2018	**2019**	**2020**
Cash flow from operations	$86.52	$145.18	$159.69
Interest expense × (1 − Tax rate)	10.98	12.08	13.28
Investment in fixed capital	(0.00)	(50.00)	(55.00)
Free cash flow to the firm	$97.50	$107.26	$117.97

4 ADDITIONAL CONSIDERATIONS IN COMPUTING FCFF

d calculate FCFF and FCFE;

e describe approaches for forecasting FCFF and FCFE;

Whether an analyst selects net income or cash flow from operations as a starting point in calculating free cash flows, some situations warrant a closer examination. In this section, we first describe classification of certain items on the statement of cash flows that merit attention when deriving free cash flow using cash flow from operations as a starting point. We then review the common adjustments for noncash charges made in deriving cash flow from net income and highlight several areas that merit additional attention from an analyst.

4.1 Classification of Certain Items on the Statement of Cash Flow

As noted above, IFRSallow the company to classify interest paid as either an operating or financing activity. Furthermore, IFRS allow dividends paid to be classified as either an operating or financing activity. In contrast, under US GAAP, interest paid to providers of debt capital must be classified as part of cash flow from operations (as are interest income and dividend income), but payment of dividends to providers of equity capital is classified as a financing activity.

Exhibit 5 summarizes IFRS and US GAAP treatment of interest and dividends.

Exhibit 5 IFRS vs. US GAAP Treatment of Interest and Dividends

	IFRS	US GAAP
Interest received	Operating or investing	Operating
Interest paid	Operating or financing	Operating
Dividends received	Operating or investing	Operating
Dividends paid	Operating or financing	Financing

To estimate FCFF by starting with CFO, it is necessary to examine the classification of these items. For example, if the after-tax interest expense was taken out of net income and out of CFO, which is required under US GAAP and allowed under IFRS, then after-tax interest must be added back to get FCFF. However, if interest paid was not classified as an operating cash outflow (i.e., it was classified as a financing cash outflow as allowed under IFRS), then it is not necessary to add interest when operating cash flow is the starting point for calculating FCFF.

4.2 Adjustments to Derive Operating Cash Flow from Net Income

The operating cash flow section of the statement of cash flows provides detail on the adjustments made in deriving operating cash flow from net income. Exhibit 6 summarizes the common adjustments (other than changes in working capital) to derive operating cash flow from net income and indicates whether each item is added to or subtracted from net income in arriving at FCFF.

Exhibit 6 Noncash Items and FCFF

Noncash Item	Adjustment to NI to Arrive at FCFF
Depreciation expense	Added back
Amortization expense and impairment of intangibles	Added back
Restructuring charges (expense)	Added back
Restructuring charges (income resulting from reversal)	Subtracted

(continued)

Exhibit 6 (Continued)

Noncash Item	Adjustment to NI to Arrive at FCFF
Amortization of long-term bond discounts	Added back
Amortization of long-term bond premiums	Subtracted
Losses on non-operating activity	Added back
Gains on non-operating activity	Subtracted
Deferred taxes	Added back but calls for special attention

An adjustment to reported net income is required for any item that was treated as an expense in calculating net income on the income statement but did not result in an equivalent cash outflow in the reporting period. For example, both depreciation and amortization expenses reduce net income, but neither involves a cash outflow in the period. Therefore, to derive operating cash flow or FCFF from net income, it is necessary to add back these amounts to net income.

Adjustments to eliminate the amount of gains and losses are made for two reasons in general. First, such transactions are typically not operating activities (e.g., a sale of fixed assets, which is an investing activity), and thus the effects must be removed from the operating section of the statement of cash flows. Second, the amount of gain or loss reported in the income statement is not necessarily equivalent to the amount of cash involved in the transaction. For example, if a company sells a piece of equipment with a book value of €60,000 for €100,000, it reports the €40,000 gain as part of net income. The €40,000 gain, however, is not equivalent to the transaction's cash flow and, therefore, must be subtracted to derive operating cash flow from net income. Further, the €100,000 *is* a cash flow, and that amount will appear as a component of the company's cash flow for investing activity. Alternatively, if the company had sold the equipment with a book value of €60,000 for €40,000 and thus reported a loss of €20,000 as part of net income, that amount would be added back in deriving operating cash flow and FCFF.

4.3 Adjustments to Derive Operating Cash Flow from Net Income That May Merit Additional Attention from an Analyst

The item "deferred taxes" in Exhibit 6 requires special attention because deferred taxes result from differences in the timing of reporting income and expenses in the company's financial statements and the company's tax return. The income tax expense deducted in arriving at net income for financial reporting purposes is not the same as the amount of cash taxes paid. Over time, these differences between book income and taxable income should offset each other and have no impact on aggregate cash flows. Generally, if the analyst's purpose is forecasting and, therefore, identifying the persistent components of FCFF, then the analyst should not add back deferred tax changes that are expected to reverse in the near future. In some circumstances, however, a company may be able to consistently defer taxes until a much later date. If a company is growing and has the ability to indefinitely defer its tax liability, adding back deferred taxes to net income is warranted. Nevertheless, an acquirer must be aware that these taxes may be payable at some time in the future.

Similarly, companies often record expenses (e.g., restructuring charges) for financial reporting purposes that are not deductible for tax purposes or record revenues that are taxable in the current period but not yet recognized for financial reporting purposes. In these cases, taxable income exceeds financial statement income, so cash outflows for current tax payments are greater than the taxes reported in the income statement. This situation results in a deferred tax *asset* and a necessary adjustment to subtract that amount in deriving operating cash flow from net income. If, however, the deferred tax asset is expected to reverse in the near future, to avoid underestimating future cash flows, the analyst should not subtract the deferred tax asset in a cash flow forecast. If the company is expected to have these charges on a continual basis, however, a subtraction that will lower the forecast of future cash flows is warranted.

A second area that may warrant an analyst's attention to the adjustments made in derivation of operating cash flow from net income pertains to employee share-based compensation (stock options). Under both IFRS and US GAAP, companies must record in the income statement an expense for options provided to employees. The granting and expensing of options themselves do not result in a cash outflow and are thus a noncash charge; however, the granting of options has long-term cash flow implications. When the employee exercises the option, the company receives some cash related to the exercise price of the option at the strike price. This cash flow is considered a financing cash flow. Also, in some cases, a company receives a tax benefit from issuing options, which could increase operating cash flow but not net income. Both IFRS and US GAAP require that a portion of the tax effect be recorded as a financing cash flow rather than an operating cash flow in the statement of cash flows. Analysts should review the statement of cash flows and footnotes to determine the impact of options on operating cash flows. If these cash flows are not expected to persist in the future, analysts should not include them in their forecasts of cash flows. Analysts should also consider the impact of stock options on the number of shares outstanding. When computing equity value, analysts may want to use the number of shares *expected* to be outstanding (based on the exercise of employee stock options) rather than the number currently outstanding.

Finally, an analyst may benefit from a careful examination of adjustments in developing expectations about the sustainability of free cash flow. When any financial forecast is developed by using historical amounts as a baseline, it is necessary to ensure that the baseline amounts are not distorted by non-recurring items. Similarly, when a forecast of free cash flows is developed using historical amounts of FCFF or FCFE as a baseline, it is necessary to ensure that the baseline amounts are not distorted by non-recurring items. Example 4 is a historical case that is adapted to illustrate issues that an analyst may face when forecasting free cash flows. Specifically, the example illustrates that when forecasting cash flows for valuation purposes, analysts should consider the sustainability of historical working capital effects on free cash flow.

EXAMPLE 4

Sustainability of Working Capital Effects on Free Cash Flow

Duplico Holdings PLC has operations in Ireland, the United Kingdom, Continental Europe, and Morocco. The operating activities section of its statement of cash flows and a portion of the investing activities section are presented in Exhibit 7. The statement of cash flows was prepared in accordance with IFRS.

Exhibit 7 Duplico Holdings PLC Excerpt from Statement of Cash Flows (Euros in Millions)

	Year Ended 31 March		
	2022	**2021**	**2020**
Operating activities			
Profit before tax	633.0	420.9	341.0
Adjustments to reconcile profits before tax to net cash provided by operating activities			
Depreciation	309.2	277.7	235.4
Increase in inventories	(0.1)	(0.2)	(0.4)
Increase in trade receivables	(0.9)	(6.3)	(2.5)
Decrease (increase) in other current assets	34.5	(20.9)	11.6
Increase (decrease) in trade payables	30.4	(3.2)	21.3
Increase in accrued expenses	11.6	135.0	189.7
Increase (decrease) in other creditors	19.7	(10.0)	30.1
Increase (decrease) in maintenance provisions	6.6	(7.9)	30.7
Gain on disposal of property, plant, and equipment	(10.4)	—	(2.0)
Loss on impairment of available-for-sale financial asset	—	—	13.5
Decrease (increase) in interest receivable	—	1.6	(1.2)
Increase (decrease) in interest payable	1.1	2.3	(0.5)
Retirement costs	(0.1)	(0.1)	(0.1)
Share-based payments	(0.7)	3.3	4.9
Income tax paid	(13.6)	(5.9)	—
Net cash provided by operating activities	1,020.3	786.3	871.5
Investing activities			
Capital expenditure (purchase of property, plant, and equipment)	(317.6)	(897.2)	(997.8)

Analysts predict that as Duplico grows in the coming years, depreciation expense will increase substantially. Based on the information given, address the following:

1 Contrast reported depreciation expense to reported capital expenditures, and describe the implications of future growth in depreciation expense (all else being equal) for future net income and future cash from operating activities.

2 Explain the effects on free cash flow to equity of changes in 2022 in working capital accounts, such as inventory, accounts receivable, and accounts payable, and comment on the long-term sustainability of such changes.

Solution to 1:

In the 2020–22 period, the amount of depreciation expense relative to the amount of capital expenditures changed significantly. For example, in 2022, capital expenditures of €317.6 million were just slightly more than the €309.2 million depreciation expense. In 2020, capital expenditures of €997.8 million were over 4 times more than depreciation charges of €235.4 million. The rate of growth in depreciation expense will be highly dependent on future capital expenditures.

In calculating net income, depreciation is a deduction. Therefore, as depreciation expense increases in the coming years, net income will decrease. Specifically, net income will be reduced by (Depreciation expense) × (1 − Tax rate). In calculating CFO, however, depreciation is added back in full to net income. The difference between depreciation expense—the amount added back to net income to calculate CFO—and the amount by which net income is reduced by depreciation expense is (Tax rate) × (Depreciation expense), which represents a positive increment to CFO. Thus, the projected increase in depreciation expense is a negative for future net income but a positive for future CFO. (At worst, if the company operates at a loss, depreciation is neutral for CFO.)

Solution to 2:

In 2022, the increases in inventory and accounts receivable ("trade receivables") resulted in negative adjustments to net income (i.e., the changes reduced cash flow relative to net income). The adjustments are negative because increases in these accounts are a use of cash. On the current liabilities side, the increase in trade payables, accrued expenses, and "other creditors" are added back to net income and are sources of cash because such increases represent increased amounts for which cash payments have yet to be made. Because CFO is a component of FCFE, the items that had a positive (negative) effect on CFO also have a positive (negative) effect on FCFE.

Although not the case here, declining balances for assets, such as inventory, or for liabilities, such as accounts payable, are not sustainable indefinitely. In the extreme case, the balance declines to zero and no further reduction is possible. Given the growth in its net income and the expansion of PP&E evidenced by capital expenditures, Duplico appears to be growing and investors should expect its working capital requirements to grow accordingly.

COMPUTING FCFE FROM FCFF

5

b explain the ownership perspective implicit in the FCFE approach;

d calculate FCFF and FCFE;

e describe approaches for forecasting FCFF and FCFE;

FCFE is cash flow available to equity holders only. To find FCFE, therefore, we must reduce FCFF by the after-tax value of interest paid to debtholders and add net borrowing (which is debt issued less debt repaid over the period for which one is calculating free cash flow):

$$\begin{aligned} \text{Free cash flow to equity} &= \text{Free cash flow to the firm} \\ \text{Less:}&\ \text{Interest expense} \times (1 - \text{Tax rate}) \\ \text{Plus:}&\ \text{Net borrowing,} \end{aligned}$$

or

$$\text{FCFE} = \text{FCFF} - \text{Int}(1 - \text{Tax rate}) + \text{Net borrowing}. \quad (9)$$

As Equation 9 shows, FCFE is found by starting from FCFF, subtracting after-tax interest expenses, and adding net new borrowing. The analyst can also find FCFF from FCFE by making the opposite adjustments—by adding after-tax interest expenses and subtracting net borrowing: FCFF = FCFE + Int(1 − Tax rate) − Net borrowing.

Exhibit 8 uses the values for FCFF for Cane Distribution calculated in Example 3 to show the calculation of FCFE when starting with FCFF. To calculate FCFE in this manner, we subtract after-tax interest expense from FCFF and then add net borrowing (equal to new debt borrowing minus debt repayment).

Exhibit 8 Calculating FCFE from FCFF

	Years Ending 31 December		
	2018	**2019**	**2020**
Free cash flow to the firm	97.50	107.26	117.97
Interest paid × (1 − Tax rate)	(10.98)	(12.08)	(13.28)
New debt borrowing	22.40	24.64	27.10
Debt repayment	(0)	(0)	(0)
Free cash flow to equity	108.92	119.82	131.79

To reiterate, FCFE is the cash flow available to common stockholders—the cash flow remaining after all operating expenses (including taxes) have been paid, capital investments have been made, and other transactions with other suppliers of capital have been carried out. The company's other capital suppliers include creditors, such as bondholders, and preferred stockholders. The cash flows (net of taxes) that arise from transactions with creditors and preferred stockholders are deducted from FCFF to arrive at FCFE.

FCFE is the amount that the company can afford to pay out as dividends. In actuality, for various reasons companies often pay out substantially more or substantially less than FCFE, so FCFE often differs from dividends paid. One reason for this difference is that the dividend decision is a discretionary decision of the board of directors. Most corporations "manage" their dividends; they prefer to raise them gradually over time, partly because they do not want to cut dividends. Many companies raise dividends slowly even when their earnings are increasing rapidly, and companies often maintain their current dividends even when their profitability has declined. Consequently, earnings are much more volatile than dividends.

In Equations 7 and 8, we showed the calculation of FCFF starting with, respectively, net income and cash flow from operations. As Equation 9 showed, FCFE = FCFF − Int(1 − Tax rate) + Net borrowing. By subtracting after-tax interest expense and adding net borrowing to Equations 7 and 8, we have equations to calculate FCFE starting with, respectively, net income and CFO:

$$\text{FCFE} = \text{NI} + \text{NCC} - \text{FCInv} - \text{WCInv} + \text{Net borrowing}. \tag{10}$$

$$\text{FCFE} = \text{CFO} - \text{FCInv} + \text{Net borrowing}. \tag{11}$$

Example 5 illustrates how to adjust net income or CFO to find FCFF and FCFE.

EXAMPLE 5

Adjusting Net Income or CFO to Find FCFF and FCFE

The balance sheet, income statement, and statement of cash flows for the Pitts Corporation are shown in Exhibit 9. Note that the statement of cash flows follows a convention according to which the positive numbers of $400 million and $85 million for "cash *used for* investing activities" and "cash *used for* financing activities," respectively, indicate outflows and thus amounts to be *subtracted*.

Analysts will also encounter a convention in which the value "(400)" for "cash provided by (used for) investing activities" would be used to indicate a subtraction of $400.

Exhibit 9 Financial Statements for Pitts Corporation (in Millions, Except for Per-Share Data)

	Year Ended 31 December	
Balance Sheet	**2019**	**2020**
Assets		
Current assets		
Cash and equivalents	$190	$200
Accounts receivable	560	600
Inventory	410	440
Total current assets	1,160	1,240
Gross fixed assets	2,200	2,600
Accumulated depreciation	(900)	(1,200)
Net fixed assets	1,300	1,400
Total assets	$2,460	$2,640

Liabilities and shareholders' equity Current liabilities		
Accounts payable	$285	$300
Notes payable	200	250
Accrued taxes and expenses	140	150
Total current liabilities	625	700
Long-term debt	865	890
Common stock	100	100
Additional paid-in capital	200	200
Retained earnings	670	750
Total shareholders' equity	970	1,050
Total liabilities and shareholders' equity	$2,460	$2,640

Statement of Income Year Ended 31 December	**2020**
Total revenues	$3,000
Operating costs and expenses	2,200
EBITDA	800
Depreciation	300
Operating income (EBIT)	500
Interest expense	100
Income before tax	400
Taxes (at 40%)	160
Net income	$ 240

(continued)

Exhibit 9 (Continued)

Statement of Income Year Ended 31 December	**2020**
Dividends	*$ 160*
Change in retained earnings (calculated as net income minus dividends)	*$ 80*
Earnings per share (EPS)	$0.48
Dividends per share	$0.32

Statement of Cash Flows Year Ended 31 December	**2020**
Operating activities	
Net income	$240
Adjustments	
Depreciation	300
Changes in working capital	
Accounts receivable	(40)
Inventories	(30)
Accounts payable	15
Accrued taxes and expenses	10
Cash provided by operating activities	$495
Investing activities	
Purchases of fixed assets	400
Cash used for investing activities	$400
Financing activities	
Notes payable	(50)
Long-term financing issuances	(25)
Common stock dividends	160
Cash used for financing activities	$85
Cash and equivalents increase (decrease)	10
Cash and equivalents at beginning of year	190
Cash and equivalents at end of year	$200
Supplemental cash flow disclosures	
Interest paid	$100
Income taxes paid	$160

Note that the Pitts Corporation had net income of $240 million in 2020. Show the calculations required to do each of the following:

1 Calculate FCFF starting with the net income figure.

2 Calculate FCFE starting from the FCFF calculated in Part 1.

3 Calculate FCFE starting with the net income figure.

4 Calculate FCFF starting with CFO.

5 Calculate FCFE starting with CFO.

Solution to 1:

The analyst can use Equation 7 to find FCFF from net income (amounts are in millions):

Net income available to common shareholders	$240
Plus: Net noncash charges	300
Plus: Interest expense × (1 − Tax rate)	60
Less: Investment in fixed capital	400
Less: Investment in working capital	45
Free cash flow to the firm	$155

In the format shown and throughout the solutions, "Less: . . . *x*" is interpreted as "subtract *x*."

This equation can also be written as

FCFF = NI + NCC + Int(1 − Tax rate) − FCInv − WCInv

= 240 + 300 + 60 − 400 − 45 = $155 million.

Some of these items need explanation. Capital spending is $400 million, which is the increase in gross fixed assets shown on the balance sheet and in capital expenditures shown as an investing activity in the statement of cash flows. The increase in working capital is $45 million, which is the increase in accounts receivable of $40 million ($600 million − $560 million) plus the increase in inventories of $30 million ($440 million − $410 million) minus the increase in accounts payable of $15 million ($300 million − $285 million) minus the increase in accrued taxes and expenses of $10 million ($150 million − $140 million). When finding the increase in working capital, we ignore cash because the change in cash is what we are calculating. We also ignore short-term debt, such as notes payable, because such debt is part of the capital provided to the company and is not considered an operating item. The after-tax interest cost is the interest expense times (1 − Tax rate): $100 million × (1 − 0.40) = $60 million. The values of the remaining items in Equation 7 can be taken directly from the financial statements.

Solution to 2:

Finding FCFE from FCFF can be done with Equation 9:

Free cash flow to the firm	$155
Less: Interest expense × (1 − Tax rate)	60
Plus: Net borrowing	75
Free cash flow to equity	$170

Or it can be done by using the equation

FCFE = FCFF − Int(1 − Tax rate) + Net borrowing

= 155 − 60 + 75 = $170 million.

Solution to 3:

The analyst can use Equation 10 to find FCFE from NI.

Net income available to common shareholders	$240
Plus: Net noncash charges	300

(continued)

Less: Investment in fixed capital	400
Less: Investment in working capital	45
Plus: Net borrowing	75
Free cash flow to equity	$170

Or the analyst can use the equation

FCFE = NI + NCC − FCInv − WCInv + Net borrowing

= 240 + 300 − 400 − 45 + 75 = $170 million.

Because notes payable increased by $50 million ($250 million − $200 million) and long-term debt increased by $25 million ($890 million − $865 million), net borrowing is $75 million.

Solution to 4:

Equation 8 can be used to find FCFF from CFO:

Cash flow from operations	$495
Plus: Interest expense × (1 − Tax rate)	60
Less: Investment in fixed capital	400
Free cash flow to the firm	$155

Or

FCFF = CFO + Int(1 − Tax rate) − FCInv

= 495 + 60 − 400 = $155 million.

Solution to 5:

Equation 11 can be used to find FCFE from CFO:

Cash flow from operations	$495
Less: Investment in fixed capital	400
Plus: Net borrowing	75
Free cash flow to equity	$170

Or

FCFE = CFO − FCInv + Net borrowing

= 495 − 400 + 75 = $170 million.

FCFE is usually less than FCFF. In this example, however, FCFE ($170 million) exceeds FCFF ($155 million) because external borrowing was large during this year.

6 FINDING FCFF AND FCFE FROM EBITA OR EBITDA

c explain the appropriate adjustments to net income, earnings before interest and taxes (EBIT), earnings before interest, taxes, depreciation, and amortization (EBITDA), and cash flow from operations (CFO) to calculate FCFF and FCFE;

d calculate FCFF and FCFE;

FCFF and FCFE are most frequently calculated from a starting basis of net income or CFO (as shown earlier). Two other starting points are EBIT and EBITDA from the income statement.

To show the relationship between EBIT and FCFF, we start with Equation 7 and assume that the only noncash charge (NCC) is depreciation (Dep):

$$\text{FCFF} = \text{NI} + \text{Dep} + \text{Int}(1 - \text{Tax rate}) - \text{FCInv} - \text{WCInv}.$$

Net income (NI) can be expressed as

$$\text{NI} = (\text{EBIT} - \text{Int})(1 - \text{Tax rate}) = \text{EBIT}(1 - \text{Tax rate}) - \text{Int}(1 - \text{Tax rate}).$$

Substituting this equation for NI in Equation 7, we have

$$\text{FCFF} = \text{EBIT}(1 - \text{Tax rate}) + \text{Dep} - \text{FCInv} - \text{WCInv}. \tag{12}$$

To get FCFF from EBIT, we multiply EBIT by (1 − Tax rate), add back depreciation, and then subtract the investments in fixed capital and working capital.

The relationship between FCFF and EBITDA can also be easily shown. Net income can be expressed as

$$\text{NI} = (\text{EBITDA} - \text{Dep} - \text{Int})(1 - \text{Tax rate})$$

$$= \text{EBITDA}(1 - \text{Tax rate}) - \text{Dep}(1 - \text{Tax rate}) - \text{Int}(1 - \text{Tax rate}).$$

Substituting this equation for NI in Equation 7 results in

$$\text{FCFF} = \text{EBITDA}(1 - \text{Tax rate}) + \text{Dep}(\text{Tax rate}) - \text{FCInv} - \text{WCInv}. \tag{13}$$

FCFF equals EBITDA times (1 − Tax rate) plus depreciation times the tax rate minus investments in fixed capital and working capital. In comparing Equations 12 and 13, note the difference in how depreciation is handled.

Many adjustments for noncash charges that are required to calculate FCFF when starting from net income are not required when starting from EBIT or EBITDA. In the calculation of net income, many noncash charges are made after computing EBIT or EBITDA, so they do not need to be added back when calculating FCFF based on EBIT or EBITDA. Another important consideration is that some noncash charges, such as depreciation, are tax deductible. A noncash charge that affects taxes must be accounted for.

In summary, in calculating FCFF from EBIT or EBITDA, whether an adjustment for a noncash charge is needed depends on where in the income statement the charge has been deducted; furthermore, the form of any needed adjustment depends on whether the noncash charge is a tax-deductible expense.

We can also calculate FCFE (instead of FCFF) from EBIT or EBITDA. An easy way to obtain FCFE based on EBIT or EBITDA is to use Equation 12 (the expression for FCFF in terms of EBIT) or Equation 13 (the expression for FCFF in terms of EBITDA), respectively, and then subtract Int(1 − Tax rate) and add net borrowing because FCFE is related to FCFF as follows (see Equation 9):

$$\text{FCFE} = \text{FCFF} - \text{Int}(1 - \text{Tax rate}) + \text{Net borrowing}.$$

Example 6 uses the Pitts Corporation financial statements to find FCFF and FCFE from EBIT and EBITDA.

EXAMPLE 6

Adjusting EBIT and EBITDA to Find FCFF and FCFE

The Pitts Corporation (financial statements provided in Example 5) had EBIT of \$500 million and EBITDA of \$800 million in 2020. Show the adjustments that would be required to find FCFF and FCFE:

1 Starting from EBIT.

2 Starting from EBITDA.

Solution to 1:

To get FCFF from EBIT using Equation 12, we carry out the following (in millions):

EBIT(1 − Tax rate) = 500(1 − 0.40)	$300
Plus: Net noncash charges	300
Less: Net investment in fixed capital	400
Less: Net increase in working capital	45
Free cash flow to the firm	$155

Or

$$\begin{aligned} \text{FCFF} &= \text{EBIT}(1 - \text{Tax rate}) + \text{Dep} - \text{FCInv} - \text{WCInv} \\ &= 500(1 - 0.40) + 300 - 400 - 45 = \$155 \text{ million.} \end{aligned}$$

To obtain FCFE, make the appropriate adjustments to FCFF:

$$\begin{aligned} \text{FCFE} &= \text{FCFF} - \text{Int}(1 - \text{Tax rate}) + \text{Net borrowing} \\ &= 155 - 100(1 - 0.40) + 75 = \$170 \text{ million.} \end{aligned}$$

Solution to 2:

To obtain FCFF from EBITDA using Equation 13, we do the following (in millions):

EBITDA(1 − Tax rate) = $800(1 − 0.40)	$480
Plus: Dep(Tax rate) = $300(0.40)	120
Less: Net investment in fixed capital	400
Less: Net increase in working capital	45
Free cash flow to the firm	$155

Or

$$\begin{aligned} \text{FCFF} &= \text{EBITDA}(1 - \text{Tax rate}) + \text{Dep}(\text{Tax rate}) - \text{FCInv} - \text{WCInv} \\ &= 800(1 - 0.40) + 300(0.40) - 400 - 45 = \$155 \text{ million.} \end{aligned}$$

Again, to obtain FCFE, make the appropriate adjustments to FCFF:

$$\begin{aligned} \text{FCFE} &= \text{FCFF} - \text{Int}(1 - \text{Tax rate}) + \text{Net borrowing} \\ &= 155 - 100(1 - 0.40) + 75 = \$170 \text{ million.} \end{aligned}$$

7 FCFF AND FCFE ON A USES-OF-FREE-CASH-FLOW BASIS

d calculate FCFF and FCFE;

g explain how dividends, share repurchases, share issues, and changes in leverage may affect future FCFF and FCFE;

Prior sections illustrated the calculation of FCFF and FCFE from various income or cash flow starting points (e.g., net income or cash flow from operations). Those approaches to calculating free cash flow can be characterized as showing the *sources* of free cash flow. An alternative perspective examines the *uses* of free cash flow. In the context of calculating FCFF and FCFE, analyzing free cash flow on a uses basis serves as a consistency check on the sources calculation and may reveal information relevant to understanding a company's capital structure policy or cash position.

In general, a firm has the following alternative uses of positive FCFF: (1) retain the cash and thus increase the firm's balances of cash and marketable securities; (2) use the cash for payments to providers of debt capital (i.e., interest payments and principal payments in excess of new borrowings); and (3) use the cash for payments to providers of equity capital (i.e., dividend payments and/or share repurchases in excess of new share issuances). Similarly, a firm has the following general alternatives for covering negative free cash flows: draw down cash balances, borrow additional cash, or issue equity.

The effects on the company's capital structure of its transactions with capital providers should be noted. For a simple example, assume that free cash flows are zero and that the company makes no change to its cash balances. Obtaining cash via net new borrowings and using the cash for dividends or net share repurchases will increase the company's leverage, whereas obtaining cash from net new share issuances and using that cash to make principal payments in excess of new borrowings will reduce leverage.

We calculate uses of FCFF as follows:

Uses of FCFF =

Increases (or minus decreases) in cash balances

Plus: Net payments to providers of debt capital, which are calculated as:

- Plus: Interest expense × (1 − Tax rate).
- Plus: Repayment of principal in excess of new borrowing (or minus new borrowing in excess of debt repayment if new borrowing is greater).

Plus: Payments to providers of equity capital, which are calculated as:

- Plus: Cash dividends.
- Plus: Share repurchases in excess of share issuance (or minus new share issuance in excess of share repurchases if share issuance is greater).

Uses of FCFF must equal sources of FCFF as previously calculated.

Free cash flows to equity reflect free cash flows to the firm net of the cash used for payments to providers of debt capital. Accordingly, we can calculate FCFE as follows:

Uses of FCFE =

Increases (or decreases) in cash balances

Plus: Payments to providers of equity capital, which are calculated as:

- Plus: Cash dividends.
- Plus: Share repurchases in excess of share issuance (or minus new share issuance in excess of share repurchases if share issuance is greater).

Again, the uses of FCFE must equal the sources of FCFE (calculated previously).

To illustrate the equivalence of sources and uses of FCFF and FCFE for the Pitts Corporation, whose financial statements are given in Exhibit 9 in Example 5, note the following for 2020:

- The increase in the balance of cash and equivalents was \$10, calculated as \$200 − \$190.
- After-tax interest expense was \$60, calculated as Interest expense × (1 − Tax rate) = \$100 × (1 − 0.40).
- Net borrowing was \$75, calculated as increase in borrowing minus repayment of debt = \$50 (increase in notes payable) + \$25 (increase in long-term debt).

- Cash dividends totaled $160.
- Share repurchases and issuance both equaled $0.

FCFF, previously calculated, was $155. Pitts Corporation used the FCFF as follows (note that payments of principal to providers of debt capital in excess of new borrowings are a use of free cash flow. Here, the corporation did not use its free cash flow to repay debt; rather, it borrowed new debt, which increased the cash flows available to be used for providers of equity capital):

Increase in balance of cash and cash equivalents		$10
Plus:	After-tax interest payments to providers of debt capital	$60
Minus:	New borrowing	($75)
Plus:	Payments of dividends to providers of equity capital	$160
Plus:	Share repurchases in excess of share issuances (or minus new share issuance in excess of share repurchases)	$0
Total uses of FCFF		$155

FCFE, previously calculated, was $170. Pitts Corporation used the FCFE as follows:

Increase in balance of cash and cash equivalents		$10
Plus:	Payments of dividends to providers of equity capital	$160
Plus:	Share repurchases in excess of share issuances (or minus new share issuance in excess of share repurchases)	$0
Total uses of FCFE		$170

In summary, an analysis of the uses of free cash flows shows that Pitts Corporation was using free cash flows to manage its capital structure by increasing debt. The additional debt was not needed to cover capital expenditures; the statement of cash flows showed that the company's operating cash flows of $495 were more than adequate to cover its capital expenditures of $400. Instead, the additional debt was used, in part, to make dividend payments to the company's shareholders.

8 FORECASTING FCFF AND FCFE

e describe approaches for forecasting FCFF and FCFE;

Computing FCFF and FCFE from historical accounting data is relatively straightforward. In some cases, these data are used directly to extrapolate free cash flow growth in a single-stage free cash flow valuation model. On other occasions, however, the analyst may expect that the future free cash flows will not bear a simple relationship to the past. The analyst who wishes to forecast future FCFF or FCFE directly for such a company must forecast the individual components of free cash flow. This section extends our previous presentation on *computing* FCFF and FCFE to the more complex task of *forecasting* FCFF and FCFE.

One method for forecasting free cash flow involves applying some constant growth rate to a current level of free cash flow (possibly adjusted, if necessary, to eliminate non-recurring components). The simplest basis for specifying the future growth rate is to assume that a historical growth rate will also apply to the future. This approach is appropriate if a company's free cash flow has tended to grow at a constant rate and if historical relationships between free cash flow and fundamental factors are expected to continue. Example 7 asks that the reader apply this approach to the Pitts Corporation based on 2020 FCFF of $155 million as calculated in Examples 5 and 6.

EXAMPLE 7

Constant Growth in FCFF

Use Pitts Corporation data to compute its FCFF for the next three years. Assume that growth in FCFF remains at the historical levels of 15% a year. The answer is as follows (in millions):

	2020 Actual	2021 Estimate	2022 Estimate	2023 Estimate
FCFF	155.00	178.25	204.99	235.74

A more complex approach is to forecast the components of free cash flow. This approach is able to capture the complex relationships among the components. One popular method is to forecast the individual components of free cash flow—EBIT(1 – Tax rate), net noncash charges, investment in fixed capital, and investment in working capital. EBIT can be forecasted directly or by forecasting sales and the company's EBIT margin based on an analysis of historical data and the current and expected economic environment. Similarly, analysts can base forecasts of capital needs on historical relationships between increases in sales and investments in fixed and working capital.

In this discussion, we illustrate a simple sales-based forecasting method for FCFF and FCFE based on the following major assumption:

> Investment in fixed capital in excess of depreciation (FCInv – Dep) and investment in working capital (WCInv) both bear a constant relationship to forecast increases in the size of the company as measured by increases in sales.

In addition, for FCFE forecasting, we assume that the capital structure represented by the debt ratio (DR)—debt as a percentage of debt plus equity—is constant. Under that assumption, DR indicates the percentage of the investment in fixed capital in excess of depreciation (also called "net new investment in fixed capital") and in working capital that will be financed by debt. This method involves a simplification because it considers depreciation as the only noncash charge, so the method does not work well when that approximation is not a good assumption.

If depreciation reflects the annual cost for maintaining the existing capital stock, the difference between fixed capital investment and depreciation—incremental FCInv—should be related to the capital expenditures required for growth. In this case, the following inputs are needed:

- forecasts of sales growth rates;
- forecasts of the after-tax operating margin (for FCFF forecasting) or profit margin (for FCFE forecasting);
- an estimate of the relationship of incremental FCInv to sales increases;
- an estimate of the relationship of WCInv to sales increases; and
- an estimate of DR.

In the case of FCFF forecasting, FCFF is calculated by forecasting EBIT(1 − Tax rate) and subtracting incremental fixed capital expenditures and incremental working capital expenditures. To estimate FCInv and WCInv, we multiply their past proportion to sales increases by the forecasted sales increases. Incremental fixed capital expenditures as a proportion of sales increases are computed as follows:

$$\frac{\text{Capital expenditures} - \text{Depreciation expense}}{\text{Increase in sales}}.$$

Similarly, incremental working capital expenditures as a proportion of sales increases are

$$\frac{\text{Increase in working capital}}{\text{Increase in sales}}.$$

When depreciation is the only significant net noncash charge, this method yields the same results as the previous equations for estimating FCFF or FCFE. Rather than adding back all depreciation and subtracting all capital expenditures when starting with EBIT(1 − Tax rate), this approach simply subtracts the net capital expenditures in excess of depreciation.

Although the recognition may not be obvious, this approach recognizes that capital expenditures have two components: those expenditures necessary to maintain existing capacity (fixed capital replacement) and those incremental expenditures necessary for growth. In forecasting, the expenditures to maintain capacity are likely to be related to the current level of sales and the expenditures for growth are likely to be related to the forecast of sales growth.

When forecasting FCFE, analysts often make an assumption that the financing of the company involves a "target" debt ratio. In this case, they assume that a specified percentage of the sum of (1) net new investment in fixed capital (new fixed capital minus depreciation expense) and (2) the increase in working capital is financed based on a target DR. This assumption leads to a simplification of FCFE calculations. If we assume that depreciation is the only noncash charge, Equation 10, which is FCFE = NI + NCC − FCInv − WCInv + Net borrowing, becomes

$$\text{FCFE} = \text{NI} - (\text{FCInv} - \text{Dep}) - \text{WCInv} + \text{Net borrowing}. \tag{14}$$

Note that FCInv − Dep represents the incremental fixed capital expenditure net of depreciation. By assuming a target DR, we eliminated the need to forecast net borrowing and can use the expression

$$\text{Net borrowing} = \text{DR}(\text{FCInv} - \text{Dep}) + \text{DR}(\text{WCInv}).$$

By using this expression, we do not need to forecast debt issuance and repayment on an annual basis to estimate net borrowing. Equation 14 then becomes

$$\text{FCFE} = \text{NI} - (\text{FCInv} - \text{Dep}) - \text{WCInv} + (\text{DR})(\text{FCInv} - \text{Dep}) + (\text{DR})(\text{WCInv})$$

or

$$\text{FCFE} = \text{NI} - (1 - \text{DR})(\text{FCInv} - \text{Dep}) - (1 - \text{DR})(\text{WCInv}). \tag{15}$$

Equation 15 says that FCFE equals NI minus the amount of fixed capital expenditure (net of depreciation) and working capital investment that is financed by equity. Again, for Equation 15, we have assumed that the only noncash charge is depreciation.

Examples 8 and 9 illustrate this sales-based method for forecasting free cash flow to the firm.

EXAMPLE 8

Free Cash Flow Tied to Sales

Carla Espinosa is an analyst following Pitts Corporation at the end of 2020. From the data in Example 5, she can see that the company's sales for 2020 were \$3,000 million, and she assumes that sales grew by \$300 million from 2019 to 2020. Espinosa expects Pitts Corporation's sales to increase by 10% a year thereafter. Pitts Corporation is a fairly stable company, so Espinosa expects it to maintain its historical EBIT margin and proportions of incremental investments in fixed and working capital. Pitts Corporation's EBIT for 2020 is \$500 million, its EBIT margin is 16.67% (500/3,000), and its tax rate is 40%.

Note from Pitts Corporation's 2020 statement of cash flows (Exhibit 9) the amount for "purchases of fixed assets" (i.e., capital expenditures) of \$400 million and depreciation of \$300 million. Thus, incremental fixed capital investment in 2020 was

$$\frac{\text{Capital expenditures} - \text{Depreciation expense}}{\text{Increase in sales}}$$

$$= \frac{400 - 300}{300} = 33.33\%.$$

Incremental working capital investment in the past year was

$$\frac{\text{Increase in working capital}}{\text{Increase in sales}} = \frac{45}{300} = 15\%.$$

So, for every \$100 increase in sales, Pitts Corporation invests \$33.33 in new equipment in addition to replacement of depreciated equipment and \$15 in working capital. Espinosa forecasts FCFF for 2013 as follows (dollars in millions):

Sales	\$3,300	Up 10%
EBIT	550	16.67% of sales
EBIT(1 – Tax rate)	330	Adjusted for 40% tax rate
Incremental FC	(100)	33.33% of sales increase
Incremental WC	(45)	15% of sales increase
FCFF	\$185	

This model can be used to forecast multiple periods and is flexible enough to allow varying sales growth rates, EBIT margins, tax rates, and rates of incremental capital increases.

EXAMPLE 9

Free Cash Flow Growth Tied to Sales Growth

Continuing her work, Espinosa decides to forecast FCFF for the next five years. She is concerned that Pitts Corporation will not be able to maintain its historical EBIT margin and that the EBIT margin will decline from the current 16.67% to 14.5% in the next five years. Exhibit 10 summarizes her forecasts.

Exhibit 10 Free Cash Flow Growth for Pitts Corporation (Dollars in Millions)

	Year 1	Year 2	Year 3	Year 4	Year 5
Sales growth	10.00%	10.00%	10.00%	10.00%	10.00%
EBIT margin	16.67%	16.00%	15.50%	15.00%	14.50%
Tax rate	40.00%	40.00%	40.00%	40.00%	40.00%
Incremental FC investment	33.33%	33.33%	33.33%	33.33%	33.33%
Incremental WC investment	15.00%	15.00%	15.00%	15.00%	15.00%
Prior-year sales	$3,000.00				
Sales forecast	$3,300.00	$3,630.00	$3,993.00	$4,392.30	$4,831.53
EBIT forecast	550.00	580.80	618.92	658.85	700.57
EBIT(1 – Tax rate)	330.00	348.48	371.35	395.31	420.34
Incremental FC	(100.00)	(110.00)	(121.00)	(133.10)	(146.41)
Incremental WC	(45.00)	(49.50)	(54.45)	(59.90)	(65.88)
FCFF	$185.00	$188.98	$195.90	$202.31	$208.05

The model need not begin with sales; it could start with net income, cash flow from operations, or EBITDA.

A similar model can be designed for FCFE, as shown in Example 10. In the case of FCFE, the analyst should begin with net income and must also forecast any net new borrowing or net preferred stock issue.

EXAMPLE 10

Finding FCFE from Sales Forecasts

Espinosa decides to forecast FCFE for the year 2021. She uses the same expectations derived in Example 8. Additionally, she expects the following:

- the profit margin will remain at 8% (= 240/3,000), and
- the company will finance incremental fixed and working capital investments with 50% debt—the target DR.

Espinosa's forecast for 2013 is as follows (dollars in millions):

Sales	$3,300	Up 10%
NI	264	8.0% of sales
Incremental FC	(100)	33.33% of sales increase
Incremental WC	(45)	15% of sales increase
Net borrowing	72.50	(100 FCInv + 45 WCInv) × 50%
FCFE	$191.50	

When the company being analyzed has significant noncash charges other than depreciation expense, the approach we have just illustrated will result in a less accurate estimate of FCFE than one obtained by forecasting all the individual components of

FCFE. In some cases, the analyst will have specific forecasts of planned components, such as capital expenditures. In other cases, the analyst will study historical relationships, such as previous capital expenditures and sales levels, to develop a forecast.

OTHER ISSUES IN FREE CASH FLOW ANALYSIS

9

f compare the FCFE model and dividend discount models;

g explain how dividends, share repurchases, share issues, and changes in leverage may affect future FCFF and FCFE;

h evaluate the use of net income and EBITDA as proxies for cash flow in valuation;

We have already presented a number of practical issues that arise in using free cash flow valuation models. Other issues relate to analyst adjustments to CFO, the relationship between free cash flow and dividends, and valuation with complicated financial structures.

9.1 Analyst Adjustments to CFO

Although many corporate financial statements are straightforward, some are not transparent (i.e., the quality of the reported numbers and of disclosures is not high). Sometimes, difficulties in analysis arise either because of lack of transparency or because the companies and their transactions are more complicated than the Pitts Corporation example we just provided.

For instance, in many corporate financial statements, the changes in balance sheet items (the increase in an asset or the decrease in a liability) differ from the changes reported in the statement of cash flows. Financial statements in which the changes in the balance sheet working capital accounts do not equal the working capital amounts reported on the statement of cash flows are described as lacking "articulation." Research on financial statement non-articulation (which is not an uncommon occurrence) identifies several reasons for these differences (Casey, Gao, Kirschenheiter, Li, and Pandit 2016; Huefner, Ketz, and Largay 1989; Bahnson, Miller, and Budge 1996; Wilkins and Loudder 2000; Hribar and Collins 2002; and Shi and Zhang 2011). Two of the factors that can cause discrepancies between changes in balance sheet accounts and the changes reported in the statement of cash flows include (1) acquisitions or divestitures (and related discontinued operations) and (2) the presence of nondomestic subsidiaries. For example, an increase in an inventory account may result from purchases from suppliers (which is an operating activity) or from an acquisition or merger with another company that has inventory on its balance sheet (which is an investing activity). Discrepancies may also occur from currency translations of the earnings of nondomestic subsidiaries.

Particularly for companies with major acquisition or divestiture activity where the CFO figure from the statement of cash flows may be distorted by cash flows related to financing and/or investing activities, an analyst may need to use greater detail in forecasting. For example, the analyst may need to adjust the amount of CFO that is used as the starting point for free cash flow calculations. Alternatively, instead of (or in addition to) developing a cash flow forecast by extrapolating from reported OCF, an analyst might forecast individual components and pay careful attention to the relation between sales forecast and forecast of specific working capital items.

9.2 Free Cash Flow versus Dividends and Other Earnings Components

Many analysts have a strong preference for free cash flow valuation models over dividend discount models. Although one type of model may have no theoretical advantage over another type, legitimate reasons to prefer one model can arise in the process of applying free cash flow models versus DDMs. First, many corporations pay no, or very low, cash dividends. Using a DDM to value these companies is difficult because they require forecasts about when dividends will be initiated, the level of dividends at initiation, and the growth rate or rates from that point forward. Second, dividend payments are at the discretion of the corporation's board of directors. Therefore, they may imperfectly signal the company's long-run profitability. Some corporations clearly pay dividends that are substantially less than their free cash flow, and others pay dividends that are substantially more. Finally, as mentioned earlier, dividends are the cash flow actually going to shareholders whereas free cash flow to equity is the cash flow available to be distributed to shareholders without impairing the company's value. If a company is being analyzed because it is a target for takeover, free cash flow is the appropriate cash flow measure; once the company is taken over, the new owners will have discretion over how free cash flow is used (including its distribution in the form of dividends).

We have defined FCFF and FCFE and presented alternative (equivalent) ways to calculate both. So, the reader should have a good idea of what is included in FCFF or FCFE but may wonder why some cash flows are not included. Specifically, what role do dividends, share repurchases, share issuance, or changes in leverage have on FCFF and FCFE? The simple answer is not much. Recall the formulas for FCFF and FCFE:

$$\text{FCFF} = \text{NI} + \text{NCC} + \text{Int}(1 - \text{Tax rate}) - \text{FCInv} - \text{WCInv},$$

and

$$\text{FCFE} = \text{NI} + \text{NCC} - \text{FCInv} - \text{WCInv} + \text{Net borrowing}.$$

Notice that dividends and share repurchases and issuance are absent from the formulas. The reason is that FCFF and FCFE are the cash flows *available* to investors or to stockholders; dividends and share repurchases are *uses* of these cash flows. So, the simple answer is that transactions between the company and its shareholders (through cash dividends, share repurchases, and share issuances) do not affect free cash flow. Leverage changes, such as the use of more debt financing, have some impact because they increase the interest tax shield (reduce corporate taxes because of the tax deductibility of interest) and reduce the cash flow available to equity. In the long run, the investing and financing decisions made today will affect future cash flows.

If all the inputs were known and mutually consistent, a DDM and an FCFE model would result in identical valuations for a stock. One possibility would be that FCFE equals cash dividends each year. Then, both cash flow streams would be discounted at the required return for equity and would have the same present value.

Generally, however, FCFE and dividends will differ, but the same economic forces that lead to low (high) dividends lead to low (high) FCFE. For example, a rapidly growing company with superior investment opportunities will retain a high proportion of earnings and pay low dividends. This same company will have high investments in fixed capital and working capital and have a low FCFE (which is clear from the expression FCFE = NI + NCC − FCInv − WCInv + Net borrowing). Conversely, a mature company that is investing relatively little might have high dividends and high FCFE. Despite this tendency, however, FCFE and dividends will usually differ.

FCFF and FCFE, as defined here, are measures of cash flow designed for valuation of the firm or its equity. Other definitions of free cash flow frequently appear in textbooks, articles, and vendor-supplied databases of financial information on public companies. In many cases, these other definitions of free cash flow are not designed for valuation purposes and thus should not be used for valuation. Using numbers supplied by others without knowing exactly how they are defined increases the likelihood of making errors in valuation. As consumers and producers of research, analysts should understand (if consumers) or make clear (if producers) the definition of free cash flow being used.

Because using free cash flow analysis requires considerable care and understanding, some practitioners erroneously use earnings components such as NI, EBIT, EBITDA, or CFO in a discounted cash flow valuation. Such mistakes may lead the practitioner to systematically overstate or understate the value of a stock. Shortcuts can be costly.

A common shortcut is to use EBITDA as a proxy for the cash flow to the firm. Equation 13 clearly shows the differences between EBITDA and FCFF:

$$\text{FCFF} = \text{EBITDA}(1 - \text{Tax rate}) + \text{Dep}(\text{Tax rate}) - \text{FCInv} - \text{WCInv}.$$

Depreciation charges as a percentage of EBITDA differ substantially for different companies and industries, as does the depreciation tax shield (the depreciation charge times the tax rate). Although FCFF captures this difference, EBITDA does not. EBITDA also does not account for the investments a company makes in fixed capital or working capital. Hence, EBITDA is a poor measure of the cash flow available to the company's investors. Using EBITDA (instead of free cash flow) in a DCF model has another important aspect as well: EBITDA is a before-tax measure, so the discount rate applied to EBITDA would be a before-tax rate. The WACC used to discount FCFF is an after-tax cost of capital.

EBITDA is a poor proxy for free cash flow to the firm because it does not account for the depreciation tax shield and the investment in fixed capital and working capital, but it is an even poorer proxy for free cash flow to equity. From a stockholder's perspective, additional defects of EBITDA include its failure to account for the after-tax interest costs or cash flows from new borrowing or debt repayments. Example 11 shows the mistakes sometimes made in discussions of cash flows.

EXAMPLE 11

The Mistakes of Using Net Income for FCFE and EBITDA for FCFF

A recent job applicant made some interesting comments about FCFE and FCFF: "I don't like the definitions for FCFE and FCFF because they are unnecessarily complicated and confusing. The best measure of FCFE, the funds available to pay dividends, is simply net income. You take the net income number straight from the income statement and don't need to make any further adjustments. Similarly, the best measure of FCFF, the funds available to the company's suppliers of capital, is EBITDA. You can take EBITDA straight from the income statement, and you don't need to consider using anything else."

How would you respond to the job applicant's definition of (1) FCFE and (2) FCFF?

Solution to 1:

The FCFE is the cash generated by the business's operations less the amount it must reinvest in additional assets plus the amounts it is borrowing. Equation 10, which starts with net income to find FCFE, shows these items:

Free cash flow to equity	=	Net income available to common shareholders
		Plus: Net noncash charges
		Less: Investment in fixed capital
		Less: Investment in working capital
		Plus: Net borrowing

Net income does not include several cash flows. So, net income tells only part of the overall story. Investments in fixed or working capital reduce the cash available to stockholders, as do loan repayments. New borrowing increases the cash available. FCFE, however, includes the cash generated from operating the business and also accounts for the investing and financing activities of the company. Of course, a special case exists in which net income and FCFE are the same. This case occurs when new investments exactly equal depreciation and the company is not investing in working capital or engaging in any net borrowing.

Solution to 2:

Assuming that EBITDA equals FCFF introduces several possible mistakes. Equation 13 highlights these mistakes:

Free cash flow to the firm	=	EBITDA(1 − Tax rate)
		Plus: Depreciation(Tax rate)
		Less: Investment in fixed capital
		Less: Investment in working capital

The applicant is ignoring taxes, which obviously reduce the cash available to the company's suppliers of capital, and is also ignoring depreciation and the investments in fixed capital and working capital.

9.3 Free Cash Flow and Complicated Capital Structures

For the most part, the discussion of FCFF and FCFE so far has assumed the company has a simple capital structure with two sources of capital—namely, debt and equity. Including preferred stock as a third source of capital requires the analyst to add terms to the equations for FCFF and FCFE to account for the dividends paid on preferred stock and for the issuance or repurchase of preferred shares. Instead of including those terms in all of the equations, we chose to leave preferred stock out because only a few corporations use preferred stock. For companies that do have preferred stock, however, the effects of the preferred stock can be incorporated in the valuation models.

For example, in Equation 7, which calculates FCFF starting with net income available to common shareholders, preferred dividends paid would be added to the cash flows to obtain FCFF. In Equation 10, which calculates FCFE starting with net income available to common shareholders, if preferred dividends were already subtracted when arriving at net income, no further adjustment for preferred dividends would be required. Issuing (redeeming) preferred stock increases (decreases) the cash flow available to common stockholders, however, so this term would have to be added in. The existence of preferred stock in the capital structure has many of the same effects as the existence of debt, except that unlike interest payments on debt, preferred stock dividends paid are not tax deductible.

Example 12 shows how to calculate WACC, FCFF, and FCFE when the company has preferred stock.

EXAMPLE 12

FCFF Valuation with Preferred Stock in the Capital Structure

Welch Corporation uses bond, preferred stock, and common stock financing. The market value of each of these sources of financing and the before-tax required rates of return for each are given in Exhibit 11:

Exhibit 11 Welch Corporation Capital Structure (Dollars in Millions)

	Market Value ($)	Required Return (%)
Bonds	400	8.0
Preferred stock	100	8.0
Common stock	500	12.0
Total	1,000	

Other financial information (dollars in millions):

- Net income available to common shareholders = \$110.
- Interest expenses = \$32.
- Preferred dividends = \$8.
- Depreciation = \$40.
- Investment in fixed capital = \$70.
- Investment in working capital = \$20.
- Net borrowing = \$25.
- Tax rate = 30%.
- Stable growth rate of FCFF = 4.0%.
- Stable growth rate of FCFE = 5.4%.

1 Calculate Welch Corporation's WACC.
2 Calculate the current value of FCFF.
3 Based on forecasted Year 1 FCFF, what is the total value of Welch Corporation and the value of its equity?
4 Calculate the current value of FCFE.
5 Based on forecasted Year 1 FCFE, what is the value of equity?

Solution to 1:

Based on the weights and after-tax costs of each source of capital, the WACC is

$$\text{WACC} = \frac{400}{1,000}8\%\left(1 - 0.30\right) + \frac{100}{1,000}8\% + \frac{500}{1,000}12\% = 9.04\%.$$

Solution to 2:

If the company did not issue preferred stock, FCFF would be

$$\text{FCFF} = \text{NI} + \text{NCC} + \text{Int}(1 - \text{Tax rate}) - \text{FCInv} - \text{WCInv}.$$

If preferred stock dividends have been paid (and net income is income available to common shareholders), the preferred dividends must be added back just as after-tax interest expenses are. The modified equation (including preferred dividends) for FCFF is

$$\text{FCFF} = \text{NI} + \text{NCC} + \text{Int}(1 - \text{Tax rate}) + \text{Preferred dividends} - \text{FCInv} - \text{WCInv}.$$

For Welch Corporation, FCFF is

$$\text{FCFF} = 110 + 40 + 32(1 - 0.30) + 8 - 70 - 20 = \$90.4 \text{ million}.$$

Solution to 3:

The total value of the firm is

$$\text{Firm value} = \frac{\text{FCFF}_1}{\text{WACC} - g} = \frac{90.4(1.04)}{0.0904 - 0.04}$$
$$= \frac{94.016}{0.0504} = \$1{,}865.40 \text{ million}.$$

The value of (common) equity is the total value of the company minus the value of debt and preferred stock:

$$\text{Equity} = 1{,}865.40 - 400 - 100 = \$1{,}365.40 \text{ million}.$$

Solution to 4:

With no preferred stock, FCFE is

$$\text{FCFE} = \text{NI} + \text{NCC} - \text{FCInv} - \text{WCInv} + \text{Net borrowing}.$$

If the company has preferred stock, the FCFE equation is essentially the same. Net borrowing in this case is the total of new debt borrowing and net issuances of new preferred stock. For Welch Corporation, FCFE is

$$\text{FCFE} = 110 + 40 - 70 - 20 + 25 = \$85 \text{ million}.$$

Solution to 5:

Valuing FCFE, which is growing at 5.4%, produces a value of equity of

$$\text{Equity} = \frac{\text{FCFE}_1}{r - g} = \frac{85(1.054)}{0.12 - 0.054} = \frac{89.59}{0.066} = \$1{,}357.42 \text{ million}.$$

Paying cash dividends on common stock does not affect FCFF or FCFE, which are the amounts of cash *available* to all investors or to common stockholders. It is simply a use of the available cash. Share repurchases of common stock also do not affect FCFF or FCFE. Share repurchases are, in many respects, a substitute for cash dividends. Similarly, issuing shares of common stock does not affect FCFF or FCFE.

Changing leverage (changing the amount of debt financing in the company's capital structure), however, does have some effects on FCFE particularly. An increase in leverage will not affect FCFF (although it might affect the calculations used to arrive at FCFF). An increase in leverage affects FCFE in two ways. In the year the debt is issued, it increases the FCFE by the amount of debt issued. After the debt is issued, FCFE is then reduced by the after-tax interest expense.

In this section, we have discussed the concepts of FCFF and FCFE and their estimation. The next section presents additional valuation models that use forecasts of FCFF or FCFE to value the firm or its equity. These free cash flow models are similar in structure to dividend discount models, although the analyst must face the reality that estimating free cash flows is more time-consuming than estimating dividends.

FREE CASH FLOW MODEL VARIATIONS: AN INTERNATIONAL APPLICATION OF THE SINGLE-STAGE MODEL AND SENSITIVITY ANALYSIS OF FCFF AND FCFE VALUATIONS

10

This section presents several extensions of the free cash flow models presented earlier. In many cases, especially when inflation rates are volatile, analysts will value real cash flows instead of nominal values. As with dividend discount models, free cash flow models are sensitive to the data inputs, so analysts routinely perform sensitivity analyses of their valuations.

Earlier, we presented the single-stage free cash flow model, which has a constant growth rate. In the following, we use the single-stage model to address selected valuation issues; we then present multistage free cash flow models.

10.1 An International Application of the Single-Stage Model

Valuation by using real (inflation-adjusted) values instead of nominal values has much appeal when inflation rates are high and volatile. Many analysts use this adaptation for both domestic and nondomestic stocks, but the use of real values is especially helpful for valuing international stocks. Special challenges to valuing equities from multiple countries include (1) incorporating economic factors—such as interest rates, inflation rates, and growth rates—that differ among countries and (2) dealing with varied accounting standards. Furthermore, performing analyses in multiple countries challenges the analyst—particularly a team of analysts—to use *consistent* assumptions for all countries.

Several securities firms have adapted the single-stage FCFE model to address some of the challenges of international valuation. They choose to analyze companies by using real cash flows and real discount rates instead of nominal values. To estimate real discount rates, they use a modification of the build-up method mentioned earlier under the topic of return concepts. Starting with a "country return," which is a real required rate of return for stocks from a particular country, they then make adjustments to the country return for the stock's industry, size, and leverage:

Country return (real)	x.xx%
+/– Industry adjustment	x.xx%
+/– Size adjustment	x.xx%
+/– Leverage adjustment	x.xx%
Required rate of return (real)	x.xx%

The adjustments in the model should have sound economic justification. They should reflect factors expected to affect the relative risk and return associated with an investment.

The securities firms making these adjustments predict the growth rate of FCFE also in real terms. The firms supply their analysts with estimates of the real economic growth rate for each country, and each analyst chooses a real growth rate for the stock being analyzed that is benchmarked against the real country growth rate. This approach is particularly useful for countries with high or variable inflation rates.

The value of the stock is found with an equation essentially like Equation 6 except that all variables in the equation are stated in real terms:

$$V_0 = \frac{\text{FCFE}_0\left(1 + g_{\text{real}}\right)}{r_{\text{real}} - g_{\text{real}}}.$$

Whenever real discount rates and real growth rates can be estimated more reliably than nominal discount rates and nominal growth rates, this method is worth using. Example 13 shows how this procedure can be applied.

EXAMPLE 13

Using Real Cash Flows and Discount Rates for International Stocks

Mukamba Ventures is a consumer staples company headquartered in Kinshasa, Democratic Republic of the Congo. Although the company's cash flows have been volatile, an analyst has estimated a per-share normalized FCFE of 1,400 Congolese francs (CDF) for the year just ended. The real country return for the Democratic Republic of the Congo is 7.30%; adjustments to the country return for Mukamba Ventures are an industry adjustment of +0.80%, a size adjustment of –0.33%, and a leverage adjustment of –0.12%. The long-term real growth rate for the Democratic Republic of the Congo is estimated to be 3.0%, and the real growth rate of Mukamba Ventures is expected to be about 0.5% below the country rate. The real required rate of return for Mukamba Ventures is calculated as follows:

Country return (real)	7.30%
Industry adjustment	+ 0.80%
Size adjustment	– 0.33%
Leverage adjustment	– 0.12%
Required rate of return	7.65%

The real growth rate of FCFE is expected to be 2.5% (3.0% – 0.5%), so the value of one share is

$$V_0 = \frac{\text{FCFE}_0\left(1 + g_{real}\right)}{r_{real} - g_{real}} = \frac{1{,}400(1.025)}{0.0765 - 0.025} = \frac{1{,}435}{0.0515} = \text{CDF}27{,}864.$$

10.2 Sensitivity Analysis of FCFF and FCFE Valuations

In large measure, growth in FCFF and in FCFE depends on a company's future profitability. Sales growth and changes in net profit margins dictate future net profits. Sales growth and profit margins depend on the growth phase of the company and the profitability of the industry. A highly profitable company in a growing industry can enjoy years of profit growth. Eventually, however, its profit margins are likely to be eroded by increased competition; sales growth is also likely to abate because of fewer opportunities for expansion of market size and market share. Growth rates and the duration of growth are difficult to forecast.

The base-year values for the FCFF and FCFE growth models are also critical. Given the same required rates of return and growth rates, the value of the firm or the value of equity will increase or decrease proportionately with the initial value of FCFF or FCFE used.

To examine how sensitive the final valuation is to changes in each of a valuation model's input variables, analysts can perform a sensitivity analysis. Some input variables have a much larger impact on stock valuation than others. Example 14 shows the sensitivity of the valuation of Petroleo Brasileiro to four input variables.

EXAMPLE 14

Sensitivity Analysis of an FCFE Valuation

Antonio Sousa is valuing the equity of Petroleo Brasileiro, commonly known as Petrobras, by using the single-stage (constant-growth) FCFE model. Estimated FCFE per share for the year just ended is 2.59 Brazilian reals (BRL). Sousa's best estimates of input values for the analysis are as follows:

- The FCFE growth rate is 7.0%.
- The risk-free rate is 8.9%.
- The equity risk premium is 5.3%.
- Beta is 1.4.

Using the capital asset pricing model (CAPM), Sousa estimates that the required rate of return for Petrobras is

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 8.9\% + 1.4(5.3\%) = 16.32\%.$$

The estimated value per share is

$$V_0 = \frac{\text{FCFE}_0(1+g)}{r-g} = \frac{2.59(1.07)}{0.1632 - 0.07} = \text{BRL}29.73.$$

Exhibit 12 shows Sousa's base case and the highest and lowest reasonable alternative estimates. The column "Valuation with Low Estimate" gives the estimated value of Petrobras based on the low estimate for the variable on the same row of the first column and the base-case estimates for the remaining three variables. "Valuation with High Estimate" gives a similar estimated value based on the high estimate for the variable at issue.

Exhibit 12 Sensitivity Analysis for Petrobras Valuation

Variable	Base-Case Estimate	Low Estimate	High Estimate	Valuation with Low Estimate	Valuation with High Estimate
Beta	1.4	1.2	1.6	BRL33.55	BRL26.70
Risk-free rate	8.9%	7.9%	9.9%	BRL33.31	BRL26.85
Equity risk premium	5.3%	4.3%	6.3%	BRL34.99	BRL25.85
FCFE growth rate	7.0%	5.0%	9.0%	BRL24.02	BRL38.57

As Exhibit 12 shows, the value of Petrobras is very sensitive to the inputs. The value is negatively related to changes in the beta, the risk-free rate, and the equity risk premium and positively related to changes in the FCFE growth rate. Of the four variables considered, the stock valuation is most sensitive to the range of estimates for the FCFE growth rate (a range from BRL24.02 to BRL38.57. The ranges of the estimates for the other three variables, while still large, are less than the range for changes in the FCFE growth rate. Of course, the variables to which a stock price is most sensitive vary from case to case. A sensitivity analysis gives the analyst a guide as to which variables are most critical to the final valuation.

11 TWO-STAGE FREE CASH FLOW MODELS

i. explain the single-stage (stable-growth), two-stage, and three-stage FCFF and FCFE models and justify the selection of the appropriate model given a company's characteristics;

j. estimate a company's value using the appropriate free cash flow model(s);

l. describe approaches for calculating the terminal value in a multistage valuation model; and

Several two-stage and multistage models exist for valuing free cash flow streams, just as several such models are available for valuing dividend streams. The free cash flow models are much more complex than the dividend discount models because to find FCFF or FCFE, the analyst usually incorporates sales, profitability, investments, financing costs, and new financing.

In two-stage free cash flow models, the growth rate in the second stage is a long-run sustainable growth rate. For a declining industry, the second-stage growth rate could be slightly below the GDP growth rate. For an industry that is expected to grow in the future faster than the overall economy, the second-stage growth rate could be slightly greater than the GDP growth rate.

The two most popular versions of the two-stage FCFF and FCFE models are distinguished by the pattern of the growth rates in Stage 1. In one version, the growth rate is constant in Stage 1 before dropping to the long-run sustainable rate in Stage 2. In the other version, the growth rate declines in Stage 1 to reach the sustainable rate at the beginning of Stage 2. This second type of model is like the H-model for discounted dividend valuation, in which dividend growth rates decline in Stage 1 and are constant in Stage 2.

Unlike multistage DDMs, in which the growth rates are consistently dividend growth rates, in free cash flow models, the "growth rate" may refer to different variables (which variables should be stated or should be clear from the context). The growth rate could be the growth rate for FCFF or FCFE, the growth rate for income (either net income or operating income), or the growth rate for sales. If the growth rate is for net income, the changes in FCFF or FCFE also depend on investments in operating assets and the financing of these investments. When the growth rate in income declines, such as between Stage 1 and Stage 2, investments in operating assets probably decline at the same time. If the growth rate is for sales, changes in net profit margins as well as investments in operating assets and financing policies will determine FCFF and FCFE.

A general expression for the two-stage FCFF valuation model is

$$\text{Firm value} = \sum_{t=1}^{n} \frac{\text{FCFF}_t}{(1+\text{WACC})^t} + \frac{\text{FCFF}_{n+1}}{(\text{WACC}-g)} \frac{1}{(1+\text{WACC})^n}. \tag{16}$$

The summation gives the present value of the first n years of FCFF. The terminal value of the FCFF from Year $n + 1$ forward is $\text{FCFF}_{n+1}/(\text{WACC} - g)$, which is discounted at the WACC for n periods to obtain its present value. Subtracting the value of outstanding debt gives the value of equity. The value per share is then found by dividing the total value of equity by the number of outstanding shares.

The general expression for the two-stage FCFE valuation model is

$$\text{Equity value} = \sum_{t=1}^{n} \frac{\text{FCFE}_t}{(1+r)^t} + \left(\frac{\text{FCFE}_{n+1}}{r-g}\right)\left[\frac{1}{(1+r)^n}\right]. \tag{17}$$

In this case, the summation is the present value of the first n years of FCFE and the terminal value of $FCFE_{n+1}/(r-g)$ is discounted at the required rate of return on equity for n years. The value per share is found by dividing the total value of equity by the number of outstanding shares.

In Equation 17, the terminal value of the stock at $t = n$, TV_n, is found by using the constant-growth FCFE model. In this case, $TV_n = FCFE_{n+1}/(r-g)$. (Of course, the analyst might choose to estimate terminal value another way, such as by using a P/E multiplied by the company's forecasted EPS.) The terminal value estimation is critical for a simple reason: The present value of the terminal value is often a substantial portion of the total value of the stock. For example, in Equation 17, when the analyst is calculating the total present value of the first n cash flows (FCFE) and the present value of the terminal value, the present value of the terminal value is often substantial. In the examples that follow, the terminal value usually represents a substantial part of total estimated value. The same is true in practice.

11.1 Fixed Growth Rates in Stage 1 and Stage 2

The simplest two-stage FCFF or FCFE growth model has a constant growth rate in each stage. Example 15 finds the value of a firm that has a 20% sales growth rate in Stage 1 and a 6% sales growth rate in Stage 2.

EXAMPLE 15

A Two-Stage FCFE Valuation Model with a Constant Growth Rate in Each Stage

Uwe Henschel is doing a valuation of TechnoSchaft on the basis of the following information:

- Year 0 sales per share = €25.
- Sales growth rate = 20% annually for three years and 6% annually thereafter.
- Net profit margin = 10% forever.
- Net investment in fixed capital (net of depreciation) = 50% of the sales increase.
- Annual increase in working capital = 20% of the sales increase.
- Debt financing = 40% of the net investments in capital equipment and working capital.
- TechnoSchaft beta = 1.20; the risk-free rate of return = 7%; the equity risk premium = 4.5%.

The required rate of return for equity is

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 7\% + 1.2(4.5\%) = 12.4\%.$$

Exhibit 13 shows the calculations for FCFE.

Exhibit 13 FCFE Estimates for TechnoSchaft (in Euros)

	Year					
	1	2	3	4	5	6
Sales growth rate	20%	20%	20%	6%	6%	6%
Sales per share	30.000	36.000	43.200	45.792	48.540	51.452
Net profit margin	10%	10%	10%	10%	10%	10%
EPS	3.000	3.600	4.320	4.579	4.854	5.145
Net FCInv per share	2.500	3.000	3.600	1.296	1.374	1.456
WCInv per share	1.000	1.200	1.440	0.518	0.550	0.582
Debt financing per share	1.400	1.680	2.016	0.726	0.769	0.815
FCFE per share	0.900	1.080	1.296	3.491	3.700	3.922
Growth rate of FCFE		20%	20%	169%	6%	6%

In Exhibit 13, sales are shown to grow at 20% annually for the first three years and then at 6% thereafter. Profits, which are 10% of sales, grow at the same rates. The net investments in fixed capital and working capital are, respectively, 50% of the increase in sales and 20% of the increase in sales. New debt financing equals 40% of the total increase in net fixed capital and working capital. FCFE is EPS minus the net investment in fixed capital per share minus the investment in working capital per share plus the debt financing per share.

Notice that FCFE grows by 20% annually for the first three years (i.e., between $t = 0$ and $t = 3$). Then, between Year 3 and Year 4, when the sales growth rate drops from 20% to 6%, FCFE increases substantially. In fact, FCFE increases by 169% from Year 3 to Year 4. This large increase in FCFE occurs because profits grow at 6% but the investments in capital equipment and working capital (and the increase in debt financing) drop substantially from the previous year. In Years 5 and 6 in Exhibit 13, sales, profit, investments, financing, and FCFE are all shown to grow at 6%.

The stock value is the present value of the first three years' FCFE plus the present value of the terminal value of the FCFE from Years 4 and later. The terminal value is

$$\mathrm{TV}_3 = \frac{\mathrm{FCFE}_4}{r - g} = \frac{3.491}{0.124 - 0.06} = €54.55.$$

The present values are

$$V_0 = \frac{0.900}{1.124} + \frac{1.080}{(1.124)^2} + \frac{1.296}{(1.124)^3} + \frac{54.55}{(1.124)^3}$$
$$= 0.801 + 0.855 + 0.913 + 38.415 = €40.98.$$

The estimated value of this stock is €40.98 per share.

As mentioned previously, the terminal value may account for a large portion of the value of a stock. In the case of TechnoSchaft, the present value of the terminal value is €38.415 out of a total value of €40.98. The present value (PV) of the terminal value is almost 94% of the total value of TechnoSchaft stock.

11.2 Declining Growth Rate in Stage 1 and Constant Growth in Stage 2

Growth rates usually do not drop precipitously as they do between the stages in the two-stage model just described, but growth rates can decline over time for many reasons. Sometimes, a small company has a high growth rate that is not sustainable as its market share increases. A highly profitable company may attract competition that makes it harder for the company to sustain its high profit margins.

In this section, we present two examples of the two-stage model with declining growth rates in Stage 1. In the first example, the growth rate of EPS declines during Stage 1. As a company's profitability declines and the company is no longer generating high returns, the company will usually reduce its net new investment in operating assets. The debt financing accompanying the new investments will also decline. Many highly profitable, growing companies have negative or low free cash flows. Later, when growth in profits slows, investments will tend to slow and the company will experience positive cash flows. Of course, the negative cash flows incurred in the high-growth stage help determine the cash flows that occur in future years.

Example 16 models FCFE per share as a function of EPS that declines constantly during Stage 1. Because of declining earnings growth rates, the company in the example also reduces its new investments over time. The value of the company depends on these free cash flows, which are substantial after the high-growth (and high-profitability) period has largely elapsed.

EXAMPLE 16

A Two-Stage FCFE Valuation Model with Declining Net Income Growth in Stage 1

Vishal Noronha needs to prepare a valuation of Sindhuh Enterprises. Noronha has assembled the following information for his analysis. It is now the first day of 2020.

- EPS for 2019 is $2.40.
- For the next five years, the growth rate in EPS is given in the following table. After 2024, the growth rate will be 7%.

	2020	2021	2022	2023	2024
Growth rate for EPS	30%	18%	12%	9%	7%

- Net investments in fixed capital (net of depreciation) for the next five years are given in the following table. After 2024, capital expenditures are expected to grow at 7% annually.

	2020	2021	2022	2023	2024
Net capital expenditure per share	$3.00	$2.50	$2.00	$1.50	$1.00

- The investment in working capital each year will equal 50% of the net investment in capital items.
- 30% of the net investment in fixed capital and investment in working capital will be financed with new debt financing.
- Current market conditions dictate a risk-free rate of 6.0%, an equity risk premium of 4.0%, and a beta of 1.10 for Sindhuh Enterprises.

- What is the per-share value of Sindhuh Enterprises on the first day of 2020?
- What should be the trailing P/E on the first day of 2020 and the first day of 2024?

Solution to 1:

The required return for Sindhuh should be

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 6\% + 1.1(4\%) = 10.4\%.$$

The FCFEs for the company for years 2020 through 2024 are given in Exhibit 14.

Exhibit 14 FCFE Estimates for Sindhuh Enterprises (Per-Share Data in US Dollars)

	Year				
	2020	2021	2022	2023	2024
Growth rate for EPS	30%	18%	12%	9%	7%
EPS	3.120	3.682	4.123	4.494	4.809
Net FCInv per share	3.000	2.500	2.000	1.500	1.000
WCInv per share	1.500	1.250	1.000	0.750	0.500
Debt financing per share[a]	1.350	1.125	0.900	0.675	0.450
FCFE per share[b]	–0.030	1.057	2.023	2.919	3.759
PV of FCFE discounted at 10.4%	–0.027	0.867	1.504	1.965	

[a]30% of (Net FCInv + WCInv).
[b]EPS – Net FCInv per share – WCInv per share + Debt financing per share.

Earnings are \$2.40 in 2019. Earnings increase each year by the growth rate given in the table. Net capital expenditures (capital expenditures minus depreciation) are the amounts that Noronha assumed. The increase in working capital each year is 50% of the increase in net capital expenditures. Debt financing is 30% of the total outlays for net capital expenditures and working capital each year. The FCFE each year is net income minus net capital expenditures minus increase in working capital plus new debt financing. Finally, for years 2020 through 2023, the present value of FCFE is found by discounting FCFE by the 10.4% required rate of return for equity.

After 2024, FCFE will grow by a constant 7% annually, so the constant-growth FCFE valuation model can be used to value this cash flow stream. At the end of 2023, the value of the future FCFE is

$$V_{2016} = \frac{FCFE_{2017}}{r-g} = \frac{3.759}{0.104-0.07} = \$110.56 \text{ per share.}$$

To find the present value of V_{2016} as of the end of 2019, V_{2019}, we discount V_{2016} at 10.4% for four years:

$$PV = 110.56/(1.104)^4 = \$74.425 \text{ per share.}$$

The total present value of the company is the present value of the first four years' FCFE plus the present value of the terminal value, or

$$V_{2012} = -0.027 + 0.867 + 1.504 + 1.965 + 74.42 = \$78.73 \text{ per share.}$$

Solution to 2:

Using the estimated $78.73 stock value, we find that the trailing P/E at the beginning of 2020 is

P/E = 78.73/2.40 = 32.8.

At the beginning of 2024, the expected stock value is $110.56, and the previous year's EPS is $4.494, so the trailing P/E at this time would be

P/E = 110.56/4.494 = 24.6.

After its high-growth phase has ended, the P/E for the company declines substantially.

The FCFE in Example 16 was based on forecasts of future EPS. Analysts often model a company by forecasting future sales and then estimating the profits, investments, and financing associated with those sales levels. For large companies, analysts may estimate the sales, profitability, investments, and financing for each division or large subsidiary. Then, they aggregate the free cash flows for all of the divisions or subsidiaries to get the free cash flow for the company as a whole.

Example 17 is a two-stage FCFE model with declining sales growth rates in Stage 1, with profits, investments, and financing keyed to sales. In Stage 1, the growth rate of sales and the profit margin on sales both decline as the company matures and faces more competition and slower growth.

EXAMPLE 17

A Two-Stage FCFE Valuation Model with Declining Sales Growth Rates

Medina Werks, a manufacturing company headquartered in Canada, has a competitive advantage that will probably deteriorate over time. Analyst Flavio Torino expects this deterioration to be reflected in declining sales growth rates as well as declining profit margins. To value the company, Torino has accumulated the following information:

- Current sales are C$600 million. Over the next six years, the annual sales growth rate and the net profit margin are projected to be as follows:

	Year 1 (%)	Year 2 (%)	Year 3 (%)	Year 4 (%)	Year 5 (%)	Year 6 (%)
Sales growth rate	20	16	12	10	8	7
Net profit margin	14	13	12	11	10.5	10

  Beginning in Year 6, the 7% sales growth rate and 10% net profit margin should persist indefinitely.

- Capital expenditures (net of depreciation) in the amount of 60% of the sales increase will be required each year.
- Investments in working capital equal to 25% of the sales increase will also be required each year.
- Debt financing will be used to fund 40% of the investments in net capital items and working capital.

- The beta for Medina Werks is 1.10; the risk-free rate of return is 6.0%; the equity risk premium is 4.5%.
- The company has 70 million outstanding shares.

1 What is the estimated total market value of equity?
2 What is the estimated value per share?

Solution to 1:

The required return for Medina is

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 6\% + 1.10(4.5\%) = 10.95\%.$$

The annual sales and net profit can be readily found as shown in Exhibit 15.

Exhibit 15 FCFE Estimates for Medina Werks (C$ in Millions)

	Year					
	1	2	3	4	5	6
Sales growth rate	20%	16%	12%	10%	8%	7%
Net profit margin	14%	13%	12%	11%	10.50%	10%
Sales	720.000	835.200	935.424	1,028.966	1,111.284	1,189.074
Net profit	100.800	108.576	112.251	113.186	116.685	118.907
Net FCInv	72.000	69.120	60.134	56.125	49.390	46.674
WCInv	30.000	28.800	25.056	23.386	20.579	19.447
Debt financing	40.800	39.168	34.076	31.804	27.988	26.449
FCFE	39.600	49.824	61.137	65.480	74.703	79.235
PV of FCFE at 10.95%	35.692	40.475	44.763	43.211	44.433	

As can be seen, sales are expected to increase each year by a declining sales growth rate. Net profit each year is the year's net profit margin times the year's sales. Capital investment (net of depreciation) equals 60% of the sales increase from the previous year. The investment in working capital is 25% of the sales increase from the previous year. The debt financing each year is equal to 40% of the total net investment in capital items and working capital for that year. FCFE is net income minus the net capital investment minus the working capital investment plus the debt financing. The present value of each year's FCFE is found by discounting FCFE at the required rate of return for equity, 10.95%.

In Year 6 and beyond, Torino predicts sales to increase at 7% annually. Net income will be 10% of sales, so net profit will also grow at a 7% annual rate. Because they are pegged to the 7% sales increase, the investments in capital items and working capital and debt financing will also grow at the same 7% rate. The amounts in Year 6 for net income, investment in capital items, investment in working capital, debt financing, and FCFE will grow at 7%.

The terminal value of FCFE in Year 6 and beyond is

$$TV_5 = \frac{FCFE_6}{r - g} = \frac{79.235}{0.1095 - 0.07} = \text{C\$2,005.95 million.}$$

The present value of this amount is

$$\text{PV of TV}_5 = \frac{2{,}005.95}{(1.1095)^5} = \text{C\$1,193.12 million.}$$

The estimated total market value of the firm is the present value of FCFE for Years 1 through 5 plus the present value of the terminal value:

MV = 35.692 + 40.475 + 44.763 + 43.211 + 44.433 + 1,193.12

= C$1,401.69 million.

Solution to 2:

Dividing C$1,401.69 million by the 70 million outstanding shares gives the estimated value per share of C$20.02.

THREE-STAGE FREE CASH FLOW MODELS

12

i. explain the single-stage (stable-growth), two-stage, and three-stage FCFF and FCFE models and justify the selection of the appropriate model given a company's characteristics;

j. estimate a company's value using the appropriate free cash flow model(s);

l. describe approaches for calculating the terminal value in a multistage valuation model; and

Three-stage models are a straightforward extension of the two-stage models. One common version of a three-stage model is to assume a constant growth rate in each of the three stages. The growth rates could be for sales, profits, and investments in fixed and working capital; external financing could be a function of the level of sales or changes in sales. A simpler model would apply the growth rate to FCFF or FCFE.

A second common model is a three-stage model with constant growth rates in Stages 1 and 3 and a declining growth rate in Stage 2. Again, the growth rates could be applied to sales or to FCFF or FCFE. Although future FCFF and FCFE are unlikely to follow the assumptions of either of these three-stage growth models, analysts often find such models to be useful approximations.

Example 18 is a three-stage FCFF valuation model with declining growth rates in Stage 2. The model directly forecasts FCFF instead of deriving FCFF from a more complicated model that estimates cash flow from operations and investments in fixed and working capital.

EXAMPLE 18

A Three-Stage FCFF Valuation Model with Declining Growth in Stage 2

Charles Jones is evaluating Reliant Home Furnishings by using a three-stage growth model. He has accumulated the following information:

- Current FCFF = $745 million.
- Outstanding shares = 309.39 million.
- Equity beta = 0.90; risk-free rate = 5.04%; equity risk premium = 5.5%.
- Cost of debt = 7.1%.

- Marginal tax rate = 34%.
- Capital structure = 20% debt, 80% equity.
- Long-term debt = \$1.518 billion.
- Growth rate of FCFF =
  - 8.8% annually in Stage 1, Years 1–4.
  - 7.4% in Year 5, 6.0% in Year 6, 4.6% in Year 7.
  - 3.2% in Year 8 and thereafter.

From the information that Jones has accumulated, estimate the following:

1 WACC.
2 Total value of the firm.
3 Total value of equity.
4 Value per share.

Solution to 1:

The required return for equity is

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 5.04\% + 0.9(5.5\%) = 9.99\%.$$

WACC is

$$\text{WACC} = 0.20(7.1\%)(1 - 0.34) + 0.80(9.99\%) = 8.93\%.$$

Solution to 2:

Exhibit 16 displays the projected FCFF for the next eight years and the present value of each FCFF discounted at 8.93%:

Exhibit 16 Forecasted FCFF for Reliant Home Furnishings

	Year							
	1	2	3	4	5	6	7	8
Growth rate	8.80%	8.80%	8.80%	8.80%	7.40%	6.00%	4.60%	3.20%
FCFF	811	882	959	1,044	1,121	1,188	1,243	1,283
PV at 8.93%	744	743	742	741	731	711	683	

The terminal value at the end of Year 7 is

$$\text{TV}_7 = \frac{\text{FCFF}_8}{\text{WACC} - g} = \frac{1{,}283}{0.0893 - 0.032} = \$22{,}391 \text{ million.}$$

The present value of this amount discounted at 8.93% for seven years is

$$\text{PV of TV}_7 = \frac{22{,}391}{(1.0893)^7} = \$12{,}304 \text{ million.}$$

The total present value of the first seven years of FCFF is \$5,097 million. The total value of the firm is 12,304 + 5,097 = \$17,401 million.

Solution to 3:

The value of equity is the value of the firm minus the market value of debt:

17,401 – 1,518 = \$15,883 million.

Solution to 4:

Dividing the equity value by the number of shares yields the value per share:

$15,883 million/309.39 million = $51.34.

INTEGRATING ESG IN FREE CASH FLOW MODELS

13

- **i.** explain the single-stage (stable-growth), two-stage, and three-stage FCFF and FCFE models and justify the selection of the appropriate model given a company's characteristics;
- **j.** estimate a company's value using the appropriate free cash flow model(s);
- **l.** describe approaches for calculating the terminal value in a multistage valuation model; and
- **m.** evaluate whether a stock is overvalued, fairly valued, or undervalued based on a free cash flow valuation model.

Integrating environmental, social, and governance (ESG) considerations in valuation models can have a material impact on valuation. ESG factors may be either quantitative or qualitative. Quantitative ESG-related information, such as the effect of a projected environmental fine on cash flows, is more straightforward to integrate in valuation models. By contrast, qualitative ESG-related information is more challenging to integrate. One approach to address this challenge is to adjust the cost of equity by adding a risk premium in a valuation model. This approach can estimate the effect of ESG-related issues that are deemed material by an analyst but are difficult to quantify. When making an adjustment to the cost of equity by adding a risk premium, the analyst relies on his or her judgment to determine what value constitutes a reasonable adjustment. Example 19 provides a case study of how an analyst may develop a multistage (three-stage, in this case) FCFF valuation model that integrates ESG considerations.

EXAMPLE 19

Integrating ESG in a Three-Stage FCFF Model

American Copper Mining Company (ACMC) is a large US-based company. Copper has many uses in manufacturing, building, and other industries. The mining of copper is resource-intensive and is highly regulated.

ACMC recently announced that it is acquiring a new copper mine in a very dry region of Latin America. After the announcement, the market welcomed the news, and ACMC's share price rose to its current level of US$110 per share. The company expects the new mine to have a useful life of approximately 15 years.

Jane Dodd is a research analyst who follows ACMC and has a "hold" rating on its shares. She is preparing a new report to determine whether ACMC's acquisition of the new copper mine changes her fundamental assessment of the company. Overall, Dodd believes that the evaluation of ESG considerations can provide critical insights into the feasibility, economics, and valuation of mining companies and mining projects.

Dodd begins her analysis by evaluating the current political, labor, and environmental situation for ACMC's new mine. She has identified three primary ESG considerations that, in her opinion, may have the greatest effects on the value of the new mine and the company:

1 Local government issues
2 Labor issues
3 Water-related issues

Dodd then assesses how each of these ESG considerations may affect ACMC's operations and cash flow.

1 *Local government issues*: To operate the new mine, ACMC must obtain a mining license from the local government in the region where the mine is located. Before obtaining the mining license, ACMC is required to submit a comprehensive rehabilitation plan indicating how the new mine's natural habitat will be restored. Dodd notices that in its other mining sites, ACMC has struggled to produce comprehensive rehabilitation plans that have been approved by government authorities in a timely manner. She concludes that ACMC is overly optimistic about the time required to get approval for the mining license. She expects that rather than three years, as management anticipates, it will likely take five years before the mine can begin operating.
2 *Labor issues*: ACMC's compensation of its employees is slightly lower than its competitors in the region of the new mine. In addition, unlike many of its competitors, ACMC does not tie executive compensation to worker safety. Some competitors in the region have experienced labor strikes (and thus production interruptions) because their employees' wages are not adjusted for inflation. Because of ACMC's compensation policies, Dodd is concerned about the potential for labor unrest and subsequent reputational risk for the company.
3 *Water-related issues*: Because a large volume of water is used for mining operations, water-related costs are typically among the largest expenditures for mining companies. Given that the development of the new mine is located in a very dry region of Latin America, Dodd believes that ACMC has significantly underestimated the required capital expenditures necessary to build water wells.

Valuation Analysis

After identifying and assessing these ESG considerations, Dodd proceeds to value ACMC's share price using a three-stage FCFF model. The three stages are as follows:

- Stage 1: the period prior to expected operation of the new mine (2020–2024)
- Stage 2: the period during expected operation of the new mine (2025–2039)
- Stage 3: the period subsequent to the expected closing of the mine (2040 and onward)

Dodd makes the following assumptions in her model.

Revenues

ACMC's total revenues during 2020 were $1 billion. Dodd expects total revenues (i.e., excluding those of the new mine) to increase 2% annually through 2024 and then remain constant during 2025–2039, when the new mine operates. When the new mine begins operations under Dodd's assumption (in 2025), Dodd expects the mine to add US$400 million to ACMC's revenues in its first year. Dodd also expects that these additional revenues from the new mine will increase by 10% annually for the next six years (2026 through 2031) and then remain constant for the remaining life of the mine (2032 through 2039). Dodd assumes that once the new mine closes in 2039, the company's total revenues will grow by 1% in perpetuity. The following is a summary of revenues for the three stages:

Stage 1 (prior to expected operation of mine):

Years 2020–2024: annual total revenue growth of 2%

Stage 2 (during expected operation of new mine):

2025: constant growth of revenues excluding the new mine; additional revenue of US$400 million from new mine

2026–2039: constant growth of revenues excluding new mine during years 2026–2039); 10% annual growth of revenue from new mine during years 2026–2031; constant growth of revenues from new mine during years 2032–2039

Stage 3 (after expected closing of new mine):

2040 and beyond: annual total revenue growth of 1%

Dodd also makes the following financial assumptions for ACMC:

EBITDA:	30% of total revenues for all three stages
Taxes:	25%
Investment in fixed capital (not including water-related investments):	50% of EBITDA for all three stages
Depreciation:	40% of capital expenditures for all three stages
Investment in working capital:	10% of total revenue for all three stages
Required return (pretax) on ACMC debt:	5%
Risk-free rate:	3%
ACMC equity beta:	1.2
Equity risk premium:	5%
Debt ratio:	50%

In addition to these "traditional" financial assumptions, Dodd also reflects ESG considerations in her analysis.

Water-related investment in fixed capital

10% of non-water-related capital expenditures, which are added to the capital expenditures noted previously.

ESG equity risk premium adjustment

Dodd concludes that the potential for labor issues discussed earlier exposes ACMC to higher financial and reputational risk compared to its peers. Dodd further believes that the ESG considerations she has identified are not recognized fully in the market price of ACMC shares. As a result, Dodd estimates that a 75 basis point premium should be added to ACMC's cost of equity.

Dodd calculates the WACC as follows:

Cost of debt = (5%)(1 – 25%) = 3.75%.

Cost of equity = 3% + (1.2)(5%) + 0.75% ESG equity risk premium adjustment = 9.75%.

WACC = (0.5)(3.75%) + (0.5)(9.75%) = 6.75%.

Exhibit 17 presents the results of Dodd's model for valuing ACMC's equity. Dodd's analysis suggests that the fair value for ACMC's equity is $97 per share. By integrating ESG considerations in a traditional valuation framework, Dodd's estimate of the fair value of ACMC's shares decreased. Given that the stock is trading at US$110, she issues a "sell" recommendation for ACMC's shares.

The next section discusses an important technical issue, the treatment of non-operating assets in valuation.

14 NON-OPERATING ASSETS AND FIRM VALUE

j estimate a company's value using the appropriate free cash flow model(s);

Free cash flow valuation focuses on the value of assets that generate or are needed to generate operating cash flows. If a company has significant non-operating assets, such as excess cash (excess in relation to what is needed for generating operating cash flows), excess marketable securities, or land held for investment, then analysts often calculate the value of the firm as the value of its operating assets (e.g., as estimated by FCFF valuation) plus the value of its non-operating assets:

$$\begin{aligned}\text{Value of firm} = &\ \text{Value of operating assets}\\ &+\text{Value of non-operating assets.}\end{aligned} \quad (18)$$

In general, if any company asset is excluded from the set of assets being considered in projecting a company's future cash flows, the analyst should add that omitted asset's estimated value to the cash flow–based value estimate. Some companies have substantial noncurrent investments in stocks and bonds that are not operating subsidiaries but, rather, financial investments. These investments should be reflected at their current market value. Those securities reported at book values on the basis of accounting conventions should be revalued to market values.

SUMMARY

Discounted cash flow models are widely used by analysts to value companies.

- Free cash flow to the firm (FCFF) and free cash flow to equity (FCFE) are the cash flows available to, respectively, all of the investors in the company and to common stockholders.
- Analysts like to use free cash flow (either FCFF or FCFE) as the return
  - if the company is not paying dividends;
  - if the company pays dividends but the dividends paid differ significantly from the company's capacity to pay dividends;

PART OF EXAMPLE 19

Exhibit 17 Estimating Fair Value of ACMC Shares (in Millions of US Dollars, Except for Per-Share Items)

	Expected Operation of New Mine													Expected Closing of Mine	
	2020	**2021**	**2022**	**2023**	**2024**	**2025**	**2026**	**2027**	**...**	**2030**	**2031**	**...**	**2034**	**2035**	**2036**
Total revenues	1,000	1,020	1,040	1,061	1,082	1,482	1,522	1,566	...	1,727	1,791	...	1,791	1,791	1,809
Revenues from new mine only						400.0	440.0	484.0		644.2	708.6		708.6	708.6	
EBITDA	300.0	306.0	312.1	318.4	324.7	444.7	456.7	469.9	...	518.0	537.3	...	537.3	537.3	542.7
EBITDA(1 – Tax rate)	225.0	229.5	234.1	238.8	243.5	333.5	342.5	352.4		388.5	403.0		403.0	403.0	407.0
Depreciation(Tax rate)	15.0	15.3	15.6	15.9	16.2	22.2	22.8	23.5	...	25.9	26.9	...	26.9	26.9	27.1
Investment in fixed capital, or FCInv	(150.0)	(153.0)	(156.1)	(159.2)	(162.4)	(222.4)	(228.4)	(235.0)	...	(259.0)	(268.7)	...	(268.7)	(268.7)	(271.3)
Investment in working capital, or WCInv*	(2.0)	(2.0)	(2.0)	(2.1)	(2.1)	(40.0)	(4.0)	(4.4)		(5.9)	(6.4)		0.0	0.0	(1.8)
Additional FCInv (water-related)	—	—	—	—	—	(22.2)	(22.8)	(23.5)	...	(25.9)	(26.9)	...	(26.9)	(26.9)	
FCFF	**88.0**	**89.8**	**91.6**	**93.4**	**95.3**	**71.2**	**110.2**	**113.1**	**...**	**123.6**	**127.9**	**...**	**134.3**	**134.3**	**161.0**
PV of FCFF up to 2039 (@ WACC of 6.75%)	82.5	78.8	75.3	71.9	68.7	48.1	69.7	67.1	...	60.3	58.4	...	38.8	36.4	
PV of FCFF Years 2016–2035	1,178														
PV of perpetual FCFF 2036 onward	758														
Total PV of future FCFF	**1,936**														
Market value of debt (50% debt ratio)	968														
Fair value of equity	968														

(continued)

PART OF EXAMPLE 19

Exhibit 17 (Continued)

							Expected Operation of New Mine						Expected Closing of Mine	
Fair value per share (10 million shares outstanding)	97													
Supplemental items:														
Depreciation	60.0	61.2	62.4	63.7	65.0	89.0	91.4	94.0	103.6	107.5	...	107.5	107.5	63.7
Working capital	100.0	102.0	104.0	106.1	108.2	148.2	152.2	156.6	172.7	179.1	...	179.1	179.1	180.9

**Note:* The 2020 investment in working capital (WCInv) reflects the change in working capital from 2019 to 2020. For simplicity, Dodd uses the same change in WCInv for 2020 as in 2021.

- if free cash flows align with profitability within a reasonable forecast period with which the analyst is comfortable; or
- if the investor takes a control perspective.

■ The FCFF valuation approach estimates the value of the firm as the present value of future FCFF discounted at the weighted average cost of capital:

$$\text{Firm value} = \sum_{t=1}^{\infty} \frac{\text{FCFF}_t}{(1 + \text{WACC})^t}.$$

The value of equity is the value of the firm minus the value of the firm's debt:

Equity value = Firm value – Market value of debt.

Dividing the total value of equity by the number of outstanding shares gives the value per share.

The WACC formula is

$$\text{WACC} = \frac{\text{MV(Debt)}}{\text{MV(Debt)} + \text{MV(Equity)}} r_d (1 - \text{Tax rate}) + \frac{\text{MV(Equity)}}{\text{MV(Debt)} + \text{MV(Equity)}} r.$$

■ The value of the firm if FCFF is growing at a constant rate is

$$\text{Firm value} = \frac{\text{FCFF}_1}{\text{WACC} - g} = \frac{\text{FCFF}_0(1 + g)}{\text{WACC} - g}.$$

■ With the FCFE valuation approach, the value of equity can be found by discounting FCFE at the required rate of return on equity, r:

$$\text{Equity value} = \sum_{t=1}^{\infty} \frac{\text{FCFE}_t}{(1 + r)^t}.$$

Dividing the total value of equity by the number of outstanding shares gives the value per share.

■ The value of equity if FCFE is growing at a constant rate is

$$\text{Equity value} = \frac{\text{FCFE}_1}{r - g} = \frac{\text{FCFE}_0(1 + g)}{r - g}.$$

■ FCFF and FCFE are frequently calculated by starting with net income:

FCFF = NI + NCC + Int(1 – Tax rate) – FCInv – WCInv.

FCFE = NI + NCC – FCInv – WCInv + Net borrowing.

■ FCFF and FCFE are related to each other as follows:

FCFE = FCFF – Int(1 – Tax rate) + Net borrowing.

■ FCFF and FCFE can be calculated by starting from cash flow from operations:

FCFF = CFO + Int(1 – Tax rate) – FCInv.

FCFE = CFO – FCInv + Net borrowing.

■ FCFF can also be calculated from EBIT or EBITDA:

FCFF = EBIT(1 – Tax rate) + Dep – FCInv – WCInv.

$$\text{FCFF} = \text{EBITDA}(1 - \text{Tax rate}) + \text{Dep}(\text{Tax rate}) - \text{FCInv} - \text{WCInv}.$$

FCFE can then be found by using FCFE = FCFF – Int(1 – Tax rate) + Net borrowing.

- Finding CFO, FCFF, and FCFE may require careful interpretation of corporate financial statements. In some cases, the necessary information may not be transparent.
- Earnings components such as net income, EBIT, EBITDA, and CFO should not be used as cash flow measures to value a firm. These earnings components either double-count or ignore parts of the cash flow stream.
- FCFF or FCFE valuation expressions can be easily adapted to accommodate complicated capital structures, such as those that include preferred stock.
- A general expression for the two-stage FCFF valuation model is

$$\text{Firm value} = \sum_{t=1}^{n}\frac{\text{FCFF}_t}{(1+\text{WACC})^t} + \frac{\text{FCFF}_{n+1}}{(\text{WACC}-g)}\frac{1}{(1+\text{WACC})^n}.$$

- A general expression for the two-stage FCFE valuation model is

$$\text{Equity value} = \sum_{t=1}^{n}\frac{\text{FCFE}_t}{(1+r)^t} + \left(\frac{\text{FCFE}_{n+1}}{r-g}\right)\left[\frac{1}{(1+r)^n}\right].$$

- One common two-stage model assumes a constant growth rate in each stage, and a second common model assumes declining growth in Stage 1 followed by a long-run sustainable growth rate in Stage 2.
- To forecast FCFF and FCFE, analysts build a variety of models of varying complexity. A common approach is to forecast sales, with profitability, investments, and financing derived from changes in sales.
- Three-stage models are often considered to be good approximations for cash flow streams that, in reality, fluctuate from year to year.
- Non-operating assets, such as excess cash and marketable securities, noncurrent investment securities, and nonperforming assets, are usually segregated from the company's operating assets. They are valued separately and then added to the value of the company's operating assets to find total firm value.

REFERENCES

Bahnson, P., P. B. Miller, and B. P. Budge. 1996. "Nonarticulation in Cash Flow Statements and Implications for Education, Research and Practice." *Accounting Horizons* 10 (4): 1–15.

Casey, Ryan J., Feng Gao, Michael T. Kirschenheiter, Siyi Li, and Shailendra Pandit. 2016. "Do Compustat Financial Statement Data Articulate?" *Journal of Financial Reporting* 1 (1): 37–59.

Gordon, Elizabeth A., Elaine Henry, Bjorn N. Jorgensen, and Cheryl L. Linthicum. 2017. "Flexibility in Cash-Flow Classification under IFRS: Determinants and Consequences." *Review of Accounting Studies* 22 (2): 839–72.

Hribar, P., and D. W. Collins. 2002. "Errors in Estimating Accruals: Implications for Empirical Research." *Journal of Accounting Research* 40 (1): 105–34.

Huefner, R. J., J. E. Ketz, and J. A. Largay. 1989. "Foreign Currency Translation and the Cash Flow Statement." *Accounting Horizons* 3 (2): 66–75.

Luehrman, T. A. 1997. "Using APV (Adjusted Present Value): A Better Tool for Valuing Operations." *Harvard Business Review* 75 (3): 145–46, 148, 150–54.

Pinto, Jerald E., Thomas R. Robinson, and John D. Stowe. 2019. "Equity Valuation: A Survey of Professional Practice." *Review of Financial Economics* 37 (2): 219–33.

Shi, Linna, and Huai Zhang. 2011. "On Alternative Measures of Accruals." *Accounting Horizons* 25 (4): 811–36.

Wilkins, M. S., and M. L. Loudder. 2000. "Articulation in Cash Flow Statements: A Resource for Financial Accounting Courses." *Journal of Accounting Education* 18:115–26.

PRACTICE PROBLEMS

1 Indicate the effect on this period's FCFF and FCFE of a change in each of the items listed here. Assume a $100 increase in each case and a 40% tax rate.

A Net income.

B Cash operating expenses.

C Depreciation.

D Interest expense.

E EBIT.

F Accounts receivable.

G Accounts payable.

H Property, plant, and equipment.

I Notes payable.

J Cash dividends paid.

K Proceeds from issuing new common shares.

L Common shares repurchased.

2 LaForge Systems, Inc., has net income of $285 million for the year 2020. Using information from the company's financial statements given here, show the adjustments to net income that would be required to find:

A FCFF.

B FCFE.

C In addition, show the adjustments to FCFF that would result in FCFE.

LaForge Systems, Inc., Balance Sheet (in Millions)

Years Ended 31 December	2019	2020
Assets		
Current assets		
Cash and equivalents	$210	$248
Accounts receivable	474	513
Inventory	520	564
Total current assets	1,204	1,325
Gross fixed assets	2,501	2,850
Accumulated depreciation	(604)	(784)
Net fixed assets	1,897	2,066
Total assets	$3,101	$3,391
Liabilities and shareholders' equity		
Current liabilities		
Accounts payable	$295	$317
Notes payable	300	310
Accrued taxes and expenses	76	99

(continued)

© 2018 CFA Institute. All rights reserved.

(Continued)

Years Ended 31 December	2019	2020
Total current liabilities	671	726
Long-term debt	1,010	1,050
Common stock	50	50
Additional paid-in capital	300	300
Retained earnings	1,070	1,265
Total shareholders' equity	1,420	1,615
Total liabilities and shareholders' equity	$3,101	$3,391

Statement of Income In Millions, except Per-Share Data	31 December 2020
Total revenues	$2,215
Operating costs and expenses	1,430
EBITDA	785
Depreciation	180
EBIT	605
Interest expense	130
Income before tax	475
Taxes (at 40%)	190
Net income	285
Dividends	90
Addition to retained earnings	195

Statement of Cash Flows In Millions	31 December 2020
Operating activities	
Net income	$285
Adjustments	
Depreciation	180
Changes in working capital	
Accounts receivable	(39)
Inventories	(44)
Accounts payable	22
Accrued taxes and expenses	23
Cash provided by operating activities	$427
Investing activities	
Purchases of fixed assets	349
Cash used for investing activities	$349
Financing activities	

(Continued)

Statement of Cash Flows In Millions	**31 December 2020**
Notes payable	$(10)
Long-term financing issuances	(40)
Common stock dividends	90
Cash used for financing activities	$40
Cash and equivalents increase (decrease)	38
Cash and equivalents at beginning of year	210
Cash and equivalents at end of year	$248
Supplemental cash flow disclosures	
Interest paid	$130
Income taxes paid	$190

Note: The statement of cash flows shows the use of a convention by which the positive numbers of $349 and $40 for cash used for investing activities and cash used for financing activities, respectively, are understood to be subtractions, because "cash used" is an outflow.

3 For LaForge Systems, whose financial statements are given in Problem 2, show the adjustments from the current levels of CFO (which is $427 million), EBIT ($605 million), and EBITDA ($785 million) to find:

A FCFF.

B FCFE.

4 The term "free cash flow" is frequently applied to cash flows that differ from the definition for FCFF that should be used to value a firm. Two such definitions of free cash flow are given below. Compare these two definitions for free cash flow with the technically correct definition of FCFF used in our coverage of the topic.

A FCF = Net income + Depreciation and amortization – Cash dividends – Capital expenditures.

B FCF = Cash flow from operations (from the statement of cash flows) – Capital expenditures.

5 Shimotsuke Co. LTD. has FCFF of 1.7 billion Japanese yen (JPY) and FCFE of JPY1.3 billion. Shimotsuke Co.'s WACC is 11%, and its required rate of return for equity is 13%. FCFF is expected to grow forever at 7%, and FCFE is expected to grow forever at 7.5%. Shimotsuke Co. has debt outstanding of JPY15 billion.

A What is the total value of Shimotsuke Co.'s equity using the FCFF valuation approach?

B What is the total value of Shimotsuke Co.'s equity using the FCFE valuation approach?

6 Hugo Dubois is evaluating NYL Manufacturing Company, Ltd. In 2020, when Dubois is performing his analysis, the company is unprofitable. Furthermore, NYL pays no dividends on its common shares. Dubois decides to value NYL Manufacturing by using his forecasts of FCFE. Dubois gathers the following facts and assumptions:

- The company has 17.0 billion shares outstanding.
- Sales will be €5.5 billion in 2021, increasing at 28% annually for the next four years (through 2025).
- Net income will be 32% of sales.
- Investment in fixed assets will be 35% of sales; investment in working capital will be 6% of sales; depreciation will be 9% of sales.
- 20% of the net investment in assets will be financed with debt.
- Interest expenses will be only 2% of sales.
- The tax rate will be 10%. NYL Manufacturing's beta is 2.1; the risk-free government bond rate is 6.4%; the equity risk premium is 5.0%.
- At the end of 2025, Dubois projects NYL terminal stock value at 18 times earnings.

What is the value of one ordinary share of NYL Manufacturing Company?

7 Do Pham is evaluating Phaneuf Accelerateur by using the FCFF and FCFE valuation approaches. Pham has collected the following information (currency in euros):

- Phaneuf has net income of €250 million, depreciation of €90 million, capital expenditures of €170 million, and an increase in working capital of €40 million.
- Phaneuf will finance 40% of the increase in net fixed assets (capital expenditures less depreciation) and 40% of the increase in working capital with debt financing.
- Interest expenses are €150 million. The current market value of Phaneuf's outstanding debt is €1,800 million.
- FCFF is expected to grow at 6.0% indefinitely, and FCFE is expected to grow at 7.0%.
- The tax rate is 30%.
- Phaneuf is financed with 40% debt and 60% equity. The before-tax cost of debt is 9%, and the before-tax cost of equity is 13%.
- Phaneuf has 10 million outstanding shares.

A Using the FCFF valuation approach, estimate the total value of the firm, the total market value of equity, and the per-share value of equity.

B Using the FCFE valuation approach, estimate the total market value of equity and the per-share value of equity.

8 PHB Company currently sells for £32.50 per share. In an attempt to determine whether PHB is fairly priced, an analyst has assembled the following information:

- The before-tax required rates of return on PHB debt, preferred stock, and common stock are, respectively, 7.0%, 6.8%, and 11.0%.
- The company's target capital structure is 30% debt, 15% preferred stock, and 55% common stock.
- The market value of the company's debt is £145 million, and its preferred stock is valued at £65 million.
- PHB's FCFF for the year just ended is £28 million. FCFF is expected to grow at a constant rate of 4% for the foreseeable future.
- The tax rate is 35%.
- PHB has 8 million outstanding common shares.

What is PHB's estimated value per share? Is PHB's stock underpriced?

9 Elina Kuznetsova is planning to value BCC Corporation, a provider of a variety of industrial metals and minerals. Kuznetsova uses a single-stage FCFF approach. The financial information Kuznetsova has assembled for her valuation is as follows:

- The company has 1,852 million shares outstanding.
- The market value of its debt is \$3.192 billion.
- The FCFF is currently \$1.1559 billion.
- The equity beta is 0.90; the equity risk premium is 5.5%; the risk-free rate is 5.5%.
- The before-tax cost of debt is 7.0%.
- The tax rate is 40%.
- To calculate WACC, he will assume the company is financed 25% with debt.
- The FCFF growth rate is 4%.

Using Kuznetsova's information, calculate the following:

A WACC.

B Value of the firm.

C Total market value of equity.

D Value per share.

10 James Smith is valuing McInish Corporation and performing a sensitivity analysis on his valuation. He uses a single-stage FCFE growth model. The base-case values for each of the parameters in the model are given, together with possible low and high estimates for each variable, in the following table.

Variable	Base-Case Value	Low Estimate	High Estimate
Normalized $FCFE_0$	£0.88	£0.70	£1.14
Risk-free rate	5.08%	5.00%	5.20%
Equity risk premium	5.50%	4.50%	6.50%
Beta	0.70	0.60	0.80
FCFE growth rate	6.40%	4.00%	7.00%

A Use the base-case values to estimate the current value of McInish Corporation.

B Calculate the range of stock prices that would occur if the base-case value for $FCFE_0$ were replaced by the low estimate and the high estimate for $FCFE_0$. Similarly, using the base-case values for all other variables, calculate the range of stock prices caused by using the low and high values for beta, the risk-free rate, the equity risk premium, and the growth rate. Based on these ranges, rank the sensitivity of the stock price to each of the five variables.

11 An aggressive financial planner who claims to have a superior method for picking undervalued stocks is trying to steal one of your clients. The planner claims that the best way to find the value of a stock is to divide EBITDA by the risk-free bond rate. The planner is urging your client to invest in NewMarket, Inc. The planner says that NewMarket's EBITDA of \$1,580 million divided by the long-term government bond rate of 7% gives a total value of \$22,571.4 million. With 318 million outstanding shares, NewMarket's value per share found by using this method is \$70.98. Shares of NewMarket currently trade for \$36.50.

A Provide your client with an alternative estimate of NewMarket's value per share based on a two-stage FCFE valuation approach. Use the following assumptions:

- Net income is currently $600 million. Net income will grow by 20% annually for the next three years.
- The net investment in operating assets (capital expenditures less depreciation plus investment in working capital) will be $1,150 million next year and grow at 15% for the following two years.
- 40% of the net investment in operating assets will be financed with net new debt financing.
- NewMarket's beta is 1.3; the risk-free bond rate is 7%; the equity risk premium is 4%.
- After three years, the growth rate of net income will be 8% and the net investment in operating assets (capital expenditures minus depreciation plus increase in working capital) each year will drop to 30% of net income.
- Debt is, and will continue to be, 40% of total assets.
- NewMarket has 318 million shares outstanding.

B Criticize the valuation approach that the aggressive financial planner used.

12 Bron has EPS of $3.00 in 2019 and expects EPS to increase by 21% in 2020. EPS are expected to grow at a decreasing rate for the following five years, as shown in the following table.

	2020	2021	2022	2023	2024	2025
Growth rate for EPS	21%	18%	15%	12%	9%	6%
Net capital expenditures per share	$5.00	$5.00	$4.50	$4.00	$3.50	$1.50

In 2025, the growth rate will be 6%, and it is expected to stay at that rate thereafter. Net capital expenditures (capital expenditures minus depreciation) will be $5.00 per share in 2019 and then follow the pattern predicted in the table. In 2025, net capital expenditures are expected to be $1.50, and they will then grow at 6% annually. The investment in working capital parallels the increase in net capital expenditures and is predicted to equal 25% of net capital expenditures each year. In 2025, investment in working capital will be $0.375, and it is predicted to grow at 6% thereafter. Bron will use debt financing to fund 40% of net capital expenditures and 40% of the investment in working capital. The required rate of return for Bron is 12%.

Estimate the value of a Bron share using a two-stage FCFE valuation approach.

13 The management of Telluride, an international diversified conglomerate, believes that the recent strong performance of its wholly owned medical supply subsidiary, Sundanci, has gone unnoticed. To realize Sundanci's full value, Telluride has announced that it will divest Sundanci in a tax-free spin-off.

Sue Carroll is director of research at Kesson and Associates. In developing an investment recommendation for Sundanci, Carroll has gathered the information shown in Exhibits 1 and 2.

Exhibit 1 Sundanci Actual 2019 and 2020 Financial Statements for Fiscal Years Ending 31 May (Dollars in Millions except Per-Share Data)

Income Statement	2019	2020
Revenue	$474	$598
Depreciation	20	23
Other operating costs	368	460
Income before taxes	86	115
Taxes	26	35
Net income	60	80
Dividends	18	24
EPS	$0.714	$0.952
Dividends per share	$0.214	$0.286
Common shares outstanding	84.0	84.0

Balance Sheet	2019	2020
Current assets (includes $5 cash in 2007 and 2008)	$201	$326
Net property, plant, and equipment	474	489
Total assets	675	815
Current liabilities (all non-interest-bearing)	57	141
Long-term debt	0	0
Total liabilities	57	141
Shareholders' equity	618	674
Total liabilities and equity	675	815
Capital expenditures	34	38

Exhibit 2 Selected Financial Information

Required rate of return on equity	14%
Industry growth rate	13%
Industry P/E	26

Abbey Naylor has been directed by Carroll to determine the value of Sundanci's stock by using the FCFE model. Naylor believes that Sundanci's FCFE will grow at 27% for two years and at 13% thereafter. Capital expenditures, depreciation, and working capital are all expected to increase proportionately with FCFE.

A Calculate the amount of FCFE per share for 2020 by using the data from Exhibit 1.

B Calculate the current value of a share of Sundanci stock based on the two-stage FCFE model.

C Describe limitations that the two-stage DDM and FCFE models have in common.

14 John Jones is head of the research department of Peninsular Research and is estimating the value of Mackinac Inc. The company has released its June 2019 financial statements, shown in Exhibits 3, 4, and 5.

Exhibit 3 Mackinac Inc. Annual Income Statement 30 June 2019 (in Thousands, except Per-Share Data)

Sales	$250,000
Cost of goods sold	125,000
Gross operating profit	125,000
Selling, general, and administrative expenses	50,000
EBITDA	75,000
Depreciation and amortization	10,500
EBIT	64,500
Interest expense	11,000
Pretax income	53,500
Income taxes	16,050
Net income	$37,450
Shares outstanding	13,000
EPS	$2.88

Exhibit 4 Mackinac Inc. Balance Sheet 30 June 2019 (in Thousands)

Current Assets			
Cash and equivalents		$20,000	
Receivables		40,000	
Inventories		29,000	
Other current assets		23,000	
Total current assets			$112,000
Noncurrent Assets			
Property, plant, and equipment	$145,000		
Less: Accumulated depreciation	43,000		
Net property, plant, and equipment		102,000	
Investments		70,000	
Other noncurrent assets		36,000	
Total noncurrent assets			208,000
Total assets			$320,000
Current Liabilities			

Exhibit 4 (Continued)

Accounts payable	$41,000	
Short-term debt	12,000	
Other current liabilities	17,000	
Total current liabilities		$ 70,000
Noncurrent Liabilities		
Long-term debt	100,000	
Total noncurrent liabilities		100,000
Total liabilities		170,000
Shareholders' Equity		
Common equity	40,000	
Retained earnings	110,000	
Total equity		150,000
Total liabilities and equity		$320,000

Exhibit 5 Mackinac Inc. Statement of Cash Flows 30 June 2019 (in Thousands)

Cash Flow from Operating Activities		
Net income		$37,450
Depreciation and amortization		10,500
Change in Working Capital		
(Increase) decrease in receivables	($5,000)	
(Increase) decrease in inventories	(8,000)	
Increase (decrease) in payables	6,000	
Increase (decrease) in other current liabilities	1,500	
Net change in working capital		(5,500)
Net cash from operating activities		$42,450
Cash Flow from Investing Activities		
Purchase of property, plant, and equipment	($15,000)	
Net cash from investing activities		($15,000)
Cash Flow from Financing Activities		
Change in debt outstanding	$4,000	
Payment of cash dividends	(22,470)	
Net cash from financing activities		(18,470)
Net change in cash and cash equivalents		$8,980
Cash at beginning of period		11,020
Cash at end of period		$20,000

Mackinac has announced that it has finalized an agreement to handle North American production of a successful product currently marketed by a company headquartered outside North America. Jones decides to value Mackinac by using the DDM and FCFE models. After reviewing Mackinac's financial statements and forecasts related to the new production agreement, Jones concludes the following:

- Mackinac's earnings and FCFE are expected to grow 17% a year over the next three years before stabilizing at an annual growth rate of 9%.
- Mackinac will maintain the current payout ratio.
- Mackinac's beta is 1.25.
- The government bond yield is 6%, and the market equity risk premium is 5%.

A Calculate the value of a share of Mackinac's common stock by using the two-stage DDM.

B Calculate the value of a share of Mackinac's common stock by using the two-stage FCFE model.

C Jones is discussing with a corporate client the possibility of that client acquiring a 70% interest in Mackinac. Discuss whether the DDM or FCFE model is more appropriate for this client's valuation purposes.

15 KMobile Telecom is an Asian mobile network operator headquartered in Seoul, South Korea. Sol Kim has estimated the normalized FCFE per share for KMobile to be 1,300 Korean won (KRW) for the year just ended. The real country return for South Korea is 6.50%. To estimate the required return for KMobile, Kim makes the following adjustments to the real country return: an industry adjustment of +0.60%, a size adjustment of –0.10%, and a leverage adjustment of +0.25%. The long-term real growth rate for South Korea is estimated to be 3.5%, and Kim expects the real growth rate of KMobile to track the country rate.

A What is the real required rate of return for KMobile Telecom?

B Using the single-stage FCFE valuation model and real values for the discount rate and FCFE growth rate, estimate the value of one share of KMobile.

16 Minsuh Park is preparing a valuation of QuickChange Auto Centers, Inc. Park has decided to use a three-stage FCFE valuation model and the following estimates. The FCFE per share for the current year is \$0.75. The FCFE is expected to grow at 10% for next year, then at 26% annually for the following three years, and then at 6% in Year 5 and thereafter. QuickChange's estimated beta is 2.00, and Park believes that current market conditions dictate a 4.5% risk-free rate of return and a 5.0% equity risk premium. Given Park's assumptions and approach, estimate the value of a share of QuickChange.

17 Astrid Nilsson has valued the operating assets of Gothenburg Extrusion AB at 720 million Swedish kronor (SEK). The company also has short-term cash and securities with a market value of SEK60 million that are not needed for Gothenburg's operations. The noncurrent investments have a book value of SEK30 million and a market value of SEK45 million. The company also has an overfunded pension plan, with plan assets of SEK210 million and plan liabilities of SEK170 million. Gothenburg Extrusion has SEK215 million of notes and bonds outstanding and 100 million outstanding shares. What is the value per share of Gothenburg Extrusion stock?

The following information relates to Questions 18–23

Ryan Leigh is preparing a presentation that analyzes the valuation of the common stock of two companies under consideration as additions to his firm's recommended list, Emerald Corporation and Holt Corporation. Leigh has prepared preliminary valuations of both companies using an FCFE model and is also preparing a value estimate for Emerald using a dividend discount model. Holt's 2019 and 2020 financial statements, contained in Exhibits 1 and 2, are prepared in accordance with US GAAP.

Exhibit 1 Holt Corporation Consolidated Balance Sheets (US$ Millions)

	As of 31 December			
		2020		2019
Assets				
Current assets				
Cash and cash equivalents		$ 372		$ 315
Accounts receivable		770		711
Inventories		846		780
Total current assets		1,988		1,806
Gross fixed assets	4,275		3,752	
Less: Accumulated depreciation	1,176	3,099	906	2,846
Total assets		$5,087		$4,652
Liabilities and shareholders' equity				
Current liabilities				
Accounts payable		$ 476		$ 443
Accrued taxes and expenses		149		114
Notes payable		465		450
Total current liabilities		1,090		1,007
Long-term debt		1,575		1,515
Common stock		525		525
Retained earnings		1,897		1,605
Total liabilities and shareholders' equity		$5,087		$4,652

Exhibit 2 Holt Corporation Consolidated Income Statement for the Year Ended 31 December 2020 (US$ Millions)

Total revenues	$3,323
Cost of goods sold	1,287
Selling, general, and administrative expenses	858
Earnings before interest, taxes, depreciation, and amortization (EBITDA)	1,178
Depreciation expense	270
Operating income	908

(continued)

Exhibit 2 (Continued)

Interest expense	195
Pretax income	713
Income tax (at 32%)	228
Net income	$ 485

Leigh presents his valuations of the common stock of Emerald and Holt to his supervisor, Alice Smith. Smith has the following questions and comments:

1 "I estimate that Emerald's long-term expected dividend payout rate is 20% and its return on equity is 10% over the long term."

2 "Why did you use an FCFE model to value Holt's common stock? Can you use a DDM instead?"

3 "How did Holt's FCFE for 2008 compare with its FCFF for the same year? I recommend you use an FCFF model to value Holt's common stock instead of using an FCFE model because Holt has had a history of leverage changes in the past."

4 "In the last three years, about 5% of Holt's growth in FCFE has come from decreases in inventory."

Leigh responds to each of Smith's points as follows:

1 "I will use your estimates and calculate Emerald's long-term, sustainable dividend growth rate."

2 "There are two reasons why I used the FCFE model to value Holt's common stock instead of using a DDM. The first reason is that Holt's dividends have differed significantly from its capacity to pay dividends. The second reason is that Holt is a takeover target and once the company is taken over, the new owners will have discretion over the uses of free cash flow."

3 "I will calculate Holt's FCFF for 2020 and estimate the value of Holt's common stock using an FCFF model."

4 "Holt is a growing company. In forecasting either Holt's FCFE or FCFF growth rates, I will not consider decreases in inventory to be a long-term source of growth."

18 Which of the following long-term FCFE growth rates is *most* consistent with the facts and stated policies of Emerald?

A 5% or lower

B 2% or higher

C 8% or higher

19 Do the reasons provided by Leigh support his use of the FCFE model to value Holt's common stock instead of using a DDM?

A Yes

B No, because Holt's dividend situation argues in favor of using the DDM

C No, because FCFE is not appropriate for investors taking a control perspective

20 Holt's FCFF (in millions) for 2020 is *closest* to:

A $308.

B $370.

C $422.

21 Holt's FCFE (in millions) for 2020 is *closest* to:

A $175.

B $250.

C $364.

22 Leigh's comment about not considering decreases in inventory to be a source of long-term growth in free cash flow for Holt is:

A inconsistent with a forecasting perspective.

B mistaken because decreases in inventory are a use rather than a source of cash.

C consistent with a forecasting perspective because inventory reduction has a limit, particularly for a growing firm.

23 Smith's recommendation to use an FCFF model to value Holt is:

A logical, given the prospect of Holt changing capital structure.

B not logical because an FCFF model is used only to value the total firm.

C not logical because FCFE represents a more direct approach to free cash flow valuation.

The following information relates to Questions 24–30

Yandie Izzo manages a dividend growth strategy for a large asset management firm. Izzo meets with her investment team to discuss potential investments in three companies: Company A, Company B, and Company C. Statements of cash flow for the three companies are presented in Exhibit 1.

Exhibit 1 Statements of Cash Flow, Most Recent Fiscal Year End (Amounts in Millions of Dollars)

	Company A	Company B	Company C
Cash Flow from Operating Activities			
Net Income	4,844	1,212	15,409
Adjustments			
Depreciation	500	288	3,746
Other noncash expenses	1,000	—	—
Changes in working capital			
(Increase) Decrease accounts receivable	(452)	(150)	(536)
(Increase) Decrease inventories	—	(200)	(803)
Increase (Decrease) accounts payable	(210)	100	(3)
Increase (Decrease) other current liabilities	540	14	350
Net cash from operating activities	6,222	1,264	18,163
Cash Flow from Investing Activities			
(Purchase) Sale of fixed assets	2,379	(1,000)	(3,463)

(continued)

Exhibit 1 (Continued)

	Company A	Company B	Company C
Net cash from investing activities	2,379	(1,000)	(3,463)
Cash Flow from Financing Activities			
Increase (Decrease) notes payable	25	3000	1,238
Increase (Decrease) long-term debt	(1,500)	(1,000)	(1,379)
Payment of common stock dividends	(1,000)	(237)	(15,000)
Net cash from financing activities	(2,475)	1,763	(15,141)
Net change in cash and cash equivalents	6,126	2,027	(441)
Cash and equivalents at beginning of year	50	100	3,000
Cash and equivalents at end of year	6,176	2,127	2,559
Supplemental Cash Flow Disclosures			
Interest	(353)	(50)	(552)
Income taxes	(1,605)	(648)	(3,787)

Izzo's team first discusses key characteristics of Company A. The company has a history of paying modest dividends relative to FCFE, has a stable capital structure, and is owned by a controlling investor.

The team also considers the impact of Company A's three noncash transactions in the most recent year on its FCFE, including the following:

Transaction 1: A \$900 million loss on a sale of equipment

Transaction 2: An impairment of intangibles of \$400 million

Transaction 3: A \$300 million reversal of a previously recorded restructuring charge

In addition, Company A's annual report indicates that the firm expects to incur additional noncash charges related to restructuring over the next few years.

To value the three companies' shares, one team member suggests valuing the companies' shares using net income as a proxy for FCFE. Another team member proposes forecasting FCFE using a sales-based methodology based on the following equation:

$$\text{FCFE} = \text{NI} - (1 - \text{DR})(\text{FCInv} - \text{Dep}) - (1 - \text{DR})(\text{WCInv}).$$

Izzo's team ultimately decides to use actual free cash flow to value the three companies' shares. Selected data and assumptions are provided in Exhibit 2.

Exhibit 2 Supplemental Data and Valuation Assumptions

	Company A	Company B	Company C
Tax rate	35%	35%	30%
Beta	1.00	0.90	1.10
Before-tax cost of debt	6%	7%	6%
Target debt ratio	50%	30%	40%
Market data:			
Risk-free rate: 3%			
Market risk premium: 7%			

The team calculates the intrinsic value of Company B using a two-stage FCFE model. FCFE growth rates for the first four years are estimated at 10%, 9%, 8%, and 7%, respectively, before declining to a constant 6% starting in the fifth year.

To calculate the intrinsic value of Company C's equity, the team uses the FCFF approach assuming a single-stage model where FCFF is expected to grow at 5% indefinitely.

24 Based on Company A's key characteristics, which discounted cash flow model would *most likely* be used by the investment team to value Company A's shares?

A DDM

B FCFE

C FCFF

25 Which noncash transaction should be subtracted from net income in arriving at Company A's FCFE?

A Transaction 1

B Transaction 2

C Transaction 3

26 Based on Exhibit 1, Company A's FCFE for the most recent year is *closest* to:

A \$5,318 million.

B \$6,126 million.

C \$7,126 million.

27 Based on Exhibit 1, using net income as a proxy for Company B's FCFE would result in an intrinsic value that is:

A lower than the intrinsic value if actual FCFE were used.

B equal to the intrinsic value if actual FCFE were used.

C higher than the intrinsic value if actual FCFE were used.

28 Based on Exhibit 1, using the proposed sales-based methodology to forecast FCFE would produce an inaccurate FCFE projection for which company?

A Company A

B Company B

C Company C

29 Based on Exhibits 1 and 2 and the proposed two-stage FCFE model, the intrinsic value of Company B's equity is *closest* to:

A \$70,602 million.

B \$73,588 million.

C \$79,596 million.

30 Based on Exhibits 1 and 2 and the proposed single-stage FCFF model, the intrinsic value of Company C's equity is *closest* to:

A \$277,907 million.

B \$295,876 million.

C \$306,595 million.

The following information relates to Questions 31–36

Gurmeet Singh, an equity portfolio manager at a wealth management company, meets with junior research analyst Cindy Ho to discuss potential investments in three companies: Sienna Limited, Colanari Manufacturing, and Bern Pharmaceutical.

Singh and Ho review key financial data from Sienna's most recent annual report, which are presented in Exhibits 1 and 2, to assess the company's ability to generate free cash flow.

Exhibit 1 Selected Data from Sienna Limited's Statement of Income for the Year Ended 31 December 2019 (Amounts in Millions of Euros)

EBITDA	4,000
Depreciation expense	800
Operating income (EBIT)	3,200
Interest expense	440
Tax rate	35%

Exhibit 2 Sienna Limited's Statement of Cash Flows for the Year Ended 31 December 2019 (Amounts in Millions of Euros)

Cash flow from operations	
Net income	1,794
Plus: Depreciation	800
Increase in accounts receivable	(2,000)
Increase in inventory	(200)
Increase in accounts payable	1,000
Cash flow from operations	1,394
Cash flow from investing activities	
Purchases of PP&E	(1,000)
Cash flow from financing activities	
Borrowing (repayment)	500
Total cash flow	894

Singh and Ho also discuss the impact of dividends, share repurchases, and leverage on Sienna's free cash flow. Ho tells Singh the following:

Statement 1 Changes in leverage do not impact free cash flow to equity.

Statement 2 Transactions between the company and its shareholders, such as the payment of dividends or share repurchases, do affect free cash flow.

Singh and Ho next analyze Colanari. Last year, Colanari had FCFF of €140 million. Singh instructs Ho to perform an FCFF sensitivity analysis of Colanari's firm value using the three sets of estimates presented in Exhibit 3. In her analysis, Ho assumes a tax rate of 35% and a stable capital structure of 30% debt and 70% equity.

Exhibit 3 Sensitivity Analysis for Colanari Valuation

Variable	Base-Case Estimate	Low Estimate	High Estimate
FCFF growth rate	4.6%	4.2%	5.0%
Before-tax cost of debt	4.9%	3.9%	5.9%
Cost of equity	11.0%	10.0%	12.0%

Finally, Singh and Ho analyze Bern. Selected financial information on Bern is presented in Exhibit 4.

Exhibit 4 Selected Financial Data on Bern Pharmaceutical

	Market Value	Required Return
Debt	€15,400 million	6.0%
Preferred stock	€4,000 million	5.5%
Common stock	€18,100 million	11.0%
FCFF, most recent year	€3,226 million	
Corporate tax rate	26.9%	

Singh notes that Bern has two new drugs that are currently in clinical trials awaiting regulatory approval. In addition to its operating assets, Bern owns a parcel of land from a decommissioned manufacturing facility with a current market value of €50 million that is being held for investment. Singh and Ho elect to value Bern under two scenarios:

Scenario 1 Value Bern assuming the two new drugs receive regulatory approval. In this scenario, FCFF is forecast to grow at 4.5% into perpetuity.

Scenario 2 Value Bern assuming the two new drugs do not receive regulatory approval. In this scenario, FCFF is forecast using a stable growth in FCFF of 1.5% for the next three years and then 0.75% thereafter into perpetuity.

31 Based on Exhibits 1 and 2, Sienna's FCFF in 2019 is:

A €680 million.

B €1,200 million.

C €3,080 million.

32 Based on Exhibits 1 and 2, Sienna's FCFE in 2019 is:

A €894 million.

B €1,466 million.

C €2,894 million.

33 Which of Ho's statements regarding free cash flow is (are) correct?

A Statement 1 only

B Statement 2 only

C Neither Statement 1 nor Statement 2

34 Based on Exhibit 3, Ho's FCFF sensitivity analysis should conclude that Colanari's value is *most* sensitive to the:

A FCFF growth rate.

B before-tax cost of debt.

C required rate of return for equity.

35 Based on Exhibit 4, Bern's firm value under Scenario 1 is *closest* to:

A €100,951.3 million.

B €105,349.1 million.

C €105,399.1 million.

36 Based on Exhibit 4, Singh and Ho should conclude that under Scenario 2, shares of Bern are:

A undervalued.

B fairly valued.

C overvalued.

SOLUTIONS

1

For a $100 increase in:	Change in FCFF (in US Dollars)	Change in FCFE (in US Dollars)
A. Net income	+100	+100
B. Cash operating expenses	–60	–60
C. Depreciation	+40	+40
D. Interest expense	0	–60
E. EBIT	+60	+60
F. Accounts receivable	–100	–100
G. Accounts payable	+100	+100
H. Property, plant, and equipment	–100	–100
I. Notes payable	0	+100
J. Cash dividends paid	0	0
K. Proceeds from issuing new common shares	0	0
L. Common shares repurchased	0	0

2 **A** Free cash flow to the firm, found with Equation 7, is

$$FCFF = NI + NCC + Int(1 - \text{Tax rate}) - FCInv - WCInv.$$

$$FCFF = 285 + 180 + 130(1 - 0.40) - 349 - (39 + 44 - 22 - 23).$$

$$FCFF = 285 + 180 + 78 - 349 - 38 = \$156 \text{ million}.$$

B Free cash flow to equity, found with Equation 10, is

$$FCFE = NI + NCC - FCInv - WFCInv + \text{Net borrowing}.$$

$$FCFE = 285 + 180 - 349 - (39 + 44 - 22 - 23) + (10 + 40).$$

$$FCFE = 285 + 180 - 349 - 38 + 50 = \$128 \text{ million}.$$

C To find FCFE from FCFF, one uses the relationship in Equation 9:

$$FCFE = FCFF - Int(1 - \text{Tax rate}) + \text{Net borrowing}.$$

$$FCFE = 156 - 130(1 - 0.40) + (10 + 40).$$

$$FCFE = 156 - 78 + 50 = \$128 \text{ million}.$$

3 **A** To find FCFF from CFO, EBIT, or EBITDA, the analyst can use Equations 8, 12, and 13.

To find FCFF from CFO:

$$FCFF = CFO + Int(1 - \text{Tax rate}) - FCInv.$$

$$FCFF = 427 + 130(1 - 0.40) - 349 = 427 + 78 - 349 = \$156 \text{ million}.$$

To find FCFF from EBIT:

$$FCFF = EBIT(1 - \text{Tax rate}) + Dep - FCInv - WCInv.$$

$$FCFF = 605(1 - 0.40) + 180 - 349 - 38.$$

$$FCFF = 363 + 180 - 349 - 38 = \$156 \text{ million}.$$

Finally, to obtain FCFF from EBITDA:

$$\text{FCFF} = \text{EBITDA}(1 - \text{Tax rate}) + \text{Dep}(\text{Tax rate}) - \text{FCInv} - \text{WCInv}.$$
$$\text{FCFF} = 785(1 - 0.40) + 180(0.40) - 349 - 38.$$
$$\text{FCFF} = 471 + 72 - 349 - 38 = \$156 \text{ million}.$$

B The simplest approach is to calculate FCFF from CFO, EBIT, or EBITDA as was done in Part A and then to find FCFE by making the appropriate adjustments to FCFF:

$$\text{FCFE} = \text{FCFF} - \text{Int}(1 - \text{Tax rate}) + \text{Net borrowing}.$$
$$\text{FCFE} = 156 - 130(1 - 0.40) + 50 = 156 - 78 + 50 = \$128 \text{ million}.$$

The analyst can also find FCFE by using CFO, EBIT, or EBITDA directly. Starting with CFO and using Equation 11, FCFE is found to be

$$\text{FCFE} = \text{CFO} - \text{FCInv} + \text{Net borrowing}.$$
$$\text{FCFE} = 427 - 349 + 50 = \$128 \text{ million}.$$

Starting with EBIT, on the basis of Equations 9 and 12, FCFE is

$$\begin{aligned}\text{FCFE} = {} & \text{EBIT}(1 - \text{Tax rate}) + \text{Dep} - \text{Int}(1 - \text{Tax rate}) - \text{FCInv} \\ & - \text{WCInv} + \text{Net borrowing}.\end{aligned}$$
$$\text{FCFE} = 605(1 - 0.40) + 180 - 130(1 - 0.40) - 349 - 38 + 50.$$
$$\text{FCFE} = 363 + 180 - 78 - 349 - 38 + 50 = \$128 \text{ million}.$$

Finally, starting with EBITDA, on the basis of Equations 9 and 13, FCFE is

$$\begin{aligned}\text{FCFE} = {} & \text{EBITDA}(1 - \text{Tax rate}) + \text{Dep}(\text{Tax rate}) \\ & - \text{Int}(1 - \text{Tax rate}) - \text{FCInv} - \text{WCInv} + \text{Net borrowing}.\end{aligned}$$
$$\text{FCFE} = 785(1 - 0.40) + 180(0.40) - 130(1 - 0.40) - 349 - 38 + 50.$$
$$\text{FCFE} = 471 + 72 - 78 - 349 - 38 + 50 = \$128 \text{ million}.$$

4 **A** FCF = Net income + Depreciation and amortization – Cash dividends – Capital expenditures. This definition of free cash flow is sometimes used to determine how much "discretionary" cash flow the management has at its disposal. Management discretion concerning dividends is limited by investor expectations that dividends will be maintained. Comparing this definition with Equation 7, FCFF = NI + NCC + Int(1 – Tax rate) – FCInv – WCInv, we find that FCFF includes a reduction for investments in working capital and the addition of after-tax interest expense. Common stock dividends are not subtracted from FCFF because dividends represent a distribution of the cash available to investors. (If a company pays preferred dividends and they were previously taken out when net income available to common shareholders was calculated, they are added back in Equation 7 to include them in FCFF.)

B FCF = Cash flow from operations (from the statement of cash flows) – Capital expenditures. Comparing this definition of free cash flow with Equation 8, FCFF = CFO + Int(1 – Tax rate) – FCInv, highlights the relationship of CFO to FCFF: The primary point is that when Equation 8 is used, after-tax interest is added back to CFO to arrive at the cash flow to all investors. Then FCInv is subtracted to arrive at the amount of that cash flow that is "free" in the sense of available for distribution to those investors after taking care of capital investment needs. If preferred dividends were subtracted to obtain net income (in CFO), they would also have to be added back in. This definition is commonly used to approximate FCFF, but it generally understates the actual FCFF by the amount of after-tax interest expense.

5 **A** The firm value is the present value of FCFF discounted at the WACC, or

$$\text{Firm value} = \frac{\text{FCFF}_1}{\text{WACC} - g} = \frac{\text{FCFF}_0(1+g)}{\text{WACC} - g} = \frac{1.7(1.07)}{0.11 - 0.07}$$
$$= \frac{1.819}{0.04} = \text{JPY}45.475 \text{ billion.}$$

The market value of equity is the value of the firm minus the value of debt:

Equity = 45.475 – 15 = JPY30.475 billion.

B Using the FCFE valuation approach, we find the present value of FCFE discounted at the required rate of return on equity to be

$$\text{PV} = \frac{\text{FCFE}_1}{r - g} = \frac{\text{FCFE}_0(1+g)}{r - g} = \frac{1.3(1.07)}{0.13 - 0.075} = \frac{1.3975}{0.055}$$
$$= \text{JPY}25.409 \text{ billion.}$$

The value of equity using this approach is JPY25.409 billion.

6 The required rate of return found with the CAPM is

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 6.4\% + 2.1(5.0\%) = 16.9\%.$$

The following table shows the values of sales, net income, capital expenditures less depreciation, and investments in working capital. FCFE equals net income less the investments financed with equity:

$$\text{FCFE} = \text{Net income} - (1 - \text{DR})(\text{Capital expenditures} - \text{Depreciation})$$
$$- (1 - \text{DR})(\text{Investment in working capital}),$$

where DR is the debt ratio (debt financing as a percentage of debt and equity). Because 20% of net new investments are financed with debt, 80% of the investments are financed with equity, which reduces FCFE by 80% of (Capital expenditures – Depreciation) and 80% of the investment in working capital.

(All Data in Billions of Euros)	2021	2022	2023	2024	2025
Sales (growing at 28%)	5.500	7.040	9.011	11.534	14.764
Net income = 32% of sales	1.760	2.253	2.884	3.691	4.724
FCInv – Dep = (35% – 9%) × Sales	1.430	1.830	2.343	2.999	3.839
WCInv = (6% of Sales)	0.330	0.422	0.541	0.692	0.886
0.80 × (FCInv – Dep + WCInv)	1.408	1.802	2.307	2.953	3.780
FCFE = NI – 0.80 × (FCInv – Dep + WCInv)	0.352	0.451	0.577	0.738	0.945
PV of FCFE discounted at 16.9%	0.301	0.330	0.361	0.395	0.433
Terminal stock value		85.032			
PV of terminal value discounted at 16.9%		38.950			
PV of FCFE (first five years)		1.820			
Total value of equity		40.770			

The terminal stock value is 18.0 times the earnings in 2025, or 18 × 4.724 = €85.03 billion. The present value of the terminal value (€38.95 billion) plus the present value of the first five years' FCFE (€1.82 billion) is €40.77 billion. Because NYL Manufacturing has 17 billion outstanding shares, the value per ordinary share is €2.398.

7 **A** The FCFF is (in euros)

$$\text{FCFF} = \text{NI} + \text{NCC} + \text{Int}(1 - \text{Tax rate}) - \text{FCInv} - \text{WCInv}.$$

$$\text{FCFF} = 250 + 90 + 150(1 - 0.30) - 170 - 40.$$

$$\text{FCFF} = 250 + 90 + 105 - 170 - 40 = 235 \text{ million}.$$

The weighted-average cost of capital is

$$\text{WACC} = 9\%(1 - 0.30)(0.40) + 13\%(0.60) = 10.32\%.$$

The value of the firm (in euros) is

$$\text{Firm value} = \frac{\text{FCFF}_1}{\text{WACC} - g} = \frac{\text{FCFF}_0(1+g)}{\text{WACC} - g} = \frac{235(1.06)}{0.1032 - 0.06}$$

$$= \frac{249.1}{0.0432} = 5{,}766.20 \text{ million}.$$

The total value of equity is the total firm value minus the value of debt: Equity = €5,766.20 million – €1,800 million = €3,966.20 million. Dividing by the number of shares gives the per-share estimate of V_0 = €3,966.20 million/10 million = €396.62 per share.

B The free cash flow to equity is

$$\text{FCFE} = \text{NI} + \text{NCC} - \text{FCInv} - \text{WCInv} + \text{Net borrowing}.$$

$$\text{FCFE} = 250 + 90 - 170 - 40 + 0.40(170 - 90 + 40).$$

$$\text{FCFE} = 250 + 90 - 170 - 40 + 48 = €178 \text{ million}.$$

Because the company is borrowing 40% of the increase in net capital expenditures (170 – 90) and working capital (40), net borrowing is €48 million.

The total value of equity is the FCFE discounted at the required rate of return of equity:

$$\text{Equity value} = \frac{\text{FCFE}_1}{r - g} = \frac{\text{FCFE}_0(1+g)}{r - g} = \frac{178(1.07)}{0.13 - 0.07}$$

$$= \frac{190.46}{0.06} = €3{,}174.33 \text{ million}.$$

The value per share is V_0 = €3,174.33 million/10 million = €317.43 per share.

8 The WACC for PHB Company is

$$\text{WACC} = 0.30(7.0\%)(1 - 0.35) + 0.15(6.8\%) + 0.55(11.0\%) = 8.435\%.$$

The firm value is

$$\text{Firm value} = \frac{\text{FCFF}_0(1+g)}{\text{WACC} - g}.$$

$$\text{Firm value} = \frac{28(1.04)}{0.08435 - 0.04}$$

$$= \frac{29.12}{0.04435}$$

$$= £656.60 \text{ million}.$$

The value of equity is the firm value minus the value of debt minus the value of preferred stock: Equity = 656.60 – 145 – 65 = £446.60 million. Dividing this amount by the number of shares gives the estimated value per share of £446.60 million/8 million shares = £55.82.

The estimated value for the stock is greater than the market price of £32.50, so the stock appears to be undervalued.

9 **A** The required return on equity is

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 5.5\% + 0.90(5.5\%) = 10.45\%.$$

The weighted-average cost of capital is

$$\text{WACC} = 0.25(7.0\%)(1 - 0.40) + 0.75(10.45\%) = 8.89\%.$$

B.

$$\text{Firm value} = \frac{\text{FCFF}_0(1+g)}{\text{WACC} - g}.$$

$$\text{Firm value} = \frac{1.1559(1.04)}{0.0889 - 0.04} = \$24.583.$$

C Equity value = Firm value – Market value of debt.

Equity value = 24.583 – 3.192 = $21.391 billion.

D Value per share = Equity value/Number of shares.

Value per share = $21.391 billion/1.852 billion = $11.55.

10 **A** The required rate of return for McInish found with the CAPM is

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 5.08\% + 0.70(5.50\%) = 8.93\%.$$

The value per share is

$$V_0 = \frac{\text{FCFE}_0(1+g)}{r - g} = \frac{0.88(1.064)}{0.0893 - 0.064} = \$37.01.$$

B The following table shows the calculated price for McInish based on the base-case values for all values except the variable being changed from the base-case value.

Variable	Estimated Price with Low Value (£)	Estimated Price with High Value (£)	Range (Rank) (£)
Normalized FCFE_0	29.44	47.94	18.50 (3)
Risk-free rate	38.22	35.33	2.89 (5)
Equity risk premium	51.17	28.99	22.18 (2)
Beta	47.29	30.40	16.89 (4)
FCFE growth rate	18.56	48.79	30.23 (1)

As the table shows, the value of McInish is most sensitive to the changes in the FCFE growth rate, with the price moving over a wide range. McInish's stock price is least sensitive to alternative values of the risk-free rate. Alternative values of beta, the equity risk premium, or the initial FCFE value also have a large impact on the value of the stock, although the effects of these variables are smaller than the effect of the growth rate.

11 A Using the CAPM, the required rate of return for NewMarket is

$$r = E(R_i) = R_F + \beta_i\left[E(R_M) - R_F\right] = 7\% + 1.3(4\%) = 12.2\%.$$

To estimate FCFE, we use Equation 15:

$$\begin{aligned}\text{FCFE} = {} & \text{Net income} - (1 - \text{DR})(\text{FCInv} - \text{Depreciation}) \\ & - (1 - \text{DR})(\text{WCInv}),\end{aligned}$$

which can be written

$$\begin{aligned}\text{FCFE} &= \text{Net income} - (1 - \text{DR})(\text{FCInv} - \text{Depreciation} + \text{WCInv}) \\ &= \text{Net income} - (1 - \text{DR})(\text{Net investment in operating assets}).\end{aligned}$$

The following table shows that net income grows at 20% annually for Years 1, 2, and 3 and then grows at 8% for Year 4. The net investment in operating assets is $1,150 million in Year 1 and grows at 15% annually for Years 2 and 3. Debt financing is 40% of this investment. FCFE is NI – Net investment in operating assets + New debt financing. Finally, the present value of FCFE for Years 1, 2, and 3 is found by discounting at 12.2%.

	Year			
(in $ Millions)	**1**	**2**	**3**	**4**
Net income	720.00	864.00	1,036.80	1,119.74
Net investment in operating assets	1,150.00	1,322.50	1,520.88	335.92
New debt financing	460.00	529.00	608.35	134.37
FCFE	30.00	70.50	124.27	918.19
PV of FCFE discounted at 12.2%	26.74	56.00	87.98	

In Year 4, net income is 8% larger than in Year 3. In Year 4, the investment in operating assets is 30% of net income and debt financing is 40% of this investment. The FCFE in Year 4 is $918.19 million. The value of FCFE after Year 3 is found by using the constant-growth model:

$$V_3 = \frac{\text{FCFE}_4}{r - g} = \frac{918.19}{0.122 - 0.08} = \$21{,}861.67 \text{ million}.$$

The present value of V_3 discounted at 12.2% is $15,477.64 million. The total value of equity, the present value of the first three years' FCFE plus the present value of V_3, is $15,648.36 million. Dividing this amount by the number of outstanding shares (318 million) gives a value per share of $49.21. For the first three years, NewMarket has a small FCFE because of the large investments it is making during the high-growth phase. In the normal-growth phase, FCFE is much larger because the investments required are much smaller.

B The planner's estimate of the share value of $70.98 is much higher than the FCFE model estimate of $49.21 for several reasons. First, taxes and interest expenses have a prior claim to the company's cash flow and should be taken out of the cash flows used in estimating the value of equity because these amounts are not available to equity holders. The planner did not do this.

Second, EBITDA does not account for the company's reinvestments in operating assets. So, EBITDA overstates the funds available to stockholders if reinvestment needs exceed depreciation charges, which is the case for growing companies such as NewMarket.

Third, EBITDA does not account for the company's capital structure. Using EBITDA to represent a benefit to stockholders (as opposed to stockholders and bondholders combined) is a mistake.

Finally, dividing EBITDA by the bond rate is a major error. The risk-free bond rate is an inappropriate discount rate for risky equity cash flows; the proper measure is the required rate of return on the company's equity. Dividing by a fixed rate also assumes, erroneously, that the cash flow stream is a fixed perpetuity. EBITDA cannot be a perpetual stream because if it were distributed, the stream would eventually decline to zero (lacking capital investments). NewMarket is actually a growing company, so assuming it to be a nongrowing perpetuity is a mistake.

12 The following table develops the information to calculate FCFE per share (amounts are in US dollars).

	2020	2021	2022	2023	2024	2025
Growth rate for EPS	21%	18%	15%	12%	9%	6%
EPS	3.630	4.283	4.926	5.517	6.014	6.374
Net capital expenditure per share	5.000	5.000	4.500	4.000	3.500	1.500
Investment in WC per share	1.250	1.250	1.125	1.000	0.875	0.375
New debt financing = 40% of (Capital expenditure + WCInv)	2.500	2.500	2.250	2.000	1.750	0.750
FCFE = NI – Net capital expenditure – WCInv + New debt financing	–0.120	0.533	1.551	2.517	3.389	5.249
PV of FCFE discounted at 12%	–0.107	0.425	1.104	1.600	1.923	

Earnings per share for 2019 are \$3.00, and the EPS estimates for 2020 through 2025 in the table are found by increasing the previous year's EPS by that year's growth rate. The net capital expenditures each year were specified by the analyst. The increase in working capital per share is equal to 25% of net capital expenditures. Finally, debt financing is 40% of that year's total net capital expenditures and investment in working capital. For example, in 2020, the per-share amount for net capital expenditures plus investment in working capital is \$5.00 + \$1.25 = \$6.25. Debt financing is 40% of \$6.25, or \$2.50. Debt financing for 2021 through 2025 is found in the same way.

FCFE equals net income minus net capital expenditures minus investment in working capital plus new debt financing. Notice that FCFE is negative in 2020 because of large capital investments and investments in working capital. As these investments decline relative to net income, FCFE becomes positive and substantial.

The present values of FCFE from 2020 through 2024 are given in the bottom row of the table. These five present values sum to \$4.944 per share. Because FCFE from 2025 onward will grow at a constant 6%, the constant-growth model can be used to value these cash flows.

$$V_{2024} = \frac{\text{FCFE}_{2025}}{r - g} = \frac{5.249}{0.12 - 0.06} = \$87.483.$$

The present value of this stream is $\$87.483/(1.12)^5 = \$49.640$. The value per share is the present value of the first five FCFEs (2020–2024) plus the present value of the FCFE after 2024, or \$4.944 + \$ 49.640 = \$54.58.

13 A FCFE is defined as the cash flow remaining after the company meets all financial obligations, including debt payment, and covers all capital expenditure and working capital needs. Sundanci's FCFE for the year 2020 is calculated as follows:

Net income	= $80 million
Plus: Depreciation expense	= 23
Less: Capital expenditures	= 38
Less: Investment in WC	= 41
Equals: FCFE	= $24 million

Thus, FCFE per share equals ($24 million)/(84 million shares) = $0.286.

B The FCFE model requires forecasts of FCFE for the high-growth years (2021 and 2022) plus a forecast for the first year of stable growth (2023) to allow for an estimate of the terminal value in 2022 based on constant perpetual growth. Because all of the components of FCFE are expected to grow at the same rate, the values can be obtained by projecting the FCFE at the common rate. (Alternatively, the components of FCFE can be projected and aggregated for each year.)

The following table provides the process for estimating Sundanci's current value on a per-share basis.

Free Cash Flow to Equity

Base assumptions:	
Shares outstanding (millions)	84
Required return on equity, r	14%

		Actual 2020	Projected 2021	Projected 2022	Projected 2023
			g = 27%	g = 27%	g = 13%
	Total	Per share			
Earnings after tax	$80	$0.952	$1.2090	$1.5355	$1.7351
Plus: Depreciation expense	$23	$0.274	$0.3480	$0.4419	$0.4994
Less: Capital expenditures	$38	$0.452	$0.5740	$0.7290	$0.8238
Less: Increase in net working capital	$41	$0.488	$0.6198	$0.7871	$0.8894
Equals: FCFE	$24	$0.286	$0.3632	$0.4613	$0.5213
Terminal value[a]				$52.1300	
Total cash flows to equity[b]			$0.3632	$52.5913	
Discounted value[c]			$0.3186	$40.4673	
Current value per share[d]	$40.7859				

[a]Projected 2022 terminal value = Projected 2023 FCFE/(r – g).
[b]Projected 2022 total cash flows to equity = Projected 2022 FCFE + Projected 2022 terminal value.
[c]Discounted values obtained by using r = 14%.
[d]Current value per share = Discounted value 2021 + Discounted value 2022.

C The following limitations of the DDM *are* addressed by the FCFE model: The DDM uses a strict definition of cash flow to equity; that is, cash flows to equity are the dividends on the common stock. The FCFE model expands the definition of cash flow to include the balance of residual cash flows after all financial obligations and investment needs have been met. Thus, the FCFE model explicitly recognizes the company's investment and financing policies as well as its dividend policy. In instances of a change of corporate control, and thus the possibility of changing dividend policy, the FCFE model provides a better estimate of value.

Both two-stage valuation models allow for two distinct phases of growth—an initial finite period when the growth is abnormal followed by a stable growth period that is expected to last forever. These two-stage models share the same limitations with respect to the growth assumptions:

First, the analyst must confront the difficulty of defining the duration of the extraordinary growth period. A long period of high growth will produce a higher valuation, and the analyst may be tempted to assume an unrealistically long period of extraordinary growth.

Second, the analyst must realize that assuming a sudden shift from high growth to lower, stable growth is unrealistic. The transformation is more likely to occur gradually over time.

Third, because value is quite sensitive to the steady-state growth assumption, overestimating or underestimating this rate can lead to large errors in value.

The two models also share other limitations—notably, difficulties in accurately estimating required rates of return.

14 A When a two-stage DDM is used, the value of a share of Mackinac, dividends per share (DPS), is calculated as follows:

$$\begin{aligned}\text{DPS}_0 &= \text{Cash dividends/Shares outstanding} = \$22{,}470/13{,}000\\ &= \$1.7285.\\ \text{DPS}_1 &= \text{DPS}_0 \times 1.17 = \$2.0223.\\ \text{DPS}_2 &= \text{DPS}_0 \times 1.17^2 = \$2.3661.\\ \text{DPS}_3 &= \text{DPS}_0 \times 1.17^3 = \$2.7683.\\ \text{DPS}_4 &= \text{DPS}_0 \times 1.17^3 \times 1.09 = \$3.0175.\end{aligned}$$

When the CAPM is used, the required return on equity, r, is

$$\begin{aligned}r &= \text{Government bond rate} + (\text{Beta} \times \text{Equity risk premium})\\ &= 0.06 + (1.25 \times 0.05) = 0.1225, \text{ or } 12.25\%.\end{aligned}$$

$$\begin{aligned}\text{Value per share} &= \text{DPS}_1/(1+r) + \text{DPS}_2/(1+r)^2 + \text{DPS}_3/(1+r)^3\\ &\quad + \left[\text{DPS}_4/(r - g_{stable})\right]/(1+r)^3.\end{aligned}$$

$$\begin{aligned}\text{Value per share} &= \$2.0223/1.1225 + \$2.3661/1.1225^2\\ &\quad + \$2.7683/1.1225^3\\ &\quad + \left[\$3.0175/(0.1225 - 0.09)\right]/1.1225^3\\ &= \$1.8016 + \$1.8778 + \$1.9573 + \$65.6450\\ &= \$71.28.\end{aligned}$$

B When the two-stage FCFE model is used, the value of a share of Mackinac is calculated as follows (in $ thousands except per-share data):

Net income = $37,450.

Depreciation = $10,500.

Capital expenditures = $15,000.

Change in working capital = $5,500.

New debt issuance − Principal repayments = Change in debt outstanding = $4,000

$$\text{FCFE}_0 = \text{Net income} + \text{Depreciation} - \text{Capital expenditures} - \text{Change in working capital} - \text{Principal repayments} + \text{New debt issues}.$$

$$\text{FCFE}_0 = \$37{,}450 + \$10{,}500 - \$15{,}000 - \$5{,}500 + \$4{,}000 = \$31{,}450.$$

$$\text{FCFE}_0 \text{ per share} = \$31{,}450/13{,}000 = \$2.4192.$$

$$\text{FCFE}_1 = \text{FCFE}_0 \times 1.17 = \$2.8305.$$

$$\text{FCFE}_2 = \text{FCFE}_0 \times 1.17^2 = \$3.3117.$$

$$\text{FCFE}_3 = \text{FCFE}_0 \times 1.17^3 = \$3.8747.$$

$$\text{FCFE}_4 = \text{FCFE}_0 \times 1.17^3 \times 1.09 = \$4.2234.$$

From the answer to A, r = 12.25%.

$$\text{Value per share} = \text{FCFE}_1/(1+r) + \text{FCFE}_2/(1+r)^2 + \text{FCFE}_3/(1+r)^3 + \left[\text{FCFE}_4/(r - g_{stable})\right]/(1+r)^3.$$

$$\begin{aligned}\text{Value per share} &= \$2.8305/1.1225 + \$3.3117/1.1225^2 \\ &\quad + \$3.8747/1.1225^3 \\ &\quad + \left[\$4.2234/(0.1225 - 0.09)\right]/1.1225^3 \\ &= \$2.5216 + \$2.6283 + \$2.7395 + \$91.8798 \\ &= \$99.77.\end{aligned}$$

C The FCFE model is best for valuing companies for takeovers or in situations that have a reasonable chance of a change in corporate control. Because controlling stockholders can change the dividend policy, they are interested in estimating the maximum residual cash flow after meeting all financial obligations and investment needs. The DDM is based on the premise that the only cash flows received by stockholders are dividends. FCFE uses a more expansive definition to measure what a company can afford to pay out as dividends.

15 A The real required rate of return for KMobile is

Country return (real)	6.50%
Industry adjustment	+0.60%
Size adjustment	–0.10%
Leverage adjustment	+0.25%
Required rate of return	7.25%

B The real growth rate of FCFE is expected to be the same as the country rate of 3.5%. The value of one share is

$$V_0 = \frac{\text{FCFE}_0(1 + g_{real})}{r_{real} - g_{real}} = \frac{1{,}300(1.035)}{0.0725 - 0.035} = \text{KRW}35{,}880.$$

16 The required return for QuickChange, found by using the CAPM, is $r = E(R_i) = R_F + \beta_i[E(R_M) - R_F] = 4.5\% + 2.0(5.0\%) = 14.5\%$. The estimated future values of FCFE per share are given in the following exhibit (amounts in US dollars):

Year *t*	Variable	Calculation	Value in Year *t*	Present Value at 14.5%
1	FCFE_1	0.75(1.10)	0.825	0.721
2	FCFE_2	0.75(1.10)(1.26)	1.040	0.793
3	FCFE_3	$0.75(1.10)(1.26)^2$	1.310	0.873
4	FCFE_4	$0.75(1.10)(1.26)^3$	1.650	0.960
4	TV_4	$\text{FCFE}_5/(r - g)$ $= 0.75(1.10)(1.26)^3(1.06)/(0.145 - 0.06)$ $= 1.749/0.085.$	20.580	11.974
0	Total value =	PV of FCFE for Years 1–4 + PV of terminal value		15.32

The FCFE grows at 10% for Year 1 and then at 26% for Years 2–4. These calculated values for FCFE are shown in the exhibit. The present values of the FCFE for the first four years discounted at the required rate of return are given in the last column of the table. After Year 4, FCFE will grow at 6% forever, so the constant-growth FCFE model is used to find the terminal value at Time 4, which is $\text{TV}_4 = \text{FCFE}_5/(r - g)$. TV_4 is discounted at the required return for four periods to find its present value, as shown in the table. Finally, the total value of the stock, \$15.32, is the sum of the present values of the first four years' FCFE per share plus the present value of the terminal value per share.

17 The total value of non-operating assets is

SEK60	million short-term securities
SEK45	million market value of noncurrent assets
SEK40	million pension fund surplus
SEK145	million non-operating assets

The total value of the firm is the value of the operating assets plus the value of the non-operating assets, or SEK720 million plus SEK145 million = SEK865 million. The equity value is the value of the firm minus the value of debt, or SEK865 million – SEK215 million = SEK650 million. The value per share is SEK650 million/100 million shares = SEK6.50 per share.

18 C is correct. The sustainable growth rate is return on equity (ROE) multiplied by the retention ratio. ROE is 10%, and the retention ratio is 1 – Payout ratio, or 1.0 – 0.2 = 0.8. The sustainable growth rate is 0.8 × 10% = 8%. FCFE growth should be at least 8% per year in the long term.

19 A is correct. Justifications for choosing the FCFE model over the DDM include the following:

- The company pays dividends, but its dividends differ significantly from the company's capacity to pay dividends (the first reason given by Leigh).
- The investor takes a control perspective (the second reason given by Leigh).

20 A is correct. FCFF = NI + NCC + Interest expense(1 – Tax rate) – FCInv – WCInv. In this case:

NI = \$485 million

NCC = Depreciation expense = \$270 million

Interest expense(1 – Tax rate) = 195(1 – 0.32) = \$132.6 million

FCInv = Net purchase of fixed assets = Increase in gross fixed assets
= 4,275 – 3,752 = \$523 million

WCInv = Increase in accounts receivable + Increase in inventory – Increase in accounts payable – Increase in accrued liabilities
= (770 – 711) + (846 – 780) – (476 – 443) – (149 – 114)
= \$57 million

FCFF = 485 + 270 + 132.6 – 523 – 57 = 307.6, or \$308 million

21 B is correct. FCFE = NI + NCC – FCInv – WCInv + Net borrowing. In this case:

$$\begin{aligned}
\text{NI} &= \$485 \text{ million.}\\
\text{NCC} &= \text{Depreciation expense} = \$270 \text{ million.}\\
\text{FCInv} &= \text{Net purchase of fixed assets} = \text{Increase in gross fixed assets}\\
&= 4{,}275 - 3{,}752 = \$523 \text{ million.}\\
\text{WCInv} &= \text{Increase in accounts receivable} + \text{Increase in inventory} - \text{Increase in accounts payable} - \text{Increase in accrued liabilities}\\
&= (770 - 711) + (846 - 780) - (476 - 443) - (149 - 114)\\
&= \$57 \text{ million.}\\
\text{Net borrowing} &= \text{Increase in notes payable} + \text{Increase in long-term debt}\\
&= (465 - 450) + (1{,}575 - 1{,}515) = \$75 \text{ million.}\\
\text{FCFE} &= 485 + 270 - 523 - 57 + 75 = \$250 \text{ million.}
\end{aligned}$$

An alternative calculation is

FCFE = FCFF – Int(1 – Tax rate) + Net borrowing.

FCFE = 307.6 – 195(1 – 0.32) + (15 +60) = \$250 million.

22 C is correct. Inventory cannot be reduced below zero. Furthermore, sales growth tends to increase inventory.

23 A is correct. The FCFF model is often selected when the capital structure is expected to change because FCFF estimation may be easier than FCFE estimation in the presence of changing financial leverage.

24 B is correct. Company A has a history of paying modest dividends relative to FCFE. An FCFF or FCFE model provides a better estimate of value over a DDM model when dividends paid differ significantly from the company's capacity to pay dividends. Also, Company A has a controlling investor; with control comes discretion over the uses of free cash flow. Therefore, there is the possibility that the controlling shareholder could change the dividend policy. Finally, Company A has a stable capital structure; using FCFE is a more direct and simpler method to value a company's equity than using FCFF when a company's capital structure is stable.

25 C is correct. The applicable noncash adjustments to net income in arriving at FCFE are as follows:

Noncash Item	Adjustment to Net Income	Amount (millions)
Transaction 1: Loss on sale of equipment	Added back	+900
Transaction 2: Impairment of intangibles	Added back	+400
Transaction 3: Reversal of restructuring charge	Subtracted	–300

In the case of Transaction 1, a loss reduces net income and thus must be added back in arriving at FCFE. Similarly, an impairment of intangibles (Transaction 2) reduces net income and thus must be added back in arriving at FCFE. Transaction 3 (reversal of a restructuring charge) would increase net income and thus must be subtracted in arriving at FCFE.

26 C is correct. FCFE for Company A for the most recent year is calculated as follows:

Net income	$4,844
Plus: Net noncash charges	1,500
Less: Investment in working capital	122
Plus: Proceeds from sale of fixed capital	2,379
Less: Net borrowing repayment	1,475
FCFE (millions)	$7,126

Net noncash charges are found by adding depreciation to other noncash expenses:

$500 million + $1,000 million = $1,500 million.

Investment in working capital is calculated by netting the increase in accounts receivable, the decrease in accounts payable, and the increase in other current liabilities:

–$452 million – $210 million + $540 million = –$122 million (outflow).

Net borrowing repayment is calculated by netting the increase in notes payable and the decrease in long-term debt:

$25 million – $1,500 million = –$1,475 million (outflow).

27 A is correct. FCFE is significantly higher than net income for Company B:

Net income = $1,212 million.

FCFE for Company B is calculated as follows:

Net income	$1,212
Plus: Net noncash charges	288
Less: Investment in WC	236
Less: Investment in fixed assets	1,000
Plus: Net borrowing	2,000
FCFE (millions)	$2,264

Investment in working capital is calculated by adding the increase in accounts receivable, the increase in inventories, the increase in accounts payable, and the increase in other current liabilities: –$150 million – $200 million + $100 million

+ \$14 million = –\$236 million. Net borrowing is calculated by adding the increase in notes payable to the decrease in long-term debt: \$3,000 million – \$1,000 million = \$2,000 million.

Therefore, using net income of \$1,212 million as a proxy for FCFE (\$2,264 million) for Company B would result in a much lower valuation estimate than if actual FCFE were used.

28 A is correct. In addition to significant noncash charges other than depreciation in the most recent year, the annual report indicates that Company A expects to recognize additional noncash charges related to restructuring over the next few years. The given equation for forecasting assumes that the only noncash charge is depreciation. When the company being analyzed has significant noncash charges other than depreciation expense, this sales-based methodology will result in a less accurate estimate of FCFE than one obtained by forecasting all the individual components of FCFE.

29 C is correct.

FCFE for the most recent year for Company B is calculated as follows:

Net income	\$1,212
Plus: Net noncash charges	288
Less: Investment in WC	236
Less: Investment in fixed assets	1,000
Plus: Net borrowing	2,000
FCFE (millions)	\$2,264

The required rate of return on equity for Company B is

$$r = E(R_i) = R_F + \beta_i[E(R_M) - R_F] = 3\% + 0.90(7\%) = 9.3\%.$$

The most recent FCFE grows for the next four years at annual growth rates of 10%, 9%, 8%, and 7%, respectively, and then 6% thereafter:

t	*g*	Calculation	FCFE (millions)
1	10%	\$2,264.00 × 1.10	\$2,490.40
2	9%	\$2,490.40 × 1.09	\$2,714.54
3	8%	\$2,714.54 × 1.08	\$2,931.70
4	7%	\$2,931.70 × 1.07	\$3,136.92
5	6%	\$3,136.92 × 1.06	\$3,325.13

The present value of FCFE for the first four years is calculated as follows:

$$\text{PV} = \frac{2{,}490.40}{1.093^1} + \frac{2{,}714.54}{1.093^2} + \frac{2{,}931.70}{1.093^3} + \frac{3{,}136.92}{1.093^4}.$$

$$\text{PV} = 2{,}278.50 + 2{,}272.25 + 2{,}245.22 + 2.197.97 = 8{,}993.94.$$

The present value of the terminal value is calculated as follows:

$$\text{PV of TV}_4 = \frac{3{,}325.13}{(0.093 - 0.06)(1.093)^4} = 70{,}601.58.$$

So, the estimated total market value of the equity is 8,993.94 + 70,601.58 = 79,595.52 ≈ \$79,596 million.

30 C is correct. Company C's firm value is calculated as follows:

The required rate of return on equity for Company C is

$r = E(R_i) = R_F + \beta_i[E(R_M) - R_F] = 3\% + 1.1(7\%) = 10.7\%.$

$$\text{WACC} = \frac{\text{MV}(\text{Debt})}{\text{MV}(\text{Debt}) + \text{MV}(\text{Equity})} r_d (1 - \text{Tax rate}) + \frac{\text{MV}(\text{Equity})}{\text{MV}(\text{Debt}) + \text{MV}(\text{Equity})} r_e.$$

WACC = 0.40(6%)(1 – 0.30) + 0.60(10.7%) = 1.68% + 6.42% = 8.10%.

FCFF for the most recent year for Company C is calculated as follows:

Net income	$15,409.00
Plus: Net noncash charges	3,746.00
Less: Investment in working capital	992.00
Less: Investment in fixed capital	3,463.00
Plus: Interest expense × (1 – Tax rate)	386.40
FCFF (in millions)	$15,086.40

Investment in working capital is found by adding the increase in accounts receivable, the increase in inventories, the decrease in accounts payable, and the increase in other current liabilities: –\$536 million – \$803 million – \$3 million + \$350 million = –\$992 million.

FCFF is expected to grow at 5.0% indefinitely. Thus,

$$\text{Firm value} = \frac{\text{FCFF}_1}{\text{WACC} - g} = \frac{\text{FCFF}_0(1 + g)}{\text{WACC} - g} = \frac{15{,}086.4(1.05)}{0.081 - 0.05} = \$510{,}990.97 \text{ million.}$$

The value of equity is the value of the firm minus the value of debt. The value of debt is found by multiplying the target debt ratio by the total firm value:

Debt value = 0.40(\$510,990.97) = \$204,396.39.

Therefore, equity value = \$510,990.97 – \$204,396.39 = \$306,594.58 million.

31 A is correct. Sienna's FCFF in 2019 is calculated as

FCFF = EBIT(1 – Tax rate) + Dep – FCInv – WCInv.

FCInv = Purchases of PP&E = 1,000 (outflow).

WCInv = Increase in accounts receivable (outflow) + Increase in inventory (outflow) + Increase in accounts payable (inflow).

WCInv = –2,000 (outflow) + –200 (outflow) + 1,000 (inflow) = –1,200 (outflow).

FCFF = 3,200(1 – 0.35) + 800 – 1,000 –1,200.

FCFF = €680 million.

FCFF can also be computed from CFO:

FCFF = CFO + Int(1 – Tax rate) – FCInv.

FCFF = 1,394 + 440(1 – 0.35) – 1,000.

FCFF = €680 million.

32 A is correct. Sienna's FCFE in 2019 is calculated as

FCFE = CFO – FCInv + Net borrowing

= 1,394 – 1,000 + 500

= €894 million.

Alternatively, FCFE may be calculated as

FCFE = FCFF – Int(1 – Tax rate) + Net borrowing.

= 680 – 440(1 – 0.35) + 500

= €894 million.

33 C is correct. Transactions between the company and its shareholders (through cash dividends, share repurchases, and share issuances) do not affect free cash flow. However, leverage changes, such as the use of more debt financing, have some impact on free cash flow because they increase the interest tax shield (reduce corporate taxes because of the tax deductibility of interest) and reduce the cash flow available to equity.

34 C is correct. Colanari's valuation is most sensitive to the cost of equity (r_e) because the range of estimated values is larger than the valuation ranges estimated from the sensitivity analysis of both the FCFF growth rate (GFCFF) and the before-tax cost of debt (r_d).

Variable	Base Case	Low Estimate	High Estimate	Valuation with Low Estimate (€ millions)	Valuation with High Estimate (€ millions)	Range (€ millions)
GFCFF	4.6%	4.2%	5.0%	3,274.16	4,021.34	747.18
r_d	4.9%	3.9%	5.9%	3,793.29	3,445.24	348.05
r_e	11.0%	10.0%	12.0%	4,364.18	3,079.38	1,284.80

WACC = $[w_d \times r_d(1 - \text{Tax rate})] + (w_e \times r_e)$.

Firm value = $\text{FCFF}_0(1 + g)/(\text{WACC} - g)$.

Cost of equity sensitivity

Using the base case estimates for the FCFF growth rate and the before-tax cost of debt and using the low estimate for the cost of equity (r_e) of 10.0%, the valuation estimate is

WACC = [(0.30)(0.049)(1 – 0.35)] + (0.70)(0.10) = 7.96%.

Firm value = 140 million(1 + 0.046)/(0.0796 – 0.046) = €4,364.18 million.

Using the base case estimates for the FCFF growth rate and the before-tax cost of debt and using the high estimate for the cost of equity (r_e) of 12.0%, the valuation estimate is

WACC = [(0.30)(0.049)(1 – 0.35)] + (0.70)(0.120) = 9.36%.

Firm value = 140 million(1 + 0.046)/(0.0936 – 0.046) = €3,079.38 million.

Therefore, the range in valuation estimates from using the highest and lowest estimates of the cost of equity is €1,284.80 million.

FCFF growth rate sensitivity

Using the base case estimates for the cost of equity and the before-tax cost of debt and using the low estimate for the FCFF growth rate (GFCFF) of 4.2%, the valuation estimate is

WACC = [(0.30)(0.049)(1 – 0.35)] + (0.70)(0.11) = 8.66%.

$$\text{Firm value} = 140 \text{ million}(1 + 0.042)/(0.0866 - 0.042) = €3{,}274.16 \text{ million.}$$

Using the base case estimates for the cost of equity and the before-tax cost of debt and using the high estimate for the FCFF growth rate (GFCFF) of 5.0%, the valuation estimate is

$$\text{WACC} = [(0.30)(0.049)(1 - 0.35)] + (0.70)(0.11) = 8.66\%.$$

$$\text{Firm value} = 140 \text{ million}(1 + 0.05)/(0.0866 - 0.05) = €4{,}021.34 \text{ million.}$$

Therefore, the range in valuation estimates from using the highest and lowest estimates of the FCFF growth rate is €747.18 million.

Before-tax cost of debt sensitivity

Using the base case estimates for the FCFF growth rate and the cost of equity and using the low estimate for the before-tax cost of debt (r_d) of 3.9%, the valuation estimate is

$$\text{WACC} = [(0.30)(0.039)(1 - 0.35)] + (0.70)(0.11) = 8.46\%.$$

$$\text{Firm value} = 140 \text{ million}(1 + 0.046)/(0.0846 - 0.046) = €3{,}793.29 \text{ million.}$$

Using the base case estimates for the FCFF growth rate and the cost of equity and using the high estimate for the before-tax cost of debt (r_d) of 5.9%, the valuation estimate is

$$\text{WACC} = [(0.30)(0.059)(1 - 0.35)] + (0.70)(0.11) = 8.85\%.$$

$$\text{Firm value} = 140 \text{ million}(1 + 0.046)/(0.0885 - 0.046) = €3{,}445.24 \text{ million.}$$

Therefore, the range in valuation estimates from using the highest and lowest estimates of the before-tax cost of debt is €348.05 million.

35 C is correct. Based on Scenario 1, where Bern receives regulatory approval for its new drugs, the growth rate in FCFF for Bern will be constant at 4.5%. Therefore, a constant-growth valuation model can be used to calculate firm value.

Bern's weighted average cost of capital is calculated as

$$\text{WACC} = [w_d \times r_d(1 - \text{Tax rate})] + (w_p \times r_p) + (w_e \times r_e).$$

The total market value of the firm is the sum of the debt, preferred stock, and common stock market values: 15,400 + 4,000 + 18,100 = 37,500.

$$\text{WACC} = [(15{,}400/37{,}500)(0.060)(1 - 0.269] + (4{,}000/37{,}500)(0.055) + (18{,}100/37{,}500)(0.11) = 7.70\%.$$

$$\text{Value of operating assets} = \text{FCFF}_0(1 + g)/(\text{WACC} - g).$$

$$\text{Value of operating assets} = 3{,}226 \text{ million}(1 + 0.045)/(0.0770 - 0.045) = €105{,}349.06 \text{ million.}$$

Total value of the company = Value of operating assets + Value of non-operating assets.

Total value of the company = 105,349.06 million + 50 million = €105,399.06 million.

36 A is correct.

The total market value of the firm is the sum of the debt, preferred stock, and common stock market values: 15,400 + 4,000 + 18,100 = 37,500 million.

$$\text{WACC} = [w_d \times r_d(1 - \text{Tax rate})] + (w_p \times r_p) + (w_e \times r_e)$$

= [(15,400/37,500)(0.060)(1 − 0.269] + (4,000/37,500)(0.055) + (18,100/37,500)(0.11)

= 7.70%.

Under the assumption that Bern has a low growth rate because it did not receive regulatory approval for its new drugs, the value of Bern can be analyzed using a two-stage valuation model.

$$\text{Company value} = \sum_{t=1}^{n}\frac{\text{FCFF}_t}{(1+\text{WACC})^t} + \frac{\text{FCFF}_{n+1}}{(\text{WACC}-g)}\frac{1}{(1+\text{WACC})^n}.$$

Year	0	1	2	3	4
g		1.50%	1.50%	1.50%	0.75%
FCFF_n (€ millions)	3,226	3,274.39	3,323.51	3,373.36	3,398.66
Present Value Factor		0.928529	0.862167	0.800547	
Present Value (€ millions)		3,040.37	2,865.42	2,700.53	

The terminal value at the end of Year 3 is $\text{TV}_3 = \text{FCFF}_4/(\text{WACC} - g_4)$.

TV_3 = 3,398.66/(0.0770 − 0.0075) = €48,901.58 million.

The total value of operating assets = (3,040.37 + 2,865.42 + 2,700.53) + 48,901.58/(1 + 0.0770)^3

= 8,606.32 + 39,144.95

= €47,751.27 million.

Value of Bern's common stock = Value of operating assets + Value of non-operating assets − Market value of debt − Preferred stock

= 47,751.27 + 50.00 − 15,400 − 4,000

= €28,401.27 million.

Since the current market value of Bern's common stock (€18,100 million) is less than the estimated value (€28,401.27 million), the shares are undervalued.

READING

25

Market-Based Valuation: Price and Enterprise Value Multiples

by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA

Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA). Thomas R. Robinson, PhD, CFA, CAIA, Robinson Global Investment Management LLC, (USA). John D. Stowe, PhD, CFA, is at Ohio University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** contrast the method of comparables and the method based on forecasted fundamentals as approaches to using price multiples in valuation and explain economic rationales for each approach;
☐	**b.** calculate and interpret a justified price multiple;
☐	**c.** describe rationales for and possible drawbacks to using alternative price multiples and dividend yield in valuation;
☐	**d.** calculate and interpret alternative price multiples and dividend yield;
☐	**e.** calculate and interpret underlying earnings, explain methods of normalizing earnings per share (EPS), and calculate normalized EPS;
☐	**f.** explain and justify the use of earnings yield (E/P);
☐	**g.** describe fundamental factors that influence alternative price multiples and dividend yield;
☐	**h.** calculate and interpret the justified price-to-earnings ratio (P/E), price-to-book ratio (P/B), and price-to-sales ratio (P/S) for a stock, based on forecasted fundamentals;
☐	**i.** calculate and interpret a predicted P/E, given a cross-sectional regression on fundamentals, and explain limitations to the cross-sectional regression methodology;
☐	**j.** evaluate a stock by the method of comparables and explain the importance of fundamentals in using the method of comparables;
☐	**k.** calculate and interpret the P/E-to-growth (PEG) ratio and explain its use in relative valuation;

(continued)

© 2020 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**l.** calculate and explain the use of price multiples in determining terminal value in a multistage discounted cash flow (DCF) model;
☐	**m.** explain alternative definitions of cash flow used in price and enterprise value (EV) multiples and describe limitations of each definition;
☐	**n.** calculate and interpret EV multiples and evaluate the use of EV/EBITDA;
☐	**o.** explain sources of differences in cross-border valuation comparisons;
☐	**p.** describe momentum indicators and their use in valuation;
☐	**q.** explain the use of the arithmetic mean, the harmonic mean, the weighted harmonic mean, and the median to describe the central tendency of a group of multiples;
☐	**r.** evaluate whether a stock is overvalued, fairly valued, or undervalued based on comparisons of multiples.

1 INTRODUCTION TO MARKET-BASED VALUATION

a contrast the method of comparables and the method based on forecasted fundamentals as approaches to using price multiples in valuation and explain economic rationales for each approach;

Among the most familiar and widely used valuation tools are price and enterprise value multiples. **Price multiples** are ratios of a stock's market price to some measure of fundamental value per share. **Enterprise value multiples**, by contrast, relate the total market value of all sources of a company's capital to a measure of fundamental value for the entire company.

The intuition behind price multiples is that investors evaluate the price of a share of stock—judge whether it is fairly valued, overvalued, or undervalued—by considering what a share buys in terms of per share earnings, net assets, cash flow, or some other measure of value (stated on a per share basis). The intuition behind enterprise value multiples is similar; investors evaluate the market value of an entire enterprise relative to the amount of earnings before interest, taxes, depreciation, and amortization (EBITDA), sales, or operating cash flow it generates. As valuation indicators (measures or indicators of value), multiples have the appealing qualities of simplicity in use and ease in communication. A multiple summarizes in a single number the relationship between the market value of a company's stock (or of its total capital) and some fundamental quantity, such as earnings, sales, or **book value** (owners' equity based on accounting values).

Among the questions we will study for answers that will help in making correct use of multiples as valuation tools are the following:

- What accounting issues affect particular price and enterprise value multiples, and how can analysts address them?

- How do price multiples relate to fundamentals, such as earnings growth rates, and how can analysts use this information when making valuation comparisons among stocks?
- For which types of valuation problems is a particular price or enterprise value multiple appropriate or inappropriate?
- What challenges arise in applying price and enterprise value multiples internationally?

Multiples may be viewed as valuation indicators relating to individual securities. Another type of valuation indicator used in security selection is **momentum indicators**. They typically relate either price or a fundamental (such as earnings) to the time series of its own past values or, in some cases, to its expected value. The logic behind the use of momentum indicators is that such indicators may provide information on future patterns of returns over some time horizon. Because the purpose of momentum indicators is to identify potentially rewarding investment opportunities, they can be viewed as a class of valuation indicators with a focus that is different from and complementary to the focus of price and enterprise value multiples.

We first put the use of price and enterprise value multiples in an economic context and present certain themes common to the use of any price or enterprise value multiple. We then present price multiples. The treatment of each multiple follows a common format: usage considerations, the relationship of the multiple to investors' expectations about fundamentals, and using the multiple in valuation based on comparables. The subsequent sections present enterprise value multiples, international considerations in using multiples, and treatment of momentum indicators. We then discuss several practical issues that arise in using valuation indicators.

1.1 Price and Enterprise Value Multiples in Valuation

In practice, two methods underpin analysts' use of price and enterprise value multiples: the method of comparables and the method based on forecasted fundamentals. Each of these methods relates to a definite economic rationale. In this section, we introduce the two methods and their associated economic rationales.

1.1.1 *The Method of Comparables*

The **method of comparables** refers to the valuation of an asset based on multiples of comparable (similar) assets—that is, valuation based on multiples benchmarked to the multiples of similar assets. The similar assets may be referred to as the **comparables**, the **comps**, or the **guideline assets** (or in the case of equity valuation, **guideline companies**). For example, multiplying a benchmark value of the price-to-earnings (P/E) multiple by an estimate of a company's earnings per share (EPS) provides a quick estimate of the value of the company's stock that can be compared with the stock's market price. Equivalently, comparing a stock's actual price multiple with a relevant benchmark multiple should lead the analyst to the same conclusion on whether the stock is relatively fairly valued, relatively undervalued, or relatively overvalued.

The idea behind price multiples is that a stock's price cannot be evaluated in isolation. Rather, it needs to be evaluated in relation to what it buys in terms of earnings, net assets, or some other measure of value. Obtained by dividing price by a measure of value per share, a price multiple gives the price to purchase one unit of value in whatever way value is measured. For example, a P/E of 20 means that it takes 20 units of currency (for example, €20) to buy one unit of earnings (for example, €1 of earnings). This scaling of price per share by value per share also makes possible comparisons among various stocks. For example, an investor pays more for a unit of earnings for a stock with a P/E of 25 than for another stock with a P/E of 20. Applying the method of comparables, the analyst would reason that if the securities are otherwise closely similar (if they have similar risk, profit margins, and growth prospects, for example), the security with the P/E of 20 is undervalued relative to the one with the P/E of 25.

The word *relative* is necessary. An asset may be undervalued relative to a comparison asset or group of assets, and an analyst may thus expect the asset to outperform the comparison asset or assets on a relative basis. If the comparison asset or assets themselves are not efficiently priced, however, the stock may not be undervalued: It could be fairly valued or even overvalued (on an absolute basis, i.e., in relation to its intrinsic value). Example 1 presents the method of comparables in its simplest application.

EXAMPLE 1

The Method of Comparables at Its Simplest

Company A's EPS is \$1.50. Its closest competitor, Company B, is trading at a P/E of 22. Assume the companies have a similar operating and financial profile.

1 If Company A's stock is trading at \$37.50, what does that indicate about its value relative to Company B?

2 If we assume that Company A's stock should trade at about the same P/E as Company B's stock, what will we estimate as an appropriate price for Company A's stock?

Solution to 1:

If Company A's stock is trading at \$37.50, its P/E will be 25 (\$37.50 divided by \$1.50). If the companies are similar, this P/E would indicate that Company A is overvalued relative to Company B.

Solution to 2:

If we assume that Company A's stock should trade at about the same P/E as Company B's stock, we will estimate that an appropriate price for Company A's stock is \$33 (\$1.50 times 22).

The method of comparables applies also to enterprise value multiples. In this application, we would evaluate the market value of an entire company in relation to some measure of value relevant to all providers of capital, not only providers of equity capital. For example, multiplying a benchmark multiple of enterprise value (EV) to earnings before interest, taxes, depreciation, and amortization (EBITDA) times an estimate of a company's EBITDA provides a quick estimate of the value of the entire company. Similarly, comparing a company's actual enterprise value multiple with a relevant benchmark multiple allows an assessment of whether the company is relatively fairly valued, relatively undervalued, or relatively overvalued.

Many choices for the benchmark value of a multiple have appeared in valuation methodologies, including the multiple of a closely matched individual stock and the average or median value of the multiple for the stock's industry peer group. The economic rationale underlying the method of comparables is the **law of one price**—the economic principle that two identical assets should sell at the same price. The method of comparables is perhaps the most widely used approach for analysts *reporting* valuation judgments on the basis of price multiples. For this reason, the use of multiples in valuation is sometimes viewed solely as a type of relative-valuation approach; however, multiples can also be derived from, and expressed in terms of, fundamentals, as discussed in the next section.

1.1.2 *The Method Based on Forecasted Fundamentals*

The **method based on forecasted fundamentals** refers to the use of multiples that are derived from forecasted fundamentals—characteristics of a business related to profitability, growth, or financial strength. For brevity, we sometimes use the phrase "based on fundamentals" in describing multiples derived using this approach. Fundamentals drive cash flows, and we can relate multiples to company fundamentals through a discounted cash flow (DCF) model. Algebraic expressions of price multiples in terms of fundamentals facilitate an examination of how valuation differences among stocks relate to different expectations for those fundamentals.

One process for relating multiples to forecasted fundamentals begins with a valuation based on a DCF model. Recall that DCF models estimate the intrinsic value of a firm or its equity as the present value of expected cash flows and that fundamentals drive cash flows. Multiples are stated with respect to a single value of a fundamental, but any price or enterprise value multiple relates to the entire future stream of expected cash flows through its DCF value.

We can illustrate this concept by first taking the present value of the stream of expected future cash flows and then expressing the result relative to a forecasted fundamental. For example, if the DCF value of a UK stock is £10.20 and its forecasted EPS is £1.2, the forward P/E multiple consistent with the DCF value is £10.20/£1.2 = 8.5. (The term **forward P/E** refers to a P/E calculated on the basis of a forecast of EPS and is discussed in further detail later in this reading.) This exercise of relating a valuation to a price multiple applies to any definition of price multiple and any DCF model or residual income model.

In summary, we can approach valuation by using multiples from two perspectives. First, we can use the method of comparables, which involves comparing an asset's multiple to a standard of comparison. Similar assets should sell at similar prices. Second, we can use the method based on forecasted fundamentals, which involves forecasting the company's fundamentals rather than making comparisons with other companies. The price multiple of an asset should be related to its expected future cash flows. We can also incorporate the insights from the method based on forecasted fundamentals in explaining valuation differences based on comparables, because we seldom (if ever) find exact comparables. In the sections covering each multiple, we will present the method based on forecasted fundamentals first so we can refer to it when using the method of comparables.

Using either method, how can an analyst communicate a view about the value of a stock? Of course, the analyst simply can offer a qualitative judgment about whether the stock appears to be fairly valued, overvalued, or undervalued (and offer specific reasons for the view). The analyst may also be more precise by communicating a **justified price multiple** for the stock. The justified price multiple is the estimated **fair value** of that multiple, which can be justified on the basis of the method of comparables or the method of forecasted fundamentals.

For an example of a justified multiple based on the method of comparables, suppose we use the price-to-book (P/B) multiple in a valuation and find that the median P/B for the company's peer group, which would be the standard of comparison, is 2.2. Note that we are using the median rather than the mean value of the peer group's multiple to avoid distortions from outliers—an important issue when dealing with peer groups that often consist of a small number of companies. The stock's justified P/B based on the method of comparables is 2.2 (without making any adjustments for differences in fundamentals). We can compare the justified P/B with the actual P/B based on market price to form an opinion about value. If the justified P/B is larger (smaller) than the actual P/B, the stock may be undervalued (overvalued). We can also, on the assumption that the comparison assets are fairly priced, translate the justified

P/B based on comparables into an estimate of absolute fair value of the stock. If the current book value per share is \$23, then the fair value of the stock is 2.2 × \$23 = \$50.60, which can be compared with its market price.

For an example of a justified multiple based on fundamentals, suppose that we are using a residual income model and estimate that the value of the stock is \$46. Then, the justified P/B based on forecasted fundamentals is \$46/\$23 = 2.0, which we can again compare with the actual value of the stock's P/B. We can also state our estimate of the stock's absolute fair value as 2 × \$23 = \$46. (Note that the analyst could report valuation judgments related to a DCF model in terms of the DCF value directly; price multiples are a familiar form, however, in which to state valuations.) Furthermore, we can incorporate the insights from the method based on fundamentals to explain differences from results based on comparables.

In the next section, we begin a discussion of specific price and enterprise value multiples used in valuation.

2 PRICE TO EARNINGS: THE BASICS

b calculate and interpret a justified price multiple;

c describe rationales for and possible drawbacks to using alternative price multiples and dividend yield in valuation;

d calculate and interpret alternative price multiples and dividend yield;

e calculate and interpret underlying earnings, explain methods of normalizing earnings per share (EPS), and calculate normalized EPS;

f explain and justify the use of earnings yield (E/P);

In this section, we first discuss the most familiar price multiple, the price-to-earnings ratio. In the context of that discussion, we introduce a variety of practical issues that have counterparts for most other multiples. These issues include analyst adjustments to the denominator of the ratio for accuracy and comparability and the use of inverse price multiples. Then, we discuss four other major price multiples from the same practical perspective.

2.1 Price to Earnings

In the first edition of *Security Analysis* (Graham and Dodd 1934, p. 351), Benjamin Graham and David L. Dodd described common stock valuation based on P/Es as the standard method of that era, and the P/E is still the most familiar valuation measure today.

We begin our discussion with rationales offered by analysts for the use of P/E and with the possible drawbacks of its use. We then define the two chief variations of the P/E: the trailing P/E and the forward P/E (also called the "leading P/E"). The multiple's numerator, market price, is (as in other multiples) definitely determinable; it presents no special problems of interpretation. But the denominator, EPS, is based on the complex rules of accrual accounting and presents significant interpretation issues. We discuss those issues and the adjustments analysts can make to obtain more-meaningful P/Es. Finally, we conclude the section by examining how analysts use P/Es to value a stock using the method of forecasted fundamentals and the method of comparables. As mentioned earlier, we discuss fundamentals first so that we can draw insights from that discussion when using comparables.

Several rationales support the use of P/E multiples in valuation:

- Earning power is a chief driver of investment value, and EPS, the denominator in the P/E ratio, is perhaps the chief focus of security analysts' attention. Surveys show that P/E ranks first among price multiples used in market-based valuation (2007 survey of CFA Institute members; for more details, see Pinto, Robinson, and Stowe 2018) and that it is the most popular valuation metric when making investment decisions (2012 BofA Merrill Lynch Institutional Factor Survey).
- The P/E ratio is widely recognized and used by investors.
- Differences in stocks' P/Es may be related to differences in long-run average returns on investments in those stocks, according to empirical research (Chan and Lakonishok 2004).

Potential drawbacks to using P/Es derive from the characteristics of EPS:

- EPS can be zero, negative, or insignificantly small relative to price, and P/E does not make economic sense with a zero, negative, or insignificantly small denominator.
- The ongoing or recurring components of earnings that are most important in determining intrinsic value can be practically difficult to distinguish from transient components.
- The application of accounting standards requires corporate managers to choose among acceptable alternatives and to use estimates in reporting. In making such choices and estimates, managers may distort EPS as an accurate reflection of economic performance. Such distortions may affect the comparability of P/Es among companies.

Methods to address these potential drawbacks will be discussed later in the reading. In the next section, we discuss alternative definitions of P/E based on alternative specifications of earnings.

2.1.1 *Alternative Definitions of P/E*

In calculating a P/E, the numerator most commonly used is the current price of the common stock, which is generally easily obtained and unambiguous for publicly traded companies. Selecting the appropriate EPS figure to be used in the denominator is not as straightforward. The following two issues must be considered:

- the time horizon over which earnings are measured, which results in alternative definitions of P/E, and
- adjustments to accounting earnings that the analyst may make so that P/Es for various companies can be compared.

Common alternative definitions of P/E are trailing P/E and forward P/E.

- A stock's **trailing P/E** (sometimes referred to as a current P/E) is its current market price divided by the most recent four quarters' EPS. In such calculations, EPS is sometimes referred to as "trailing 12-month (TTM) EPS." Note, however, that the Value Line Investment Survey uses "current P/E" to mean a P/E based on EPS for the most recent six months plus the projected EPS for the coming six months. That calculation blends historical and forward-looking elements.
- The **forward P/E** (also called the **leading P/E** or **prospective P/E**) is a stock's current price divided by next year's expected earnings. Trailing P/E is the P/E usually presented first in stock profiles that appear in financial databases, but

most databases also provide the forward P/E. In practice, the forward P/E has a number of important variations that depend on how "next year" is defined, as we discuss later.

Other names and time-horizon definitions for P/E exist. For example, Thomson First Call (part of Refinitiv) provides various P/Es, including ratios that have as the denominator a stock's trailing 12-month EPS, last reported annual EPS, and EPS forecasted for one year to three years ahead. Another example is Value Line's company reports which display a median P/E, which is a rounded average of the four middle values of the range of annual average P/Es over the past 10 years.

In using the P/E, an analyst should apply the same definition to all companies and time periods under examination. Otherwise, the P/Es are not comparable, for a given company over time or for various companies at a specific point in time. One reason is that the differences in P/Es calculated by different methods may be systematic (as opposed to random). For example, for companies with rising earnings, the forward P/E will be smaller than the trailing P/E because the denominator in the forward P/E calculation will be larger.

Valuation is a forward-looking process, so analysts usually focus on the forward P/E when earnings forecasts are available. For large public companies, an analyst can develop earnings forecasts and/or obtain consensus earnings forecasts from a commercial database. When earnings are not readily predictable, however, a trailing P/E (or another valuation metric) may be more appropriate than a forward P/E. Furthermore, logic sometimes indicates that a particular definition of the P/E is not relevant. For example, a major acquisition or divestiture or a significant change in financial leverage may change a company's operating or financial risk so much that the trailing P/E based on past EPS is not informative about the future and thus not relevant to a valuation. In such a case, the forward P/E is the appropriate measure. In the following sections, we address issues that arise in calculating trailing and forward P/Es.

Trailing P/Es and forward P/Es are based on a single year's EPS. If that number is negative or viewed as unrepresentative of a company's earning power, however, an analyst may base the P/E calculation on a longer-run expected average EPS value. P/Es based on such normalized EPS data may be called **normalized P/Es**. Because the denominators in normalized P/Es are typically based on historical information, they are covered in the next section on calculating the trailing P/E.

2.1.2 *Calculating the Trailing P/E*

When using trailing earnings to calculate a P/E, the analyst must take care in determining the EPS to be used in the denominator. The analyst must consider the following:

- potential **dilution** of EPS (a reduction in proportional ownership interest as a result of the issuance of new shares.);
- transitory, nonrecurring components of earnings that are company specific;
- transitory components of earnings ascribable to cyclicality (business or industry cyclicality); and
- differences in accounting methods (when different companies' stocks are being compared).

Among the considerations mentioned, potential dilution of EPS generally makes the least demands on analysts' accounting expertise because companies are themselves required to present both basic EPS and diluted EPS. **Basic earnings per share** data reflect total earnings divided by the weighted average number of shares actually outstanding during the period. **Diluted earnings per share** reflects division by the number of shares that would be outstanding if holders of securities such as executive stock options, equity warrants, and convertible bonds exercised their options to obtain common stock. The diluted EPS measure also reflects the effect of such conversion on the numerator, earnings. For example, conversion of a convertible bond affects both the numerator (earnings) and the denominator (number of shares) in the EPS calculation. Because companies present both EPS numbers, the analyst does not need to make the computation. Companies also typically report details of the EPS computation in a footnote to the financial statements. Example 2, illustrating the first bullet point, shows the typical case in which the P/E based on diluted EPS is higher than the P/E based on basic EPS.

EXAMPLE 2

Basic versus Diluted EPS

For the fiscal year ended 30 September 2018, Siemens AG (SIE-DE) reported basic EPS of €7.12 and diluted EPS of €7.01. Based on a closing stock price of €95.94 on 29 March 2019, the trailing P/E for Siemens is 13.47 if basic EPS is used and 13.69 if diluted EPS is used.

When comparing companies, analysts generally prefer to use diluted EPS so that the EPS of companies with differing amounts of dilutive securities are on a comparable basis. The other bulleted considerations frequently lead to analyst adjustments to reported earnings numbers and are discussed in order below.

2.1.2.1 Analyst Adjustments for Nonrecurring Items Items in earnings that are not expected to recur in the future are generally removed by analysts because valuation concentrates on future cash flows. The analyst's focus is on estimating **underlying earnings** (other names for this concept include **persistent earnings**, **continuing earnings**, and **core earnings**)—that is, earnings that exclude nonrecurring items. An increase in underlying earnings reflects an increase in earnings that the analyst expects to persist into the future. Companies may disclose adjusted earnings, which may be called non-IFRS earnings (because they differ, as a result of adjustments, from earnings as reportable under International Financial Reporting Standards), non-GAAP earnings (because they differ, as a result of adjustments, from earnings as reportable under US generally accepted accounting principles), pro forma earnings, adjusted earnings, or, as in Example 3, core earnings. All of these terms indicate that the earnings number differs in some way from that presented in conformity with accounting standards. Example 3 shows the calculation of EPS and P/E before and after analyst adjustments for nonrecurring items.

EXAMPLE 3

Calculating Trailing 12-Month EPS and Adjusting EPS for Nonrecurring Items

You are calculating a trailing P/E for Evergreen PLC as of 31 May 20X9, when the share price closed at £50.11 in London. In its first quarter of 20X9, ended 31 March, Evergreen reported basic and diluted EPS according to IFRS of £0.81, which included £0.34 of restructuring costs and £0.26 of amortization of intangibles arising from acquisitions. Adjusting for all of these items, Evergreen reported "core EPS" of £1.41 for the first quarter of 20X9, compared with core EPS of £1.87 for the first quarter of 20X8. Because the core EPS differed from the EPS calculated under IFRS, the company provided a reconciliation of the two EPS figures.

Other data for Evergreen as of 31 March 20X9 are given below. The trailing 12-month diluted EPS for 31 March 20X9 includes one quarter in 20X9 and three quarters in 20X8.

Measure	Full Year 20X8 (a)	Less 1st Quarter 20X8 (b)	Three Quarters of 20X8 (c = a – b)	Plus 1st Quarter 20X9 (d)	Trailing 12-Month EPS (e = c + d)
Reported diluted EPS	£4.98	£1.27	£3.71	£0.81	£4.52
Core EPS	£6.41	£1.81	£4.60	£1.41	£6.01
EPS excluding 20X8 legal provisions	£5.07	£1.28	£3.79	£0.81	£4.60

Based on the table and information about Evergreen, address the following:

1 Based on the company's reported EPS, determine the trailing P/E of Evergreen as of 31 March 20X9.

2 Determine the trailing P/E of Evergreen as of 31 March 20X9 using core earnings as determined by Evergreen.

Suppose you expect the amortization charges to continue for some years and note that, although Evergreen excluded restructuring charges from its core earnings calculation, Evergreen has reported restructuring charges in previous years. After reviewing all relevant data, you conclude that, in this instance, only the legal provision related to a previously disclosed legal matter should be viewed as clearly nonrecurring.

3 Determine the trailing P/E based on your adjustment to EPS.

Solution to 1:

Based on reported EPS and without any adjustments for nonrecurring items, the trailing P/E is £50.11/£4.52 = 11.1.

Solution to 2:

Using the company's reported core earnings, you find that the trailing EPS would be £6.01 and the trailing P/E would be £50.11/£6.01 = 8.3.

Solution to 3:

The trailing EPS excluding only what you consider to be nonrecurring items is £4.60, and the trailing P/E on that basis is £50.11/£4.60 = 10.9.

Example 3 makes several important points:

- By any of its various names, underlying earnings, or core earnings, is a non-IFRS concept without prescribed rules for its calculation.
- An analyst's calculation of underlying earnings may well differ from that of the company supplying the earnings numbers. Company-reported core earnings may not be comparable among companies because of differing bases of calculation. Analysts should thus always carefully examine the calculation and, generally, should not rely on such company-reported core earnings numbers.
- In general, the P/E that an analyst uses in valuation should reflect the analyst's judgment about the company's underlying earnings and should be calculated on a consistent basis among all stocks under review.

The identification of nonrecurring items often requires detailed work—in particular, examination of the income statement, the footnotes to the income statement, and the management discussion and analysis section. The analyst cannot rely on income statement classifications alone to identify nonrecurring components of earnings. Nonrecurring items (for example, gains and losses from the sale of assets, asset **write-downs**, goodwill impairment, provisions for future losses, and changes in **accounting estimates**) often appear in the income from continuing operations portion of a business's income statement. An analyst may decide not to exclude income/loss from discontinued operations when assets released from discontinued operations are redirected back into the company's earnings base. An analyst who takes income statement classifications at face value may draw incorrect conclusions in a valuation.

This discussion does not exhaust the analysis that may be necessary to distinguish earnings components that are expected to persist into the future from those that are not. For example, earnings may be decomposed into cash flow and accrual components (where the accrual component of earnings is the difference between a cash measure of earnings and a measure of earnings under the relevant set of accounting standards). Some research indicates that the cash flow component of earnings should receive a greater weight than the accrual component of earnings in valuation, and analysts may attempt to reflect that conclusion in the earnings used in calculating P/Es.

2.1.2.2 Analyst Adjustments for Business-Cycle Influences In addition to company-specific effects, such as restructuring costs, transitory effects on earnings can come from business-cycle or industry-cycle influences. These effects are somewhat different from company-specific effects. Because business cycles repeat, business-cycle effects, although transitory, can be expected to recur in subsequent cycles.

Because of cyclical effects, the most recent four quarters of earnings may not accurately reflect the average or long-term earning power of the business, particularly for **cyclical businesses**—those with high sensitivity to business- or industry-cycle influences, such as automobile and steel manufacturers. Trailing EPS for such stocks is often depressed or negative at the bottom of a cycle and unusually high at the top of a cycle. Empirically, P/Es for cyclical companies are often highly volatile over a cycle even without any change in business prospects: High P/Es on depressed EPS at the bottom of the cycle and low P/Es on unusually high EPS at the top of the cycle reflect the countercyclical property of P/Es known as the **Molodovsky effect**, named after Nicholas Molodovsky, who wrote on this subject in the 1950s and referred to using average earnings as a simple starting point for understanding a company's underlying earnings power. Analysts address this problem by normalizing EPS—that is, estimating the level of EPS that the business could be expected to achieve under mid-cyclical conditions (**normalized EPS** or **normal EPS**). Please note that we are using the term "normalized earnings" to refer to earnings adjusted for the effects of a business cycle. Some sources use the term "normalized earnings" also to refer to earnings adjusted for nonrecurring items.

Two of several available methods to calculate normalized EPS are as follows:

- The method of *historical average EPS*, in which normalized EPS is calculated as average EPS over the most recent full cycle
- The method of *average return on equity*, in which normalized EPS is calculated as the average return on equity (ROE) from the most recent full cycle, multiplied by current book value per share

The first method is one of several possible statistical approaches to the problem of cyclical earnings; however, this method does not account for changes in a business's size. The second alternative, by using recent book value per share, reflects more accurately the effect on EPS of growth or shrinkage in the company's size. For that reason, the method of average ROE is sometimes preferred. When reported current book value does not adequately reflect company size in relation to past values (because of items such as large write-downs), the analyst can make appropriate accounting adjustments. The analyst can also estimate normalized earnings by multiplying total assets by an estimate of the long-run return on total assets or by multiplying shareholders' equity by an estimate of the long-run return on total shareholders' equity. These methods are particularly useful for a period in which a cyclical company has reported a loss.

Example 4 illustrates this concept. The example uses data for an **American Depositary Receipt** (ADR) but is applicable to any equity security. An ADR is intended to facilitate US investment in non-US companies. It is a negotiable certificate issued by a depositary bank that represents ownership in a non-US company's deposited equity (i.e., equity held in custody by the depositary bank in the company's home market). One ADR may represent one, more than one, or less than one deposited share. The number of or fraction of deposited securities represented by one ADR is referred to as the "ADR ratio."

EXAMPLE 4

Normalizing EPS for Business-Cycle Effects

You are researching the valuation of Zenlandia Chemical Company, a large (fictitious) manufacturer of specialty chemicals. Your research is for a US investor who is interested in the company's ADRs rather than the company's shares listed on the Zenlandia Stock Exchange. On 5 July 2021, the closing price of the US-listed ADR was $18.21. The chemical industry is notably cyclical, so you decide to normalize earnings as part of your analysis. You believe that data from 2014 reasonably capture the beginning of the most recent business cycle, and you want to evaluate a normalized P/E. Exhibit 1 supplies data on EPS (based on Zenlandia GAAP) for one ADR, book value per share (BVPS) for one ADR, and the company's ROE.

Exhibit 1 Zenlandia Chemical Company (Currency in US Dollars)

Measure	2014	2015	2016	2017	2018	2019	2020
EPS (ADR)	\$0.74	\$0.63	\$0.61	\$0.54	\$1.07	\$0.88	\$1.08
BVPS (ADR)	\$3.00	\$2.93	\$2.85	\$2.99	\$3.80	\$4.03	\$4.82
ROE	24.7%	21.5%	21.4%	18.1%	28.2%	21.8%	22.4%

Note: This example involves a single company. When the analyst compares multiple companies on the basis of P/Es based on normalized EPS and uses this normalization approach, the analyst should be sure that the ROEs are being calculated consistently by the subject companies. In this example, ROE for each year is being calculated by using ending BVPS and, essentially, trailing earnings are being normalized.

Using the data in Exhibit 1:

1 Calculate a normalized EPS by the method of historical average EPS and then calculate the P/E based on that estimate of normalized EPS.

2 Calculate a normalized EPS by the method of average ROE and the P/E based on that estimate of normalized EPS.

3 Explain the source of the differences in the normalized EPS calculated by the two methods, and contrast the impact on the estimate of a normalized P/E.

Solution to 1:

Averaging EPS over the 2014–20 period, you would find it to be ($0.74 + $0.63 + $0.61 + $0.54 + $1.07 + $0.88 + $1.08)/7 = $0.79. Thus, according to the method of historical average EPS, normalized EPS is $0.79. The P/E based on this estimate is $18.21/$0.79 = 23.1.

Solution to 2:

Average ROE over the 2014–20 period is (24.7% + 21.5% + 21.4% + 18.1% + 28.2% + 21.8% + 22.4%)/7 = 22.6%. Based on the current BVPS of $4.82, the method of average ROE gives 0.226 × $4.82 = $1.09 as normalized EPS. The P/E based on this estimate is $18.21/$1.09 = 16.7.

Solution to 3:

From 2014 to 2020, BVPS increased from $3.00 to $4.82, an increase of about 61%. The estimate of normalized EPS of $1.09 from the average ROE method reflects the use of information on the current size of the company better than does the $0.79 calculated from the historical average EPS method. Because of that difference, the company appears more conservatively valued (as indicated by a lower P/E) when the method based on average ROE is used.

2.1.2.3 Analyst Adjustments for Comparability with Other Companies Analysts adjust EPS for differences in accounting methods between the company and companies it is being compared with so that the P/Es will be comparable. For example, if an analyst is comparing a company that uses the last-in, first-out (LIFO) method of inventory accounting as permitted by US GAAP (but not by IFRS) with another company that uses the first-in, first-out (FIFO) method, the analyst should adjust earnings to provide comparability in all ratio and valuation analyses. In general, any adjustment made to a company's reported financials for purposes of financial statement analysis should be incorporated into an analysis of P/E and other multiples.

2.1.2.4 Dealing with Extremely Low, Zero, or Negative Earnings Having addressed the challenges that arise in calculating P/E because of nonrecurring items and business-cycle influences and for comparability among companies, we present in this section the methods analysts have developed for dealing with extremely low, zero, or negative earnings.

Stock selection disciplines that use P/Es or other price multiples often involve ranking stocks from highest value of the multiple to lowest value of the multiple. The security with the lowest positive P/E has the lowest purchase cost per currency unit of earnings among the securities ranked. Zero earnings and negative earnings pose a problem if the analyst wishes to use P/E as the valuation metric. Because division by zero is undefined, P/Es cannot be calculated for zero earnings.

A P/E can technically be calculated in the case of negative earnings. Negative earnings, however, result in a negative P/E. A negative-P/E security will rank below the lowest positive-P/E security, but because earnings are negative, the negative-P/E security is actually the most costly in terms of earnings purchased. Thus, negative P/Es are not meaningful.

In some cases, an analyst might handle negative EPS by using normalized EPS instead. Also, when trailing EPS is negative, the year-ahead EPS and thus the forward P/E may be positive. An argument in favor of either of these approaches based on positive earnings is that if a company is appropriately treated as a going concern, losses cannot be the usual operating result.

If the analyst is interested in a ranking, however, one solution (applicable to any ratio involving a quantity that can be negative or zero) is the use of an **inverse price ratio**—that is, the reciprocal of the original ratio, which places price in the denominator. The use of inverse price multiples addresses the issue of consistent ranking because price is never negative. In the case of the P/E, the inverse price ratio is earnings to price (E/P), known as the **earnings yield**. Ranked by earnings yield from highest to lowest, the securities are correctly ranked from cheapest to most costly in terms of the amount of earnings one unit of currency buys. Earnings yield can be based on normalized EPS, expected next-year EPS, or trailing EPS. In these cases also, earnings yield provides a consistent ranking.

Exhibit 2 illustrates these points for a group of automobile companies, one of which has a negative EPS. When reporting a P/E based on negative earnings, analysts should report such P/Es as "NM" (not meaningful).

Exhibit 2 P/E and E/P for Five Automobile Companies (as of 28 June 2019; in US Dollars)

Company	Current Price	Diluted EPS (TTM)	P/E (TTM)	E/P (%)
Ford Motor Co. (F)	10.28	0.78	13.2	7.59
Honda Motor Co.	25.85	3.12	8.3	12.06
Fiat Chrysler	13.88	2.32	6.0	16.71
General Motors	38.57	6.29	11.72	8.53
Tesla Inc.	224.45	–7.72	NM	–2.51

Source: Yahoo! Finance.

In addition to zero and negative earnings, extremely low earnings can pose problems when using P/Es—particularly for evaluating the distribution of P/Es of a group of stocks under review. In this case, again, inverse price ratios can be useful. The P/E of a stock with extremely low earnings may, nevertheless, be extremely high because an earnings rebound is anticipated. An extremely high P/E—an outlier P/E—can overwhelm the effect of the other P/Es in the calculation of the mean P/E. Although the use of median P/Es and other techniques can mitigate the problem of skewness caused by outliers, the distribution of inverse price ratios is inherently less susceptible to outlier-induced skewness.

As mentioned, earnings yield is but one example of an inverse price ratio—that is, the reciprocal of a price ratio. Exhibit 3 summarizes inverse price ratios for all the price ratios we discuss in this reading.

Exhibit 3 Summary of Price and Inverse Price Ratios

Price Ratio	Inverse Price Ratio	Comments
Price to earnings (P/E)	Earnings yield (E/P)	Both forms commonly used.
Price to book (P/B)	Book to market (B/P)*	Book value is less commonly negative than EPS. Book to market is favored in research but not common in practitioner usage.
Price to sales (P/S)	Sales to price (S/P)	S/P is rarely used except when all other ratios are being stated in the form of inverse price ratios; sales is not zero or negative in practice for going concerns.

Exhibit 3 (Continued)

Price Ratio	Inverse Price Ratio	Comments
Price to cash flow (P/CF)	Cash flow yield (CF/P)	Both forms are commonly used.
Price to dividends (P/D)	Dividend yield (D/P)	Dividend yield is much more commonly used because P/D is not calculable for non-dividend-paying stocks, but both D/P and P/D are used in discussing index valuation.

*"Book to *market*" is probably more common usage than "book to *price*." Book to market is variously abbreviated B/M, BV/MV (for "book value" and "market value"), or B/P.
Note: B, S, CF, and D are in per-share terms.

2.1.3 *Forward P/E*

The forward P/E is a major and logical alternative to the trailing P/E because valuation is naturally forward looking. In the definition of forward P/E, analysts have interpreted "next year's expected earnings" as expected EPS for

- the next four quarters,
- the next 12 months, or
- the next fiscal year.

In this section, unless otherwise stated, we use the first definition of forward P/E (i.e., the next four quarters), which is closest to how cash flows are dated in our discussion of DCF valuation. To illustrate the calculation, suppose the current market price of a stock is $15 as of 1 March 2020 and the most recently reported quarterly EPS (for the quarter ended 31 December 2019) is $0.22. Our forecasts of EPS are as follows:

- $0.15 for the quarter ending 31 March 2020,
- $0.18 for the quarter ending 30 June 2020,
- $0.18 for the quarter ending 30 September 2020, and
- $0.24 for the quarter ending 31 December 2020.

The sum of the forecasts for the next four quarters is $0.15 + $0.18 + $0.18 + $0.24 = $0.75, and the forward P/E for this stock is $15/$0.75 = 20.0.

Another important concept related to the forward P/E is the next 12-month (NTM) P/E, which corresponds in a forward-looking sense to the TTM P/E concept of trailing P/E. A stock's **NTM P/E** is its current market price divided by an estimated next 12-month EPS, which typically combines the annual EPS estimates from two fiscal years, weighted to reflect the relative proximity of the fiscal year. For example, assume that in late August 2020, an analyst is looking at Microsoft Corporation. Microsoft has a June 30 fiscal year end, so at the time of the analyst's scrutiny, there were 10 months remaining until the end of the company's 2021 fiscal year (i.e., September 2020 through June 2021, inclusive). The estimated next 12-month EPS for Microsoft would be calculated as [(10/12) × FY21E EPS] + [(2/12) × FY22E EPS]. NTM P/E is useful because it facilitates comparison of companies with different fiscal year ends without the need to use quarterly estimates, which for many companies are not available.

Applying the fiscal year concept, Thomson First Call reports a stock's "forward P/E" in two ways: first, based on the mean of analysts' *current fiscal year* (FY1 = Fiscal Year 1) forecasts, for which analysts may have actual EPS in hand for some quarters, and second, based on analysts' *following fiscal year* (FY2 = Fiscal Year 2) forecasts, which must be based entirely on forecasts. For Thomson First Call, "forward P/E" contrasts with "current P/E," which is based on the last reported annual EPS.

Clearly, analysts must be consistent in the definition of forward P/E when comparing stocks. Examples 5 and 6 illustrate two ways of calculating forward P/E.

EXAMPLE 5

Calculating a Forward P/E (1)

A market price for the common stock of IBM in late June 2019 was \$137.90. IBM's fiscal year coincides with the calendar year. At that time, the consensus EPS forecast of the 22 analysts covering IBM was \$13.91 for 2019 (FY1), and the consensus EPS forecast of 20 analysts covering IBM was \$14.17 for 2020 (FY2).

1 Calculate IBM's forward P/E based on the fiscal year consensus forecasted EPS for FY1.

2 Calculate IBM's forward P/E based on a fiscal year definition and the FY2 consensus forecasted EPS.

Solution to 1:

IBM's forward P/E is \$137.90/\$13.91 = 9.9 based on FY1 forecasted EPS. Note that this EPS number includes the reported first quarter earnings and a forecast of the three remaining quarters as of late June 2019.

Solution to 2:

IBM's forward P/E is \$137.90/\$14.17 = 9.7 based on FY2 forecasted EPS.

In Example 5, the company's EPS was expected to increase by slightly less than 2%, so the forward P/Es based on the two different EPS specifications differed from one another somewhat but not significantly. Example 6 presents the calculation of forward P/Es for a company with volatile earnings.

EXAMPLE 6

Calculating a Forward P/E (2)

In this example, we use alternative definitions of "forward" to compute forward P/Es. Exhibit 4 presents actual and forecasted EPS for Selene Gaming Corp. (Selene), which owns and operates gaming entertainment properties.

Exhibit 4 Quarterly EPS for Selene (in US Dollars; Excluding Nonrecurring Items and Discontinued Operations)

Year	31 March	30 June	30 September	31 December	Annual Estimate
2020	0.10	0.00	E(0.10)	E(0.50)	(0.50)
2021	E0.70	E0.80	E0.30	E(0.30)	1.50

Source: The Value Line Investment Survey.

On 9 August 2020, Selene closed at \$12.20. Selene's fiscal year ends on 31 December. As of 9 August 2020, solve the following problems by using the information in Exhibit 4:

1 Calculate Selene's forward P/E based on the next four quarters of forecasted EPS.

2 Calculate Selene's NTM P/E.

3 Calculate Selene's forward P/E based on a fiscal year definition and current fiscal year (2020) forecasted EPS.

4 Calculate Selene's forward P/E based on a fiscal year definition and next fiscal year (2021) forecasted EPS.

Solution to 1:

We sum forecasted EPS as follows:

3Q:2020 EPS (estimate)	(\$0.10)
4Q:2020 EPS (estimate)	(\$0.50)
1Q:2021 EPS (estimate)	\$0.70
2Q:2021 EPS (estimate)	\$0.80
Sum	\$0.90

The forward P/E by this definition is \$12.20/\$0.90 = 13.6.

Solution to 2:

As of 9 August 2020, approximately five months remained in FY2020. Therefore, the estimated next 12-month EPS for Selene would be based on annual estimates in the last column of Exhibit 4: $[(5/12) \times \text{FY20E EPS}] + [(7/12) \times \text{FY21E EPS}] = (5/12)(-0.50) + (7/12)(1.50) = 0.67$. The NTM P/E would be \$12.20/\$0.67 = 18.2.

Solution to 3:

We sum EPS as follows:

1Q:2020 EPS (actual)	\$0.10
2Q:2020 EPS (actual)	\$0.00
3Q:2020 EPS (estimate)	(\$0.10)
4Q:2020 EPS (estimate)	(\$0.50)
Sum	(\$0.50)

The forward P/E is \$12.20/(\$0.50) = −24.4, which is not meaningful.

Solution to 4:

We sum EPS as follows:

1Q:2021 EPS (estimate)	\$0.70
2Q:2021 EPS (estimate)	\$0.80
3Q:2021 EPS (estimate)	\$0.30
4Q:2021 EPS (estimate)	(\$0.30)
Sum	\$1.50

The forward P/E by this definition is \$12.20/\$1.50 = 8.1.

As illustrated in Example 6, for companies with volatile earnings, forward P/Es and thus valuations based on forward P/Es can vary dramatically depending on the definition of earnings. The analyst would probably be justified in normalizing EPS for Selene. The gaming industry is highly sensitive to discretionary spending; thus, Selene's earnings are strongly procyclical.

Having explored the issues involved in calculating P/Es, we turn to using them in valuation.

3 PRICE TO EARNINGS: VALUATION BASED ON FORECASTED FUNDAMENTALS

g describe fundamental factors that influence alternative price multiples and dividend yield;

i calculate and interpret a predicted P/E, given a cross-sectional regression on fundamentals, and explain limitations to the cross-sectional regression methodology;

The analyst who understands DCF valuation models can use them not only in developing an estimate of the justified P/E for a stock but also to gain insight into possible sources of valuation differences when the method of comparables is used. Linking P/Es to a DCF model helps us address what value the market should place on a dollar of EPS when we are given a particular set of expectations about the company's profitability, growth, and cost of capital.

3.1 Justified P/E

The simplest of all DCF models is the Gordon (constant) growth form of the dividend discount model (DDM). Presentations of discounted dividend valuation commonly show that the P/E of a share can be related to the value of a stock as calculated in the Gordon growth model through the expressions

$$\frac{P_0}{E_1} = \frac{D_1/E_1}{r-g} = \frac{1-b}{r-g} \tag{1}$$

for the forward P/E and

$$\frac{P_0}{E_0} = \frac{D_0(1+g)/E_0}{r-g} = \frac{(1-b)(1+g)}{r-g} \tag{2}$$

for the trailing P/E, where

P = price

E = earnings

D = dividends

r = required rate of return

g = dividend growth rate

b = retention rate

Under the assumption of constant dividend growth, the first expression gives the justified forward P/E and the second gives the justified trailing P/E. Note that both expressions state P/E as a function of two fundamentals: the stock's required rate of return, r, which reflects its risk, and the expected (stable) dividend growth rate, g. The dividend payout ratio, $1 - b$, also enters into the expressions.

A particular value of the P/E is associated with a set of forecasts of the fundamentals and the dividend payout ratio. This value is the stock's **justified (fundamental) P/E** based on forecasted fundamentals (that is, the P/E justified by fundamentals). All else being equal, the higher the expected dividend growth rate or the lower the stock's required rate of return, the higher the stock's intrinsic value and the higher its justified P/E.

This intuition carries over to more-complex DCF models. Using any DCF model, all else being equal, justified P/E is

- inversely related to the stock's required rate of return and
- positively related to the growth rate(s) of future expected cash flows, however defined.

We illustrate the calculation of a justified forward P/E in Example 7.

EXAMPLE 7

Forward P/E Based on Fundamental Forecasts (1)

BP p.l.c. (London: BP) is one of the world's largest integrated oil producers. The company has continued to deal with litigation concerns surrounding its role in a 2010 drilling rig accident. Jan Unger, an energy analyst, forecasts a long-term earnings retention rate, b, for BP of 40% and a long-term growth rate of 3.5%. Given the significant legal uncertainties still facing BP shareholders, Unger estimates a required rate of return of 7.6%. Based on Unger's forecasts of fundamentals and Equation 1, BP's justified forward P/E is

$$\frac{P_0}{E_1} = \frac{1-b}{\mathrm{r}-g} = \frac{1-0.40}{0.076-0.035} = 14.6.$$

When using a complex DCF model to value the stock (e.g., a model with varying growth rates and varying assumptions about dividends), the analyst may not be able to express the P/E as a function of fundamental, constant variables. In such cases, the analyst can still calculate a justified P/E by dividing the value per share (that results from a DCF model) by estimated EPS, as illustrated in Example 8. Approaches similar to this one can be used to develop other justified multiples.

EXAMPLE 8

Forward P/E Based on Fundamental Forecasts (2)

Toyota Motor Corporation is one of the world's largest vehicle manufacturers. The company's most recent fiscal year ended on 31 March 2019. In late June 2019, you are valuing Toyota stock, which closed at ¥6,688 on the previous day. You have used a free cash flow to equity (FCFE) model to value the company stock and have obtained a value of ¥6,980 for the stock. For ease of communication, you want to express your valuation in terms of a forward P/E based on your forecasted fiscal year 2020 EPS of ¥720. Toyota's fiscal year 2020 is from 1 April 2019 through 31 March 2020.

1 What is Toyota's justified P/E based on forecasted fundamentals?

2 Based on a comparison of the current price of ¥6,688 with your estimated intrinsic value of ¥6,980, the stock appears to be undervalued by approximately 4%. Use your answer to Part 1 to state this evaluation in terms of P/Es.

Solution to 1:

Value of the stock derived from FCFE = ¥6,980.

Forecasted 2014 EPS = ¥720.

¥6,980/¥720 = 9.7 is the justified forward P/E.

Solution to 2:

The justified P/E of 9.7 is about 4% higher than the forward P/E based on current market price, ¥6,688/¥720 = 9.3.

The next section illustrates another, but less commonly used, approach to relating price multiples to fundamentals.

3.2 Predicted P/E Based on Cross-Sectional Regression

A predicted P/E, which is conceptually similar to a justified P/E, can be estimated from cross-sectional regressions of P/E on the fundamentals believed to drive security valuation. Kisor and Whitbeck (1963) and Malkiel and Cragg (1970) pioneered this approach. Their studies measured P/Es for a group of stocks and the characteristics thought to determine P/E: growth rate in earnings, payout ratio, and a measure of volatility, such as standard deviation of earnings changes or beta. An analyst can conduct such cross-sectional regressions by using any set of explanatory variables considered to determine investment value; the analyst must bear in mind, however, potential distortions that can be introduced by multicollinearity among independent variables. Example 9 illustrates the prediction of P/E using cross-sectional regression.

EXAMPLE 9

Predicted P/E Based on a Cross-Sectional Regression

You are valuing a food company with a beta of 0.9, a dividend payout ratio of 0.45, and an earnings growth rate of 0.08. The estimated regression for a group of other stocks in the same industry is

$$\text{Predicted P/E} = 12.12 + (2.25 \times \text{DPR}) - (0.20 \times \text{Beta}) + (14.43 \times \text{EGR}),$$

where DPR is the dividend payout ratio and EGR is the five-year earnings growth rate.

1 Based on this cross-sectional regression, what is the predicted P/E for the food company?

2 If the stock's actual trailing P/E is 18, is the stock fairly valued, overvalued, or undervalued?

Solution to 1:

Predicted P/E = 12.12 + (2.25 × 0.45) − (0.20 × 0.9) + (14.43 × 0.08) = 14.1. The predicted P/E is 14.1.

Solution to 2:

Because the predicted P/E of 14.1 is less than the actual P/E of 18, the stock appears to be overvalued. That is, it is selling at a higher multiple than is justified by its fundamentals.

A cross-sectional regression summarizes a large amount of data in a single equation and can provide a useful additional perspective on a valuation. It is not frequently used as a main tool, however, because it is subject to at least three limitations:

- The method captures valuation relationships only for the specific stock (or sample of stocks) over a particular time period. The predictive power of the regression for a different stock and different time period is not known.

- The regression coefficients and explanatory power of the regressions tend to change substantially over a number of years. The relationships between P/E and fundamentals may thus change over time. Empirical evidence suggests that the relationships between P/Es and such characteristics as earnings growth, dividend payout, and beta are not stable over time (Damodaran 2012). Furthermore, because distributions of multiples change over time, the predictive power of results from a regression at any point in time can be expected to diminish with the passage of time (Damodaran 2012).
- Because regressions based on this method are prone to the problem of multicollinearity (correlation within linear combinations of the independent variables), interpreting individual regression coefficients is difficult.

Overall, rather than examining the relationship between a stock's P/E multiple and economic variables, the bulk of capital market research examines the relationship between companies' stock prices (and returns on the stock) and explanatory variables, one of which is often earnings (or unexpected earnings). A classic example of such research is the Fama and French (1992) study showing that, used alone, a number of factors explained cross-sectional stock returns in the 1963–90 period; the factors were E/P, size, leverage, and the book-to-market multiples. When these variables were used in combination, however, size and book to market had explanatory power that absorbed the roles of the other variables in explaining cross-sectional stock returns. Research building on that study eventually resulted in the Fama–French three-factor model (with the factors of size, book to market, and beta). Another classic academic study providing evidence that accounting variables appear to have predictive power for stock returns is Lakonishok, Shleifer, and Vishny (1994), which also provided evidence that value strategies—buying stocks with low prices relative to earnings, book value, cash flow, and sales growth—produced superior five-year buy-and-hold returns in the 1968–90 period without involving greater fundamental risk than a strategy of buying growth stocks.

4 PRICE-EARNINGS: USING THE P/E IN VALUATION

h calculate and interpret the justified price-to-earnings ratio (P/E), price-to-book ratio (P/B), and price-to-sales ratio (P/S) for a stock, based on forecasted fundamentals;

k calculate and interpret the P/E-to-growth (PEG) ratio and explain its use in relative valuation;

l calculate and explain the use of price multiples in determining terminal value in a multistage discounted cash flow (DCF) model;

r evaluate whether a stock is overvalued, fairly valued, or undervalued based on comparisons of multiples.

The most common application of the P/E approach to valuation is to estimate the value of a company's stock by applying a benchmark multiple to the company's actual or forecasted earnings. An essentially equivalent approach is to compare a stock's actual price multiple with a benchmark value of the multiple. This section explores these comparisons for P/Es. Using any multiple in the method of comparables involves the following steps:

- Select and calculate the price multiple that will be used in the comparison.

- Select the comparison asset or assets and calculate the value of the multiple for the comparison asset(s). For a group of comparison assets, calculate a median or mean value of the multiple for the assets. The result in either case is the **benchmark value of the multiple**.
- Use the benchmark value of the multiple, possibly subjectively adjusted for differences in fundamentals, to estimate the value of a company's stock. (Equivalently, compare the subject stock's actual multiple with the benchmark value.)
- When feasible, assess whether differences between the estimated value of the company's stock and the current price of the company's stock are explained by differences in the fundamental determinants of the price multiple and modify conclusions about relative valuation accordingly. (An essentially equivalent approach is to assess whether differences between a company's actual multiple and the benchmark value of the multiple can be explained by differences in fundamentals.)

These bullet points provide the structure for this reading's presentation of the method of comparables. The first price multiple that will be used in the comparison is the P/E. Practitioners' choices for the comparison assets and the benchmark value of the P/E derived from these assets include the following:

- the average or median value of the P/E for the company's peer group of companies within an industry, including an average past value of the P/E for the stock relative to this peer group;
- the average or median value of the P/E for the company's industry or sector, including an average past value of the P/E for the stock relative to the industry or sector;
- the P/E for a representative equity index, including an average past value of the P/E for the stock relative to the equity index; and
- an average past value of the P/E for the stock.

To illustrate the first bullet point, the company's P/E (say, 15) may be compared to the median P/E for the peer companies currently (say, 10), or the ratio 15/10 = 1.5 may be compared to its average past value. The P/E of the most closely matched individual stock can also be used as a benchmark; because of averaging, however, using a group of stocks or an equity index is typically expected to generate less valuation error than using a single stock. We later illustrate a comparison with a single closely matched individual stock.

Economists and investment analysts have long attempted to group companies by similarities and differences in their business operations. A country's economy overall is typically grouped most broadly into **economic sectors** or large industry groupings. These groupings differ depending on the source of the financial information, and an analyst should be aware of differences among data sources. Classifications often attempt to group companies by what they supply (e.g., energy, consumer goods), by demand characteristics (e.g., consumer discretionary), or by financial market or economic "theme" (e.g., consumer cyclical, consumer noncyclical).

Two classification systems that are widely used in equity analysis are the Global Industry Classification System (GICS) sponsored by Standard & Poor's and MSCI and the Industrial Classification Benchmark (ICB). Many other classification schemes developed by commercial and governmental organizations and by academics are also in use.

The GICS structure assigns each company to one of 158 subindustries, an industry (69 in total), an industry group (24 in total), and an economic sector (11 in total: consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, real estate, telecommunication services, and utilities). The assignment is made by a judgment as to the company's principal business activity, which is based primarily on sales. Because a company is classified on the basis of one business activity, a given company appears in just one group at each level of the classification. A classification ("industrial conglomerates") is available under the capital goods sector of industrials for companies that cannot be assigned to a principal business activity.

The ICB, like GICS, has four levels, but the terminology of ICB uses "sector" and "industry" in nearly opposite senses. The ICB is managed by FTSE Russell. At the bottom of the four levels are 173 subsectors, each of which belongs to one of 45 sectors; each sector belongs to one of 20 supersectors; and each supersector belongs to one of 11 industries at the highest level of classification. (The numbers in the groups were changed effective 1 July 2019; changes are made to the classification from time to time. See www.ftserussell.com/data/industry-classification-benchmark-icbwww.icbenchmark.com for updates.) The industries are technology, telecommunications, health care, financials, real estate, consumer discretionary, consumer staples, industrials, basic materials, energy, and utilities.

For these classification systems, analysts often choose the narrowest grouping (i.e., subindustry for GICS and subsector for ICB) as an appropriate starting point for comparison asset identification. To narrow the list of comparables in the subsector, an analyst might use information on company size (as measured by revenue or market value of equity) and information on the specific markets served.

Analysts should be aware that, although different organizations often group companies in a broadly similar fashion, sometimes they differ sharply. The lists of peer companies or competitors given by each of these organizations can be, as a result, quite distinct.

The comparable companies—selected by using any of the choices described previously—provide the basis for calculating a benchmark value of the multiple. In analyzing differences between the subject company's multiple and the benchmark value of the multiple, financial ratio analysis serves as a useful tool. Financial ratios can point out

- a company's ability to meet short-term financial obligations (liquidity ratios);
- the efficiency with which assets are being used to generate sales (asset turnover ratios);
- the use of debt in financing the business (leverage ratios);
- the degree to which fixed charges, such as interest on debt, are being met by earnings or cash flow (coverage ratios); and
- profitability (profitability ratios).

With this understanding of terms in hand, we turn to using the method of comparables. We begin with cross-sectional P/Es derived from industry peer groups and move to P/Es derived from comparison assets that are progressively less closely matched to the stock. We then turn to using historical P/Es—that is, P/Es derived from the company's own history. Finally, we sketch how both fundamentals- and comparables-driven models for P/Es can be used to calculate the terminal value in a multistage DCF valuation.

4.1 Peer-Company Multiples

Companies operating in the same industry as the subject company (i.e., its peer group) are frequently used as comparison assets. The advantage of using a peer group is that the constituent companies are typically similar in their business mix to the company being analyzed. This approach is consistent with the idea underlying the method of comparables—that similar assets should sell at similar prices. The subject stock's P/E is compared with the median or mean P/E for the peer group to arrive at a relative valuation. Equivalently, multiplying the benchmark P/E by the company's EPS provides an estimate of the stock's value that can be compared with the stock's market price. The value estimated in this way represents an estimate of intrinsic value if the comparison assets are efficiently (fairly) priced.

In practice, analysts often find that the stock being valued has some significant differences from the median or mean fundamental characteristics of the comparison assets. In applying the method of comparables, analysts usually attempt to judge whether differences from the benchmark value of the multiple can be explained by differences in the fundamental factors believed to influence the multiple. The following relationships for P/E hold, all else being equal:

- If the subject stock has higher-than-average (or higher-than-median) expected earnings growth, a higher P/E than the benchmark P/E is justified.
- If the subject stock has higher-than-average (or higher-than-median) risk (operating or financial), a lower P/E than the benchmark P/E is justified.

Another perspective on these two points is that for a group of stocks with comparable relative valuations, the stock with the greatest expected growth rate (or the lowest risk) is, all else equal, the most attractively valued. Example 10 illustrates a simple comparison of a company with its peer group.

EXAMPLE 10

A Simple Peer-Group Comparison

As a telecommunication industry analyst at a brokerage firm, you are valuing Verizon Communications, Inc., a telecommunication company. The valuation metric that you have selected is the trailing P/E. You are evaluating the P/E using the median trailing P/E of peer-group companies as the benchmark value. According to GICS, Verizon is in the telecommunication services sector and, within it, the integrated telecommunication services subindustry. Exhibit 5 presents the relevant data.

Exhibit 5 Trailing P/Es of Telecommunication Services Companies

Company	Trailing P/E
AT&T	13.20
Comcast Corporation	16.23
CenturyLink	NMF
China Telecom	13.14
Charter Communications Corp.	70.67
Verizon Communications	15.03
Windstream Holdings	19.01
Mean*	24.55
Median	15.03

**Mean, six firms excluding CenturyLink.*
NMF = not meaningful.

Based on the data in Exhibit 5, address the following:

1 Given the definition of the benchmark stated above, determine the most appropriate benchmark value of the P/E for Verizon.
2 State whether Verizon is relatively fairly valued, relatively overvalued, or relatively undervalued, assuming no differences in fundamentals among the peer group companies. Justify your answer.
3 Identify the stocks in this group of telecommunication companies that appear to be relatively undervalued when the median trailing P/E is used as a benchmark. Explain what further analysis might be appropriate to confirm your answer.

Solution to 1:

As stated earlier, the use of median values mitigates the effect of outliers on the valuation conclusion. In this instance, the P/Es for CenturyLink and Charter Communications are clearly outliers. Therefore, the median trailing P/E for the group, 15.03, is more appropriate than the mean trailing P/E of 24.55 for use as the benchmark value of the P/E. Note that when a group includes an odd number of companies, as here, the median value will be the middle value when the values are ranked (in either ascending or descending order). When the group includes an even number of companies, the median value will be the average of the two middle values.

Solution to 2:

If you assume no differences in fundamentals among the peer group companies, Verizon appears to be fairly valued because its P/E is identical to the median P/E of 15.03.

Solution to 3:

AT&T, China Telecom, and CenturyLink appear to be undervalued relative to their peers because their trailing P/Es are lower than the median P/E. Verizon appears to be relatively fairly valued because its P/E equals the median P/E. Charter Communications, Comcast Corporation, and Windstream appear to be overvalued.

To confirm this valuation conclusion, you should look at other metrics. One issue for this particular industry is that earnings may differ significantly from cash flow. These companies invest considerable amounts of money to build out their networks—whether it be landlines or increasing bandwidth capacity for mobile users. Because telecommunication service providers are frequently required to take large noncash charges on their infrastructure, reported earnings are typically very volatile and frequently much lower than cash flow.

A metric that appears to address the impact of earnings growth on P/E is the P/E-to-growth (PEG) ratio. The **PEG ratio** is calculated as the stock's P/E divided by the expected earnings growth rate (in percentage terms). The ratio, in effect, is a calculation of a stock's P/E per percentage point of expected growth. Stocks with lower PEG ratios are more attractive than stocks with higher PEG ratios, all else being equal. Some consider that a PEG ratio less than 1 is an indicator of an attractive value level. The PEG ratio is useful but must be used with care for several reasons:

- The PEG ratio assumes a linear relationship between P/E and growth. The model for P/E in terms of the DDM shows that, in theory, the relationship is not linear.

- The PEG ratio does not factor in differences in risk, an important determinant of P/E.
- The PEG ratio does not account for differences in the duration of growth. For example, dividing P/Es by short-term (five-year) growth forecasts may not capture differences in long-term growth prospects.

The way in which fundamentals can add insight to comparables is illustrated in Example 11.

EXAMPLE 11

A Peer-Group Comparison Modified by Fundamentals

Continuing with the valuation of telecommunication service providers, you gather information on selected fundamentals related to risk (beta), profitability (five-year earnings growth forecast), and valuation (trailing and forward P/Es). Analysts may also use other measures of risk in comparables work. These data are reported in Exhibit 6, which lists companies in order of descending earnings growth forecast. The use of forward P/Es recognizes that differences in trailing P/Es could be the result of transitory effects on earnings.

Exhibit 6 Valuation Data for Telecommunication Services Companies (as of 11 September 2013)

Company	Trailing P/E	Forward P/E	Five-Year EPS Growth Forecast	Forward PEG Ratio	Beta
AT&T	13.20	9.36	1.83%	7.20	0.56
Comcast Corporation	16.23	12.92	11.20	1.45	1.09
CenturyLink	NMF	8.89	8.52	1.04	0.81
China Telecom	13.14	10.31	6.90	1.90	0.81
Charter Communications	70.67	30.32	45.30	1.56	1.24
Verizon	15.03	11.99	2.51	5.99	0.50
Windstream Holdings	19.01	16.29	3.19	5.96	0.45
Mean	24.55	14.30	11.30	3.59	0.78
Median	15.03	11.99	6.90	1.90	0.78

Notes: NMF = not meaningful. The trailing P/E for CenturyLink is a negative number, which would result in a P/E that is not meaningful.
Source: www.finviz.com.

Based on the data in Exhibit 6, answer the following questions:

1 In Example 10, Part 3, AT&T, China Telecom, and CenturyLink were identified as possibly relatively undervalued compared with the peer group as a whole, and Verizon was identified as relatively fairly valued. What does the additional information relating to profitability and risk suggest about the relative valuation of the stocks in Exhibit 6?

2 AT&T has a consensus year-ahead EPS forecast of $3.63. Suppose the median P/E of 11.99 for the peer group is subjectively adjusted upward to 13.00 to reflect AT&T's superior profitability and below-average risk. Estimate AT&T's intrinsic value.

3 AT&T's current market price is $33.98. State whether AT&T appears to be fairly valued, overvalued, or undervalued when compared with the intrinsic value estimated in the answer to Part 2.

Solution to 1:

Among the three companies identified as underpriced (based on their low trailing P/Es), CenturyLink has the highest five-year EPS growth forecast and the lowest PEG ratio. AT&T and China Telecom have lower growth rates and higher PEG ratios than CenturyLink. Among the other companies in Exhibit 6, Comcast and Charter Communications had the highest EPS growth forecasts and the second and third lowest PEG ratios. The three stocks with the lowest trailing P/Es (AT&T, CenturyLink, and China Telecom) also had the lowest forward P/Es.

The two stocks with the highest growth forecasts, Comcast and Charter Communications, also had the highest betas, which is consistent with studies that have shown that growth stocks tend to have higher beta values than those of value stocks. Based on the high trailing and forward P/Es, it appears that investors in Charter Communications have high expectations concerning the company's future earnings potential. However, the high beta value is likely reflective of the uncertainty surrounding the earnings forecast and the possibility that actual future earnings may be less than expected.

Some analysts consider a PEG ratio below 1 to be a signal of undervaluation. However, one limitation of the PEG ratio is that it does not account for the overall growth rate of an industry or the economy as a whole. Hence, it is typically a good idea for an investor to compare a stock's PEG ratio to an average or median PEG ratio for the industry, as well as the entire market, to get an accurate sense of how fairly valued a stock is. The PEG ratio of CenturyLink is not below 1, but it is significantly lower than the PEG ratios for the other telecommunication companies—further indicating that this company is relatively undervalued.

Solution to 2:

$3.63 × 13.0 = $47.19 is an estimate of intrinsic value.

Solution to 3:

Because the estimated intrinsic value of $47.19 is greater than the current market price of $33.98, AT&T appears to be undervalued by the market on an absolute basis.

In Problem 2 of the Example 11, a peer median P/E of 11.99 was subjectively adjusted upward to 13.00. Depending on the context, the justification for using the specific value of 13.00 as the relevant benchmark rather than some other value could be raised. To avoid that issue, one way to express the analysis and results would be as follows: Given its modest growth and lower risk, AT&T should trade at a premium to the median P/E (11.99) of its peer group. Of course, this is a bullish outlook for AT&T because its forward P/E is only 9.36.

Analysts frequently compare a stock's multiple with the median or mean value of the multiple for larger sets of assets than a company's peer group. The next sections examine comparisons with these larger groups.

4.2 Industry and Sector Multiples

Median or mean P/Es for industries and for economic sectors are frequently used in relative valuations. Although median P/Es have the advantage that they are insensitive to outliers, some databases report only mean values of multiples for industries.

The mechanics of using industry multiples are identical to those used for peer-group comparisons. Taking account of relevant fundamental information, we compare a stock's multiple with the median or mean multiple for the company's industry.

Using industry and sector data can help an analyst explore whether the peer-group comparison assets are themselves appropriately priced. Comparisons with broader segments of the economy can potentially provide insight about whether the relative valuation based on comparables accurately reflects intrinsic value. For example, Value Line reports a relative P/E that is calculated as the stock's current P/E divided by the median P/E of all issues under Value Line review. The less closely matched the stock is to the comparison assets, the more dissimilarities are likely to be present to complicate the analyst's interpretation of the data. Arguably, however, the larger the number of comparison assets, the more likely that mispricings of individual assets cancel out. In some cases, we may be able to draw inferences about an industry or sector overall. For example, during the 1998–2000 internet bubble, comparisons of an individual internet stock's value with the overall market would have been more likely to point to overvaluation than comparisons of relative valuation only among internet stocks.

4.3 Overall Market Multiple

Although the logic of the comparables approach suggests the use of industry and peer companies as comparison assets, equity market indexes also have been used as comparison assets. The mechanics of using the method of comparables do not change in such an approach, although the user should be cognizant of any size differences between the subject stock and the stocks in the selected index.

The question of whether the overall market is fairly priced has captured analyst interest throughout the entire history of investing. We mentioned one approach to market valuation (using a DDM) in an earlier reading.

Example 12 shows a valuation comparison to the broad equity market on the basis of P/E.

EXAMPLE 12

Valuation Relative to the Market

You are analyzing three large-cap US stock issues with approximately equal earnings growth prospects and risk. As one step in your analysis, you have decided to check valuations relative to the S&P 500 Index. Exhibit 7 provides the data.

Exhibit 7 Comparison with an Index Multiple (Prices and EPS in US Dollars; as of 28 June 2019)

Measure	Stock A	Stock B	Stock C	S&P 500
Current price	23	50	80	2,941.76
P/E	15.2	30.0	15.2	21.8
Five-year average P/E (as a % of S&P 500 P/E)	80	120	105	

Source: www.us.spindices.com for S&P 500 data.

Based only on the data in Exhibit 7, address the following:

1 Explain which stock appears relatively undervalued when compared with the S&P 500.

2 State the assumption underlying the use of five-year average P/E comparisons.

Solution to 1:

Stock C appears to be undervalued when compared to the S&P 500. Stock A and Stock C are both trading at a P/E of 15.2 relative to trailing earnings, versus a P/E of 21.8 for the S&P 500. But the last row of Exhibit 7 indicates that Stock A has historically traded at a P/E reflecting a 20% discount to the S&P 500 (which, based on the current level of the S&P 500, would imply a P/E of 0.8 × 21.8 = 17.4). In contrast, Stock C has usually traded at a premium to the S&P 500 P/E but now trades at a discount to it. Stock B is trading at a high P/E, even higher than its historical relationship to the S&P 500's P/E (1.2 × 21.8 = 16.1).

Solution to 2:

Using historical relative-value information in investment decisions relies on an assumption of stable underlying economic relationships (that is, that the past is relevant for the future).

Because many equity indexes are market-capitalization weighted, financial databases often report the average market P/E with the individual P/Es weighted by the company's market capitalization. As a consequence, the largest constituent stocks heavily influence the calculated P/E. If P/Es differ systematically by market capitalization, however, differences in a company's P/E multiple from the index's multiple may be explained by that effect. Therefore, particularly for stocks in the middle-cap range, the analyst should favor using the median P/E for the index as the benchmark value of the multiple.

As with other comparison assets, the analyst may be interested in whether the equity index itself is efficiently priced. A common comparison is the index's P/E in relation to historical values. Siegel (2014) noted that recent P/Es were more than twice as high as the average P/E for US stocks over a long time period. Potential justifications for a higher-than-average P/E include lower-than-average interest rates and/or higher-than-average expected growth rates. An alternative hypothesis in a situation (historical high P/Es) is that the market as a whole is overvalued or, alternatively, that earnings are abnormally low.

The time frame for comparing average multiples is important. For example, at the end of the fourth quarter of 2008, the P/E for the S&P 500 was 60.70. That value is much higher than the 15.8 historical average since 1935. From 2006 through 2018, the highest quarterly P/E was 122.4 (30 June 2009) and the lowest was 13.0 (30 September 2011), and the quarterly P/E ranged between 18.9 and 24.1 over the five years ending in 2018. The use of past data relies on the key assumption that the past (sometimes the distant past) is relevant for the future.

We end this section with an introduction to valuation of the equity market itself on the basis of P/E. A well-known comparison is the earnings yield (the E/P) on a group of stocks and the interest yield on a bond. The so-called Fed model, based on a paper written by three analysts at the US Federal Reserve, predicts the return on the S&P 500 on the basis of the relationship between forecasted earnings yields and yields on bonds (Lander, Orphanides, and Douvogiannis 1997). Example 13 illustrates the Fed model.

EXAMPLE 13

The Fed Model

One of the main drivers of P/E for the market as a whole is the level of interest rates. The inverse relationship between value and interest rates can be seen from the expression of P/E in terms of fundamentals, because the risk-free rate is one component of the required rate of return that is inversely related to value. The Fed model relates the earnings yield on the S&P 500 to the yield to maturity on 10-year US Treasury bonds. As we have defined it, the earnings yield (E/P) is the inverse of the P/E; the Fed model uses expected earnings for the next 12 months in calculating the ratio.

Based on the premise that the two yields should be closely linked, on average, the trading rule based on the Fed model considers the stock market to be overvalued when the market's current earnings yield is less than the 10-year Treasury bond (T-bond) yield. The intuition is that when risk-free T-bonds offer a yield that is higher than that of stocks—which are a riskier investment—stocks are an unattractive investment.

According to the model, the justified or fair value P/E for the S&P 500 is the reciprocal of the 10-year T-bond yield. As of 28 December 2018, according to the model, with a 10-year T-bond yielding 2.72%, the justified P/E on the S&P 500 was 1/0.0272 = 36.8. The trailing P/E based for 31 December 2018 was 18.9.

We previously presented an expression for the justified P/E in terms of the Gordon growth model. That expression indicates that the expected growth rate in dividends or earnings is a variable that enters into the intrinsic value of a stock (or an index of stocks). A concern in considering the Fed model is that this variable is lacking in the model. Please note that the earnings yield is, in fact, the expected rate of return on a no-growth stock (under the assumption that price equals value). With the PVGO (present value of growth opportunities) and setting price equal to value, we obtain $P_0 = E_1/r + \text{PVGO}$. Setting the present value of growth opportunities equal to zero and rearranging, we obtain $r = E_1/P_0$. Example 14 presents a valuation model for the equity market that incorporates the expected growth rate in earnings.

EXAMPLE 14

The Yardeni Model

Yardeni (2000) developed a model that incorporates the expected growth rate in earnings—a variable that is missing in the Fed model. This model is presented as one example of more-complex models than the Fed model. Yardeni's model is

$$\text{CEY} = \text{CBY} - b \times \text{LTEG} + \text{Residual},$$

where CEY is the current earnings yield on the market index, CBY is the current Moody's Investors Service A rated corporate bond yield, and LTEG is the consensus five-year earnings growth rate forecast for the market index. The coefficient b measures the weight the market gives to five-year earnings projections. (Recall that the expression for P/E in terms of the Gordon growth model is based on the long-term sustainable growth rate and that five-year forecasts of growth may not be sustainable.) Although CBY incorporates a default risk premium relative to T-bonds, it does not incorporate an equity risk premium per se. For example, in the bond yield plus risk premium model for the cost of equity, an analyst typically adds 300–400 basis points to a corporate bond yield.

Yardeni found that, prior to publication of the model in 2000, the coefficient b had averaged 0.10. In recent years, he has reported valuations based on growth weights of 0.10, 0.20, and 0.25. Noting that CEY is E/P and taking the inverse of both sides of this equation, Yardeni obtained the following expression for the justified P/E on the market:

$$\frac{P}{E} = \frac{1}{\text{CBY} - b \times \text{LTEG}}.$$

Consistent with valuation theory, in Yardeni's model, higher current corporate bond yields imply a lower justified P/E and higher expected long-term growth results in a higher justified P/E.

Critics of the Fed model point out that it ignores the equity risk premium (Stimes and Wilcox 2011). The model also inadequately reflects the effects of inflation and incorrectly incorporates the differential effects of inflation on earnings and interest payments (e.g., Siegel 2014). Some empirical evidence has shown that prediction of future returns based on simple P/E outperforms prediction based on the Fed model's differential with bond yields (for the US market, see Arnott and Asness 2003; for nine other markets, see Aubert and Giot 2007).

Another drawback to the Fed model is that the relationship between interest rates and earnings yields is not a linear one. This drawback is most noticeable at low interest rates; Example 13 provided an example of this limitation of the model. Furthermore, small changes in interest rates and/or corporate profits can significantly alter the justified P/E predicted by the model. Overall, an analyst should look to the Fed model only as one tool for calibrating the overall value of the stock market and should avoid overreliance on the model as a predictive method, particularly in periods of low inflation and low interest rates.

4.4 Own Historical P/E

As an alternative to comparing a stock's valuation with that of other stocks, one traditional approach uses past values of the stock's own P/E as a basis for comparison. Underlying this approach is the idea that a stock's P/E may regress to historical average levels.

An analyst can obtain a benchmark value in a variety of ways with this approach. Value Line reports as a "P/E median" a rounded average of four middle values of a stock's average annual P/E for the previous 10 years. The five-year average trailing P/E is another reasonable metric. In general, trailing P/Es are more commonly used than forward P/Es in such computations. In addition to "higher" and "lower" comparisons with this benchmark, justified price based on this approach may be calculated as follows:

$$\text{Justified price} = (\text{Benchmark value of own historical P/Es}) \times (\text{Most recent EPS}). \quad (3)$$

Normalized EPS replaces most recent EPS in this equation when EPS is negative and whenever otherwise appropriate.

Example 15 illustrates the use of past values of the stock's own P/E as a basis for reaching a valuation conclusion.

EXAMPLE 15

Valuation Relative to Own Historical P/Es

As of June 2019, you are valuing Honda Motor Company, among the market leaders in Japan's auto manufacturing industry. You are applying the method of comparables using Honda's five-year average P/E as the benchmark value of the multiple. Exhibit 8 presents the data.

Exhibit 8 Historical P/Es for Honda Motor Company

2018	2017	2016	2015	2014	Mean	Median
6.9	10.0	10.9	10.8	9.7	9.7	10.0

Sources: The Value Line Investment Survey for average annual P/Es; calculations for mean and median P/Es.

1 State a benchmark value for Honda's P/E.

2 Given forecasted EPS for fiscal year 2019 (ended 31 December) of ¥381.93, calculate and interpret a justified price for Honda.

3 Compare the justified price with the stock's recent price of ¥2,837.

Solution to 1:

From Exhibit 8, the benchmark value based on the median P/E value is 10.0 and based on the mean P/E value is 9.7.

Solution to 2:

The calculation is 10.0 × ¥381.93 = ¥3,819 when the median-based benchmark P/E is used and 9.7 × ¥381.93 = ¥3,704 when the mean-based benchmark P/E is used.

Solution to 3:

The stock's recent price is 26.2% (calculated as 2,817/3,819 − 1) less than the justified price of the stock based on median historical P/E but 23.9% (calculated as 2,817/3,704 − 1) less than the justified price of the stock based on mean historical P/E. The stock may be undervalued, and misvaluation, if present, appears significant.

In using historical P/Es for comparisons, analysts should be alert to the impact on P/E levels of changes in a company's business mix and leverage over time. If the company's business has changed substantially within the time period being examined, the method based on a company's own past P/Es is prone to error. Shifts in the use of financial leverage may also impair comparability based on average own past P/E.

Changes in the interest rate environment and economic fundamentals over different time periods can be another limitation to using an average past value of P/E for a stock as a benchmark. A specific caution is that inflation can distort the economic meaning of reported earnings. Consequently, if the inflationary environments reflected in current P/E and average own past P/E are different, a comparison between the two P/Es may be misleading. Changes in a company's ability to pass through cost inflation to higher prices over time may also affect the reliability of such comparisons, as illustrated in Example 16 in the next section.

4.5 P/Es in Cross-Country Comparisons

When comparing the P/Es of companies in different countries, the analyst should be aware of the following effects that may influence the comparison:

- The effect on EPS of differences in accounting standards: Comparisons (without analyst adjustments) among companies preparing financial statements based on different accounting standards may be distorted. Such distortions may occur when, for example, the accounting standards differ as to permissible recognition of revenues, expenses, or gains.
- The effect on market-wide benchmarks of differences in their macroeconomic contexts: Differences in macroeconomic contexts may distort comparisons of benchmark P/E levels among companies operating in different markets.

A specific case of the second bullet point is differences in inflation rates and in the ability of companies to pass through inflation in their costs in the form of higher prices to their customers. For two companies with the same pass-through ability, the company operating in the environment with higher inflation will have a lower justified P/E; if the inflation rates are equal but pass-through rates differ, the justified P/E should be lower for the company with the lower pass-through rate. Example 16 provides analysis in support of these conclusions.

EXAMPLE 16

An Analysis of P/Es and Inflation

Assume a company with no real earnings growth, such that its earnings growth can result only from inflation, will pay out all its earnings as dividends. Based on the Gordon (constant growth) DDM, the value of a share is

$$P_0 = \frac{E_0(1 + I)}{r - I},$$

where

P_0 = current price, which is substituted for the intrinsic value, V_0, for purposes of analyzing a justified P/E

E_0 = current EPS, which is substituted for current dividends per share, D_0, because the assumption in this example is that all earnings are paid out as dividends

I = rate of inflation, which is substituted for expected growth, g, because of the assumption in this example that the company's only growth is from inflation

r = required return

Suppose the company has the ability to pass on some or all inflation to its customers, and let λ represent the percentage of inflation in costs that the company can pass through to earnings. The company's earnings growth may then be expressed as λI, and the equation becomes

$$P_0 = \frac{E_0(1+\lambda I)}{r-\lambda I} = \frac{E_1}{r-\lambda I}.$$

Now, introduce a real rate of return, defined here as r minus I and represented as ρ. The value of a share and the justified forward P/E can now be expressed, respectively, as follows:

$$P_0 = \frac{E_1}{\rho + (1-\lambda)I},$$

and

$$\frac{P_0}{E_1} = \frac{1}{\rho + (1-\lambda)I}.$$

(Note that the denominator of this equation is derived from the previous equation as follows: $r - \lambda I = r - I + I - I\lambda = (r - I) + (1 - \lambda)I = \rho + (1 - \lambda)I$.

If a company can pass through all inflation, such that λ = 1 (100%), then the P/E is equal to 1/ρ. But if the company can pass through no inflation, such that λ = 0, then the P/E is equal to $1/(\rho + I)$—that is, $1/r$.

You are analyzing two companies, Company M and Company P. The real rate of return required on the shares of Company M and Company P is 3% per year. Using the analytic framework provided, address the following:

1 Suppose both Company M and Company P can pass through 75% of cost increases. Cost inflation is 6% for Company M but only 2% for Company P.

A Estimate the justified P/E for each company.

B Interpret your answer to Part A.

2 Suppose both Company M and Company P face 6% a year inflation. Company M can pass through 90% of cost increases, but Company P can pass through only 70%.

A Estimate the justified P/E for each company.

B Interpret your answer to Part A.

Solution to 1:

A For Company M, $\dfrac{1}{0.03 + (1 - 0.75)0.06} = 22.2.$

For Company P, $\dfrac{1}{0.03 + (1 - 0.75)0.02} = 28.6.$

B With less than 100% cost pass-through, the justified P/E is inversely related to the inflation rate.

Solution to 2:

A For Company M, $\dfrac{1}{0.03 + (1 - 0.90)0.06} = 27.8.$

For Company P, $\dfrac{1}{0.03 + (1 - 0.70)0.06} = 20.8.$

B For equal inflation rates, the company with the higher pass-through rate has a higher justified P/E.

Note that this example follows the analysis of Solnik and McLeavey (2004, pp. 289–290).

Example 16 illustrates that with less than 100% cost pass-through, the justified P/E is inversely related to the inflation rate (with complete cost pass-through, the justified P/E should not be affected by inflation). The higher the inflation rate, the greater the impact of incomplete cost pass-through on P/E. From Example 16, one can also infer that the higher the inflation rate, the more serious the effect on justified P/E of a pass-through rate that is less than 100%.

4.6 Using P/Es to Obtain Terminal Value in Multistage Dividend Discount Models

In using a DDM to value a stock, whether applying a multistage model or modeling within a spreadsheet (forecasting specific cash flows individually up to some horizon), estimation of the terminal value of the stock is important. The key condition that must be satisfied is that terminal value reflects earnings growth that the company can sustain in the long run. Analysts frequently use price multiples—in particular, P/Es and P/Bs—to estimate terminal value. We can call such multiples **terminal price multiples**. Choices for the terminal multiple, with a terminal P/E multiple used as the example, include the following two types:

Terminal price multiple based on fundamentals: As illustrated earlier, analysts can restate the Gordon growth model as a multiple by, for example, dividing both sides of the model by EPS. For terminal P/E multiples, dividing both sides of the Gordon growth model by EPS at time n, where n is the point in time at which the final stage begins (i.e., E_n), gives a trailing terminal price multiple; dividing both sides by EPS at time $n + 1$ (i.e., E_{n+1}) gives a leading terminal price multiple. Of course, an analyst can use the Gordon growth model to estimate terminal value and need not go through the process of deriving a terminal price multiple and then multiplying by the same value of the fundamental to estimate terminal value. Because of their familiarity, however, multiples may be useful in communicating an estimate of terminal value.

Terminal price multiple based on comparables: Analysts have used various choices for the benchmark value, including:

- median industry P/E,
- average industry P/E, and
- average of own past P/Es.

Having selected a terminal multiple, the expression for terminal value when using a terminal P/E multiple is

V_n = Benchmark value of trailing terminal P/E × E_n

or

V_n = Benchmark value of forward terminal P/E × E_{n+1},

where V_n = Terminal value at time n.

The use of a comparables approach has the strength that it is entirely grounded in market data. In contrast, the Gordon growth model calls for specific estimates (the required rate of return, the dividend payout ratio, and the expected mature growth rate), and the model's output is very sensitive to changes in those estimates. A possible disadvantage to the comparables approach is that when the benchmark value reflects mispricing (over- or undervaluation), so will the estimate of terminal value. Example 17 illustrates the use of P/Es and the Gordon growth model to estimate terminal value.

EXAMPLE 17

Using P/Es and the Gordon Growth Model to Value the Mature Growth Phase

As an energy analyst, you are valuing the stock of an oil exploration company. You have projected earnings and dividends three years out (to t = 3), and you have gathered the following data and estimates:

- Required rate of return = 0.10.
- Average dividend payout rate for mature companies in the market = 0.45.
- Industry average ROE = 0.13.
- E_3 = \$3.00.
- Industry average P/E = 14.3.

On the basis of this information, carry out the following:

1. Calculate terminal value based on comparables, using your estimated industry average P/E as the benchmark.
2. Contrast your answer in Part 1 to an estimate of terminal value using the Gordon growth model.

Solution to 1:

V_n = Benchmark value of P/E × E_n = 14.3 × \$3.00 = \$42.90.

Solution to 2:

Recall that the Gordon growth model expresses intrinsic value, V, as the present value of dividends divided by the required rate of return, r, minus the growth rate, g: $V_0 = D_0(1 + g)/(r - g)$. Here we are estimating terminal value, so the relevant expression is $V_n = D_n(1 + g)/(r - g)$. You would estimate that the dividend at t = 3 will equal earnings in Year 3 of \$3.00 times the average payout ratio of 0.45, or D_n = \$3.00 × 0.45 = \$1.35. Recall also the sustainable growth rate expression—that is, $g = b \times$ ROE, where b is the retention rate and equivalent to 1 minus the dividend payout ratio. In this example, b = (1 – 0.45) = 0.55, and you can use ROE = 0.13 (the industry average). Therefore, $g = b \times$ ROE = 0.55 × 0.13 = 0.0715. Given the required rate of return of 0.10, you obtain the estimate

V_n = (\$1.35)(1 + 0.0715)/(0.10 – 0.0715) = \$50.76. In this example, therefore, the Gordon growth model estimate of terminal value is 18.3% higher than the estimate based on comparables calculated in Part 1 (i.e., 0.1832 = \$50.76/\$42.90 – 1).

PRICE TO BOOK VALUE

5

- **b** calculate and interpret a justified price multiple;
- **c** describe rationales for and possible drawbacks to using alternative price multiples and dividend yield in valuation;
- **d** calculate and interpret alternative price multiples and dividend yield;
- **g** describe fundamental factors that influence alternative price multiples and dividend yield;
- **h** calculate and interpret the justified price-to-earnings ratio (P/E), price-to-book ratio (P/B), and price-to-sales ratio (P/S) for a stock, based on forecasted fundamentals;
- **j** evaluate a stock by the method of comparables and explain the importance of fundamentals in using the method of comparables;
- **r** evaluate whether a stock is overvalued, fairly valued, or undervalued based on comparisons of multiples.

The ratio of market price per share to book value per share (P/B), like P/E, has a long history of use in valuation practice. According to the 2012 BofA Merrill Lynch Institutional Factor Survey, 53% of respondents considered P/B when making investment decisions.

In the P/E multiple, the measure of value (EPS) in the denominator is a flow variable relating to the income statement. In contrast, the measure of value in the P/B's denominator (book value per share) is a stock or level variable coming from the balance sheet. (*Book* refers to the fact that the measurement of value comes from accounting records or books, in contrast to market value.) Intuitively, therefore, we note that book value per share attempts to represent, on a per-share basis, the investment that common shareholders have made in the company. To define book value per share more precisely, we first find **shareholders' equity** (total assets minus total liabilities). Because our purpose is to value common stock (as opposed to valuing the company as a whole), we subtract from shareholders' equity any value attributable to preferred stock to obtain common shareholders' equity, or the **book value of equity** (often called simply book value). Dividing book value by the number of common stock shares outstanding, we obtain **book value per share**, the denominator in P/B.

In the remainder of this section, we present the reasons analysts have offered for using P/B and possible drawbacks to its use. We then illustrate the calculation of P/B and discuss the fundamental factors that drive P/B. We end the section by showing the use of P/B based on the method of comparables.

Analysts have offered several rationales for the use of P/B; some specifically compare P/B with P/E:

- Because book value is a cumulative balance sheet amount, book value is generally positive even when EPS is zero or negative. An analyst can generally use P/B when EPS is zero or negative, whereas P/E based on a zero or negative EPS is not meaningful.
- Because book value per share is more stable than EPS, P/B may be more meaningful than P/E when EPS is abnormally high or low or is highly variable.

- As a measure of net asset value per share, book value per share has been viewed as appropriate for valuing companies composed chiefly of liquid assets, such as finance, investment, insurance, and banking institutions (Wild, Bernstein, and Subramanyam 2001, p. 233). For such companies, book values of assets may approximate market values. When information on individual corporate assets is available, analysts may adjust reported book values to market values where they differ.
- Book value has also been used in the valuation of companies that are not expected to continue as a going concern (Martin 1998, p. 22).
- Differences in P/Bs may be related to differences in long-run average returns, according to empirical research (Bodie, Kane, and Marcus 2008).

Possible drawbacks of P/Bs in practice include the following:

- Assets in addition to those recognized in financial statements may be critical operating factors. For example, in many service companies, **human capital**—the value of skills and knowledge possessed by the workforce—is more important than physical capital as an operating factor, but it is not reflected as an asset on the balance sheet. Similarly, the good reputation that a company develops by consistently providing high-quality goods and services is not reflected as an asset on the balance sheet.
- P/B may be misleading as a valuation indicator when the levels of assets used by the companies under examination differ significantly. Such differences may reflect differences in business models.
- Accounting effects on book value may compromise how useful book value is as a measure of the shareholders' investment in the company. In general, intangible assets that are generated internally (as opposed to being acquired) are not shown as assets on a company's balance sheet. For example, companies account for advertising and marketing as expenses, so the value of internally generated brands, which are created and maintained by advertising and marketing activities, do not appear as assets on a company's balance sheet under IFRS or US GAAP. Similarly, when accounting standards require that research and development (R&D) expenditures be treated as expenses, the value of internally developed patents does not appear as assets. Certain R&D expenditures can be capitalized, although rules vary among accounting standards. Accounting effects such as these may impair the comparability of P/B among companies and countries unless appropriate analyst adjustments are made.
- Book value reflects the reported value of assets and liabilities. Some assets and liabilities, such as some financial instruments, may be reported at fair value as of the balance sheet date; other assets, such as property, plant, and equipment, are generally reported at historical cost, net of accumulated depreciation, amortization, depletion, and/or impairment. It is important to examine the notes to the financial statements to identify how assets and liabilities are measured and reported. For assets measured at net historical cost, inflation and technological change can eventually result in significant divergence between the book value and the market value of assets. As a result, book value per share often does not accurately reflect the value of shareholders' investments. When comparing companies, significant differences in the average age of assets may lessen the comparability of P/Bs.
- Share repurchases or issuances may distort historical comparisons.

As an example of the effects of share repurchases, consider Colgate-Palmolive Company. As of 13 September 2013, Colgate-Palmolive's trailing P/E and P/B were, respectively, 24.84 and 36.01. Five years earlier, Colgate-Palmolive's trailing P/E and P/B were 23.55 and 15.94. In other words, the company's P/E widened by 5.5% (= 24.84/23.55 – 1) while its P/B widened by 125.9% (= 36.01/15.94 – 1). The majority of the difference in changes in these two multiples can be attributed to the substantial amount of shares that Colgate-Palmolive repurchased over those five years, as reflected by book value (i.e., total common equity) declining from $2.48 billion as of 30 June 2008 to $1.53 billion as of 30 June 2013. Because of those share repurchases, Colgate-Palmolive's book value declined at an annual rate of 9.2%. In summary, when a company repurchases shares at a price higher than the current book value per share, it lowers the overall book value per share for the company. All else being equal, the effect is to make the stock appear more expensive if the current P/B is compared to its historical values.

Example 18 illustrates another potential limitation to using P/B in valuation.

EXAMPLE 18

Differences in Business Models Reflected in Differences in P/Bs

The US banking industry has a wide range of P/Bs. Much of these differences in P/Bs can be attributed to differences in company-specific business models. Exhibit 9 presents P/Bs for three major US banks as of 31 December 2018.

Exhibit 9 P/Bs for Selected US Banks

Entity	P/B
Citigroup, Inc.	0.69
Wells Fargo & Company	1.21
US Bancorp	1.63

Source: S&P Capital IQ

Citigroup's low P/B versus its peers is a reflection of the "one-stop shopping" business model it and some other mega-banks pursued in the 1990s. Citigroup suffered huge losses during the global financial crisis and had to be rescued in November 2008 by the US government.

Wells Fargo derives most of its revenue from loans and service fees. Its business model focuses on cross-selling multiple products, and in 2012 it was responsible for originating close to a third of all US home loans. Wells Fargo is also predominantly a domestic business, whereas other large banks are much more exposed to overseas markets.

US Bancorp's relatively risk-averse business model is focused on consumer and business banking as well as trusts and payment processing. Compared with other mega-banks, US Bancorp has a much smaller presence in investment banking and capital markets. Another reason for the bank's relatively high P/B was its acquisition activity, which has helped it grow its business considerably.

5.1 Determining Book Value

In this section, we illustrate how to calculate book value and how to adjust book value to improve the comparability of P/Bs among companies. To compute book value per share, we need to refer to the business's balance sheet, which has a shareholders' (or stockholders') equity section. The computation of book value is as follows:

- (Shareholders' equity) − (Total value of equity claims that are senior to common stock) = Common shareholders' equity.
- (Common shareholders' equity)/(Number of common stock shares outstanding) = Book value per share.

Possible claims senior to the claims of common stock, which would be subtracted from shareholders' equity, include the value of preferred stock and the dividends in arrears on preferred stock. Example 19 illustrates the calculation.

EXAMPLE 19

Computing Book Value per Share

Headquartered in Toronto, Canada, the Toronto-Dominion Bank and its subsidiaries are collectively known as TD Bank Group (TD). With operations organized into four segments (Canadian Personal and Commercial Banking, US Personal and Commercial Banking, Wholesale Banking, and Wealth and Insurance), in 2018 TD provided financial products and services to approximately 26 million customers. Exhibit 10 presents data from the equity section of TD's consolidated balance sheets for the years 2016–2018. TD's fiscal years end on 31 October.

Exhibit 10 Equity Data for TD Bank Group (Millions of Canadian Dollars)

	31 October 2018	31 October 2017	31 October 2016
Equity			
Common shares	CAD21,221	CAD20,931	CAD20,711
Millions of shares issued and outstanding:			
2018: 1,830.4			
2017: 1,842.5			
2016: 1,857.6			
Preferred shares	5,000	4,750	4,400
Millions of shares issued and outstanding:			
2018: 200.0			
2017: 190.0			
2016: 176.0			
Treasury shares—common	(151)	(183)	(36)
Millions of shares held:			
2018: 2.1			
2017: 2.9			
2016: 0.4			
Treasury shares—preferred	(1)	—	(1)

Exhibit 10 (Continued)

	31 October 2018	31 October 2017	31 October 2016
2018: nil			
2017: nil			
2016: nil			
Contributed surplus	193	214	203
Retained earnings	46,145	40,489	35,452
Accumulated and other comprehensive income	6,639	8,006	11,834
	79,047	74,207	72,564
Non-controlling interests in subsidiaries	993	983	1,650
Total equity	**CAD80,040**	**CAD75,190**	**CAD74,214**

Source: TD Bank Group 2018 annual report.

1 Using the data in Exhibit 10, calculate book value per share for 2016, 2017, and 2018.

2 Given a closing price of CAD73.03 on 31 October 2018, calculate TD's 2018 P/B.

Solution to 1:

Because preferred shareholders have a claim on income and assets that is senior to that of the common shareholders, total equity must be adjusted by the value of outstanding and repurchased preferred shares. The divisor is the number of common shares outstanding.

2018: Book value per share = (80,040 – 5,000)/1,830.4 = CAD41.00.

2017: Book value per share = (75,190 – 4,750)/1,842.5 = CAD38.23.

2016: Book value per share = (74,214 – 4,400)/1,857.6 = CAD37.58.

Solution to 2:

P/B = CAD73.03/CAD41.00 = 1.78.

Example 19 illustrated the calculation of book value per share without any adjustments. Adjusting P/B has two purposes: (1) to make the book value per share more accurately reflect the value of shareholders' investment and (2) to make P/B more useful for making comparisons among different stocks. Some adjustments are as follows:

- Some services and analysts report a **tangible book value per share**. Computing tangible book value per share involves subtracting reported intangible assets on the balance sheet from common shareholders' equity. The analyst should be familiar with the calculation. From the viewpoint of financial theory, however, the general exclusion of all intangibles may not be warranted. In the case of individual intangible assets, such as patents, which can be separated from the entity and sold, exclusion may not be justified. Exclusion may be appropriate, however, for goodwill from acquisitions, particularly for comparative purposes. **Goodwill** represents the excess of the purchase price of an acquisition beyond the fair value of acquired tangible assets and specifically identifiable intangible assets. Many analysts believe that goodwill does not represent an asset because it is not separable and may reflect overpayment for an acquisition.

- Certain adjustments may be appropriate for enhancing comparability. For example, one company may use FIFO whereas a peer company uses LIFO, which in an inflationary environment will generally understate inventory values. To accurately assess the relative valuation of the two companies, the analyst should restate the book value of the company using LIFO to what it would be based on FIFO. For a more complete discussion of adjustments to balance sheet amounts, refer to readings on financial statement analysis.
- For book value per share to most accurately reflect current values, the balance sheet should be adjusted for significant off-balance-sheet assets and liabilities. An example of an off-balance-sheet liability is a guarantee to pay a debt of another company in the event of that company's default. US accounting standards require companies to disclose off-balance-sheet liabilities.

Example 20 illustrates adjustments an analyst might make to a financial firm's P/B to obtain an accurate firm value.

EXAMPLE 20

Adjusting Book Value (Historical Example)

Edward Stavos is a junior analyst at a major US pension fund. Stavos is researching Barclays PLC for his fund's Credit Services Portfolio and is preparing background information prior to an upcoming meeting with the company. Headquartered in London, United Kingdom, Barclays is a major global financial services provider engaged in personal banking, credit cards, corporate and investment banking, and wealth and investment management with an extensive international presence in Europe, the Americas, Africa, and Asia.

Stavos is particularly interested in Barclays' P/B and how adjusting asset and liability accounts to their current fair value impacts the ratio. He gathers the condensed 2012 balance sheet (as of 31 December) and footnote data from Barclay's website as shown in Exhibit 11.

Exhibit 11 Barclays PLC 2012 Condensed Consolidated Balance Sheet and Footnote Data (£ in Millions)

	2012
Assets	
Cash and balances at central banks	£86,175
Items in the course of collection from other banks	1,456
Trading portfolio assets	145,030
Financial assets designated at fair value	46,061
Derivative financial instruments	469,146
Available for sale investments	75,109
Loans and advances to banks	40,489
Loans and advances to customers	425,729
Reverse repurchase agreements and other similar secured lending	176,956
Prepayments, accrued income, and other assets	4,360

Exhibit 11 (Continued)

	2012
Investments in associates and joint ventures	570
Property, plant, and equipment	5,754
Goodwill and intangible assets	7,915
Current tax assets	252
Deferred tax assets	3,016
Retirement benefit assets	2,303
Total assets	£1,490,321
Liabilities	
Deposits from banks	77,010
Items in the course of collection due to other banks	1,573
Customer accounts	385,707
Repurchase agreements and other similar secured borrowing	217,342
Trading portfolio liabilities	44,794
Financial liabilities designated at fair value	78,280
Derivative financial instruments	462,468
Debt securities in issue	119,581
Subordinated liabilities	24,018
Accruals, deferred income, and other liabilities	12,232
Provisions	2,766
Current tax liabilities	621
Deferred tax liabilities	719
Retirement benefit liabilities	253
Total liabilities	1,427,364
Shareholders' equity	
Shareholders' equity excluding non-controlling interests	53,586
Non-controlling interests	9,371
Total shareholders' equity	62,957
Total liabilities and shareholders' equity	£1,490,321

Excerpt from Footnotes to the Barclays Financial Statements: Financial Assets and Liabilities at Carrying Amount and Fair Value

	2012	
	Carrying amount	Fair value
Financial assets		
Loans and advances to banks	£40,489	£40,489

(continued)

(Continued)

	2012	
	Carrying amount	**Fair value**
Loans and advances to customers:		
—Home loans	174,988	164,608
—Credit cards, unsecured and other retail lending	66,414	65,357
—Corporate loans	184,327	178,492
Reverse repurchase agreements and other similar secured lending	176,956	176,895
	£643,174	£625,841
Financial liabilities		
Deposits from banks	77,010	77,023
Customer accounts:		
—Current and demand accounts	127,819	127,819
—Savings accounts	99,875	99,875
—Other time deposits	158,013	158,008
Debt securities in issue	119,581	119,725
Repurchase agreements and other similar secured borrowing	217,342	217,342
Subordinated liabilities	24,018	23,467
	£823,658	£823,259

Source: Barclays' 2012 annual report.

The 31 December 2012 share price for Barclays was £2.4239, and the diluted weighted average number of shares was 12,614 million. Stavos computes book value per share initially by dividing total shareholders' equity by the share count and arrives at a book value per share of £4.9910 (£62,957/12,614) and a P/B of 0.49 (£2.4239/£4.9910).

Stavos then computes tangible book value per share as £4.3636 (calculated as £62,957 minus £7,915 of goodwill and intangible assets, which is then divided by 12,614 shares). The P/B based on tangible book value per share is 0.56 (£2.4239/£4.3636).

Stavos then turns to the footnotes to examine the fair value data. He notes the fair value of financial assets is £17,333 million less than their carrying amount (£643,174 – £625,841) and the fair value of financial liabilities is £399 million less than their carrying amount (£823,658 – £823,259). Including these adjustments to tangible book value results in an adjusted book value per share of £3.0211 [(£62,957 – £7,915 – £17,333 + £399)/12,614]. Stavos' adjusted P/B is 0.80 (£2.4239/£3.0211).

Stavos is concerned about the wide range in his computed P/Bs. He knows that if quoted prices are not available for financial assets and liabilities, IAS 39 allows for the use of valuation models to estimate fair value. He decides to question management regarding the use of models to value assets, liabilities, and derivatives and the sensitivity of these accounts to changes in interest rates and currency values.

An analyst should also be aware of differences in accounting standards related to how assets and liabilities are valued in financial statements. Accounting standards currently require companies to report some assets and liabilities at fair value and others at historical cost (with some adjustments).

Financial assets, such as investments in marketable securities, are usually reported at fair value. Investments classified as "held to maturity" and reported on a historical cost basis are an exception. (Instead of the term "held-to-maturity," IFRS refer to this category of investments as financial assets measured at amortized cost.) Some financial liabilities also are reported at fair value.

Nonfinancial assets, such as land and equipment, are generally reported at their historical acquisition costs, and in the case of equipment, the assets are depreciated over their useful lives. The value of these assets may have increased over time, however, or the value may have decreased more than is reflected in the accumulated depreciation. When the reported amount of an asset—that is, its carrying value—exceeds its recoverable amount, both international accounting standards (IFRS) and US accounting standards (GAAP) require companies to reduce the reported amount of the asset and show the reduction as an impairment loss (the two sets of standards differ in the measurement of impairment losses). US GAAP, however, prohibit subsequent reversal of impairment losses, whereas IFRS permit subsequent reversals. In addition, as mentioned above, IFRS allow companies to measure fixed assets using either the historical cost model or a revaluation model, under which the assets are reported at their current value. When assets are reported at fair value, P/Bs become more comparable among companies; for this reason, P/Bs are considered to be more comparable for companies with significant amounts of financial assets.

5.2 Valuation Based on Forecasted Fundamentals

We can use forecasts of a company's fundamentals to estimate a stock's justified P/B. For example, assuming the Gordon growth model and using the expression $g = b \times \text{ROE}$ for the sustainable growth rate, the expression for the justified P/B based on the most recent book value (B_0) is

$$\frac{P_0}{B_0} = \frac{\text{ROE} - g}{r - g}. \tag{4}$$

For example, if a business's ROE is 12%, its required rate of return is 10%, and its expected growth rate is 7%, then its justified P/B based on fundamentals is (0.12 – 0.07)/(0.10 – 0.07) = 1.67.

Deriving the Justified P/B Expression

According to the Gordon growth model, $V_0 = E_1 \times (1 - b)/(r - g)$. Defining ROE as E_1/B_0 so that $E_1 = B_0 \times \text{ROE}$ and substituting for E_1 into the prior expression, we have $V_0 = B_0 \times \text{ROE} \times (1 - b)/(r - g)$, giving $V_0/B_0 = \text{ROE} \times (1 - b)/(r - g)$. The sustainable growth rate expression is $g = b \times \text{ROE}$. Substituting $b = g/\text{ROE}$ into the expression just given for V_0/B_0, we have $V_0/B_0 = (\text{ROE} - g)/(r - g)$. Because justified price is intrinsic value, V_0, we obtain Equation 4.

Equation 4 states that the justified P/B is an increasing function of ROE, all else equal. Because the numerator and denominator are differences of, respectively, ROE and r from the same quantity, g, what determines the justified P/B in Equation 4 is

ROE in relation to the required rate of return, r. The larger ROE is in relation to r, the higher is the justified P/B based on fundamentals. This relationship can be seen clearly if we set g equal to 0 (the no-growth case): P_0/B_0 = ROE/r.

A practical insight from Equation 4 is that we cannot conclude whether a particular value of the P/B reflects undervaluation without taking into account the business's profitability. Equation 4 also suggests that if we are evaluating two stocks with the same P/B, the one with the higher ROE is relatively undervalued, all else equal. These relationships have been confirmed through cross-sectional regression analyses (Harris and Marston 1994; Fairfield, 1994).

Further insight into P/B comes from the residual income model, which is discussed in detail in another reading. The expression for the justified P/B based on the residual income valuation is

$$\frac{P_0}{B_0} = 1 + \frac{\text{Present value of expected future residual earnings}}{B_0}. \tag{5}$$

Equation 5, which makes no special assumptions about growth, states the following:

- If the present value of expected future residual earnings is zero—for example, if the business just earns its required return on investment in every period—the justified P/B is 1.
- If the present value of expected future residual earnings is positive (negative), the justified P/B is greater than (less than) 1.

Justified P/B Expression Based on Residual Income

Noting that (ROE – r) × B_0 would define a level residual income stream, we can show that Equation 4 is consistent with Equation 5 (a general expression) as follows. In P_0/B_0 = (ROE – g)/(r – g), we can successively rewrite the numerator (ROE – g) + r – r = (r – g) + (ROE – r), so P_0/B_0 = [(r – g) + (ROE – r)]/(r – g) = 1 + (ROE – r)/(r – g), which can be written P_0/B_0 = 1 + [(ROE – r)/(r – g)] × B_0/B_0 = 1 + [(ROE – r) × B_0/(r – g)]/B_0; the second term in the final expression is the present value of residual income divided by B_0 as in Equation 5.

5.3 Valuation Based on Comparables

To use the method of comparables for valuing stocks using a P/B, we follow the steps given earlier. In contrast to EPS, however, analysts' forecasts of book value are not aggregated and widely disseminated by financial data vendors; in practice, most analysts use trailing book value in calculating P/Bs. Evaluation of relative P/Bs should consider differences in ROE, risk, and expected earnings growth. The use of P/Bs in the method of comparables is illustrated in Example 21.

EXAMPLE 21 (HISTORICAL EXAMPLE)

P/B Comparables Approach

You are working on a project to value an independent securities brokerage firm. You know the industry had a significant decline in valuations during the 2007–09 financial crisis. You decide to perform a time series analysis on three firms: E*TRADE Financial Corp. (ETFC), the Charles Schwab Corporation (SCHW), and TD Ameritrade Holding Corp. (AMTD). Exhibit 12 presents information on these firms.

Exhibit 12 Price-to-Book Comparables

	Price-to-Book Value Ratio								
Entity	2006	2007	2008	2009	2010	2011	2012	As of 19 July 2013	Mean
ETFC	2.37	2.38	0.68	0.88	0.84	0.74	0.54	0.65	1.14
Forecasted growth in book value: 1.5%									
Forecasted growth in revenues: −1.0%									
Beta: 1.65									
SCHW	4.23	6.69	6.14	3.54	3.15	2.50	1.96	2.31	3.81
Forecasted growth in book value: 10.5%									
Forecasted growth in revenues: 5.0%									
Beta: 1.20									
AMTD	6.96	4.85	3.33	2.60	2.68	2.44	2.20	2.53	3.45
Forecasted growth in book value: 9.0%									
Forecasted growth in revenues: 3.5%									
Beta: 1.10									

Source: The Value Line Investment Survey. The price-to-book value ratio is based on the average of the annual high and low prices and end-of-year book value.

Based only on the information in Exhibit 12, discuss the relative valuation of ETFC relative to the other two companies.

Solution:

ETFC is currently selling at a P/B that is less than 30% of the P/B for either SCHW or AMTD. It is also selling at a P/B that is less than 60% of its average P/B for the time period noted in the exhibit. The likely explanation for ETFC's low P/B is that its growth forecasts for book value and revenues are lower and its beta is higher than those for SCHW and AMTD. In deciding whether ETFC is overvalued or undervalued, an analyst would likely decide how his or her growth forecast and the uncertainty surrounding that forecast compare to the market consensus.

PRICE TO SALES

6

b calculate and interpret a justified price multiple;

c describe rationales for and possible drawbacks to using alternative price multiples and dividend yield in valuation;

d calculate and interpret alternative price multiples and dividend yield;

g describe fundamental factors that influence alternative price multiples and dividend yield;

h calculate and interpret the justified price-to-earnings ratio (P/E), price-to-book ratio (P/B), and price-to-sales ratio (P/S) for a stock, based on forecasted fundamentals;

j evaluate a stock by the method of comparables and explain the importance of fundamentals in using the method of comparables;

r evaluate whether a stock is overvalued, fairly valued, or undervalued based on comparisons of multiples.

Certain types of privately held companies, including investment management companies and many types of companies in partnership form, have long been valued by a multiple of annual revenues. In recent decades, the ratio of price to sales has become well known as a valuation indicator for the equity of publicly traded companies as well. Based on US data, O'Shaughnessy (2005) characterized P/S as the best ratio for selecting undervalued stocks.

According to the 2012 BofA *Merrill Lynch Institutional Factor Survey*, about 30% of respondents consistently used P/S in their investment process. Analysts have offered the following rationales for using P/S:

- Sales are generally less subject to distortion or manipulation than are other fundamentals, such as EPS or book value. For example, through discretionary accounting decisions about expenses, company managers can distort EPS as a reflection of economic performance. In contrast, total sales, as the top line in the income statement, is prior to any expenses.
- Sales are positive even when EPS is negative. Therefore, analysts can use P/S when EPS is negative, whereas the P/E based on a zero or negative EPS is not meaningful.
- Because sales are generally more stable than EPS, which reflects operating and financial leverage, P/S is generally more stable than P/E. P/S may be more meaningful than P/E when EPS is abnormally high or low.
- P/S has been viewed as appropriate for valuing the stocks of mature, cyclical, and zero-income companies (Martin 1998).
- Differences in P/S multiples may be related to differences in long-run average returns, according to empirical research (Nathan, Sivakumar and Vijayakumar, 2001; O'Shaughnessy, 2005).

Possible drawbacks of using P/S in practice include the following:

- A business may show high growth in sales even when it is not operating profitably as judged by earnings and cash flow from operations. To have value as a going concern, a business must ultimately generate earnings and cash.
- Share price reflects the effect of debt financing on profitability and risk. In the P/S multiple, however, price is compared with sales, which is a prefinancing income measure—a logical mismatch. For this reason, some experts use a ratio of enterprise value to sales because enterprise value incorporates the value of debt.
- P/S does not reflect differences in cost structures among different companies.
- Although P/S is relatively robust with respect to manipulation, revenue recognition practices have the potential to distort P/S.

Despite the contrasts between P/S to P/E, the ratios have a relationship with which analysts should be familiar. The fact that (Sales) × (Net profit margin) = Net income means that (P/E) × (Net profit margin) = P/S. For two stocks with the same positive P/E, the stock with the higher P/S has a higher (actual or forecasted) net profit margin, calculated as the ratio of P/S to P/E.

6.1 Determining Sales

P/S is calculated as price per share divided by annual net sales per share (net sales is total sales minus returns and customer discounts). Analysts usually use annual sales from the company's most recent fiscal year in the calculation, as illustrated in Example 22. Because valuation is forward looking in principle, the analyst may also develop and use P/S multiples based on forecasts of next year's sales.

EXAMPLE 22

Calculating P/S

Stora Enso Oyj (Helsinki Stock Exchange: STERV) is an integrated paper, packaging, and forest products company headquartered in Finland. In its fiscal year ended 31 December 2018, Stora Enso reported net sales of €10,486 million and had 788.4 million shares outstanding. Calculate the P/S for Stora Enso based on a closing price of €10.34 on 28 June 2019.

Solution:

Sales per share = €10,486 million/788.6 million shares = €13.30. So, P/S = €10.34/€13.30 = 0.778.

Although the determination of sales is more straightforward than the determination of earnings, the analyst should evaluate a company's revenue recognition practices—in particular those tending to speed up the recognition of revenues—before relying on the P/S multiple. An analyst using a P/S approach who does not also assess the quality of accounting for sales may place too high a value on the company's shares. Example 23 illustrates the problem.

6.2 Valuation Based on Forecasted Fundamentals

Like other multiples, P/S can be linked to DCF models. In terms of the Gordon growth model, we can state P/S as

$$\frac{P_0}{S_1} = \frac{\left(\frac{E_1}{S_1}\right)(1-b)}{r-g}. \qquad (6)$$

where E_1/S_1 is the business's forward-looking profit margin (the equation can be obtained from the Gordon Growth model $P_0=D_1/(r\text{-}g)$, by substituting $D_1=E_1(1\text{-}b)$ into the numerator and then dividing both sides by S_1). Equation 6 states that the justified P/S is an increasing function of the profit margin and earnings growth rate, and the intuition behind Equation 6 generalizes to more-complex DCF models.

EXAMPLE 23

Revenue Recognition Practices (1)

Analysts label stock markets "bubbles" when market prices appear to lose contact with intrinsic values. To many analysts, the run-up in the prices of internet stocks in the US market in the 1998–2000 period represented a bubble. During that period, many analysts adopted P/S as a metric for valuing the many internet stocks that had negative earnings and cash flow. Perhaps at least partly as a result of this practice, some internet companies engaged in questionable revenue recognition practices to justify their high valuations. To increase sales, some companies engaged in bartering website advertising with other internet companies. For example, InternetRevenue.com might barter $1,000,000 worth of banner advertising with RevenueIsUs.com. Each could then show $1,000,000 of revenue and $1,000,000 of expenses. Although neither had any net income or cash flow, each company's revenue growth and market valuation was enhanced (at least temporarily). In addition, the value placed on the advertising was frequently questionable.

As a result of these and other questionable activities, the US SEC issued a stern warning to companies and formalized revenue recognition practices for barter in Staff Accounting Bulletin No. 101. Similarly, international accounting standard setters issued Standing Interpretations Committee Interpretation 31 to define revenue recognition principles for barter transactions involving advertising services. The analyst should review footnote disclosures to assess whether a company may be recognizing revenue prematurely or otherwise aggressively.

Example 24 illustrates another classic instance in which an analyst should look behind the accounting numbers.

EXAMPLE 24

Revenue Recognition Practices (2)

Sales on a **bill-and-hold basis** involve selling products but not delivering those products until a later date. Sales on this basis have the effect of accelerating the recognition of those sales into an earlier reporting period. In its form 10-K filed 30 September 2008, Diebold, a provider of bank security systems and ATMs, provided the following note:

> **Revenues**
>
> Bill and Hold—The largest of the revenue recognition adjustments relates to the Company's previous long-standing method of accounting for bill and hold transactions under Staff Accounting Bulletin 104, Revenue Recognition in Financial Statements (SAB 104), in its North America and International businesses. On January 15, 2008, the Company announced that it had concluded its discussions with the OCA in regard to its practice of recognizing certain revenue on a bill and hold basis in its North America business segment. As a result of those discussions, the Company determined that its previous, long-standing method of accounting for bill and hold transactions was in error, representing a misapplication of GAAP. To correct for this error, the Company announced it would discontinue the use of bill and hold as a method of revenue recognition in its North America and International businesses and restate its financial statements for this change.
>
> The Company completed an analysis of transactions and recorded adjusting journal entries related to revenue and costs recognized previously under a bill and hold basis that is now recognized upon customer acceptance of products at a customer location. Within the North America business segment, when the Company is contractually responsible for installation, customer acceptance will be upon completion of the installation of all of the items at a job site and the Company's demonstration that the items are in operable condition.

> Where items are contractually only delivered to a customer, revenue recognition of these items will continue upon shipment or delivery to a customer location depending on the terms in the contract. Within the International business segment, customer acceptance is upon either delivery or completion of the installation depending on the terms in the contract with the customer. The Company restated for transactions affecting both product revenue for hardware sales and service revenue for installation and other services that had been previously recognized on a bill and hold basis.
>
> Other Revenue Adjustments—The Company also adjusted for other specific revenue transactions in both its North America and International businesses related to transactions largely where the Company recognized revenue in incorrect periods. The majority of these adjustments were related to misapplication of GAAP related to revenue recognition requirements as defined within SAB 104. Generally, the Company recorded adjustments for transactions when the Company previously recognized revenue prior to title and/or risk of loss transferring to the customer.

In 2010, Diebold agreed to pay $25 million to settle Securities and Exchange Commission charges that it manipulated its earnings from at least 2002 through 2007. During that period, the company misstated the company's reported pre-tax earnings by at least $127 million.

According to the SEC, Diebold's financial management received reports, sometimes on a daily basis, comparing the company's actual earnings to analyst earnings forecasts. Diebold's management would prepare "opportunity lists" of ways to close the gap between the company's actual financial results and analyst forecasts. Many of the methods were fraudulent accounting transactions designed to improperly recognize revenue or otherwise inflate Diebold's financial performance. Among the fraudulent practices identified by the SEC were the following: improper use of bill and hold accounting, recognition of revenue on a lease agreement subject to a side buy-back agreement, manipulating reserves and accruals, improperly delaying and capitalizing expenses, and writing up the value of used inventory.

Example 25 briefly summarizes another example of aggressive revenue recognition practices.

EXAMPLE 25

Revenue Recognition Practices (3)

Groupon is a deal-of-the-day website that features discounted gift certificates usable at local or national companies. Before going public in November 2011, Groupon amended its registration statement eight times. One SEC-mandated restatement forced it to change an auditor-sanctioned method of reporting revenue, reducing sales by more than 50%. Essentially, Groupon had initially counted the gross amount its members paid for coupons or certificates as revenue, without deducting the share (typically half or more) that it sends to local merchants. The SEC also demanded Groupon remove from its offering document a non-GAAP metric it had invented called "adjusted consolidated segment operating income." This measure was considered misleading because it ignored marketing expenses, which are one of the major risks of Groupon's business model.

Even when a company discloses its revenue recognition practices, the analyst cannot always determine precisely by how much sales may be overstated. If a company is engaging in questionable revenue recognition practices and the amount being manipulated is unknown, the analyst might do well to suggest avoiding investment in that company's securities. At the very least, the analyst should be skeptical and assign the company a higher risk premium than otherwise, which would result in a lower justified P/S.

6.3 Valuation Based on Forecasted Fundamentals

Like other multiples, P/S can be linked to DCF models. In terms of the Gordon growth model, we can state P/S as

$$\frac{P_0}{S_0} = \frac{(E_0/S_0)\,(1-b)(1+g)}{r-g}, \tag{7}$$

where E_0/S_0 is the business's profit margin (the equation can be obtained from the Gordon growth model, $P_0 = D_0(1 + g)/(r - g)$, by substituting $D_0 = E_0(1 - b)$ into the numerator and then dividing both sides by S_0). Although the profit margin is stated in terms of trailing sales and earnings, the analyst may use a long-term forecasted profit margin in Equation 7. Equation 7 states that the justified P/S is an increasing function of the profit margin and earnings growth rate, and the intuition behind Equation 7 generalizes to more-complex DCF models.

Profit margin is a determinant of the justified P/S not only directly but also through its effect on g. We can illustrate this concept by restating the equation for the sustainable growth rate [g = (Retention rate, b) × ROE], as follows:

$$g = b \times \text{PM}_0 \times \frac{\text{Sales}}{\text{Total assets}} \times \frac{\text{Total assets}}{\text{Shareholders' equity}},$$

where PM_0 is profit margin and the last three terms come from the DuPont analysis of ROE. An increase (decrease) in the profit margin produces a higher (lower) sustainable growth rate as long as sales do not decrease (increase) proportionately. Example 26 illustrates the use of justified P/S and how to apply it in valuation.

EXAMPLE 26

Justified P/S Based on Forecasted Fundamentals

As a health care analyst, you are valuing the stocks of three medical equipment manufacturers, including the Swedish company Getinge AB (GETI) in March 2019. Based on an average of estimates obtained from capital asset pricing model (CAPM) and bond yield plus risk premium approaches, you estimate that GETI's required rate of return is 9%. You have gathered the following data from GETI's annual reports (amounts in millions of Swedish krona, or SEK):

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Net sales	22,816	22,712	21,854	24,248	25,287	26,669	30,235	29,756	22,496	24,172
Growth rates (geometric)										
2009–2018	0.6%									

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
2014–2018	−2.4%									
Year / Year		−0.5%	−3.8%	11.0%	4.3%	5.5%	13.4%	−1.6%	−24.4%	7.5%
Net profit	1,914	2,280	2,537	2,531	2,285	1,433	1,390	1,188	1,376	−967
Growth rates (geometric)										
2009–2018	NMF									
2014–2018	NMF									
Year / Year		19.1%	11.3%	−0.2%	−9.7%	−37.3%	−3.0%	−14.5%	15.8%	−170.3%
Net profit margin	8.4%	10.0%	11.6%	10.4%	9.0%	5.4%	4.6%	4.0%	6.1%	−4.0%
Averages										
2009–2018	6.6%									
2014–2018	3.2%									
Dividend payout ratio	0.3%	34.0%	35.3%	39.2%	43.3%	69.3%	49.7%	57.7%	36.0%	−43.8%
Averages										
2009–2018	32.1%									
2014–2018	33.8%									

Sales growth and profitability have been quite variable in recent years, particularly in 2017 and 2018, making it difficult to extrapolate future trends. Based on further research on the company and its industry, , you make the following long-term forecasts:

Profit margin = 9.0%

Dividend payout ratio = 35.0%

Earnings growth rate = 7.0%

1. Based on these data, calculate GETI's justified P/S.
2. Given a forecast of GETI's sales per share (in Swedish krona) for 2019 of SEK94.3, estimate the intrinsic value of GETI stock.
3. Given a market price for GETI of SEK133.70 on 26 August 2019 and your answer to Part 2, determine whether GETI stock appears to be fairly valued, overvalued, or undervalued.

Solution to 1:

From Equation 7, GETI's justified P/S is calculated as follows:

$$\frac{P_0}{S_1} = \frac{(E_1/S_1)(1-b)}{r-g} = \frac{0.09 \times 0.35}{0.09 - 0.07} = 1.575$$

Solution to 2:

An estimate of the intrinsic value of GETI stock is 1.575 × SEK94.3 = SEK148.52.

Solution to 3:

GETI stock appears to be undervalued because its current market value of SEK133.70 is less than its estimated intrinsic value of SEK148.52.

6.4 Valuation Based on Comparables

Using P/S in the method of comparables to value stocks follows the steps given in Section 4. As mentioned earlier, P/Ss are usually reported on the basis of trailing sales. Analysts may also base relative valuations on P/S multiples calculated on forecasted sales. In doing so, analysts may make their own sales forecasts or may use forecasts supplied by data vendors. In valuing stocks using the method of comparables, analysts should also gather information on profit margins, expected earnings growth, and risk. As always, the quality of accounting also merits investigation. Example 27 illustrates the use of P/S in the comparables approach.

EXAMPLE 27

P/S Comparables Approach

Continuing with the project to value Getinge AB, you have compiled the information on GETI and peer companies Cantel Medical Corporation (CMD) and New Genomics (NEO) given in Exhibit 13.

Exhibit 13 P/S Comparables (as of 26 October 2019)

Measure	GETI	CMD	NEO
Price/Sales (TTM)	1.54	3.96	8.79
Profit Margin (TTM)	−2.49%	6.95%	14.53%
Quarterly Revenue Growth (YoYy)	9.50%	5.20%	1.50%
Total Debt/Equity (mrq)	58.43	35.58	28.50
Enterprise Value/Revenue (TTM)	1.88	4.14	8.23

Source: Yahoo! Finance.

Use the data in Exhibit 13 to address the following:

1 Based on the P/S but referring to no other information, assess GETI's relative valuation.

2 State whether GETI is more closely comparable to CMD or to NEO. Justify your answer.

Solution to 1:

Because the P/S for GETI, 1.54, is the lowest of the three P/S multiples, if no other information is referenced, GETI appears to be relatively undervalued.

Solution to 2:

On the basis of the information given, GETI appears to be more closely matched to CMD than to NEO. NEO's P/S is significantly higher than the P/S for GETI and CMD. The profit margin and revenue growth are key fundamentals in the P/S approach, and NEO's higher P/S reflects its high profit margin. GETI's funding (Total debt/Equity) is higher than that of CMD and NEO, and its Enterprise value/Revenue is low and much closer to CMD's ratio than to that of NEO. Overall, GETI's valuation seems to be more like that of CMD than that of NEO. GETI's low P/S is consistent with its other relative-valuation metrics in Exhibit 13.

7 PRICE TO CASH FLOW

b calculate and interpret a justified price multiple;

c describe rationales for and possible drawbacks to using alternative price multiples and dividend yield in valuation;

d calculate and interpret alternative price multiples and dividend yield;

g describe fundamental factors that influence alternative price multiples and dividend yield;

h calculate and interpret the justified price-to-earnings ratio (P/E), price-to-book ratio (P/B), and price-to-sales ratio (P/S) for a stock, based on forecasted fundamentals;

j evaluate a stock by the method of comparables and explain the importance of fundamentals in using the method of comparables;

r evaluate whether a stock is overvalued, fairly valued, or undervalued based on comparisons of multiples.

Price to cash flow is a widely reported valuation indicator. According to the 2012 BofA Merrill Lynch Institutional Factor Survey, price to free cash flow trailed only P/E, beta, enterprise value/EBITDA, ROE, size, and P/B in popularity as a valuation factor and was used as a valuation metric by approximately half of the institutions surveyed.

In this section, we present price to cash flow based on alternative major cash flow concepts. Note that "price to cash flow" is used to refer to the ratio of share price to any one of these definitions of cash flow whereas "P/CF" is reserved for the ratio of price to the earnings-plus-noncash-charges definition of cash flow, explained later. Because of the wide variety of cash flow concepts in use, the analyst should be especially careful to understand (and communicate) the exact definition of "cash flow" that is the basis for the analysis.

Analysts have offered the following rationales for the use of price to cash flow:

- Cash flow is less subject to manipulation by management than earnings.
- Because cash flow is generally more stable than earnings, price to cash flow is generally more stable than P/E.
- Using price to cash flow rather than P/E addresses the issue of differences in accounting conservatism between companies (differences in the quality of earnings).
- Differences in price to cash flow may be related to differences in long-run average returns, according to empirical research (O'Shaughnessy 2005).

Possible drawbacks to the use of price to cash flow include the following:

- When cash flow from operations is defined as EPS plus noncash charges, items affecting actual cash flow from operations, such as noncash revenue and net changes in working capital, are ignored. So, for example, aggressive recognition of revenue (front-end loading) would not be accurately captured in the earnings-plus-noncash-charges definition because the measure would not reflect the divergence between revenues as reported and actual cash collections related to that revenue.
- Theory views free cash flow to equity (FCFE) rather than cash flow as the appropriate variable for price-based valuation multiples. We can use P/FCFE, but FCFE does have the possible drawback of being more volatile than cash flow for many businesses. FCFE is also more frequently negative than cash flow.
- As analysts' use of cash flow has increased over time, some companies have increased their use of accounting methods that enhance cash flow measures. Operating cash flow, for example, can be enhanced by securitizing accounts receivable to speed up a company's operating cash inflow or by outsourcing the payment of accounts payable to slow down the company's operating cash outflow (while the outsource company continues to make timely payments and

provides financing to cover any timing differences). Mulford and Comiskey (2005) described a number of opportunistic accounting choices that companies can make to increase their reported operating cash flow.

- Operating cash flow from the statement of cash flows under IFRS may not be comparable to operating cash flow under US GAAP because IFRS allow more flexibility in classification of interest paid, interest received, and dividends received. Under US GAAP, all three of these items are classified in operating cash flow, but under IFRS, companies have the option to classify them as operating or investing (for interest and dividends received) and as operating or financing (for interest paid).

One approximation of cash flow in practical use is EPS plus per-share depreciation, amortization, and depletion. This simple approximation is used in Example 28 to highlight issues of interest to the analyst in valuation.

EXAMPLE 28

Accounting Methods and Cash Flow

Consider two hypothetical companies, Company A and Company B, that have constant cash revenues and cash expenses (as well as a constant number of shares outstanding) in 2018, 2019, and 2020. In addition, both companies incur total depreciation of $15.00 per share during the three-year period, and both use the same depreciation method for tax purposes. The two companies use different depreciation methods, however, for financial reporting. Company A spreads the depreciation expense evenly over the three years (straight-line depreciation, or SLD). Because its revenues, expenses, and depreciation are constant over the period, Company A's EPS is also constant. In this example, Company A's EPS is assumed to be $10 each year, as shown in Column 1 in Exhibit 14.

Company B is identical to Company A except that it uses accelerated depreciation. Company B's depreciation is 150% of SLD in 2018 and declines to 50% of SLD in 2020, as shown in Column 5.

Exhibit 14 Earnings Growth Rates and Cash Flow (All Amounts per Share)

	Company A			Company B		
Year	Earnings (1)	Depreciation (2)	Cash Flow (3)	Earnings (4)	Depreciation (5)	Cash Flow (6)
2018	$10.00	$5.00	$15.00	$7.50	$7.50	$15.00
2019	10.00	5.00	15.00	10.00	5.00	15.00
2020	10.00	5.00	15.00	12.50	2.50	15.00
Total		$15.00			$15.00	

Because of the different depreciation methods used by Company A and Company B for financial reporting purposes, Company A's EPS (Column 1) is flat at $10.00 whereas Company B's EPS (Column 4) shows 29% compound growth: $(\$12.50/\$7.50)^{1/2} - 1.00 = 0.29$. Thus, Company B appears to have positive earnings momentum. Analysts comparing Companies A and B might be misled by using the EPS numbers as reported instead of putting EPS on a comparable basis. For both companies, however, cash flow per share is level at $15.

Depreciation may be the simplest noncash charge to understand; write-offs and other noncash charges may offer more latitude for the management of earnings.

7.1 Determining Cash Flow

In practice, analysts and data vendors often use simple *approximations* of cash flow from operations in calculating cash flow for price-to-cash-flow analysis. For many companies, depreciation and amortization are the major noncash charges regularly added to net income in the process of calculating cash flow from operations by the add-back method, so the approximation focuses on them. A representative approximation specifies cash flow per share as EPS plus per-share depreciation, amortization, and depletion. We call this estimation the "earnings-plus-noncash-charges" definition and in this section use the acronym CF for it. Keep in mind, however, that this definition is only one commonly used in calculating price to cash flow, not a technically accurate definition from an accounting perspective. We will also describe more technically accurate cash flow concepts: cash flow from operations, free cash flow to equity, and EBITDA (an estimate of pre-interest, pretax operating cash flow).

Most frequently, trailing price to cash flow is reported. A trailing price to cash flow is calculated as the current market price divided by the sum of the most recent four quarters' cash flow per share. A fiscal year definition is also possible, as in the case of EPS.

Example 29 illustrates the calculation of P/CF with cash flow defined as earnings plus noncash charges.

EXAMPLE 29

Calculating Price to Cash Flow with Cash Flow Defined as Earnings plus Noncash Charges

In 2018, Koninklijke Philips Electronics N.V. (PHIA) reported net income from continuing operations of €1,310 million, equal to EPS of €1.41. The company's depreciation and amortization was €1,089 million, or €1.17 per share. An AEX price for PHIA as of 29 March 2019 was €36.31. Calculate the P/CF for PHIA.

Solution:

CF (defined as EPS plus per-share depreciation, amortization, and depletion) is €1.41 + €1.17 = €2.58 per share. Thus, P/CF = €36.31/€2.58 = 14.1.

Rather than use an approximate EPS-plus-noncash-charges concept of cash flow, analysts can use cash flow from operations (CFO) in a price multiple. CFO is found in the statement of cash flows. Similar to the adjustments to normalize earnings, adjustments to CFO for components not expected to persist into future time periods may also be appropriate. In addition, adjustments to CFO may be required when comparing companies that use different accounting standards. For example, as noted above, under IFRS, companies have flexibility in classifying interest payments, interest receipts, and dividend receipts across operating, investing, and financing. US GAAP require companies to classify interest payments, interest receipts, and dividend receipts as operating cash flows.

As an alternative to CF and CFO, the analyst can relate price to FCFE, the cash flow concept with the strongest link to valuation theory. Because the amounts of capital expenditures in proportion to CFO generally differ among companies being compared, the analyst may find that rankings by price to cash flow from operations (P/CFO) and by P/CF will differ from rankings by P/FCFE. Period-by-period FCFE may be more volatile than CFO (or CF), however, so a trailing P/FCFE is not necessarily more informative in a valuation. For example, consider two similar businesses with the same CFO and capital expenditures over a two-year period. If the first company times its capital expenditures to fall toward the beginning of the period and the second times its capital expenditures to fall toward the end of the period, the P/FCFEs for the two stocks may differ sharply without representing a meaningful economic difference. The analyst could, however, appropriately use the FCFE discounted cash flow model value, which incorporates all expected future free cash flows to equity. This concern can be addressed, at least in part, by using price to average free cash flow, as in Hackel, Livnat, and Rai (1994).

Another cash flow concept used in multiples is EBITDA (earnings before interest, taxes, depreciation, and amortization). To forecast EBITDA, analysts usually start with their projections of EBIT and simply add depreciation and amortization to arrive at an estimate for EBITDA. In calculating EBITDA from historical numbers, one can start with earnings from continuing operations, excluding nonrecurring items. To that earnings number, interest, taxes, depreciation, and amortization are added.

In practice, both EV/EBITDA and P/EBITDA have been used by analysts as valuation metrics. EV/EBITDA has been the preferred metric, however, because its numerator includes the value of debt; therefore, it is the more appropriate method because EBITDA is pre-interest and is thus a flow to both debt and equity. EV/EBITDA is discussed in detail in a later section.

7.2 Valuation Based on Forecasted Fundamentals

The relationship between the justified price to cash flow and fundamentals follows from the familiar mathematics of the present value model. The justified price to cash flow, all else being equal, is inversely related to the stock's required rate of return and positively related to the growth rate(s) of expected future cash flows (however defined). We can find a justified price to cash flow based on fundamentals by finding the value of a stock using the most suitable DCF model and dividing that number by cash flow (based on our chosen definition of cash flow). Example 30 illustrates the process.

EXAMPLE 30

Justified Price to Cash Flow Based on Forecasted Fundamentals

As a consumer staples analyst, you are working on the valuation of Colgate-Palmolive (CL), a global consumer products supplier. As a first estimate of value, you are applying an FCFE model under the assumption of a stable long-term growth rate in FCFE:

$$V_0 = \frac{(1 + g)\text{FCFE}_0}{r - g},$$

where g is the expected growth rate of FCFE. You estimate trailing FCFE at \$2.66 per share and trailing CF (based on the earnings-plus-noncash-charges definition) at \$3.26. Your other estimates are a 7.4% required rate of return and a 3.2% expected growth rate of FCFE.

1 What is the intrinsic value of CL according to a constant growth FCFE model?

2 What is the justified P/CF based on forecasted fundamentals?

3 What is the justified P/FCFE based on forecasted fundamentals?

Solution to 1:

Calculate intrinsic value as (1.032 × \$2.66)/(0.074 − 0.032) = \$58.41.

Solution to 2:

Calculate a justified P/CF based on forecasted fundamentals as \$58.41/\$3.26 = 24.6.

Solution to 3:

The justified P/FCFE is \$58.41/\$2.66 = 22.0.

7.3 Valuation Based on Comparables

The method of comparables for valuing stocks based on price to cash flow follows the steps given previously and illustrated for P/E, P/B, and P/S. Example 31 is a simple exercise in the comparables method based on price-to-cash-flow measures.

EXAMPLE 31

Price to Cash Flow and Comparables

Exhibit 15 provides information on P/CF, P/FCFE, and selected fundamentals as of 16 April 2020 for two hypothetical companies. Using the information in Exhibit 15, compare the valuations of the two companies.

Exhibit 15 Comparison of Two Companies (All Amounts per Share)

Company	Current Price (£)	Trailing CF per Share (£)	P/CF	Trailing FCFE per Share (£)	P/FCFE	Consensus Five-Year CF Growth Forecast (%)	Beta
Company A	17.98	1.84	9.8	0.29	62	13.4	1.50
Company B	15.65	1.37	11.4	−0.99	NMF	10.6	1.50

Company A is selling at a P/CF (9.8) approximately 14% smaller than the P/CF of Company B (11.4). Based on that comparison, we expect that, all else equal, investors would anticipate a higher growth rate for Company B. Contrary to that expectation, however, the consensus five-year earnings growth forecast for Company A is 280 basis points higher than it is for Company B. As of the date of the comparison, Company A appears to be relatively undervalued compared with Company B, as judged by P/CF and expected growth. The information in Exhibit 15 on FCFE supports the proposition that Company A may be relatively undervalued. The positive FCFE for Company A indicates that operating cash flows and new debt borrowing are more than sufficient to cover capital expenditures. Negative FCFE for Company B suggests the need for external funding of growth.

8 PRICE TO DIVIDENDS AND DIVIDEND YIELD

- **b** calculate and interpret a justified price multiple;
- **c** describe rationales for and possible drawbacks to using alternative price multiples and dividend yield in valuation;
- **d** calculate and interpret alternative price multiples and dividend yield;
- **g** describe fundamental factors that influence alternative price multiples and dividend yield;
- **h** calculate and interpret the justified price-to-earnings ratio (P/E), price-to-book ratio (P/B), and price-to-sales ratio (P/S) for a stock, based on forecasted fundamentals;
- **j** evaluate a stock by the method of comparables and explain the importance of fundamentals in using the method of comparables;
- **r** evaluate whether a stock is overvalued, fairly valued, or undervalued based on comparisons of multiples.

The total return on an equity investment has a capital appreciation component and a dividend yield component. Dividend yield data are frequently reported to provide investors with an estimate of the dividend yield component in total return. Dividend yield is also used as a valuation indicator. Although the 2012 BofA Merrill Lynch Institutional Factor Survey did not survey this metric, in its surveys from 1989 to 2006 slightly more than one-quarter of respondents on average reported using dividend yield as a factor in the investment process.

Analysts have offered the following rationales for using dividend yields in valuation:

- Dividend yield is a component of total return.
- Dividends are a less risky component of total return than capital appreciation.

Possible drawbacks of using dividend yields include the following:

- Dividend yield is only one component of total return; not using all information related to expected return is suboptimal.
- Investors may trade off future earnings growth to receive higher current dividends. That is, holding return on equity constant, dividends paid now displace earnings in all future periods (a concept known as the **dividend displacement of earnings**). Arnott and Asness (2003) and Zhou and Ruland (2006), however, showed that caution must be exercised in assuming that dividends displace future earnings in practice, because dividend payout may be correlated with future profitability.
- The argument about the relative safety of dividends presupposes that market prices reflect in a biased way differences in the relative risk of the components of return.

8.1 Calculation of Dividend Yield

This reading so far has presented multiples with market price (or market capitalization) in the numerator. P/Ds have sometimes appeared in valuation, particularly with respect to indexes. Many stocks, however, do not pay dividends, and P/D is undefined with zero in the denominator. For such non-dividend-paying stocks, dividend yield (D/P) *is* defined: It is equal to zero. For practical purposes, then, dividend yield is the preferred way to present this multiple.

Trailing dividend yield is generally calculated by using the dividend rate divided by the current market price per share. The annualized amount of the most recent dividend is known as the **dividend rate**. For companies paying quarterly dividends, the dividend rate is calculated as four times the most recent quarterly per-share dividend. (Some data sources use the dividends in the last four quarters as the dividend rate for purposes of a trailing dividend yield.) For companies that pay semiannual dividends comprising an interim dividend that typically differs in magnitude from the final dividend, the dividend rate is usually calculated as the most recent annual per-share dividend.

The dividend rate indicates the annual amount of dividends per share under the assumption of no increase or decrease over the year. The analyst's forecast of leading dividends could be higher or lower and is the basis of the leading dividend yield. The **leading dividend yield** is calculated as forecasted dividends per share over the next year divided by the current market price per share. Example 32 illustrates the calculation of dividend yield.

EXAMPLE 32

Calculating Dividend Yield

Exhibit 16 gives quarterly dividend data for Canadian telecommunications company BCE Inc. (BCE) and semiannual dividend data for the ADRs of BT Group (BT), formerly British Telecom.

Exhibit 16 Dividends Paid per Share for BCE Inc. and for BT Group ADRs

Period	BCE ($)	BT ADR ($)
4Q:2016	0.51	
1Q:2017	0.54	0.685
2Q:2017	0.53	
3Q:2017	0.57	0.339
Total	2.15	1.024
4Q:2017	0.56	
1Q:2018	0.60	0.675
2Q:2018	0.58	
3Q:2018	0.58	0.301
Total	2.32	0.976

Source: Value Line.

1 Given a price per share for BCE of $39.53 during 4Q:2018, calculate this company's trailing dividend yield.

2 Given a price per ADR for BT of $15.20 during 4Q:2018, calculate the trailing dividend yield for the ADRs.

Solution to 1:

The dividend rate for BCE is \$0.58 × 4 = \$2.32. The dividend yield is \$2.32/\$39.53 = 0.0587, or 5.87%.

Solution to 2:

Because BT pays semiannual dividends that differ in magnitude between the interim and final dividends, the dividend rate for BT's ADR is the total dividend in the most recent year, \$0.976. The dividend yield is \$0.976/\$15.20 = 0.0642, or 6.52%.

8.2 Valuation Based on Forecasted Fundamentals

The relationship of dividend yield to fundamentals can be illustrated in the context of the Gordon growth model. From that model, we obtain the expression

$$\frac{D_0}{P_0} = \frac{r - g}{1 + g}. \tag{8}$$

Equation 8 shows that dividend yield is negatively related to the expected rate of growth in dividends and positively related to the stock's required rate of return. The first point implies that the selection of stocks with relatively high dividend yields is consistent with an orientation to a value rather than growth investment style.

8.3 Valuation Based on Comparables

Using dividend yield with comparables is similar to the process that has been illustrated for other multiples. An analyst compares a company with its peers to determine whether it is attractively priced, considering its dividend yield and risk. The analyst should examine whether differences in expected growth explain the differences in dividend yield. Another consideration used by some investors is the security of the dividend (the probability that it will be reduced or eliminated). A useful metric in assessing the safety of the dividend is the payout ratio: A high payout relative to other companies operating in the same industry may indicate a less secure dividend because the dividend is less well covered by earnings. Balance sheet metrics are equally important in assessing the safety of the dividend, and relevant ratios to consider include the interest coverage ratio and the ratio of net debt to EBITDA. Example 33 illustrates use of the dividend yield in the method of comparables.

EXAMPLE 33

Dividend Yield Comparables

William Leiderman is a portfolio manager for a US pension fund's domestic equity portfolio. The portfolio is exempt from taxes, so any differences in the taxation of dividends and capital gains are not relevant. Leiderman's client requires high current income. Leiderman is considering the purchase of utility stocks for the fund in August 2019. In the course of his review, he considers the four large-cap US electric utilities shown in Exhibit 17.

Exhibit 17 Using Dividend Yield to Compare Stocks

Company	Consensus Earnings Growth Forecast (%)	Beta	Dividend Yield (%)	Payout Ratio (%)
Duke Energy	7.20	0.18	4.24	89
NiSource Inc.	4.63	0.22	2.70	NMF
Portland General Electric Co.	5.20	0.24	2.76	59
PPL Corp.	0.60	0.55	5.37	63

Sources: www.finviz.com and Yahoo! Finance.

All of the securities exhibit similar low market risk; they each have a beta substantially less than 1.00. The dividend payout ratio for NiSource is not meaningful due to a negative EPS. Duke Energy's dividend payout ratio of 89%, the highest of the group, also suggests that its dividend may be subject to greater risk. Leiderman notes that PPL Corp.'s relatively low payout ratio means that the dividend is well supported; however, the expected low earnings growth rate is a negative factor. Summing Portland General Electric's dividend yield and expected earnings growth rate, Leiderman estimates Portland General Electric's expected total return is about 7.96%; because the total return estimate is relatively attractive and because Portland General Electric does not appear to have any strong negatives, Leiderman decides to focus his further analysis on Portland General Electric.

ENTERPRISE VALUE TO EBITDA

9

- **m** explain alternative definitions of cash flow used in price and enterprise value (EV) multiples and describe limitations of each definition;
- **n** calculate and interpret EV multiples and evaluate the use of EV/EBITDA;
- **r** evaluate whether a stock is overvalued, fairly valued, or undervalued based on comparisons of multiples.

Enterprise value multiples are multiples that relate the enterprise value of a company to some measure of value (typically, a pre-interest income measure). Perhaps the most frequently advanced argument for using enterprise value multiples rather than price multiples in valuation is that enterprise value multiples are relatively less sensitive to the effects of financial leverage than price multiples when one is comparing companies that use differing amounts of leverage. Enterprise value multiples, in defining the numerator as they do, take a control perspective (discussed in more detail later). Thus, even where leverage differences are not an issue, enterprise value multiples may complement the perspective of price multiples. Indeed, although some analysts strictly favor one type of multiple, other analysts report both price and enterprise value multiples.

9.1 Enterprise Value to EBITDA

Enterprise value to EBITDA is by far the most widely used enterprise value multiple.

Earlier, EBITDA was introduced as an estimate of pre-interest, pretax operating cash flow. Because EBITDA is a flow to both debt and equity, as noted, defining an EBITDA multiple by using a measure of total company value in the numerator, such as EV, is appropriate. Recall that **enterprise value** is total company value (the market value of debt, common equity, and preferred equity) minus the value of cash and short-term investments. Thus, EV/EBITDA is a valuation indicator for the overall company rather than solely its common stock. If, however, the analyst can assume that the business's debt and preferred stock (if any) are efficiently priced, the analyst can use EV/EBITDA to draw an inference about the valuation of common equity. Such an inference is often reasonable.

Analysts have offered the following rationales for using EV/EBITDA:

- EV/EBITDA is usually more appropriate than P/E alone for comparing companies with different financial leverage (debt), because EBITDA is a pre-interest earnings figure, in contrast to EPS, which is postinterest.
- By adding back depreciation and amortization, EBITDA controls for differences in depreciation and amortization among businesses, in contrast to net income, which is postdepreciation and postamortization. For this reason, EV/EBITDA is frequently used in the valuation of capital-intensive businesses (for example, cable companies and steel companies). Such businesses typically have substantial depreciation and amortization expenses.
- EBITDA is frequently positive when EPS is negative.

Possible drawbacks to using EV/EBITDA include the following (Moody's 2000; Grant and Parker 2001):

- EBITDA will overestimate cash flow from operations if working capital is growing. EBITDA also ignores the effects of differences in revenue recognition policy on cash flow from operations.
- Free cash flow to the firm (FCFF), which directly reflects the amount of the company's required capital expenditures, has a stronger link to valuation theory than does EBITDA. Only if depreciation expenses match capital expenditures do we expect EBITDA to reflect differences in businesses' capital programs. This qualification to EBITDA comparisons may be particularly meaningful for the capital-intensive businesses to which EV/EBITDA is often applied.

9.1.1 *Determining Enterprise Value*

We illustrated the calculation of EBITDA previously. As discussed, analysts commonly define enterprise value as follows:

Market value of common equity (Number of shares outstanding × Price per share)

Plus: Market value of preferred stock (if any) and any minority interest (unless included elsewhere)

Plus: Market value of debt

Less: Cash and investments (specifically, cash, cash equivalents, and short-term investments)

Equals: Enterprise value.

Cash and investments (sometimes termed **nonearning assets**) are subtracted because EV is designed to measure the net price an acquirer would pay for the company as a whole. The acquirer must buy out current equity and debt providers but then receives access to the cash and investments, which lower the net cost of the acquisition. (For example, cash and investments can be used to pay off debt or loans used to finance the purchase.) The same logic explains the use of market values: In repurchasing debt, an acquirer has to pay market prices. Some debt, however, may be private and does not trade; some debt may be publicly traded but may trade infrequently. When analysts do not have market values, they often use book values obtained from the balance sheet. Alternatively, they may use so-called matrix price estimates of debt market values in such cases; where they are available, they may be more accurate. Matrix price estimates are based on characteristics of the debt issue and information on how the marketplace prices those characteristics. Example 34 illustrates the calculation of EV/EBITDA.

EXAMPLE 34

Calculating EV/EBITDA

Colgate-Palmolive (CL) provides a variety of household products. Exhibit 18 presents the company's consolidated balance sheet as of 31 December 2018.

Exhibit 18 Colgate-Palmolive Condensed Consolidated Balance Sheet (in Millions except Par Values; Unaudited)

Assets	
Current assets:	
Cash and cash equivalents	$726
Accounts receivable, net	1,400
Inventories	1,250
Other current assets	417
Total current assets	3,793
Property and equipment, net	3,881
Goodwill and other intangible assets, net	4,167
Other non-current assets	320
Total assets	$12,161

Liabilities and Shareholders' Equity	
Current liabilities:	
Accounts payable	$1,222
Accrued income taxes	411
Other accruals	1,696
Current portion of long-term debt	0
Notes and loans payable	12
Total current liabilities	3,341
Long-term debt	6,354
Other non-current liabilities	2,269
Total liabilities	$11,964
Shareholders' equity:	
Preference stock	—

(continued)

Exhibit 18 (Continued)

Liabilities and Shareholders' Equity	
Common stock outstanding—863 million shares	1,466
Additional paid-in capital	2,204
Accumulated comprehensive income (loss)	(4,191)
Retained earnings	21,615
Treasury stock—common shares at cost	(21,196)
Noncontrolling interests	299
Total shareholders' equity	197
Total liabilities and shareholders' equity	$12,161

Source: Company financial report.

This financial statement is audited because US companies are required to have audits only for their annual financial statements. Quarterly statements are labeled as unaudited.

From CL's financial statements, the income statement and statement of cash flows for the year ended 31 December 2018 provided the following items (in millions):

Item	Source	Year Ended 31 December 2018
Net income	Income statement	$2,400
Interest expense (net of interest income)	Income statement	143
Income tax provision	Income statement	906
Depreciation and amortization	Statement of cash flows	511

The company's share price as of 15 February 2019 was $66.48. Based on the above information, calculate EV/EBITDA.

Solution:

- For EV, we first calculate the total value of CL's equity: 863 million shares outstanding times $66.48 price per share equals $57,372 million market capitalization.

  CL has only one class of common stock, no preferred shares, and no **minority interest**. For companies that have multiple classes of common stock, market capitalization includes the total value of all classes of common stock. Similarly, for companies that have preferred stock and/or minority interest, the market value of preferred stock and the amount of minority interest are added to market capitalization.

  EV also includes the value of long-term debt obligations. Per CL's balance sheet, this is the sum of long-term debt ($6,354 million), the current portion of long-term debt ($0 million), and other non-current liabilities ($2,034 million), or $8,388 million. Typically, the book value of long-term debt is used in EV. If, however, the market value of the debt is readily available and materially different from the book value, the market value should be used.

EV excludes cash, cash equivalents, and short-term investments. Per CL's balance sheet, the total of cash and cash equivalents is $726 million.

So, CL's EV is $57,372 million + $8,388 million − $720 million = $65,040 million.

- For EBITDA, we use the trailing 12-month (TTM) data, which are shown in the table above for the year ending 31 December 2018. The EBITDA calculation is

  EBITDA = Net income + Interest + Income taxes + Depreciation and amortization.

  EBITDA = $2,400 + $143 + $906 + $511 = $3,960 million.

CL does not have preferred equity. Companies that do have preferred equity typically present in their financial statement net income available to common shareholders. In those cases, the EBITDA calculation uses net income available to *both* preferred and common equity holders.

For CL, we conclude that EV/EBITDA = ($65,040 million)/($3,960 million) = 16.4.

9.1.2 *Valuation Based on Forecasted Fundamentals*

As with other multiples, intuition about the fundamental drivers of enterprise value to EBITDA can help when applying the method of comparables. All else being equal, the justified EV/EBITDA based on fundamentals should be positively related to the expected growth rate in free cash flow to the firm, positively related to expected profitability as measured by return on invested capital, and negatively related to the business's weighted average cost of capital. **Return on invested capital** (ROIC) is calculated as operating profit after tax divided by invested capital. In analyzing ratios such as EV/EBITDA, ROIC is the relevant measure of profitability because EBITDA flows to all providers of capital.

9.1.3 *Valuation Based on Comparables*

All else equal, a lower EV/EBITDA value relative to peers indicates that a company is relatively undervalued. An analyst's recommendations, however, are usually not completely determined by relative EV/EBITDA; from an analyst's perspective, EV/EBITDA is simply one piece of information to consider.

Example 35 presents a comparison of enterprise value multiples for four peer companies. The example includes a measure of total firm value—**total invested capital** (TIC), sometimes also known as the **market value of invested capital**—that is an alternative to enterprise value. Similar to EV, TIC includes the market value of equity and debt but does not deduct cash and investments.

EXAMPLE 35

Comparable Enterprise Value Multiples

Exhibit 19 presents EV multiples on 27 August 2019 for four companies in the household products industry: Colgate-Palmolive (CL), Kimberly Clark Corp. (KMB), Clorox Co. (CLX), and Church & Dwight Co. (CHD).

Exhibit 19 Enterprise Value Multiples for Industry Peers (Amounts in $ Millions, Except Where Indicated Otherwise)

Measure	CL	KMB	CLX	CHD
Price	$72.60	$140.25	$156.96	$79.15
Times: Shares outstanding (millions)	860	344	127	247
Equals: Equity market cap	62.44	48.25	19.93	19.55
Plus: Debt (most recent quarter)	7.33	8.46	2.69	2.38
Plus: Preferred stock	—	—	—	—
Equals: Market value of TIC	69.77	56.71	22.62	21.93
Less: Cash	0.93	0.53	0.11	0.10
Equals: Enterprise value (EV)	$68.84	$56.18	$22.51	$21.83
EBITDA (TTM)	$4.07	$3.81	$1.28	$0.97
TIC/EBITDA	17.1	14.9	17.7	22.6
EV/EBITDA	16.9	14.7	17.6	22.5
Profit margin (TTM)	14.8%	9.8%	13.2%	5.0%
Quarterly revenue growth (year over year)	–0.5%	–0.2%	–3.8%	13.8%

Sources: Yahoo! Finance; authors' calculations.

1 Exhibit 19 provides two alternative enterprise value multiples, TIC/EBITDA and EV/EBITDA. The ranking of the companies' multiples is identical by both multiples. In general, what could cause the rankings to vary?

2 Each EBITDA multiple incorporates a comparison with enterprise value. How do these multiples differ from price-to-cash-flow multiples?

3 Based solely on the information in Exhibit 19, how does the valuation of CL compare with that of the other three companies?

Solution to 1:

The difference between TIC and EV is that EV excludes cash, cash equivalents, and marketable securities. So, a material variation among companies in cash, cash equivalents, or marketable securities relative to EBITDA could cause the rankings to vary.

Solution to 2:

These multiples differ from price-to-cash-flow multiples in that the numerator is a measure of firm value rather than share price, to match the denominator, which is a pre-interest measure of earnings. These multiples thus provide a more appropriate comparison than price to cash flow when companies have significantly different capital structures.

Solution to 3:

Based on its lower TIC/EBITDA and EV/EBITDA multiples of 17.1 and 16.9, respectively, CL appears undervalued relative to CLX and CHD and overvalued relative to KMB. These valuation ratios may be warranted given differences in profitability and growth rates. Compared with CHD, CL has a similar profit margin and lower revenue growth, which may explain CL's lower valuation multiples. Compared with KMB, the enterprise value multiples of CL are higher, which is consistent with CL being more profitable than KMB (profit margin of 14.8% versus 9.8%).

OTHER ENTERPRISE VALUE MULTIPLES

10

m explain alternative definitions of cash flow used in price and enterprise value (EV) multiples and describe limitations of each definition;

n calculate and interpret EV multiples and evaluate the use of EV/EBITDA;

r evaluate whether a stock is overvalued, fairly valued, or undervalued based on comparisons of multiples.

Although EV/EBITDA is the most widely known and used enterprise value multiple, other enterprise value multiples are used together with or in place of EV/EBITDA—either in a broad range of applications or for valuations in a specific industry. EV/FCFF is an example of a broadly used multiple; an example of a special-purpose multiple is EV/EBITDAR (where R stands for rent expense), which is favored by airline industry analysts. Here we review the most common such multiples (except EV/sales, which is covered in the next section). In each case, a valuation metric could be formulated in terms of TIC rather than EV.

Major alternatives to using EBITDA in the denominator of enterprise value multiples include FCFF (free cash flow to the firm), EBITA (earnings before interest, taxes, and amortization), and EBIT (earnings before interest and taxes). Exhibit 20 summarizes the components of each of these measurements and how they relate to net income. Note that, in practice, analysts typically forecast EBITDA by forecasting EBIT and adding depreciation and amortization.

Exhibit 20 Alternative Denominators in Enterprise Value Multiples

Free Cash Flow to the Firm =	Net Income	plus Interest Expense	minus Tax Savings on Interest	plus Depreciation	plus Amortization	less Investment in Working Capital	less Investment in Fixed Capital
EBITDA =	Net Income	plus Interest Expense	plus Taxes	plus Depreciation	plus Amortization		
EBITA =	Net Income	plus Interest Expense	plus Taxes		plus Amortization		
EBIT =	Net Income	plus Interest Expense	plus Taxes				

Note that the calculation of all the measures given in Exhibit 20 add interest back to net income, which reflects that these measures are flows relevant to all providers of both debt and equity capital. As one moves down the rows of Exhibit 20, the measures incorporate increasingly less precise information about a company's tax position and its capital investments, although each measure has a rationale. For example, EBITA may be chosen in cases in which amortization (associated with intangibles) but not depreciation (associated with tangibles) is a major expense for companies being compared. EBIT may be chosen where neither depreciation nor amortization is a major item.

In addition to enterprise value multiples based on financial measures, in some industries or sectors, the analyst may find it appropriate to examine enterprise value multiples based on a nonfinancial measurement that is specific to that industry or sector. For example, for satellite and cable TV broadcasters, an analyst might usefully examine EV to subscribers. For a resource-based company, a multiple based on reserves of the resource may be appropriate.

Regardless of the specific denominator used in an enterprise value multiple, the concept remains the same—namely, to relate the market value of the total company to some fundamental financial or nonfinancial measure of the company's value.

10.1 Enterprise Value to Sales

Enterprise value to sales is a major alternative to the price-to-sales ratio. The P/S multiple has the conceptual weakness that it fails to recognize that for a debt-financed company, not all sales belong to a company's equity investors. Some of the proceeds from the company's sales will be used to pay interest and principal to the providers of the company's debt capital. For example, a P/S for a company with little or no debt would not be comparable to a P/S for a company that is largely financed with debt. EV/S would be the basis for a valid comparison in such a case. In summary, EV/S is an alternative sales-based ratio that is particularly useful when comparing companies with diverse capital structures. Example 36 illustrates the calculation of EV/S multiples.

EXAMPLE 36

Calculating Enterprise Value to Sales

As described in Example 22, Stora Enso Oyj (Helsinki Stock Exchange: STERV) reported net sales of €10,486 million for 2018. Based on 788.6 million shares outstanding and a stock price of €10.34 on 28 June 2019, the total market value of the company's equity was €8,154 million. The company reported non-current debt of €2,970 million and cash of €1,130 million. Assume that the market value of the company's debt is equal to the amount reported. Calculate the company's EV/S.

Solution:

Enterprise value = €8,145 million + €2,970 million – €1,130 million = €9,994 million. So, EV/S = €9,994 million/€10,486 million = 0.953.

10.2 Price and Enterprise Value Multiples in a Comparable Analysis: Some Illustrative Data

In previous sections, we explained the major price and enterprise value multiples. Analysts using multiples and a benchmark based on closely similar companies should be aware of the range of values for multiples for peer companies and should track the fundamentals that may explain differences. For the sake of illustration, Exhibit 21 shows the median value of various multiples by GICS economic sector, the median dividend payout ratio, and median values of selected fundamentals:

- ROE and its determinants (net profit margin, asset turnover, and financial leverage)
- The compound average growth rate in operating margin for the three years ending with FY2007 (shown in the last column under "3-Year CAGR Operating Margin")

Exhibit 21 is based on the S&P 1500 Composite Index for US equities, consisting of the S&P 500, the S&P MidCap 400 Index, and the S&P SmallCap 600 Index. GICS was described earlier.

At the level of aggregation shown in Exhibit 21, the data are, arguably, most relevant to relative sector valuation. For the purposes of valuing individual companies, analysts would most likely use more narrowly defined industry or sector classification.

INTERNATIONAL CONSIDERATIONS WHEN USING MULTIPLES 11

o explain sources of differences in cross-border valuation comparisons;

Clearly, to perform a relative-value analysis, an analyst must use comparable companies and underlying financial data prepared by applying comparable methods. Therefore, using relative-valuation methods in an international setting is difficult. Comparing companies across borders frequently involves differences in accounting methods, cultural differences, economic differences, and resulting differences in risk and growth opportunities. P/Es for individual companies in the same industry but in different countries have been found to vary widely. Furthermore, P/Es of different national markets often vary substantially at any single point in time.

Although international accounting standards are converging, significant differences still exist across borders, sometimes making comparisons difficult. Even when harmonization of accounting principles is achieved, the need to adjust accounting data for comparability will remain. As we showed earlier, even within a single country's accounting standards, differences between companies result from accounting choices (e.g., FIFO versus average cost for inventory valuation). Prior to 2008, the US SEC required non-US companies whose securities trade in US markets to provide a reconciliation between their earnings from home-country accounting principles to US GAAP. This requirement not only assisted the analyst in making necessary adjustments but also provided some insight into appropriate adjustments for other companies not required to provide this data. In December 2007, however, the SEC eliminated the reconciliation requirement for non-US companies that use IFRS. Research analyzing reconciliations by EU companies with US listings shows that most of those companies reported net income under IFRS that was higher than they would have reported under US GAAP and lower shareholders' equity than they would have under US GAAP, with a result that more of the sample companies reported higher ROE under IFRS than under US GAAP.

In a study of companies filing such reconciliations to US GAAP, Harris and Muller (1999) classified common differences into seven categories, as shown in Exhibit 22.

Exhibit 21 Fundamental and Valuation Statistics by GICS Economic Sector: Median Values from S&P 1500, FY2007

	Valuation Statistics							Fundamental Statistics					
GICS Sector (count)	Trailing P/E	P/B	P/S	P/CF	Dividend Yield (%)	EV/ EBITDA	EV/S	Net Profit Margin (%)	Asset Turnover	Financial Leverage	ROE (%)	Dividend Payout Ratio (%)	3-Year CAGR Operating Margin (%)
Energy (85)	14.406	2.531	2.186	8.622	0.4	7.733	2.64	13.942	0.573	2.103	19.688	4.024	12.035
Materials (85)	15.343	2.254	0.888	9.588	1.4	7.686	1.095	5.568	0.995	2.465	15.728	17.874	4.157
Industrials (207)	17.275	2.578	1.045	11.642	1.0	8.979	1.209	6.089	1.139	2.143	15.262	16.066	5.337
Consumer Discretionary (279)	15.417	2.254	0.789	9.986	0.7	7.634	0.928	4.777	1.383	2.12	13.289	0	–2.682
Consumer Staples (80)	19.522	3.048	1.122	13.379	1.4	10.66	1.237	5.306	1.351	2.208	17.264	23.133	–0.88
Health Care (167)	23.027	3.088	2.061	15.762	0	11.623	2.274	6.637	0.83	1.854	12.399	0	–1.708
Financials (257)	14.648	1.559	1.888	11.186	3.1	9.482	4.017	13.113	0.113	5.848	10.348	41.691	–4.124
Information Technology (252)	20.205	2.444	2.162	45.073	0	11.594	1.811	7.929	0.743	1.587	10.444	0	1.524
Telecommunication Services (13)	19.585	2.485	1.527	5.266	0.8	6.681	2.345	7.109	0.471	2.367	5.43	6.862	–2.421
Utilities (75)	16.682	1.784	1.151	8.405	3.1	9.056	1.903	7.21	0.439	3.52	11.853	52.738	0.361
Overall (1,500)	17.148	2.246	1.398	11.328	0.8	9.108	1.626	7.318	0.839	2.227	12.701	8.051	0.181

Source: Standard & Poor's Research Insight.

Exhibit 22 Reconciliation of IFRS to US GAAP: Average Adjustment

Category	Earnings	Equity
Differences in the treatment of goodwill	Minus	Plus
Deferred income taxes	Plus	Plus
Foreign exchange adjustments	Plus	Minus
Research and development costs	Minus	Minus
Pension expense	Minus	Plus
Tangible asset revaluations	Plus	Minus
Other	Minus	Minus

In a more recent study of reconciliation data, Henry, Lin, and Yang (2009) found that among 20 categories of reconciliations, the most frequently occurring adjustments are in the pension category (including post-retirement benefits) and the largest value of adjustments are in the goodwill category.

Although the SEC's decision to eliminate the requirement for reconciliation has eliminated an important resource for analysts, accounting research can provide some insight into areas where differences between IFRS and US GAAP have commonly arisen. Going forward, analysts must be aware of differences between standards and make adjustments when disclosures provide sufficient data to do so.

International accounting differences affect the comparability of all price multiples. Of the price multiples we examined, P/CFO and P/FCFE will generally be least affected by accounting differences. P/B, P/E, and multiples based on such concepts as EBITDA, which start from accounting earnings, will generally be the most affected.

12 MOMENTUM VALUATION INDICATORS

p describe momentum indicators and their use in valuation;

The valuation indicators we call momentum indicators relate either price or a fundamental, such as earnings, to the time series of their own past values or, in some cases, to the fundamental's expected value. One style of growth investing uses positive momentum in various senses as a selection criterion, and practitioners sometimes refer to such strategies as "growth/momentum investment strategies." Momentum indicators based on price, such as the relative-strength indicator we will discuss here, have also been referred to as **technical indicators**. According to the BofA Merrill Lynch Institutional Factor Survey, various momentum indicators were used by many institutional investors. In this section, we review three representative momentum indicators: earnings surprise, standardized unexpected earnings, and relative strength.

To define standardized unexpected earnings, we define **unexpected earnings** (also called **earnings surprise**) as the difference between reported earnings and expected earnings:

$$UE_t = \text{EPS}_t - E(\text{EPS}_t),$$

where UE_t is the unexpected earnings for quarter t, EPS_t is the reported EPS for quarter t, and $E(\text{EPS}_t)$ is the expected EPS for the quarter.

For example, a stock with reported quarterly earnings of \$1.05 and expected earnings of \$1.00 would have a positive earnings surprise of \$0.05. Often, the percentage earnings surprise (i.e., earnings surprise divided by expected EPS) is reported by data providers; in this example, the percentage earning surprise would be $0.05/$1.00 = 0.05, or 5%. When used directly as a valuation indicator, earnings surprise is generally scaled by a measure reflecting the variability or range in analysts' EPS estimates. The principle is that the less disagreement among analysts' forecasts, the more meaningful the EPS forecast error of a given size in relation to the mean. A way to accomplish such scaling is to divide unexpected earnings by the standard deviation of analysts' earnings forecasts, which we refer to as the **scaled earnings surprise**. Example 37 illustrates the calculation of such a scaled earnings surprise.

EXAMPLE 37

Calculating Scaled Earnings Surprise by Using Analysts' Forecasts

During the third quarter of 2019, the mean consensus earnings forecast for BP plc for the fiscal year ending December 2019 was $3.26. Of the 11 estimates, the low forecast was $2.76, the high forecast was $3.74, and the standard deviation was $0.29. If actual reported earnings for 2019 come in equal to the high forecast, what would be the measure of the earnings surprise for BP scaled to reflect the dispersion in analysts' forecasts?

Solution:

In this case, scaled earnings surprise would be ($3.74 – $3.26)/$0.29 = $0.48/$0.29 = 1.66.

The rationale behind using earnings surprise is the thesis that positive surprises may be associated with persistent positive abnormal returns, or alpha. The same rationale lies behind a momentum indicator that is closely related to earnings surprise but more highly researched—namely, **standardized unexpected earnings** (SUE). The SUE measure is defined as

$$\mathrm{SUE}_t = \frac{\mathrm{EPS}_t - E(\mathrm{EPS}_t)}{\sigma[\mathrm{EPS}_t - E(\mathrm{EPS}_t)]},$$

where

EPS_t = Actual EPS for time t

$E(\mathrm{EPS}_t)$ = Expected EPS for time t

$\sigma[\mathrm{EPS}_t - E(\mathrm{EPS}_t)]$ = Standard deviation of $[\mathrm{EPS}_t - E(\mathrm{EPS}_t)]$ over some historical time period

In words, the numerator is the unexpected earnings at time t and the denominator is the standard deviation of past unexpected earnings over some period prior to time t—for example, the 20 quarters prior to t, as in Latané and Jones (1979), the article that introduced the SUE concept (for a summary of the research on SUE, see Brown 1997). In SUE, the magnitude of unexpected earnings is scaled by a measure of the size of historical forecast errors or surprises. The principle is that the smaller (larger) the historical size of forecast errors, the more (less) meaningful a given size of EPS forecast error.

Suppose that for a stock with a $0.05 earnings surprise, the standard deviation of past surprises is $0.20. The $0.05 surprise is relatively small compared with past forecast errors, which would be reflected in a SUE score of $0.05/$0.20 = 0.25. If the standard error of past surprises were smaller—say, $0.07—the SUE score would be $0.05/$0.07 = 0.71. Example 38 applies analysis of SUE to two companies.

EXAMPLE 38

Unexpected Earnings (Historical Example)

Exhibits 23 and 24 provide information about the earnings surprise history for two companies: Exxon Mobil Corporation and Volkswagen AG (VW).

Exhibit 23 Earnings Surprise History for Exxon Mobil Corporation (in US$)

Quarter Ending	EPS Release Date	Mean Consensus EPS Forecast	Actual EPS	% Surprise	Std. Dev.	SUE Score
Sep 2013	31 Oct 2013	1.77	1.79	0.88	0.1250	0.16
Jun 2013	1 Aug 2013	1.90	1.55	−18.39	0.0997	−3.51
Mar 2013	25 Apr 2013	2.05	2.12	3.59	0.0745	0.94
Dec 2012	1 Feb 2013	2.00	2.20	10.20	0.0463	4.32

Exhibit 24 Earnings Surprise History for Volkswagen AG (in Euros)

Quarter Ending	EPS Release Date	Mean Consensus EPS Forecast	Actual EPS	% Surprise	Std. Dev.	SUE Score
Sep 2013	30 Oct 2013	4.53	3.79	−16.37	0.2846	−2.60
Jun 2013	30 Jul 2013	5.10	5.86	14.99	0.3858	1.97
Mar 2013	24 Apr 2013	4.15	4.24	2.17	1.1250	0.08
Dec 2012	22 Feb 2013	5.56	3.54	−36.33	0.5658	−3.57

Source: Thomson Surprise Report.

1 Explain how Exxon's SUE score of 0.16 for the quarter ending September 2013 is calculated.

2 Based on these exhibits, for which company were the consensus forecasts less accurate over the past four quarters?

3 Was the consensus forecast more accurate for Exxon or VW for the quarter ending March 2013?

Solution to 1:

The amount of Exxon's unexpected earnings (i.e., its earnings surprise) for the quarter ending September 2013 was \$1.79 − \$1.77 = \$0.02. Dividing by the standard deviation of \$0.1250 gives a SUE score of 0.16.

Solution to 2:

The answer depends on whether accuracy is measured by the percentage surprise or by the SUE score. If accuracy is measured by the percentage surprise, then VW's consensus forecasts were less accurate: Percentage surprise varied from −36.33% to +14.99% for VW versus −18.39% to +10.20% for Exxon. Using SUE, Exxon's consensus forecasts were less accurate: SUE varied from −3.51 to +4.32

for Exxon versus −3.57 to +1.13 for VW. The reason for these differing results is that the standard deviation of the earnings estimates is relatively smaller for Exxon than it is for VW.

Solution to 3:

For the quarter ending March 2013, the consensus forecast was more accurate for VW than Exxon. Both the percentage surprise and SUE were lower for VW in this quarter.

Another set of indicators, **relative-strength indicators**, compares a stock's performance during a particular period either with its own past performance or with the performance of some group of stocks. The simplest relative-strength indicator that compares a stock's performance during a period with its past performance is the stock's compound rate of return over some specified time horizon, such as six months or one year. This indicator has also been referred to as **price momentum** in the academic literature. Despite its simplicity, this measure has been used in numerous studies. The rationale behind its use is the thesis that patterns of persistence or reversal exist in stock returns that may be shown empirically to depend on the investor's time horizon (Lee and Swaminathan 2000).

Other definitions of relative strength relate a stock's return over a recent period to its return over a longer period that includes the more recent period. For example, a classic study of technical momentum indicators (Brock, Lakonishok, and LeBaron 1992) examined trading strategies based on two technical rules—namely, a moving-average oscillator and a trading-range break (i.e., resistance and support levels)—in which buy and sell signals are determined by the relationship between a short period's moving average and a longer period's moving average (and bands around those averages). The reader should keep in mind that research on patterns of historical stock returns is notoriously vulnerable to data snooping and hindsight biases. Furthermore, investing strategies based purely on technical momentum indicators are viewed as inherently self-destructing, in that "once a useful technical rule (or price pattern) is discovered, it ought to be invalidated when the mass of traders attempts to exploit it" (Bodie, Kane, and Marcus 2008, p. 377). Yet, the possibility of discovering a profitable trading rule and exploiting it prior to mass use continues to motivate research.

A simple relative-strength indicator of the second type (i.e., the stock's performance relative to the performance of some group of stocks) is the stock's performance divided by the performance of an equity index. If the value of this ratio increases, the stock price increases relative to the index and displays positive relative strength. Often, the relative-strength indicator is scaled to 1.0 at the beginning of the study period. If the stock goes up at a higher (lower) rate than the index, then relative strength will be above (below) 1.0. Relative strength in this sense is often calculated for industries and individual stocks. Example 39 explores this indicator.

EXAMPLE 39

Relative Strength in Relation to an Equity Index

Exhibit 25 shows the values of the S&P 500 and three exchange-traded funds (ETFs) for the end of each of 18 months from March 2018 through August 2019. The ETFs are for long-term US Treasury securities, for the STOXX Europe 50 Index, and for emerging markets. SPDRs and iShares are families of exchange-traded funds managed by State Street Global Advisors and by Blackrock, Inc.

Exhibit 25 A Relative-Strength Comparison

First Day of	S&P 500 Index	iShares 20+ Year Treasury Bond ETF (TLT)	SPDR STOXX Europe 50 ETF (FEU)	iShares Emerging Markets ETF (EEM)
Mar-18	2,640.87	121.90	34.64	48.28
Apr-18	2,648.05	119.10	35.36	46.92
May-18	2,705.27	121.22	34.29	45.69
Jun-18	2,718.37	121.72	33.43	43.33
Jul-18	2,816.29	119.70	34.94	44.86
Aug-18	2,901.52	121.00	33.53	43.17
Sep-18	2,913.98	117.27	33.60	42.92
Oct-18	2,711.74	113.58	31.51	39.16
Nov-18	2,760.17	115.33	31.61	41.08
Dec-18	2,506.85	121.51	29.89	39.06
Jan-19	2,704.10	121.97	31.38	43.10
Feb-19	2,784.49	120.02	32.61	42.44
Mar-19	2,834.40	126.44	33.09	42.92
Apr-19	2,945.83	123.65	34.14	43.93
May-19	2,752.06	131.83	32.71	40.71
Jun-19	2,941.76	132.81	34.17	42.91
Jul-19	2,980.38	132.89	33.22	41.77
Aug-19	2,923.65	144.04	32.47	39.70

To produce the information for Exhibit 26, we divided each ETF value by the S&P 500 value for the same month and then scaled those results so that the value of the relative-strength indicator (RSTR) for March 2018 would equal 1.0. To illustrate, on 1 March 2018, the value of TLT divided by the S&P 500 was 121.90/2,640.87 = 0.04616. The RSTR for TLT on that date, by design, is then 0.04616/0.04616 = 1.0. In April, the value of TLT divided by the S&P 500 was 119.10/2,648.05 = 0.04498, which we scaled by the April number. The RSTR for 1 April 2018 for TLT is 0.04498/0.04616 = 0.9744, shown in Exhibit 26 as 0.974.

Exhibit 26 Relative-Strength Indicators

First Day of	RSTR iShares 20+ Year Treasury Bond ETF (TLT)	RSTR SPDR STOXX Europe 50 ETF (FEU)	RSTR iShares Emerging Markets ETF (EEM)
Mar-18	1.000	1.000	1.000
Apr-18	0.974	1.018	0.969
May-18	0.971	0.966	0.924
Jun-18	0.970	0.938	0.872
Jul-18	0.921	0.946	0.871
Aug-18	0.903	0.881	0.814
Sep-18	0.872	0.879	0.806
Oct-18	0.907	0.886	0.790

(continued)

Exhibit 26 (Continued)

First Day of	RSTR iShares 20+ Year Treasury Bond ETF (TLT)	RSTR SPDR STOXX Europe 50 ETF (FEU)	RSTR iShares Emerging Markets ETF (EEM)
Nov-18	0.905	0.873	0.814
Dec-18	1.050	0.909	0.852
Jan-19	0.977	0.885	0.872
Feb-19	0.934	0.893	0.834
Mar-19	0.966	0.890	0.828
Apr-19	0.909	0.884	0.816
May-19	1.038	0.906	0.809
Jun-19	0.978	0.886	0.798
Jul-19	0.966	0.850	0.767
Aug-19	1.067	0.847	0.743

On the basis of Exhibits 25 and 26, address the following:

1 State the relative strength of long-term US Treasury securities, the STOXX Europe 50 Index, and emerging market stocks over the entire time period March 2018 through August 2019. Interpret the relative strength for each sector over that period.

2 Discuss the relative performance of the STOXX Europe 50 Index ETF and the emerging market ETF in the month of December 2018.

Solution to 1:

The relative-strength indicator for long-term US Treasuries is 1.067. This number represents 1.067 − 1.000 = 0.067, or 6.7% overperformance relative to the S&P 500 over the time period. The relative-strength indicator for the STOXX Europe 50 Index is 0.995. This number represents 0.847 − 1.000 = −0.153, or 15.3% underperformance relative to the S&P 500 over the time period. The relative-strength indicator for the emerging market ETF is 0.743, indicating that it underperformed the S&P 500 by 25.7% over the time frame.

Solution to 2:

The December 2018 performance is found by comparing the RSTR at 1 December 2018 and 1 January 2019. The December 2019 RSTR for the STOXX Europe 50 Index ends at 0.885, which is 2.7% lower than its value for the prior month (0.909). The emerging market RSTR, at 0.872, is higher than the prior month value of 0.852 by 2.3%. In December 2018, the emerging market ETF outperformed the STOXX Europe 50 Index ETF. The relative performance for that one month differs from the relative performance over the entire period, during which the STOXX Europe 50 Index significantly outperformed the emerging market ETF.

Overall, momentum indicators have a substantial following among professional investors. Some view momentum indicators as signals that should prompt an analyst to consider whether a stock price is moving successively *farther from* or successively *closer to* the fundamental valuations derived from models and multiples. In other words, an analyst might be correct about the intrinsic value of a firm, and the momentum indicators might provide a clue about when the market price will converge with that intrinsic value. The use of such indicators continues to be a subject of active research in industry and in business schools.

VALUATION INDICATORS: ISSUES IN PRACTICE

13

q explain the use of the arithmetic mean, the harmonic mean, the weighted harmonic mean, and the median to describe the central tendency of a group of multiples;

All the valuation indicators discussed are quantitative aids but not necessarily solutions to the problem of security selection. In this section, we discuss some issues that arise in practice when averages are used to establish benchmark multiples and then illustrate the use of multiple valuation indicators.

13.1 Averaging Multiples: The Harmonic Mean

The harmonic mean and the weighted harmonic mean are often applied to average a group of price multiples.

Consider a hypothetical portfolio that contains two stocks. For simplicity, assume the portfolio owns 100% of the shares of each stock. One stock has a market capitalization of €715 million and earnings of €71.5 million, giving it a P/E of 10. The other stock has a market capitalization of €585 million and earnings of €29.25 million, for a P/E of 20. Note that the P/E for the portfolio is calculated directly by aggregating the companies' market capitalizations and earnings: (€715 + €585)/(€71.50 + €29.25) = €1,300/€100.75 = 12.90. The question that will be addressed is, What calculation of portfolio P/E, based on the individual stock P/Es, best reflects the value of 12.90?

If the ratio of an individual holding is represented by X_i, the expression for the simple **harmonic mean** of the ratio is

$$X_H = \frac{n}{\sum_{i=1}^{n}(1/X_i)}, \tag{9}$$

which is the reciprocal of the arithmetic mean of the reciprocals.

The expression for the **weighted harmonic mean** is

$$X_{WH} = \frac{1}{\sum_{i=1}^{n}(w_i/X_i)}, \tag{10}$$

where the w_i are portfolio value weights (summing to 1) and $X_i > 0$ for $i = 1, 2, \ldots, n$.

Exhibit 27 displays the calculation of the hypothetical portfolio's simple arithmetic mean P/E, weighted mean P/E, (simple) harmonic mean P/E, and weighted harmonic mean P/E.

Exhibit 27 Alternative Mean P/Es

	Market Cap		Earnings	Stock				
Security	(€ Millions)	Percent	(€ Millions)	P/E	(1)	(2)	(3)	(4)
Stock 1	715	55	71.50	10	0.5 × 10	0.55 × 10	0.5 × 0.1	0.55 × 0.1
Stock 2	585	45	29.25	20	0.5 × 20	0.45 × 20	0.5 × 0.05	0.45 × 0.05
					15	14.5	0.075	0.0775
Arithmetic mean P/E (1)					**15**			
Weighted mean P/E (2)						**14.5**		
Harmonic mean P/E (3)							1/0.075 = **13.33**	
Weighted harmonic mean P/E (4)								1/0.0775 = **12.90**

The weighted harmonic mean P/E precisely corresponds to the portfolio P/E value of 12.90. This example explains why index fund vendors frequently use the weighted harmonic mean to calculate the "average" P/E or average value of other price multiples for indexes. In some applications, an analyst might not want or be able to incorporate the market value weight information needed to calculate the weighted harmonic mean. In such cases, the simple harmonic mean can still be calculated.

Note that the simple harmonic mean P/E is smaller than the arithmetic mean and closer to the directly calculated value of 12.90 in this example. The harmonic mean inherently gives less weight to higher P/Es and more weight to lower P/Es. In general, unless all the observations in a data set have the same value, the harmonic mean is less than the arithmetic mean.

As explained and illustrated earlier, using the median rather than the arithmetic mean to derive an average multiple mitigates the effect of outliers. The harmonic mean is sometimes also used to reduce the impact of large outliers—which are typically the major concern in using the arithmetic mean multiple—but not the impact of small outliers (i.e., those close to zero). The harmonic mean tends to mitigate the impact of large outliers. The harmonic mean may aggravate the impact of small outliers, but such outliers are bounded by zero on the downside.

We can use the group of telecommunications companies examined earlier (see Exhibit 5) to illustrate differences between the arithmetic mean and the harmonic mean. This group includes two large outliers for P/E: CenturyLink, with a P/E that is not meaningful, and Charter Communications, with a P/E of 70.67. Exhibit 28 shows mean values excluding CenturyLink and excluding both CenturyLink and Charter Communications (two outliers).

Exhibit 28 Arithmetic versus Harmonic Mean

Company	Trailing P/E (without CenturyLink)	Trailing P/E (No Outliers)
AT&T	13.20	13.20
Comcast Corporation	16.23	16.23
CenturyLink	NMF	
China Telecom	13.14	13.14

Exhibit 28 (Continued)

Company	Trailing P/E (without CenturyLink)	Trailing P/E (No Outliers)
Charter Communications Corp.	70.67	
Verizon Communications	15.03	15.03
Windstream Holdings	24.55	24.55
Arithmetic mean	25.30	16.43
Median	15.23	15.03
Harmonic mean	17.70	15.39

Note that for the entire group, the arithmetic mean (25.30) is far higher than the median (15.23) because of the high P/E of Charter Communications (CenturyLink was not included). The harmonic mean (17.70) is much closer to the median and more plausible as representing central tendency. Once the outliers are eliminated, the values for the arithmetic mean (16.43), median (15.03), and harmonic mean (15.39) are more tightly grouped. The lower value for the harmonic mean reflects the fact that this approach mitigates the effect of the relatively high P/E for Charter Communications.

This example illustrates the importance for the analyst of understanding how an average has been calculated, particularly when the analyst is reviewing information prepared by another analyst, and the usefulness of examining several summary statistics.

13.2 Using Multiple Valuation Indicators

Because each carefully selected and calculated price multiple, momentum indicator, or fundamental may supply some piece of the puzzle of stock valuation, many investors and analysts use more than one valuation indicator (in addition to other criteria) in stock valuation and selection. Example 40 illustrates the use of multiple indicators.

EXAMPLE 40

Multiple Indicators in Stock Valuation

Analysts may use more valuation indicators than they describe in their company reports. The two following excerpts, adapted from past equity analyst reports, illustrate the use of multiple ratios in communicating views about a stock's value. In the first excerpt, from a report on Aussie Beverage Ltd. (ABEV), the analyst has used a discounted cash flow valuation as the preferred methodology but notes that the stock is also attractive when a price-to-earnings ratio (PER in the report) is used. In the second excerpt, from a report on Südliche Logistik (SLOG), an analyst evaluates the stock price (then trading at 42.80) by using two multiples, price to earnings (P/E) and EV/EBITDA, in relation to revised forecasts.

> **Aussie Beverage**
>
> Our DCF for ABEV is A$0.82ps, which represents a 44% prem. to the current price. Whilst the DCF valuation is our preferred methodology, we recognise that ABEV also looks attractive on different metrics.

Applying a mid-cycle PER multiple of 10.5 × (30% disc to mkt) to FY08 EPS of 7.6cps, we derive a valuation of A$0.80. Importantly, were the stock to reach our target of A$0.75ps in 12mths, ABEV would be trading on a fwd PER of 9.1×, which we do not view as demanding. At current levels, the stock is also offering an attractive dividend yield of 5.7% (fully franked). [*Note*: "Fully franked" is a concept specific to the Australian market and refers to tax treatment of the dividend.]

Südliche Logistik

Based on our slightly increased estimates, the shares are valued at a P/E and EV/EBITDA 2012 of 12.4x and 9x, slightly below the valuation of peer companies. Given its stronger profit growth, SLOG could command a premium. We raise our target price from EUR52 to EUR53, implying a 24% upside. Buy.

In selecting stocks, institutional investors surveyed in the BofA Merrill Lynch Institutional Factor Surveys from 1989 to 2012 used an average of 9.3 factors in selecting stocks (does not include 2008–2010 due to a lack of sufficient responses). The survey factors included not only price multiples, momentum indicators, and the DDM but also the fundamentals ROE, debt to equity, projected five-year EPS growth, EPS variability, EPS estimate dispersion, size, beta, foreign exposure, low price, and neglect. Exhibit 29 lists the factors classified by percentage of investors indicating that they use that factor in making investment decisions, out of 137 responders in 2012.

Exhibit 29 Frequency of Investor Usage of Factors in Making Investment Decisions

High (●) >50%; Med (♦) >30% <50%; Low (○) <30%

Factor	Frequency
P/E	●
Beta	●
EV/EBITDA	●
ROE	●
Size	●
P/B	●
P/FCF	♦
Share Repurchase	♦
Earnings Estimate Revision	♦
Margins	♦
Relative Strength	♦
EPS Momentum	♦
D/E	♦
EPS Variability	♦
DDM/DCF	♦
PEG Ratio	♦
Long-Term Price Trend	♦
P/CF	♦
Analyst Neglect	♦

Exhibit 29 (Continued)

Factor	High (●) >50%; Med (♦) >30% <50%; Low (○) <30% Frequency
Dividend Growth	♦
Projected 5-Year EPS Growth	♦
Mean Reversion	♦
Normalized P/E	♦
P/S	♦
Net Debt/EBITDA	○
EPS Surprise	○
ROC	○
ROA	○
EPS Estimate Dispersion	○
Analyst Rating Revisions	○
Foreign Exposure	○
Long-Term Price Trend w/ Short-Term Reversal	○
Trading Volume	○
Price Target	○
Ownership	○
Short-Term Price Trend	○
EV/Sales	○
Low Price	○
Altman Z-Score	○
Equity Duration	○

Source: 2012 BofA Merrill Lynch Institutional Factor Survey.

An issue concerning the use of ratios in an investing strategy is look-ahead bias. **Look-ahead bias** is the use of information that was not contemporaneously available in computing a quantity. Investment analysts often use historical data to back test an investment strategy that involves stock selection based on price multiples or other factors. When back testing, an analyst should be aware that time lags in the reporting of financial results create the potential for look-ahead bias in such research. For example, as of early January 2019, most companies had not reported EPS for the last quarter of 2018, so at that time, a company's trailing P/E would be based on EPS for the first, second, and third quarters of 2018 and the last quarter of 2017. Any investment strategy based on a trailing P/E that used actual EPS for the last quarter of 2018 could be implemented only after the data became available. Thus, if an analysis assumed that an investment was made in early January 2019 based on full-year 2018 data, the analysis would involve look-ahead bias. To avoid this bias, an analyst would calculate the trailing P/E based on the most recent four quarters of EPS then being reported. The same principle applies to other multiples calculated on a trailing basis.

The application of a set of criteria to reduce an investment universe to a smaller set of investments is called **screening**. Stock screens often include not only criteria based on the valuation measures that featured in our discussion but also on fundamental criteria that may explain differences in such measures. Computerized stock screening is an efficient way to narrow a search for investments and is a part of many stock selection disciplines. The limitations to many commercial databases and screening tools usually include lack of control by the user of the calculation of important inputs (such as EPS); the absence of qualitative factors in most databases is another important limitation. Example 41 illustrates the use of a screen in stock selection.

EXAMPLE 41

Using Screens to Find Stocks for a Portfolio

Janet Larsen manages an institutional portfolio and is currently looking for new stocks to add to the portfolio. Larsen has a commercial database with information on US stocks. She has designed several screens to select stocks with low P/Es and low P/B multiples. Because Larsen is aware that screening for low P/E and low P/B multiples may identify stocks with low expected growth, she also wants stocks that have a PEG ratio less than 1.0. She decides to screen for stocks with a dividend yield of at least 3.0% and a total market capitalization over $10 billion. Exhibit 30 shows the number of stocks that successively met each of the five criteria as of 17 July 2019 (so, the number of stocks that met all five criteria is 10).

Exhibit 30 Stock Screen

Criterion	Stocks Meeting Each Criterion Successively
P/E < 20.0	2,096
P/B < 2.0	1,384
PEG ratio < 1.0	89
Dividend yield ≥ 3.0%	23
Market capitalization over $10 billion	10

Other information:

- The screening database indicates that the trailing P/E was 22.3, P/B was 3.5, and the dividend yield was 1.9% for the S&P 500 as of the date of the screen.
- The "S&P U.S. Style Indices Methodology" (June 2019) indicates that the style indexes measure growth and value by the following six factors, which S&P standardizes and uses to compute growth and value scores for each company:

    Three Growth Factors

    Three-year change in EPS over price per share

    Three-year sales per-share growth rate

    Momentum (12-month percentage price change)

    Three Value Factors

    Book value-to-price ratio

    Earnings-to-price ratio

    Sales-to-price ratio

- In February of 2019, the S&P Dow Jones US Index Committee raised the market cap guidelines used when selecting companies for the S&P 500, S&P MidCap 400 and S&P SmallCap 600. The new guidelines are as follows:

  S&P 500: Over $8.2 billion

  S&P MidCap 400: $2.4 billion to $8.2 billion

  S&P SmallCap 600: $600 million to $2.4 billion

Using the information supplied, answer the following questions:

1. What type of valuation indicators does Larsen *not* include in her stock screen?
2. Characterize the overall orientation of Larsen as to investment style.
3. State two limitations of Larsen's stock screen.

Solution to 1:

Larsen has not included momentum indicators in the screen.

Solution to 2:

Larsen can be characterized as a large-cap value investor, based on the specified market capitalization. Although her screen does include a PEG ratio, it excludes explicit growth rate criteria, such as those used by S&P, and it excludes momentum indicators usually associated with a growth orientation, such as positive earnings surprise. Larsen also uses a cutoff for P/B that is less than the average P/B for the S&P 500. Note that her criteria for multiples are all "less than" criteria.

Solution to 3:

Larsen does not include any profitability criteria or risk measurements. These omissions are a limitation because a stock's expected low profitability or high risk may explain its low P/E. Another limitation of her screen is that the computations of the value indicators in a commercial database may not reflect the appropriate adjustments to inputs. The absence of qualitative criteria is also a possible limitation.

Investors also apply all the metrics that we have illustrated in terms of individual stocks to industries and economic sectors. For example, average price multiples and momentum indicators can be used in sector rotation strategies to determine relatively under- or overvalued sectors. A sector rotation strategy is an investment strategy that overweights economic sectors that are anticipated to outperform or lead the overall market.

SUMMARY

We have defined and explained the most important valuation indicators in professional use and illustrated their application to a variety of valuation problems.

- Price multiples are ratios of a stock's price to some measure of value per share.
- Price multiples are most frequently applied to valuation in the method of comparables. This method involves using a price multiple to evaluate whether an asset is relatively undervalued, fairly valued, or overvalued in relation to a benchmark value of the multiple.

- The benchmark value of the multiple may be the multiple of a similar company or the median or average value of the multiple for a peer group of companies, an industry, an economic sector, an equity index, or the company's own median or average past values of the multiple.
- The economic rationale for the method of comparables is the law of one price.
- Price multiples may also be applied to valuation in the method based on forecasted fundamentals. Discounted cash flow (DCF) models provide the basis and rationale for this method. Fundamentals also interest analysts who use the method of comparables because differences between a price multiple and its benchmark value may be explained by differences in fundamentals.
- The key idea behind the use of price-to-earnings ratios (P/Es) is that earning power is a chief driver of investment value and earnings per share (EPS) is probably the primary focus of security analysts' attention. The EPS figure, however, is frequently subject to distortion, often volatile, and sometimes negative.
- The two alternative definitions of P/E are trailing P/E, based on the most recent four quarters of EPS, and forward P/E, based on next year's expected earnings.
- Analysts address the problem of cyclicality by normalizing EPS—that is, calculating the level of EPS that the business could achieve currently under mid-cyclical conditions (normalized EPS).
- Two methods to normalize EPS are the method of historical average EPS (calculated over the most recent full cycle) and the method of average return on equity (EPS = average ROE multiplied by current book value per share).
- Earnings yield (E/P) is the reciprocal of the P/E. When stocks have zero or negative EPS, a ranking by earnings yield is meaningful whereas a ranking by P/E is not.
- Historical trailing P/Es should be calculated with EPS lagged a sufficient amount of time to avoid look-ahead bias. The same principle applies to other multiples calculated on a trailing basis.
- The fundamental drivers of P/E are the expected earnings growth rate and the required rate of return. The justified P/E based on fundamentals bears a positive relationship to the first factor and an inverse relationship to the second factor.
- The PEG (P/E-to-growth) ratio is a tool to incorporate the impact of earnings growth on P/E. The PEG ratio is calculated as the ratio of the P/E to the consensus growth forecast. Stocks with low PEG ratios are, all else equal, more attractive than stocks with high PEG ratios.
- We can estimate terminal value in multistage DCF models by using price multiples based on comparables. The expression for terminal value, V_n, is (using P/E as the example)

  $$V_n = \text{Benchmark value of trailing P/E} \times E_n$$

  or

  $$V_n = \text{Benchmark value of forward P/E} \times E_{n+1}.$$

- Book value per share is intended to represent, on a per-share basis, the investment that common shareholders have in the company. Inflation, technological change, and accounting distortions, however, may impair the use of book value for this purpose.
- Book value is calculated as common shareholders' equity divided by the number of shares outstanding. Analysts adjust book value to accurately reflect the value of the shareholders' investment and to make P/B (the price-to-book ratio) more useful for comparing different stocks.

- The fundamental drivers of P/B are ROE and the required rate of return. The justified P/B based on fundamentals bears a positive relationship to the first factor and an inverse relationship to the second factor.
- An important rationale for using the price-to-sales ratio (P/S) is that sales, as the top line in an income statement, are generally less subject to distortion or manipulation than other fundamentals, such as EPS or book value. Sales are also more stable than earnings and are never negative.
- P/S fails to take into account differences in cost structure between businesses, may not properly reflect the situation of companies losing money, and may be subject to manipulation through revenue recognition practices.
- The fundamental drivers of P/S are profit margin, growth rate, and the required rate of return. The justified P/S based on fundamentals bears a positive relationship to the first two factors and an inverse relationship to the third factor.
- Enterprise value (EV) is total company value (the market value of debt, common equity, and preferred equity) minus the value of cash and investments.
- The ratio of EV to total sales is conceptually preferable to P/S because EV/S facilitates comparisons among companies with varying capital structures.
- A key idea behind the use of price to cash flow is that cash flow is less subject to manipulation than are earnings. Price-to-cash-flow multiples are often more stable than P/Es. Some common approximations to cash flow from operations have limitations, however, because they ignore items that may be subject to manipulation.
- The major cash flow (and related) concepts used in multiples are earnings plus noncash charges (CF), cash flow from operations (CFO), free cash flow to equity (FCFE), and earnings before interest, taxes, depreciation, and amortization (EBITDA).
- In calculating price to cash flow, the earnings-plus-noncash-charges concept is traditionally used, although FCFE has the strongest link to financial theory.
- CF and EBITDA are not strictly cash flow numbers because they do not account for noncash revenue and net changes in working capital.
- The fundamental drivers of price to cash flow, however defined, are the expected growth rate of future cash flow and the required rate of return. The justified price to cash flow based on fundamentals bears a positive relationship to the first factor and an inverse relationship to the second.
- EV/EBITDA is preferred to P/EBITDA because EBITDA, as a pre-interest number, is a flow to all providers of capital.
- EV/EBITDA may be more appropriate than P/E for comparing companies with different amounts of financial leverage (debt).
- EV/EBITDA is frequently used in the valuation of capital-intensive businesses.
- The fundamental drivers of EV/EBITDA are the expected growth rate in free cash flow to the firm, profitability, and the weighted average cost of capital. The justified EV/EBITDA based on fundamentals bears a positive relationship to the first two factors and an inverse relationship to the third.
- Dividend yield has been used as a valuation indicator because it is a component of total return and is less risky than capital appreciation.
- Trailing dividend yield is calculated as four times the most recent quarterly per-share dividend divided by the current market price.
- The fundamental drivers of dividend yield are the expected growth rate in dividends and the required rate of return.

- Comparing companies across borders frequently involves dealing with differences in accounting standards, cultural differences, economic differences, and resulting differences in risk and growth opportunities.
- Momentum indicators relate either price or a fundamental to the time series of the price's or fundamental's own past values (in some cases, to their expected values).
- Momentum valuation indicators include earnings surprise, standardized unexpected earnings (SUE), and relative strength.
- Unexpected earnings (or earnings surprise) equals the difference between reported earnings and expected earnings.
- SUE is unexpected earnings divided by the standard deviation in past unexpected earnings.
- Relative-strength indicators allow comparison of a stock's performance during a period either with its own past performance (first type) or with the performance of some group of stocks (second type). The rationale for using relative strength is the thesis that patterns of persistence or reversal in returns exist.
- Screening is the application of a set of criteria to reduce an investment universe to a smaller set of investments and is a part of many stock selection disciplines. In general, limitations of such screens include the lack of control in vendor-provided data of the calculation of important inputs and the absence of qualitative factors.

REFERENCES

Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal* 59 (1): 70–87.

Aubert, Samuel, and Pierre Giot. 2007. "An International Test of the Fed Model." *Journal of Asset Management* 8 (2): 86–100.

Bodie, Zvi, Alex Kane, and Alan J. Marcus. 2008. *Investments*. 7th ed. New York: McGraw-Hill.

Brock, William, Joseph Lakonishok, and Blake LeBaron. 1992. "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns." *Journal of Finance* 47 (5): 1731–64.

Brown, Lawrence D. 1997. "Earnings Surprise Research: Synthesis and Perspectives." *Financial Analysts Journal* 53 (2): 13–20.

Chan, L. K. C., and J. Lakonishok. 2004. "Value and Growth Investing: Review and Update." *Financial Analysts Journal* 60 (1): 71–86.

Damodaran, Aswath. 2012. *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset*. 3rd ed. Hoboken, NJ: John Wiley & Co.

Fairfield, Patricia M. 1994. "P/E, P/B and the Present Value of Future Dividends." *Financial Analysts Journal* 50 (4): 23–31.

Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47 (2): 427–65.

Graham, Benjamin, and David L. Dodd. 1934. *Security Analysis*. McGraw-Hill Professional Publishing.

Grant, Julia, and Larry Parker. 2001. "EBITDA!" *Research in Accounting Regulation* 15:205–11.

Hackel, Kenneth S., Joshua Livnat, and Atul Rai. 1994. "The Free Cash Flow/Small-Cap Anomaly." *Financial Analysts Journal* 50 (5): 33–42.

Harris, Robert S., and Felicia C. Marston. 1994. "Value versus Growth Stocks: Book-to-Market, Growth, and Beta." *Financial Analysts Journal* 50 (5): 18–24.

Harris, Mary, and Karl A. Muller III. 1999. "The Market Valuation of IAS versus U.S. GAAP Accounting Measures Using Form 20-F Reconciliations." *Journal of Accounting and Economics* 26 (1–3): 285–312.

Henry, E., S. Lin, and Y. Yang. 2009. "The European–U.S. GAAP Gap: Amount, Type, Homogeneity, and Value Relevance of IFRS to U.S. GAAP Form 20-F Reconciliations." *Accounting Horizons* 23 (2): 121–50.

Kisor, Manown, Jr, and Volkert S. Whitbeck. 1963. "A New Tool in Investment Decision-Making." *Financial Analysts Journal* 19 (3): 55–62.

Lakonishok, J., A. Shleifer, and R. W. Vishny. 1994. "Contrarian Investment, Extrapolation and Risk." *Journal of Finance* 49 (5): 1541–78.

Lander, Joel, Athanasios Orphanides, and Martha Douvogiannis. 1997. "Earnings Forecasts and the Predictability of Stock Returns: Evidence from Trading the S&P." *Journal of Portfolio Management* 23 (4): 24–35.

Latané, Henry A., and Charles P. Jones. 1979. "Standardized Unexpected Earnings—1971–77." *Journal of Finance* 34 (3): 717–24.

Lee, Charles M.C., and Bhaskaran Swaminathan. 2000. "Price Momentum and Trading Volume." *Journal of Finance* 55 (5): 2017–69.

Malkiel, Burton, and John Cragg. 1970. "Expectations and the Structure of Share Prices." *American Economic Review* 60 (4): 601–17.

Martin, Thomas A., Jr 1998. "Traditional Equity Valuation Methods." In *Equity Research and Valuation Techniques*. Charlottesville, VA: AIMR.

Moody's. 2000. *Putting EBITDA in Perspective*. Moody's Investors Service Global Credit Research.

Mulford, Charles W., and Eugene E. Comiskey. 2005. *Creative Cash Flow Reporting: Uncovering Sustainable Financial Performance*. Hoboken, NJ: John Wiley & Sons.

Nathan, Siva, Kumar Sivakumar, and Jayaraman Vijayakumar. 2001. "Returns to Trading Strategies Based on Price-to-Earnings and Price-to-Sales Ratios." *Journal of Investing* 10 (2): 17–28.

O'Shaughnessy, James P. 2005. *What Works on Wall Street: A Guide to the Best-Performing Investment Strategies of All Time*. 3rd ed. New York: McGraw-Hill Professional Publishing.

Pinto, Jerald, Thomas Robinson, and John Stowe. 2018. "Equity Valuation: A Survey of Professional Practice." *Review of Financial Economics* 37:219–33.

Siegel, Jeremy. 2014. *Stocks for the Long Run*. 5th ed. New York: McGraw-Hill.

Solnik, Bruno, and Dennis McLeavey. 2004. *International Investments*. 5th ed. Boston: Pearson Addison-Wesley.

Stimes, Peter C., and Stephen E. Wilcox. 2011. "Equity Market Valuation." Chap. 11 in *Investments: Principles of Portfolio and Equity Analysis*. Hoboken, NJ: John Wiley & Sons.

Wild, John J., Leopold A. Bernstein, and K. R. Subramanyam. 2001. *Financial Statement Analysis*. 7th ed. New York: McGraw-Hill Irwin.

Yardeni, Edward. 2000. *"How to Value Earnings Growth." Topical Study #49*. Deutsche Banc Alex Brown.

Zhou, Ping, and William Ruland. 2006. "Dividend Payout and Future Earnings Growth." *Financial Analysts Journal* 62 (3): 58–69.

PRACTICE PROBLEMS

1 As of February 2020, you are researching Jonash International, a hypothetical company subject to cyclical demand for its services. Jonash shares closed at $57.98 on 2 February 2019. You believe the 2015–18 period reasonably captures average profitability:

Measure	2019	2018	2017	2016	2015
EPS	E$3.03	$1.45	$0.23	$2.13	$2.55
BV per share	E$19.20	$16.21	$14.52	$13.17	$11.84
ROE	E16.0%	8.9%	1.6%	16.3%	21.8%

A Define normalized EPS.

B Calculate a normalized EPS for Jonash based on the method of historical average EPS, and then calculate the P/E based on normalized EPS.

C Calculate a normalized EPS for Jonash based on the method of average ROE and the P/E based on normalized EPS.

2 An analyst plans to use P/E and the method of comparables as a basis for recommending purchasing shares of one of two peer-group companies in the business of manufacturing personal digital assistants. Neither company has been profitable to date, and neither is expected to have positive EPS over the next year. Data on the companies' prices, trailing EPS, and expected growth rates in sales (five-year compounded rates) are given in the following table:

Company	Price	Trailing EPS	P/E	Expected Growth (Sales)
Hand	$22	–$2.20	NMF	45%
Somersault	$10	–$1.25	NMF	40%

Unfortunately, because the earnings for both companies have been negative, their P/Es are not meaningful. On the basis of this information, address the following:

A Discuss how the analyst might make a relative valuation in this case.

B State which stock the analyst should recommend.

3 May Stewart, CFA, a retail analyst, is performing a P/E-based comparison of two hypothetical jewelry stores as of early 2020. She has the following data for Hallwhite Stores (HS) and Ruffany (RUF).

- HS is priced at $44. RUF is priced at $22.50.
- HS has a simple capital structure, earned $2.00 per share (basic and diluted) in 2019, and is expected to earn $2.20 (basic and diluted) in 2020.
- RUF has a complex capital structure as a result of its outstanding stock options. Moreover, it had several unusual items that reduced its basic EPS in 2019 to $0.50 (versus the $0.75 that it earned in 2018).
- For 2020, Stewart expects RUF to achieve net income of $30 million. RUF has 30 million shares outstanding and options outstanding for an additional 33,333,333 shares.

A Which P/E (trailing or forward) should Stewart use to compare the two companies' valuation?

© 2016 CFA Institute. All rights reserved.

B Which of the two stocks is relatively more attractive when valued on the basis of P/Es (assuming that all other factors are approximately the same for both stocks)?

4 You are researching the valuation of the stock of a company in the food-processing industry. Suppose you intend to use the mean value of the forward P/Es for the food-processing industry stocks as the benchmark value of the multiple. This mean P/E is 18.0. The forward or expected EPS for the next year for the stock you are studying is $2.00. You calculate 18.0 × $2.00 = $36, which you take to be the intrinsic value of the stock based only on the information given here. Comparing $36 with the stock's current market price of $30, you conclude the stock is undervalued.

A Give two reasons why your conclusion that the stock is undervalued may be in error.

B What additional information about the stock and the peer group would support your original conclusion?

5 Suppose an analyst uses an equity index as a comparison asset in valuing a stock. In making a decision to recommend purchase of an individual stock, which price multiple(s) would cause concern about the impact of potential overvaluation of the equity index?

6 Christie Johnson, CFA, has been assigned to analyze Sundanci. Johnson assumes that Sundanci's earnings and dividends will grow at a constant rate of 13%. Exhibits 1 and 2 provide financial statements for the most recent two years (2020 and 2021) and other information for Sundanci.

Exhibit 1 Sundanci Actual 2020 and 2021 Financial Statements for Fiscal Years Ending 31 May (in Millions except Per-Share Data)

Income Statement	**2020**	**2021**
Revenue	$474	$598
Depreciation	20	23
Other operating costs	368	460
Income before taxes	86	115
Taxes	26	35
Net income	60	80
Dividends	18	24
Earnings per share	$0.714	$0.952
Dividends per share	$0.214	$0.286
Common shares outstanding	84.0	84.0
Balance Sheet	**2020**	**2021**
Current assets	$201	$326
Net property, plant, and equipment	474	489
Total assets	675	815
Current liabilities	57	141
Long-term debt	0	0

(continued)

Exhibit 1 (Continued)

Income Statement	2020	2021
Total liabilities	57	141
Shareholders' equity	618	674
Total liabilities and equity	675	815
Other Information		
Capital expenditures	34	38

Exhibit 2 Selected Financial Information

Required rate of ROE	14%
Growth rate of industry	13%
Industry P/E	26

A Based on information in Exhibits 1 and 2 and on Johnson's assumptions for Sundanci, calculate justified trailing and forward P/Es for this company.

B Identify, within the context of the constant dividend growth model, how *each* of the following fundamental factors would affect the P/E:

i. The risk (beta) of Sundanci increases substantially.

ii. The estimated growth rate of Sundanci's earnings and dividends increases.

iii. The equity risk premium increases.

Note: A change in a fundamental factor is assumed to happen in isolation; interactive effects between factors are ignored. That is, every other item of the company is unchanged.

7 Tom Smithfield is valuing the stock of a food-processing business. He feels confident explicitly projecting earnings and dividends to three years (to $t = 3$). Other information and estimates are as follows:

- Required rate of return = 0.09.
- Average dividend payout rate for mature companies in the market = 0.45.
- Industry average ROE = 0.10.
- E_3 = \$3.00.
- Industry average P/E = 12.

On the basis of this information, answer the following questions:

A Compute terminal value (V_3) based on comparables.

B Contrast your answer in Part A to an estimate of terminal value based on the Gordon growth model.

8 Discuss three types of stocks or investment situations for which an analyst could appropriately use P/B in valuation.

9 Aratatech is a multinational distributor of semiconductor chips and related products to businesses. Its leading competitor around the world is Trymye Electronics. Aratatech has a current market price of $10.00, 20 million shares outstanding, annual sales of $1 billion, and a 5% profit margin. Trymye has a market price of $20.00, 30 million shares outstanding, annual sales of $1.6 billion, and a profit margin of 4.9%. Based on the information given, answer the following questions:

A Which of the two companies has a more attractive valuation based on P/S?

B Identify and explain one advantage of P/S over P/E as a valuation tool.

10 Wilhelm Müller, CFA, has organized the selected data on four food companies that appear below (TTM stands for trailing 12-month):

Measure	Hoppelli Foods	Telli Foods	Drisket Co.	Whiteline Foods
Stock price	€25.70	€11.77	€23.65	€24.61
Shares outstanding (thousands)	138,923	220,662	108,170	103,803
Market cap (€ millions)	3,570	2,597	2,558	2,555
Enterprise value (€ millions)	3,779	4,056	3,846	4,258
Sales (€ millions)	4,124	10,751	17,388	6,354
Operating income (€ millions)	285	135	186	396
Operating profit margin	6.91%	1.26%	1.07%	6.23%
Net income (€ millions)	182	88	122	252
TTM EPS	€1.30	€0.40	€1.14	€2.43
Return on equity	19.20%	4.10%	6.40%	23.00%
Net profit margin	4.41%	0.82%	0.70%	3.97%

On the basis of the data given, answer the following questions:

A Calculate the trailing P/E and EV/sales for each company.

B Explain, on the basis of fundamentals, why these stocks have different EV/S multiples.

11 John Jones, CFA, is head of the research department at Peninsular Research. Peninsular has a client who has inquired about the valuation method best suited for comparing companies in an industry with the following characteristics:

- Principal competitors within the industry are located in the United States, France, Japan, and Brazil.
- The industry is currently operating at a cyclical low, with many companies reporting losses.

Jones recommends that the client consider the following valuation ratios:

1 P/E

2 P/B

3 EV/S

Determine which *one* of the three valuation ratios is most appropriate for comparing companies in this industry. Support your answer with *one* reason that makes that ratio superior to either of the other two ratios in this case.

12 GN Growing AG (GG) is currently selling for €240, with TTM EPS and dividends per share of €1.5 and €0.9, respectively. The company's trailing P/E is 16.0, P/B is 3.2. P/Sales based on forecast sales, is 1.5. ROE is 20%, and for the profit margin on sales is 10.0%. The Treasury bond rate is 4.9%, the equity risk premium is 5.5%, and GG's beta is 1.2.

A What is GG's required rate of return, based on the capital asset pricing model (CAPM)?

B Assume that the dividend and earnings growth rates are 8%. What trailing P/E and P/B multiples would be justified in light of the required rate of return in Part A and current values of the dividend payout ratio and ROE ?

C Calculate the justified P/Sales ratio based on the forward-looking margin of 10% and current values of dividend payout.

D Given that the assumptions and constant growth model are appropriate, state and justify whether GG, based on fundamentals, appears to be fairly valued, overvalued, or undervalued.

13 Jorge Zaldys, CFA, is researching the relative valuation of two companies in the aerospace/defense industry, NCI Heavy Industries (NCI) and Relay Group International (RGI). He has gathered relevant information on the companies in the following table.

EBITDA Comparisons (in € Millions except Per-Share and Share-Count Data)

Company	RGI	NCI
Price per share	150	100
Shares outstanding	5 million	2 million
Market value of debt	50	100
Book value of debt	52	112
Cash and investments	5	2
Net income	49.5	12
Net income from continuing operations	49.5	8
Interest expense	3	5
Depreciation and amortization	8	4
Taxes	2	3

Using the information in the table, answer the following questions:

A Calculate P/EBITDA for NCI and RGI.

B Calculate EV/EBITDA for NCI and RGI.

C Which company should Zaldys recommend as relatively undervalued? Justify the selection.

14 Define the major alternative cash flow concepts, and state one limitation of each.

15 Data for two hypothetical companies in the pharmaceutical industry, DriveMed and MAT Technology, are given in the following table. For both companies, expenditures on fixed capital and working capital during the previous year reflect anticipated average expenditures over the foreseeable horizon.

Measure	DriveMed	MAT Technology
Current price	$46.00	$78.00
Trailing CF per share	$3.60	$6.00
P/CF	12.8	13.0
Trailing FCFE per share	$1.00	$5.00
P/FCFE	46.0	15.6
Consensus five-year growth forecast	15%	20%
Beta	1.25	1.25

On the basis of the information supplied, discuss the valuation of MAT Technology relative to DriveMed. Justify your conclusion.

16 Your value-oriented investment management firm recently hired a new analyst, Bob Westard, because of his expertise in the life sciences and biotechnology areas. At the firm's weekly meeting, during which each analyst proposes a stock idea for inclusion in the firm's approved list, Westard recommends Hitech Clothing International (HCI). He bases his recommendation on two considerations. First, HCI has pending patent applications but a P/E that he judges to be low in light of the potential earnings from the patented products. Second, HCI has had high relative strength versus the S&P 500 over the past month.

A Explain the difference between Westard's two approaches—that is, the use of price multiples and the relative-strength approach.

B State which, if any, of the bases for Westard's recommendation is consistent with the investment orientation of your firm.

17 Kirstin Kruse, a portfolio manager, has an important client who wants to alter the composition of her equity portfolio, which is currently a diversified portfolio of 60 global common stocks. Because of concerns about the economy and based on the thesis that the consumer staples sector will be less hurt than others in a recession, the client wants to add stocks trading in the United States (including ADRs) from the consumer staples sector. In addition, the client wants the stocks to meet the following criteria:

- Stocks must be considered large cap (i.e., have a large market capitalization).
- Stocks must have a dividend yield of at least 4.0%.
- Stocks must have a forward P/E no greater than 15.

The following table shows how many stocks satisfied each screen, which was run in June 2019.

Screen	Number Satisfying
Consumer staples sector	424
Large cap	361
Dividend yield of at least 4.0%	887
P/E less than 15	5,409
All four screens	3

The stocks meeting all four screens were Altria Group, Inc.; British American Tobacco PLC (the company's ADR); and Kraft Heinz Co.

A Critique the construction of the screen.

B Do these criteria identify appropriate additions to this client's portfolio?

Questions 18–24 relate to Mark Cannan

Mark Cannan is updating research reports on two well-established consumer companies before first quarter 2021 earnings reports are released. His supervisor, Sharolyn Ritter, has asked Cannan to use market-based valuations when updating the reports.

Delite Beverage is a manufacturer and distributor of soft drinks and recently acquired a major water bottling company in order to offer a broader product line. The acquisition will have a significant impact on Delite's future results.

You Fix It is a US retail distributor of products for home improvement, primarily for those consumers who choose to do the work themselves. The home improvement industry is cyclical; the industry was adversely affected by the recent downturn in the economy, the level of foreclosures, and slow home sales. Although sales and earnings at You Fix It weakened, same store sales are beginning to improve as consumers undertake more home improvement projects. Poor performing stores were closed, resulting in significant restructuring charges in 2020.

Before approving Cannan's work, Ritter wants to discuss the calculations and choices of ratios used in the valuation of Delite and You Fix It. The data used by Cannan in his analysis are summarized in Exhibit 1.

Exhibit 1 Select Financial Data for Delite Beverage and You Fix It

	Delite Beverage	You Fix It
2020 earnings per share (EPS)	\$3.44	\$1.77
2021 estimated EPS	\$3.50	\$1.99
Book value per share end of year	\$62.05	\$11.64
Current share price	\$65.50	\$37.23
Sales (billions)	\$32.13	\$67.44
Free cash flow per share	\$2.68	\$0.21
Shares outstanding end of year	2,322,034,000	1,638,821,000

Cannan advises Ritter that he is considering three different approaches to value the shares of You Fix It:

Approach 1 Price-to-book ratio (P/B)

Approach 2 Price-to-earnings ratio (P/E) using trailing earnings

Approach 3 Price-to-earnings ratio using normalized earnings

Cannan tells Ritter that he calculated the price-to-sales ratio (P/S) for You Fix It but chose not to use it in the valuation of the shares. Cannan states to Ritter that it is more appropriate to use the P/E than the P/S because

Reason 1 Earnings are more stable than sales.

Reason 2 Earnings are less easily manipulated than sales.

Reason 3 The P/E reflects financial leverage, whereas the P/S does not.

Cannan also informs Ritter that he did not use a price-to-cash-flow multiple in valuing the shares of Delite or You Fix It. The reason is that he could not identify a cash flow measure that would both account for working capital and noncash revenues and be after interest expense and thus not be mismatched with share price. Ritter advises Cannan that such a cash flow measure does exist.

Ritter provides Cannan with financial data on three close competitors as well as the overall beverage sector, which includes other competitors, in Exhibit 2. She asks Cannan to determine, based on the P/E-to-growth (PEG) ratio, whether Delite shares are overvalued, fairly valued, or undervalued.

Exhibit 2 Beverage Sector Data

	Forward P/E	Earnings Growth
Delite	—	12.41%
Fresh Iced Tea Company	16.59	9.52%
Nonutter Soda	15.64	11.94%
Tasty Root Beer	44.10	20%
Beverage sector average	16.40	10.80%

After providing Ritter his answer, Cannan is concerned about the inclusion of Tasty Root Beer in the comparables analysis. Specifically, Cannan says to Ritter:

> "I feel we should mitigate the effect of large outliers but not the impact of small outliers (i.e., those close to zero) when calculating the beverage sector P/E. What measure of central tendency would you suggest we use to address this concern?"

Ritter requests that Cannan incorporate their discussion points before submitting the reports for final approval.

18 Based on the information in Exhibit 1, the *most appropriate* price-to-earnings ratio to use in the valuation of Delite is *closest* to:

- **A** 18.71.
- **B** 19.04.
- **C** 24.44.

19 Based on the information in Exhibit 1, the price-to-sales ratio for You Fix It is *closest* to:

- **A** 0.28.
- **B** 0.55.
- **C** 0.90.

20 Which valuation approach would be *most* appropriate in valuing shares of You Fix It?

- **A** Approach 1
- **B** Approach 2
- **C** Approach 3

21 Cannan's preference to use the P/E over the P/S is *best* supported by:

- **A** Reason 1.
- **B** Reason 2.
- **C** Reason 3.

22 The cash flow measure that Ritter would *most likely* recommend to address Cannan's concern is:

- **A** free cash flow to equity.
- **B** earnings plus noncash charges.

C earnings before interest, tax, depreciation, and amortization.

23 Based on the information in Exhibits 1 and 2, Cannan would most likely conclude that Delite's shares are:

A overvalued.

B undervalued.

C fairly valued.

24 The measure of central tendency that Ritter will *most likely* recommend is the:

A median.

B harmonic mean.

C arithmetic mean.

The following information relates to Questions 25–30

Andrea Risso is a junior analyst with AquistareFianco, an independent equity research firm. Risso's supervisor asks her to update, as of 1 January 2020, a quarterly research report for Centralino S.p.A., a telecommunications company headquartered in Italy. On that date, Centralino's common share price is €50 and its preferred shares trade for €5.25 per share.

Risso gathers information on Centralino. Exhibit 1 presents earnings and dividend data, and Exhibit 2 presents balance sheet data. Net sales were €3.182 billion in 2019. Risso estimates a required return of 15% for Centralino and forecasts growth in dividends of 6% into perpetuity.

Exhibit 1 Earnings and Dividends for Centralino, 2016–2020

	2016	2017	2018	2019	2020(E)
Earnings per share (EPS, €)	4.93	5.25	4.46	5.64	6.00
Dividends per share (DPS, €)	2.45	2.60	2.60	2.75	2.91
Return on equity (ROE)	13.01%	13.71%	11.58%	14.21%	14.96%

Note: The data for 2016–2019 are actual and for 2020 are estimated.

Exhibit 2 Summary Balance Sheet for Centralino, Year Ended 31 December 2019

Assets (€ millions)		Liabilities and Shareholders' Equity (€ millions)	
Cash and cash equivalents	102	Current liabilities	259
Accounts receivable	305	Long-term debt	367
Inventory	333	**Total liabilities**	626
Total current assets	740	Preferred shares	80
Property and equipment, net	913	Common shares	826
Total assets	1,653	Retained earnings	121

Exhibit 2 (Continued)

Assets (€ millions)		Liabilities and Shareholders' Equity (€ millions)	
		Total shareholders' equity	1,027
		Total liabilities and shareholders' equity	1,653

Notes: The market value of long-term debt is equal to its book value. Shares outstanding are 41.94 million common shares and 16.00 million preferred shares.

Exhibit 3 presents forward price-to-earnings ratios (P/Es) for Centralino's peer group. Risso assumes no differences in fundamentals among the peer-group companies.

Exhibit 3 Peer Group Forward P/Es

Company	Forward P/E
Brinaregalo	5.9
Camporio	8.3
Esperto	3.0
Fornodissione	15.0
Radoresto	4.6

Risso also wants to calculate normalized EPS using the average return on equity method. She determines that the 2016–19 time period in Exhibit 1 represents a full business cycle for Centralino.

25 Based on Exhibit 1, the trailing P/E for Centralino as of 1 January 2020, ignoring any business-cycle influence, is *closest to*:

A 8.3.

B 8.9.

C 9.9.

26 Based on Exhibit 1 and Risso's estimates of return and dividend growth, Centralino's justified forward P/E based on the Gordon growth dividend discount model is *closest* to:

A 5.4.

B 5.7.

C 8.3.

27 Based on Exhibit 2, the price-to-book multiple for Centralino is *closest* to:

A 2.0.

B 2.2.

C 2.5.

28 Based on Exhibit 2, the multiple of enterprise value to sales for Centralino as of 31 December 2019 is *closest* to:

A 0.67.

B 0.74.

C 0.77.

29 Based on Exhibit 1 and using the harmonic mean of the peer group forward P/Es shown in Exhibit 3 as a valuation indicator, the common shares of Centralino are:

- **A** undervalued.
- **B** fairly valued.
- **C** overvalued.

30 Based on Exhibits 1 and 2, the normalized earnings per share for Centralino as calculated by Risso should be *closest* to:

- **A** €2.96.
- **B** €3.21.
- **C** €5.07.

The following information relates to Questions 31–37

Cátia Pinho is a supervisor in the equity research division of Suite Securities. Pinho asks Flávia Silveira, a junior analyst, to complete an analysis of Adesivo S.A., Enviado S.A., and Gesticular S.A.

Pinho directs Silveira to use a valuation metric that would allow for a meaningful ranking of relative value of the three companies' shares. Exhibit 1 provides selected financial information for the three companies.

Exhibit 1 Selected Financial Information for Adesivo, Enviado, and Gesticular (Brazilian Real, BRL)

	Adesivo	Enviado	Gesticular
Stock's current price	14.72	72.20	132.16
Diluted EPS (last four quarters)	0.81	2.92	–0.05
Diluted EPS (next four quarters)	0.91	3.10	2.85
Dividend rate (annualized most recent dividend)	0.44	1.24	0.00

Silveira reviews underlying trailing EPS for Adesivo. Adesivo has basic trailing EPS of BRL0.84. Silveira finds the following note in Adesivo's financial statements:

> "On a per share basis, Adesivo incurred in the last four quarters
> i. from a lawsuit, a nonrecurring gain of BRL0.04; and
> ii. from factory integration, a nonrecurring cost of BRL0.03 and a recurring cost of BRL0.01 in increased depreciation."

Silveira notes that Adesivo is forecasted to pay semiannual dividends of BRL0.24 next year. Silveira estimates five-year earnings growth rates for the three companies, which are presented in Exhibit 2.

Exhibit 2	Earnings Growth Rate Estimates over Five Years
Company	**Earnings Growth Rate Estimate (%)**
Adesivo	16.67
Enviado	21.91
Gesticular	32.33

Pinho asks Silveira about the possible use of the price-to-sales ratio (P/S) in assessing the relative value of the three companies. Silveira tells Pinho:

Statement 1 The P/S is not affected by revenue recognition practices.

Statement 2 The P/S is less subject to distortion from expense accounting than is the P/E.

Pinho asks Silveira about using the Fed and Yardeni models to assess the value of the equity market. Silveira states:

Statement 3 The Fed model concludes that the market is undervalued when the market's current earnings yield is greater than the 10-year Treasury bond yield.

Statement 4 The Yardeni model includes the consensus five-year earnings growth rate forecast for the market index.

Silveira also analyzes the three companies using the enterprising value (EV)-to-EBITDA multiple. Silveira notes that the EBITDA for Gesticular for the most recent year is BRL560 million and gathers other selected information on Gesticular, which is presented in Exhibit 4.

Exhibit 4 Selected Information on Gesticular at Year End (BRL Millions)

Market Value of Debt	Market Value of Common Equity	Market Value of Preferred Equity	Cash	Short-Term Investments
1,733	6,766	275	581	495

Pinho asks Silveira about the use of momentum indicators in assessing the shares of the three companies. Silveira states:

Statement 5 Relative-strength indicators compare an equity's performance during a period with the performance of some group of equities or its own past performance.

Statement 6 In the calculation of standardized unexpected earnings (SUE), the magnitude of unexpected earnings is typically scaled by the standard deviation of analysts' earnings forecasts.

31 Based on Pinho's directive and the data from the last four quarters presented in Exhibit 1, the valuation metric that Silveira should use is the:

A price-to-earnings ratio (P/E).

B production-to-demand ratio (P/D).

C earnings-to-price ratio (E/P).

32 Based on Exhibit 1 and the note to Adesivo's financial statements, the trailing P/E for Adesivo using underlying EPS is *closest* to:

A 17.7.

B 18.2.

C 18.4.

33 Based on Exhibits 1 and 2, which company's shares are the most attractively priced based on the five-year forward P/E-to-growth (PEG) ratio?

A Adesivo

B Enviado

C Gesticular

34 Which of Silveira's statements concerning the use of the P/S is correct?

A Statement 1 only

B Statement 2 only

C Both Statement 1 and Statement 2

35 Which of Silveira's statements concerning the Fed and Yardeni models is correct?

A Statement 3 only

B Statement 4 only

C Both Statement 3 and Statement 4

36 Based on Exhibit 4, Gesticular's EV/EBITDA multiple is *closest* to:

A 11.4.

B 13.7.

C 14.6.

37 Which of Silveira's statements concerning momentum indicators is correct?

A Statement 5 only

B Statement 6 only

C Both Statement 5 and Statement 6

SOLUTIONS

1 **A** Normalized EPS is the level of earnings per share that the company could currently achieve under mid-cyclical conditions.

B Averaging EPS over the 2015–18 period, we find that (\$2.55 + \$2.13 + \$0.23 + \$1.45)/4 = \$1.59. According to the method of historical average EPS, Jonash's normalized EPS is \$1.59. The P/E based on this estimate is \$57.98/\$1.59 = 36.5.

C Averaging ROE over the 2015–18 period, we find that (0.218 + 0.163 + 0.016 + 0.089)/4 = 0.1215. For current BV per share, you would use the estimated value of \$19.20 for year end 2019. According to the method of average ROE, 0.1215 × \$19.20 = \$2.33 is the normalized EPS. The P/E based on this estimate is \$57.98/\$2.33 = 24.9.

2 **A** The analyst can rank the two stocks by earnings yield (E/P). Whether EPS is positive or negative, a lower E/P reflects a richer (higher) valuation and a ranking from high to low E/P has a meaningful interpretation.

In some cases, an analyst might handle negative EPS by using normalized EPS in its place. Neither business, however, has a history of profitability. When year-ahead EPS is expected to be positive, forward P/E is positive. Thus, the use of forward P/Es sometimes addresses the problem of trailing negative EPS. Forward P/E is not meaningful in this case, however, because next year's earnings are expected to be negative.

B Hand has an E/P of –0.100, and Somersault has an E/P of –0.125. A higher earnings yield has an interpretation that is similar to that of a lower P/E, so Hand appears to be relatively undervalued. The difference in earnings yield cannot be explained by differences in sales growth forecasts. In fact, Hand has a higher expected sales growth rate than Somersault. Therefore, the analyst should recommend Hand.

3 **A** Because investing looks to the future, analysts often favor forward P/E when earnings forecasts are available, as they are here. A specific reason to use forward P/Es is the fact given that RUF had some unusual items affecting EPS for 2008. The data to make appropriate adjustments to RUF's 2008 EPS are not given. In summary, Stewart should use forward P/Es.

B Because RUF has a complex capital structure, the P/Es of the two companies must be compared on the basis of diluted EPS.

For HS, forward P/E = \$44/2.20 = 20.

For RUF, forward P/E per diluted share

= \$22.50/(\$30,000,000/33,333,333) = \$22.50/\$0.90 = 25.

Therefore, HS has the more attractive valuation at present.

The problem illustrates some of the considerations that should be taken into account in using P/Es and the method of comparables.

4 **A** Your conclusion may be in error because of the following:

- The peer-group stocks themselves may be overvalued; that is, the mean P/E of 18.0 may be too high in terms of intrinsic value. If so, using 18.0 as a multiplier of the stock's expected EPS will lead to an estimate of stock value in excess of intrinsic value.

- The stock's fundamentals may differ from those of the mean food-processing industry stock. For example, if the stock's expected growth rate is lower than the mean industry growth rate and its risk is higher than the mean, the stock may deserve a lower P/E than the industry mean.

In addition, mean P/E may be influenced by outliers.

B The following additional evidence would support the original conclusion:

- Evidence that stocks in the industry are, at least on average, fairly valued (that stock prices reflect fundamentals)
- Evidence that no significant differences exist in the fundamental drivers of P/E for the stock being compared and the average industry stock

5 In principle, the use of any price multiple for valuation is subject to the concern stated. If the stock market is overvalued, an asset that appears to be fairly or even undervalued in relation to an equity index may also be overvalued.

6 **A** The formula for calculating the justified forward P/E for a stable-growth company is the payout ratio divided by the difference between the required rate of return and the growth rate of dividends. If the P/E is being calculated on trailing earnings (Year 0), the payout ratio is increased by 1 plus the growth rate. According to the 2020 income statement, the payout ratio is 18/60 = 0.30; the 2021 income statement gives the same number (24/80 = 0.30). Thus, we can find the following:

P/E based on trailing earnings:

$$\begin{aligned}\text{P/E} &= \left[\text{Payout ratio} \times (1+g)\right]/(r-g)\\ &= (0.30 \times 1.13)/(0.14-0.13) = 33.9.\end{aligned}$$

P/E based on next year's earnings:

$$\begin{aligned}\text{P/E} &= \text{Payout ratio}/(r-g)\\ &= 0.30/(0.14-0.13) = 30.\end{aligned}$$

B

Fundamental Factor	Effect on P/E	Explanation (Not Required in Question)
The risk (beta) of Sundanci increases substantially.	Decrease	P/E is a decreasing function of risk; that is, as risk increases, P/E decreases. Increases in the risk of Sundanci stock would be expected to lower its P/E.
The estimated growth rate of Sundanci's earnings and dividends increases.	Increase	P/E is an increasing function of the growth rate of the company; that is, the higher the expected growth, the higher the P/E. Sundanci would command a higher P/E if the market price were to incorporate expectations of a higher growth rate.
The equity risk premium increases.	Decrease	P/E is a decreasing function of the equity risk premium. An increased equity risk premium increases the required rate of return, which lowers the price of a stock relative to its earnings. A higher equity risk premium would be expected to lower Sundanci's P/E.

7 **A** V_n = Benchmark value of P/E × E_n = 12 × \$3.00 = \$36.0.

B In the expression for the sustainable growth rate, $g = b \times \text{ROE}$, you can use $(1 - 0.45) = 0.55 = b$ and ROE = 0.10 (the industry average), obtaining $0.55 \times 0.10 = 0.055$. Given the required rate of return of 0.09, you obtain the estimate $3.00(0.45)(1.055)/(0.09 − 0.055) = $40.69. In this case, the estimate of terminal value obtained from the Gordon growth model is higher than the estimate based on multiples. The two estimates may differ for a number of reasons, including the sensitivity of the Gordon growth model to the values of the inputs.

8 Although the measurement of book value has a number of widely recognized shortcomings, P/B may still be applied fruitfully in several circumstances:

- The company is not expected to continue as a going concern. When a company is likely to be liquidated (so ongoing earnings and cash flow are not relevant), the value of its assets less its liabilities is of utmost importance. Naturally, the analyst must establish the fair value of these assets.
- The company is composed mainly of liquid assets, which is the case for finance, investment, insurance, and banking institutions.
- The company's EPS is highly variable or negative.

9 A Aratatech: P/S = ($10 price per share)/[($1 billion sales)/(20 million shares)] = $10/($1,000,000,000/20,000,000) = 0.2.

Trymye: P/S = ($20 price per share)/[($1.6 billion sales)/(30 million shares)] = $20/($1,600,000,000/30,000,000) = 0.375.

Aratatech has a more attractive valuation than Trymye based on its lower P/S but a comparable profit margin.

B One advantage of P/S over P/E is that companies' accounting decisions typically have a much greater impact on reported earnings than they are likely to have on reported sales. Although companies are able to make a number of legitimate business and accounting decisions that affect earnings, their discretion over reported sales (revenue recognition) is limited. Another advantage is that sales are almost always positive, so using P/S eliminates issues that arise when EPS is zero or negative.

10 A The P/Es are as follows:

Hoppelli	25.70/1.30 = 19.8.
Telli	11.77/0.40 = 29.4.
Drisket	23.65/1.14 = 20.7.
Whiteline	24.61/2.43 = 10.1.

The EV/S multiples for each company are as follows:

Hoppelli	3,779/4,124 = 0.916.
Telli	4,056/10,751 = 0.377.
Drisket	3,846/17,388 = 0.221.
Whiteline	4,258/6,354 = 0.670.

B The data for the problem include measures of profitability, such as operating profit margin, ROE, and net profit margin. Because EV includes the market values of both debt and equity, logically the ranking based on EV/S should be compared with a pre-interest measure of profitability—namely, operating profit margin. The ranking of the stocks by EV/S from highest to lowest and the companies' operating margins are shown below:

Company	EV/S	Operating Profit Margin (%)
Hoppelli	0.916	6.91
Whiteline	0.670	6.23
Telli	0.377	1.26
Drisket	0.221	1.07

The differences in EV/S appear to be explained, at least in part, by differences in cost structure as measured by operating profit margin.

11 For companies in the industry described, EV/S would be superior to either of the other two ratios. Among other considerations, EV/S is:

- more useful than P/E in valuing companies with negative earnings;
- better than either P/E or P/B for comparing companies in different countries that are likely to use different accounting standards (a consequence of the multinational nature of the industry);
- less subject to manipulation than earnings (i.e., through aggressive accounting decisions by management, who may be more motivated to manage earnings when a company is in a cyclical low, rather than in a high, and thus likely to report losses).

12 A Based on the CAPM, the required rate of return is 4.9% + 1.2 × 5.5% = 11.5%.

B The dividend payout ratio is €0.9/€1.50 = 0.6. The justified values for the trailing P/E and P/BV ratios should be

$$\frac{P_0}{E_0} = \frac{(1-b)\times(1+g)}{r-g} = \frac{0.6\times(1+0.08)}{0.115-0.08} = 18.5$$

$$\frac{P_0}{E_0} = \frac{ROE-g}{r-g} = \frac{0.20-0.08}{0.115-0.08} = 3.4$$

C The justified P/S ratio based on assumed profit margin of 10% should be

$$\frac{P_0}{S_1} = \frac{\left(\frac{E_1}{S_1}\right)(1-b)}{r-g} = \frac{0.10\times 0.6}{0.115-0.08} = 1.7$$

D The justified trailing P/E is higher than the trailing P/E (18.5 versus 16), the justified trailing P/B is higher than the actual trailing P/B (3.4 versus 3.2). The justified P/S based on forward looking margin assumptions is higher than the actual P/S based of forecast sales (1.7 versus 1.5). Therefore, based on these three measures, GG appears to be slightly undervalued.

13 A EBITDA = Net income (from continuing operations) + Interest expense + Taxes + Depreciation + Amortization.

EBITDA for RGI = €49.5 million + €3 million + €2 million + €8 million = €62.5 million.

Per-share EBITDA = (€62.5 million)/(5 million shares) = €12.5.

P/EBITDA for RGI = €150/€12.5 = 12.

EBITDA for NCI = €8 million + €5 million + €3 million + €4 million = €20 million.

Per-share EBITDA = (€20 million)/(2 million shares) = €10.

P/EBITDA for NCI = €100/€10 = 10.

B For RGI:

Market value of equity = €150 × 5 million = €750 million.

Market value of debt = €50 million.

Total market value = €750 million + €50 million = €800 million.

EV = €800 million – €5 million (cash and investments) = €795 million.

Now, Zaldys would divide EV by total (as opposed to per-share) EBITDA:

EV/EBITDA for RGI = (€795 million)/(€62.5 million) = 12.72.

For NCI:

Market value of equity = €100 × 2 million = €200 million.

Market value of debt = €100 million.

Total market value = €200 million + €100 million = €300 million.

EV = €300 million – €2 million (cash and investments) = €298 million.

Now, Zaldys would divide EV by total (as opposed to per-share) EBITDA:

EV/EBITDA for NCI = (€298 million)/(€20 million) = 14.9.

C Zaldys should select RGI as relatively undervalued.

First, it is correct that NCI *appears* to be relatively undervalued based on P/EBITDA, because NCI has a lower P/EBITDA multiple:

- P/EBITDA = €150/€12.5 = 12 for RGI.
- P/EBITDA = €100/€10 = 10 for NCI.

RGI is relatively undervalued on the basis of EV/EBITDA; however, because RGI has the lower EV/EBITDA multiple,

- EV/EBITDA = (€795 million)/(€62.5 million) = 12.72 for RGI.
- EV/EBITDA = (€298 million)/(€20 million) = 14.9 for NCI.

EBITDA is a pre-interest flow; therefore, it is a flow to both debt and equity and the EV/EBITDA multiple is more appropriate than the P/EBITDA multiple. Zaldys would rely on EV/EBITDA to reach his decision if the two ratios conflicted. Note that P/EBITDA does not take into account differences in the use of financial leverage. Substantial differences in leverage exist in this case (NCI uses much more debt), so the preference for using EV/EBITDA rather than P/EBITDA is supported.

14 The major concepts are as follows:

- EPS plus per-share depreciation, amortization, and depletion (CF)

  Limitation: Ignores changes in working capital and noncash revenue; not a free cash flow concept.

- Cash flow from operations (CFO)

  Limitation: Not a free cash flow concept, so not directly linked to theory.

- Free cash flow to equity (FCFE)

  Limitation: Often more variable and more frequently negative than other cash flow concepts.

- Earnings before interest, taxes, depreciation, and amortization (EBITDA)

  Limitation: Ignores changes in working capital and noncash revenue; not a free cash flow concept. Relative to its use in P/EBITDA, EBITDA is mismatched with the numerator because it is a pre-interest concept.

15 MAT Technology is relatively undervalued compared with DriveMed on the basis of P/FCFE. MAT Technology's P/FCFE multiple is 34% the size of DriveMed's FCFE multiple (15.6/46 = 0.34, or 34%). The only comparison slightly in DriveMed's favor, or approximately equal for both companies, is the comparison based on P/CF (i.e., 12.8 for DriveMed versus 13.0 for MAT Technology). However, FCFE is more strongly grounded in valuation theory than P/CF. Because DriveMed's and MAT Technology's expenditures for fixed capital and working capital during the previous year reflected anticipated average expenditures over the foreseeable horizon, you would have additional confidence in the P/FCFE comparison.

16 A Relative strength is based strictly on price movement (a technical indicator). As used by Westard, the comparison is between the returns on HCI and the returns on the S&P 500. In contrast, the price multiple approaches are based on the relationship of current price not to past prices but to some measure of value, such as EPS, book value, sales, or cash flow.

B Only the reference to the P/E in relationship to the pending patent applications in Westard's recommendation is consistent with the company's value orientation. High relative strength would be relevant for a portfolio managed with a growth/momentum investment style.

17 A As a rule, a screen that includes a maximum P/E should include criteria requiring positive earnings; otherwise, the screen could select companies with negative P/Es. The screen may be too narrowly focused on value measures. It did not include criteria related to expected growth, required rate of return, risk, or financial strength.

B The screen results in a very concentrated portfolio. The screen selected only three companies, including two tobacco companies, which typically pay high dividends. Owning these three stocks would provide little diversification.

18 A is correct. The forward P/E should be used given the recent significant acquisition of the water bottling company. Since a major change such as an acquisition or divestiture can affect results, the forward P/E, also known as the leading P/E or prospective P/E, is the most appropriate P/E to use for Delite. Earnings estimates for 2021 should incorporate the performance of the water bottling company. The forward P/E is calculated as the current price divided by the projected earnings per share, or \$65.50/\$3.50 = 18.71.

19 C is correct. The price-to-sales ratio is calculated as price per share divided by annual net sales per share.

Price per share = \$37.23.

Annual net sales per share = \$67.44 billion/1.638821 billion shares = \$41.15.

Price-to-sales ratio (P/S) = \$37.23/\$41.15 = 0.90.

20 C is correct. You Fix It is in the cyclical home improvement industry. The use of normalized earnings should address the problem of cyclicality in You Fix It earnings by estimating the level of earnings per share that the company could achieve currently under mid-cyclical conditions.

21 C is correct. The price to sales ratio (P/S) fails to consider differences in cost structures. Also, while share price reflects the effect of debt financing on profitability and risk, sales is a pre-financing income measure and does not incorporate the impact of debt in the firm's capital structure. Earnings reflect operating and financial leverage, and thus the price-to-earnings ratio (P/E) incorporates the impact of debt in the firm's capital structure.

22 A is correct. Free cash flow to equity (FCFE) is defined as cash flow available to shareholders after deducting all operating expenses, interest and debt payments, and investments in working and fixed capital. Cannan's requirement that the cash flows include interest expense, working capital, and noncash revenue is satisfied by FCFE.

23 C is correct. The P/E-to-growth (PEG) ratio is calculated by dividing a stock's P/E by the expected earnings growth rate, expressed as a percentage. To calculate Delite's PEG ratio, first calculate the P/E: \$65.50/\$3.50 = 18.71. In this case, the forward earnings should be used given the recent acquisition of the water bottling company. Next, calculate Delite's PEG ratio: 18.71/12.41 = 1.51.

Comparing Delite's PEG ratio of 1.51 with the PEG ratios of 1.74 (16.59/9.52) for Fresh Iced Tea and 1.31 (15.64/11.94) for Nonutter Soda and with the beverage sector average of 1.52 (16.40/10.80), it appears that Delite's shares are fairly valued. This is determined by the fact that Delite's PEG ratio is in the middle of the range of PEG ratios and very close to the sector average. Therefore, the shares appear to be fairly valued.

24 B is correct. The harmonic mean is sometimes used to reduce the impact of large outliers—which are typically the major concern in using the arithmetic mean multiple—but not the impact of small outliers (i.e., those close to zero). The harmonic mean may aggravate the impact of small outliers, but such outliers are bounded by zero on the downside.

25 B is correct. The trailing P/E is calculated as follows:

Stock's current price/Most recent four quarters' EPS =

€50/€5.64 = 8.9.

26 A is correct. The justified forward P/E is calculated as follows:

$$\frac{P_0}{E_1} = \frac{D_1/E_1}{r-g}$$
$$= \frac{(2.91/6.00)}{(0.15-0.06)} = 5.4.$$

27 B is correct. Price to book is calculated as the current market price per share divided by book value per share. Book value per share is common shareholders' equity divided by the number of common shares outstanding. Common shareholders' equity is calculated as total shareholders' equity minus the value of preferred stock.

Thus,

Common shareholders' equity = €1,027 – €80 = €947 million.

Book value per share = €947 million/41.94 million = €22.58.

Price-to-book ratio (P/B) for Centralino = €50/€22.58 = 2.2.

28 C is correct. Enterprise value (EV) is calculated as follows:

EV = Market value of common equity +
Market value of preferred stock + Market value of debt
– Cash, cash equivalents, and short-term investments
= (€50 × 41.94 million) + (€5.25 × 16.00 million) + €367 million
– €102 million
= €2,446 million (or €2.446 billion).

So, EV/sales = €2.446 billion/€3.182 billion = 0.77.

29 C is correct. The harmonic mean is calculated as follows:

$$x_H = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)} = \frac{5}{\left(\frac{1}{5.9}\right)+\left(\frac{1}{8.3}\right)+\left(\frac{1}{3.0}\right)+\left(\frac{1}{15.0}\right)+\left(\frac{1}{4.6}\right)} = 5.5.$$

The forward P/E for Centralino is €50/€6.00 = 8.3. Because Centralino's forward P/E is higher than the harmonic mean of the peer group, the shares of Centralino appear relatively overvalued.

30 A is correct. Based on the method of average ROE, normalized EPS is calculated as the average ROE from the most recent full business cycle multiplied by current book value per share. The most recent business cycle was 2016–2019, and the average ROE over that period was

$$\frac{0.1301 + 0.1371 + 0.1158 + 0.1421}{4} = 0.131.$$

The book value of (common) equity, or simply book value, is the value of shareholders' equity less any value attributable to the preferred stock: €1,027 million – €80 million = €947 million.

Current book value per share (BVPS) is calculated as €947 million/41.94 million = €22.58.

So, normalized EPS is calculated as

Average ROE × BVPS = 0.131 × €22.58 = €2.96.

31 C is correct. The E/P based on trailing earnings would offer the most meaningful ranking of the shares. Using E/P places Gesticular's negative EPS in the numerator rather than the denominator, leading to a more meaningful ranking.

32 C is correct. The EPS figure that Silveira should use is diluted trailing EPS of BRL0.81, adjusted as follows:

1 Subtract the BRL0.04 nonrecurring legal gain.

2 Add BRL0.03 for the nonrecurring factory integration charge.

No adjustment needs to be made for the BRL0.01 charge related to depreciation because it is a recurring charge.

Therefore, underlying trailing EPS = BRL0.81 – BRL0.04 + BRL0.03 = BRL0.80 and trailing P/E using underlying trailing EPS = BRL14.72/BRL0.80 = 18.4.

33 A is correct. The forward PEG ratios for the three companies are calculated as follows:

Forward P/E = Stock's current price/Forecasted EPS.

Forward PEG ratio = Forward P/E ÷ Expected earnings growth rate (in percentage terms).

Adesivo forward P/E = BRL14.72/BRL0.91 = 16.18.

Adesivo forward PEG ratio = 16.18/16.67 = 0.97.

Enviado forward P/E = BRL72.20/BRL3.10 = 23.29.

Enviado forward PEG ratio = 23.29/21.91 = 1.06.

Gesticular forward P/E = BRL132.16/BRL2.85 = 46.37.

Gesticular forward PEG ratio = 46.37/32.33 = 1.43.

Adesivo has the lowest forward PEG ratio, 0.97, indicating that it is the most undervalued of the three equities based on the forward PEG ratio.

34 B is correct. Statement 2 is correct because sales, as the top line of the income statement, are less subject to accounting distortion or manipulation than are other fundamentals, such as earnings. Statement 1 is incorrect because sales figures can be distorted by revenue recognition practices, in particular those tending to speed up the recognition of revenues.

35 C is correct. The Fed model considers the equity market to be undervalued when the market's current earnings yield is greater than the 10-year Treasury bond yield. The Yardeni model incorporates the consensus five-year earnings growth rate forecast for the market index, a variable missing in the Fed model.

36 B is correct. The EV for Gesticular is calculated as follows:

$$\begin{aligned} EV &= \text{Market value of debt} + \text{Market value of common equity} + \text{Market value of preferred equity} - \text{Cash and short-term investments.} \\ EV &= \text{BRL1,733 million} + \text{BRL6,766 million} + \text{BRL275 million} - \text{BRL581 million} - \text{BRL495 million} \\ &= \text{BRL7,698 million.} \\ EV/EBITDA &= \text{BRL7,698 million/BRL560 million} = 13.7. \end{aligned}$$

37 A is correct. Relative-strength indicators compare an equity's performance with the performance of a group of equities or with its own past performance. SUE is unexpected earnings scaled by the standard deviation in past unexpected earnings (not the standard deviation of analysts' earnings forecasts, which is used in the calculation of the scaled earnings surprise).

READING

26

Residual Income Valuation

by Jerald E. Pinto, PhD, CFA, Elaine Henry, PhD, CFA, Thomas R. Robinson, PhD, CFA, CAIA, and John D. Stowe, PhD, CFA

Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). Elaine Henry, PhD, CFA, is at Stevens Institute of Technology (USA). Thomas R. Robinson, PhD, CFA, CAIA, Robinson Global Investment Management LLC, (USA). John D. Stowe, PhD, CFA, is at Ohio University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** calculate and interpret residual income, economic value added, and market value added;
☐	**b.** describe the uses of residual income models;
☐	**c.** calculate the intrinsic value of a common stock using the residual income model and compare value recognition in residual income and other present value models;
☐	**d.** explain fundamental determinants of residual income;
☐	**e.** explain the relation between residual income valuation and the justified price-to-book ratio based on forecasted fundamentals;
☐	**f.** calculate and interpret the intrinsic value of a common stock using single-stage (constant-growth) and multistage residual income models;
☐	**g.** calculate the implied growth rate in residual income, given the market price-to-book ratio and an estimate of the required rate of return on equity;
☐	**h.** explain continuing residual income and justify an estimate of continuing residual income at the forecast horizon, given company and industry prospects;
☐	**i.** compare residual income models to dividend discount and free cash flow models;

(continued)

© 2020 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**j.** explain strengths and weaknesses of residual income models and justify the selection of a residual income model to value a company's common stock;
☐	**k.** describe accounting issues in applying residual income models;
☐	**l.** evaluate whether a stock is overvalued, fairly valued, or undervalued based on a residual income model.

1 INTRODUCTION AND RESIDUAL INCOME

a calculate and interpret residual income, economic value added, and market value added;

b describe the uses of residual income models;

Residual income models of equity value have become widely recognized tools in both investment practice and research. Conceptually, residual income is net income less a charge (deduction) for common shareholders' opportunity cost in generating net income. It is the residual or remaining income after considering the costs of all of a company's capital. The appeal of residual income models stems from a shortcoming of traditional accounting. Specifically, although a company's income statement includes a charge for the cost of debt capital in the form of interest expense, it does not include a charge for the cost of equity capital. A company can have positive net income but may still not be adding value for shareholders if it does not earn more than its cost of equity capital. Residual income models explicitly recognize the costs of all the capital used in generating income.

As an economic concept, residual income has a long history, dating back to Alfred Marshall in the late 1800s (Alfred Marshall, 1890). As far back as the 1920s, General Motors used the concept in evaluating business segments. More recently, residual income has received renewed attention and interest, sometimes under names such as economic profit, abnormal earnings, or economic value added. Although residual income concepts have been used in a variety of contexts, including the measurement of internal corporate performance, we will focus on the residual income model for estimating the intrinsic value of common stock. Among the questions we will study to help us apply residual income models are the following:

- How is residual income measured, and how can an analyst use residual income in valuation?
- How does residual income relate to fundamentals, such as return on equity and earnings growth rates?
- How is residual income linked to other valuation methods, such as a price-multiple approach?
- What accounting-based challenges arise in applying residual income valuation?

The following section develops the concept of residual income, introduces the use of residual income in valuation, and briefly presents alternative measures used in practice. The subsequent sections present the residual income model and illustrate its use in valuing common stock, show practical applications, and describe the relative strengths and weaknesses of residual income valuation compared with other valuation methods. The last section addresses accounting issues in the use of residual income valuation. We then conclude with a summary.

1.1 Residual Income

Traditional financial statements, particularly the income statement, are prepared to reflect earnings available to owners. As a result, the income statement shows net income after deducting an expense for the cost of debt capital (i.e., interest expense). The income statement does not, however, deduct dividends or other charges for equity capital. Thus, traditional financial statements essentially let the owners decide whether earnings cover their opportunity costs. Conversely, the economic concept of residual income explicitly deducts the estimated cost of equity capital, the finance concept that measures shareholders' opportunity costs. The cost of equity is the marginal cost of equity, also referred to as the required rate of return on equity. The cost of equity is a marginal cost because it represents the cost of additional equity, whether generated internally or by selling more equity interests. Example 1 illustrates, in a stylized setting, the calculation and interpretation of residual income. To simplify this introduction, we assume that net income accurately reflects clean surplus accounting, a condition that income (earnings) reflects all changes in the book value of equity other than ownership transactions. This concept will be explained later. Our discussions also assume that companies' financing consists only of common equity and debt. In the case of a company that also has preferred stock financing, the residual income calculation would reflect the deduction of preferred stock dividends from net income.

EXAMPLE 1

Calculation of Residual Income

Axis Manufacturing Company, Inc. (AXCI), a very small company in terms of market capitalization, has total assets of €2 million financed 50% with debt and 50% with equity capital. The cost of debt is 7% before taxes; this example assumes that interest is tax deductible, so the after-tax cost of debt is 4.9%. Note that in countries where corporate interest is not tax deductible, the after-tax cost of debt equals the pretax cost of debt. The cost of equity capital is 12%. The company has earnings before interest and taxes (EBIT) of €200,000 and a tax rate of 30%. Net income for AXCI can be determined as follows:

EBIT	€200,000
Less: Interest Expense	€70,000
Pretax Income	€130,000
Less: Income Tax Expense	€39,000
Net Income	€91,000

With earnings of €91,000, AXCI is clearly profitable in an accounting sense. But was the company's profitability adequate return for its owners? Unfortunately, it was not. To incorporate the cost of equity capital, compute residual income. One approach to calculating residual income is to deduct an **equity charge** (the estimated cost of equity capital in money terms) from net income. Compute the equity charge as follows:

$$\begin{aligned}\text{Equity charge} &= \text{Equity capital} \times \text{Cost of equity capital}\\ &= €1{,}000{,}000 \times 12\%\\ &= €120{,}000.\end{aligned}$$

As stated, residual income is equal to net income minus the equity charge:

Net Income	€91,000
Less: Equity Charge	€120,000
Residual Income	(€29,000)

AXCI did not earn enough to cover the cost of equity capital. As a result, it has negative residual income. Although AXCI is profitable in an accounting sense, it is not profitable in an economic sense.

In Example 1, residual income is calculated based on net income and a charge for the cost of equity capital. Analysts will also encounter another approach to calculating residual income that yields the same results under certain assumptions. In this second approach, which takes the perspective of all providers of capital (both debt and equity), a **capital charge** (the company's total cost of capital in money terms) is subtracted from the company's after-tax operating profit. In the case of AXCI in Example 1, the capital charge is €169,000:

Equity charge	0.12 × €1,000,000 =	€120,000
Debt charge	0.07(1 – 0.30) × €1,000,000 =	€49,000
Total capital charge		€169,000

The company's net operating profit after taxes (NOPAT) is €140,000 (€200,000 – 30% taxes). The capital charge of €169,000 is higher than the after-tax operating profit of €140,000 by €29,000, the same figure obtained in Example 1.

As the following table illustrates, both approaches yield the same results in this case because of two assumptions. First, this example assumes that the marginal cost of debt equals the current cost of debt—that is, the cost used to determine net income. Specifically, in this instance, the after-tax interest expense incorporated in net income [€49,000 = €70,000 × (1 – 30%)] is equal to the after-tax cost of debt incorporated into the capital charge. Second, this example assumes that the weights used to calculate the capital charge are derived from the book value of debt and equity. Specifically, it uses the weights of 50% debt and 50% equity.

Approach 1		**Reconciliation**	**Approach 2**	
Net income	€91,000	Plus the after-tax interest expense of €49,000	Net operating profit after tax	€140,000
Less: Equity charge	€120,000	Plus the after-tax capital charge for debt of €49,000	Less: Capital charge	€169,000
Residual income	(€29,000)		Residual income	(€29,000)

That the company is not profitable in an economic sense can also be seen by comparing the company's cost of capital with its return on capital. Specifically, the company's capital charge is greater than its after-tax return on total assets or capital. The after-tax net operating return on total assets or capital is calculated as profits divided by total assets (or total capital). In this example, the after-tax net operating return on total assets is 7% (€140,000/€2,000,000), which is 1.45 percentage points less than the company's effective capital charge of 8.45% (€169,000/€2,000,000). The amount of after-tax net operating profits as a percentage of total assets or capital has been called **return on invested capital** (ROIC). Residual income can also be calculated as (ROIC – Effective capital charge) × Beginning capital.

1.1.1 *The Use of Residual Income in Equity Valuation*

A company that is generating more income than its cost of obtaining capital—that is, one with positive residual income—is creating value. Conversely, a company that is not generating enough income to cover its cost of capital—that is, a company with negative residual income—is destroying value. Thus, all else equal, higher (lower) residual income should be associated with higher (lower) valuations.

To illustrate the effect of residual income on equity valuation using the case of AXCI presented in Example 1, assume the following:

- Initially, AXCI equity is selling for book value or €1 million with 100,000 shares outstanding. Thus, AXCI's book value per share and initial share price are both €10.
- Earnings per share (EPS) is €0.91 (€91,000/100,000 shares).
- Earnings will continue at the current level indefinitely.
- All net income is distributed as dividends.

Because AXCI is not earning its cost of equity, as shown in Example 1, the company's share price should fall. Given the information, AXCI is destroying €29,000 of value per year, which equals €0.29 per share (€29,000/100,000 shares). Discounted at 12% cost of equity, the present value of the perpetuity is €2.42 (€0.29/12%). The current share price minus the present value of the value being destroyed equals €7.58 (€10 – €2.42).

Another way to look at these data is to note that the earnings yield (E/P) for a no-growth company is an estimate of the expected rate of return. Therefore, when price reaches the point at which E/P equals the required rate of return on equity, an investment in the stock is expected to just cover the stock's required rate of return. With EPS of €0.91, the earnings yield is exactly 12% (AXCI's cost of equity) when its share price is €7.58333 (i.e., €0.91/€7.58333 = 12%). At a share price of €7.58333, the total market value of AXCI's equity is €758,333. When a company has negative residual income, shares are expected to sell at a discount to book value. In this example, AXCI's price-to-book ratio (P/B) at this level of discount from book value would be 0.7583. In contrast, if AXCI were earning positive residual income, then its shares should sell at a premium to book value. In summary, higher residual income is expected to be associated with higher market prices (and higher P/Bs), all else being equal.

Residual income (RI) models have been used to value both individual stocks and stock indexes such as the Dow Jones Industrial Average (see Fleck, Craig, Bodenstab, Harris, and Huh 2001; and Lee, Myers, and Swaminathan 1999). Recall that **impairment** in an accounting context means downward adjustment, and **goodwill** is an intangible asset that may appear on a company's balance sheet as a result of its purchase of another company.

Residual income and residual income models have been referred to by a variety of names. Residual income has sometimes been called **economic profit** because it estimates the company's profit after deducting the cost of all capital: debt and equity. In forecasting future residual income, the term **abnormal earnings** is also used. Under the assumption that in the long term the company is expected to earn its cost of capital (from all sources), any earnings in excess of the cost of capital can be termed abnormal earnings. The residual income model has also been called the **discounted abnormal earnings model** and the **Edwards–Bell–Ohlson model** after the names of researchers in the field. Our focus is on a general residual income model that analysts can apply using publicly available data and nonproprietary accounting adjustments. A number of commercial implementations of the approach, however, are also very well known. Before returning to the general residual income model we briefly discuss one such commercial implementation and the related concept of market value added.

1.1.2 *Commercial Implementations*

One example of several competing commercial implementations of the residual income concept is **economic value added** (EVA, an acronym trademarked by Stern Stewart & Co. and generally associated with a specific set of adjustments proposed by Stern Stewart & Co.). EVA aims to produce a value that is a good approximation of economic profit (see Stewart 1991 and Peterson and Peterson 1996). The previous section illustrated a calculation of residual income starting from net operating profit after taxes, and EVA takes the same broad approach. Specifically, economic value added is computed as

$$\text{EVA} = \text{NOPAT} - (\text{C\%} \times \text{TC}), \quad (1)$$

where NOPAT is the company's net operating profit after taxes, C% is the cost of capital, and TC is total capital. In this model, both NOPAT and TC are determined under generally accepted accounting principles and adjusted for a number of items. Some of the more common adjustments include the following:

- Research and development (R&D) expenses are capitalized and amortized rather than expensed (i.e., R&D expense, net of estimated amortization, is added back to earnings to compute NOPAT).
- In the case of strategic investments that are not expected to generate an immediate return, a charge for capital is suspended until a later date.
- Deferred taxes are eliminated such that only cash taxes are treated as an expense.
- Any inventory LIFO (last in, first out) reserve is added back to capital, and any increase in the LIFO reserve is added in when calculating NOPAT.
- Operating leases are treated as capital leases, and non-recurring items are adjusted.

Because of the adjustments made in calculating EVA, a different numerical result will be obtained, in general, than that resulting from the use of the simple computation presented in Example 1. In practice, general (nonbranded) residual income valuation also considers the effect of accounting methods on reported results. Analysts' adjustments to reported accounting results in estimating residual income, however, will generally reflect some differences from the set specified for EVA. A later section will explore accounting considerations in more detail.

Over time, a company must generate economic profit for its market value to increase. A concept related to economic profit (and EVA) is market value added (MVA):

$$\begin{aligned}\text{MVA} = &\ \text{Market value of the company} \\ &- \text{Accounting book value of total capital}\end{aligned} \quad (2)$$

A company that generates positive economic profit should have a market value in excess of the accounting book value of its capital.

Research on the ability of value-added concepts to explain equity value and stock returns has reached mixed conclusions. Peterson and Peterson (1996) found that value-added measures are slightly more highly correlated with stock returns than traditional measures, such as return on assets and return on equity. Bernstein and Pigler (1997) and Bernstein, Bayer, and Pigler (1998) found that value-added measures are no better at predicting stock performance than are such measures as earnings growth.

A variety of commercial models related to the residual income concept have been marketed by other major accounting and consulting firms. Interestingly, the application focus of these models is not, in general, equity valuation. Rather, these implementations of the residual income concept are marketed primarily for measuring internal corporate performance and determining executive compensation.

THE RESIDUAL INCOME MODEL AND THE GENERAL RESIDUAL INCOME MODEL, AND FUNDAMENTAL DETERMINANTS OF RESIDUAL INCOME

2

c calculate the intrinsic value of a common stock using the residual income model and compare value recognition in residual income and other present value models;

d explain fundamental determinants of residual income;

e explain the relation between residual income valuation and the justified price-to-book ratio based on forecasted fundamentals;

In the previous section, we discussed the concept of residual income and briefly introduced the relationship of residual income to equity value. In the long term, companies that earn more than the cost of capital should sell for more than book value, and companies that earn less than the cost of capital should sell for less than book value. The **residual income model** of valuation analyzes the intrinsic value of equity as the sum of two components:

- the current book value of equity, and
- the present value of expected future residual income.

Note that when the change is made from valuing total shareholders' equity to directly valuing an individual common share, earnings per share rather than net income is used. According to the residual income model, the intrinsic value of common stock can be expressed as follows:

$$V_0 = B_0 + \sum_{t=1}^{\infty}\frac{\mathrm{RI}_t}{(1+r)^t} = B_0 + \sum_{t=1}^{\infty}\frac{E_t - rB_{t-1}}{(1+r)^t} \quad (3)$$

where

V_0 = value of a share of stock today (t = 0)
B_0 = current per-share book value of equity
B_t = expected per-share book value of equity at any time t
r = required rate of return on equity investment (cost of equity)
E_t = expected EPS for period t
RI_t = expected per-share residual income, equal to $E_t - rB_{t-1}$

The per-share residual income in period t, RI_t, is the EPS for the period, E_t, minus the per-share equity charge for the period, which is the required rate of return on equity multiplied by the book value per share at the beginning of the period, or rB_{t-1}. Whenever earnings per share exceed the per-share cost of equity, per-share residual income is positive; and whenever earnings are less, per-share residual income is negative. Example 2 illustrates the calculation of per-share residual income.

EXAMPLE 2

Per-Share Residual Income Forecasts

David Smith is evaluating the expected residual income as of the end of January 2019 of the Canadian Railway Company (CNR). Using an adjusted beta of 1.02 relative to the TSX 300 Index, a 10-year government bond yield of 1.75%, and an estimated equity risk premium of 7.5%, Smith uses the capital asset pricing model (CAPM) to estimate CNR's required rate of return, r, at 9.40% [1.75% + (1.02 × 7.5%)]. Smith obtains the following (in Canadian dollars, CAD) as of the close on 1 February 2019:

Current market price	109.12
Book value per share as of 31 December 2018	24.32
Consensus annual earnings estimates	
FY 2019 (ending December)	6.23
FY 2020	6.96
Annualized dividend per share forecast	
FY 2019	2.15
FY 2020	2.32

What is the forecast residual income for fiscal years ended December 2019 and December 2020?

Solution:

Forecasted residual income and calculations are shown in Exhibit 1.

Exhibit 1 Canadian National Railway Company (all data in CAD)

Year	2019		2020	
Forecasting book value per share				
Beginning book value (B_{t-1})		24.32		28.40
Earnings per share forecast (E_t)	6.23		6.96	
Less dividend forecast (D_t)	2.15		2.31	
Add Change in retained earnings ($E_t - D_t$)		4.08		4.65
Forecast ending book value per share ($B_{t-1} + E_t - D_t$)		28.40		33.05
Calculating the equity charge				
Beginning book value per share	24.32		28.40	
Multiply cost of equity	× 0.094		× 0.094	
Per-share equity charge ($r \times B_{t-1}$)		2.29		2.67
Estimating per share residual income				
EPS forecast	6.23		6.96	
Less equity charge	2.29		2.67	
Per-share residual income		3.94		4.29

The use of Equation 3, the expression for the estimated intrinsic value of common stock, is illustrated in Example 3.

EXAMPLE 3

Using the Residual Income Model (1)

Bugg Properties' expected EPS is \$2.00, \$2.50, and \$4.00 for the next three years. Analysts expect that Bugg will pay dividends of \$1.00, \$1.25, and \$12.25 for the three years. The last dividend is anticipated to be a liquidating dividend; analysts expect Bugg will cease operations after Year 3. Bugg's current book value is \$6.00 per share, and its required rate of return on equity is 10%.

1 Calculate per-share book value and residual income for the next three years.

2 Estimate the stock's value using the residual income model given in Equation 3

$$V_0 = B_0 + \sum_{t=1}^{\infty} \frac{E_t - rB_{t-1}}{(1+r)^t}$$

3 Confirm your valuation estimate in Part 2 using the discounted dividend approach (i.e., estimating the value of a share as the present value of expected future dividends).

Solution to 1:

The book value and residual income for the next three years are shown in Exhibit 2.

Exhibit 2

Year	1	2	3
Beginning book value per share (B_{t-1})	\$6.00	\$7.00	\$8.25
Net income per share (EPS)	2.00	2.50	4.00
Less dividends per share (D)	1.00	1.25	12.25
Change in retained earnings ($EPS - D$)	1.00	1.25	–8.25
Ending book value per share (B_{t-1} + EPS – D)	\$7.00	\$8.25	\$0.00
Net income per share (EPS)	2.00	2.50	4.000
Less per-share equity charge (rB_{t-1})	0.60	0.70	0.825
Residual income (EPS – Equity charge)	\$1.40	\$1.80	\$3.175

Solution to 2:

The value using the residual income model is

$$\begin{aligned} V_0 &= 6.00 + \frac{1.40}{(1.10)} + \frac{1.80}{(1.10)^2} + \frac{3.175}{(1.10)^3} \\ &= 6.00 + 1.2727 + 1.4876 + 2.3854 \\ &= \$11.15 \end{aligned}$$

Solution to 3:

The value using a discounted dividend approach is

$$\begin{aligned} V_0 &= \frac{1.00}{(1.10)} + \frac{1.25}{(1.10)^2} + \frac{12.25}{(1.10)^3} \\ &= 0.9091 + 1.0331 + 9.2036 \\ &= \$11.15 \end{aligned}$$

Example 3 illustrates two important points about residual income models. First, the RI model is fundamentally similar to other valuation models, such as the dividend discount model (DDM), and given consistent assumptions will yield equivalent results. Second, recognition of value typically occurs earlier in RI models than in the DDM. In Example 3, the RI model attributes \$6.00 of the \$11.15 total value to the beginning of the *first* period. In contrast, the DDM attributes \$9.2036 of the \$11.15 total value to the present value of the *final* period. The rest of this section develops the most familiar general expression for the RI model and illustrates the model's application.

2.1 The General Residual Income Model

The residual income model has a clear relationship to other valuation models, such as the DDM. In fact, the residual income model given in Equation 3 can be derived from the DDM. The general expression for the DDM is

$$V_0 = \frac{D_1}{(1+r)^1} + \frac{D_2}{(1+r)^2} + \frac{D_3}{(1+r)^3} + \ldots$$

The **clean surplus relation** states the relationship among earnings, dividends, and book value as follows:

$$B_t = B_{t-1} + E_t - D_t$$

In other words, the ending book value of equity equals the beginning book value plus earnings minus dividends, apart from ownership transactions. The condition that income (earnings) reflects all changes in the book value of equity other than ownership transactions is known as clean surplus accounting. By rearranging the clean surplus relation, the dividend for each period can be viewed as the net income minus the earnings retained for the period, or net income minus the increase in book value:

$$D_t = E_t - (B_t - B_{t-1}) = E_t + B_{t-1} - B_t$$

Substituting $E_t + B_{t-1} - B_t$ for D_t in the expression for V_0 results in:

$$V_0 = \frac{E_1 + B_0 - B_1}{(1+r)^1} + \frac{E_2 + B_1 - B_2}{(1+r)^2} + \frac{E_3 + B_2 - B_3}{(1+r)^3} + \ldots$$

This equation can be rewritten as follows:

$$V_0 = B_0 + \frac{E_1 - rB_0}{(1+r)^1} + \frac{E_2 - rB_1}{(1+r)^2} + \frac{E_3 - rB_2}{(1+r)^3} + \ldots$$

Expressed with summation notation, the following equation restates the residual income model given in Equation 3:

$$V_0 = B_0 + \sum_{t=1}^{\infty} \frac{\mathrm{RI}_t}{(1+r)^t} = B_0 + \sum_{t=1}^{\infty} \frac{E_t - rB_{t-1}}{(1+r)^t}$$

According to the expression, the value of a stock equals its book value per share plus the present value of expected future per-share residual income. Note that when the present value of expected future per-share residual income is positive (negative), intrinsic value, V_0, is greater (smaller) than book value per share, B_0.

The residual income model used in practice today has its origins largely in the academic work of Ohlson (1995) and Feltham and Ohlson (1995) along with the earlier work of Edwards and Bell (1961), although in the United States this method has been used to value small businesses in tax cases since the 1920s. In tax valuation, the approach is known as the **excess earnings method** (Hitchner 2017 and US IRS Revenue Ruling 68-609). The general expression for the residual income model based on this work (Hirst and Hopkins 2000) can also be stated as:

$$V_0 = B_0 + \sum_{t=1}^{\infty} \frac{(\text{ROE}_t - r)B_{t-1}}{(1+r)^t} \quad (4)$$

Equation 4 is equivalent to the expressions for V_0 given earlier because in any year, t, $\text{RI}_t = (\text{ROE}_t - r)B_{t-1}$. Other than the required rate of return on common stock, the inputs to the residual income model come from accounting data. Note that return on equity (ROE) in this context uses beginning book value of equity in the denominator, whereas in financial statement analysis ROE is frequently calculated using the average book value of equity in the denominator. Example 4 illustrates the estimation of value using Equation 4.

EXAMPLE 4

Using the Residual Income Model (2)

To recap the data from Example 3, Bugg Properties has expected earnings per share of \$2.00, \$2.50, and \$4.00 and expected dividends per share of \$1.00, \$1.25, and \$12.25 for the next three years. Analysts expect that the last dividend will be a liquidating dividend and that Bugg will cease operating after Year 3. Bugg's current book value per share is \$6.00, and its estimated required rate of return on equity is 10%.

Using this data, estimate the value of Bugg Properties' stock using a residual income model of the form:

$$V_0 = B_0 + \sum_{t=1}^{\infty} \frac{(\text{ROE}_t - r)B_{t-1}}{(1+r)^t}$$

Solution:

To value the stock, forecast residual income. Exhibit 3 illustrates the calculation of residual income. (Note that Exhibit 3 arrives at the same estimates of residual income as Exhibit 2 in Example 3.)

Exhibit 3

Year	1	2	3
Earnings per share	\$2.00	\$2.50	\$4.00
Divided by beginning book value per share	÷ 6.00	÷ 7.00	÷ 8.25
ROE	0.3333	0.3571	0.4848
Less required rate of return on equity	– 0.1000	– 0.1000	– 0.1000

(continued)

Exhibit 3 (Continued)

Year	1	2	3
Abnormal rate of return (ROE − r)	0.2333	0.2571	0.3848
Multiply by beginning book value per share	× 6.00	× 7.00	× 8.25
Residual income (ROE − r) × Beginning BV	$1.400	$1.800	$3.175

Estimate the stock value as follows:

$$
\begin{aligned}
V_0 &= 6.00 + \frac{1.40}{(1.10)} + \frac{1.80}{(1.10)^2} + \frac{3.175}{(1.10)^3} \\
&= 6.00 + 1.2727 + 1.4876 + 2.3854 \\
&= \$11.15
\end{aligned}
$$

Note that the value is identical to the estimate obtained using Equation 3, as illustrated in Example 3, because the assumptions are the same and Equations 3 and 4 are equivalent expressions:

$$V_0 = \underbrace{B_0 + \sum_{t=1}^{\infty} \frac{E_t - rB_{t-1}}{(1+r)^t}}_{\text{Equation 3}} = \underbrace{B_0 + \sum_{t=1}^{\infty} \frac{(ROE_t - r)B_{t-1}}{(1+r)^t}}_{\text{Equation 4}}$$

Example 4 showed that residual income value can be estimated using current book value, forecasts of earnings, forecasts of book value, and an estimate of the required rate of return on equity. The forecasts of earnings and book value translate into ROE forecasts.

EXAMPLE 5

Valuing a Company Using the General Residual Income Model

Robert Sumargo, an equity analyst, is considering the valuation of Alphabet Inc. Class C shares (GOOG), in mid 2019 when a recent closing price is $1,037.39. (Alphabet Inc. is the parent company of Google.) Sumargo notes that in general, Alphabet had a fairly high ROE during the past 10 years and that consensus analyst forecasts for EPS for the next two fiscal years reflect a fairly high expected ROE percentage. He expects that a high ROE may not be sustainable in the future. Sumargo usually takes a present value approach to valuation. As of the date of the valuation, Alphabet does not pay dividends; although a discounted dividend valuation is possible, Sumargo does not feel confident about predicting the date of a dividend initiation. He decides to apply the residual income model to value Alphabet and uses the following data and assumptions:

- According to the CAPM, Alphabet has a required rate of return of approximately 8.2%.
- Alphabet's book value per share on 31 December 2018 was $255.40.

- ROE is expected to be 20.2% for 2019. Because of competitive pressures, Sumargo expects Google's ROE to decline in the following years and incorporates an assumed decline of 0.5% each year until it reaches the CAPM required rate of return. In 2043, the ROE will be 8.2%, and residual income that year and after will be zero.
- Google does not currently pay a dividend. Sumargo does not expect the company to pay a dividend in the foreseeable future, so all earnings will be reinvested. In addition, Sumargo expects that share repurchases will approximately offset new share issuances.

Compute the value of Google using the residual income model (Equation 4).

Solution:

Book value per share is initially \$255.40. Based on a ROE forecast of 20.2% in the first year, the forecast EPS would be \$51.59. Because no dividends are paid and the clean surplus relation is assumed to hold, book value at the end of the period is forecast to be \$306.99 (\$255.40 + \$51.59). For 2019, residual income is measured as projected EPS of \$51.59 minus an equity charge of \$20.94, or \$30.65. This amount is equivalent to the beginning book value per share of \$255.40 multiplied by the difference between ROE of 20.2% and r of 8.2% [i.e., \$255.40 × (0.20.2 − 0.082) = \$30.65]. The present value of \$30.65 at 8.2% for one year is \$28.33. This process is continued year by year as presented in Exhibit 4. The value of Alphabet using this residual income model would be the present value of each year's residual income plus the current book value per share. Because residual income is zero starting in 2043, no forecast is required beyond that period. The estimated value under this model is \$972.25, as shown in Exhibit 4.

Exhibit 4 Valuation of Alphabet Using the Residual Income Model

Year	Projected Income EPS	Projected Dividend per Share	Book Value per Share	Forecast ROE (Based on Beginning Book Value)	Cost of Equity	Equity Charge	Residual Income (RI)	PV of BV and RI
	[Plus]	[Minus]	255.40					255.40
2019	$51.59	$0.00	$306.99	20.20%	8.20%	$20.94	$30.65	28.33
2020	60.48	0.00	367.47	19.70%	8.20%	25.17	35.30	30.16
2021	70.55	0.00	438.02	19.20%	8.20%	30.13	40.42	31.91
2022	81.91	0.00	519.93	18.70%	8.20%	35.92	45.99	33.56
2023	94.63	0.00	614.56	18.20%	8.20%	42.63	51.99	35.06
2024	108.78	0.00	723.34	17.70%	8.20%	50.39	58.38	36.39
2025	124.41	0.00	847.75	17.20%	8.20%	59.31	65.10	37.50
2026	141.57	0.00	989.32	16.70%	8.20%	69.52	72.06	38.36
2027	160.27	0.00	1,149.60	16.20%	8.20%	81.12	79.15	38.94
2028	180.49	0.00	1,330.08	15.70%	8.20%	94.27	86.22	39.20
2029	202.17	0.00	1,532.25	15.20%	8.20%	109.07	93.11	39.13
2030	225.24	0.00	1,757.50	14.70%	8.20%	125.64	99.60	38.68
2031	249.56	0.00	2,007.06	14.20%	8.20%	144.11	105.45	37.85
2032	274.97	0.00	2,282.03	13.70%	8.20%	164.58	110.39	36.62
2033	301.23	0.00	2,583.25	13.20%	8.20%	187.13	114.10	34.99
2034	328.07	0.00	2,911.33	12.70%	8.20%	211.83	116.25	32.94

(continued)

Exhibit 4 (Continued)

Year	Projected Income EPS	Projected Dividend per Share	Book Value per Share	Forecast ROE (Based on Beginning Book Value)	Cost of Equity	Equity Charge	Residual Income (RI)	PV of BV and RI
2035	355.18	0.00	3,266.51	12.20%	8.20%	238.73	116.45	30.50
2036	382.18	0.00	3,648.69	11.70%	8.20%	267.85	114.33	27.67
2037	408.65	0.00	4,057.35	11.20%	8.20%	299.19	109.46	24.49
2038	434.14	0.00	4,491.48	10.70%	8.20%	332.70	101.43	20.97
2039	458.13	0.00	4,949.61	10.20%	8.20%	368.30	89.83	17.17
2040	480.11	0.00	5,429.73	9.70%	8.20%	405.87	74.24	13.11
2041	499.53	0.00	5,929.26	9.20%	8.20%	445.24	54.30	8.86
2042	515.85	0.00	6,445.11	8.70%	8.20%	486.20	29.65	4.47
Total								972.25

Note: PV is present value and BV is book value. This table was created in Excel, so numbers may differ from what will be obtained using a calculator, because of rounding.

Example 5 refers to the assumption of clean surplus accounting. The residual income model, as stated earlier, assumes clean surplus accounting. The clean surplus accounting assumption is illustrated in Exhibit 4, for example, in which ending book value per share is computed as beginning book value plus net income minus dividends. Under International Financial Reporting Standards (IFRS) and US generally accepted accounting principles (US GAAP), several items of income and expense occurring during a period, such as changes in the market value of certain securities, bypass the income statement and affect a company's book value of equity directly. Items that bypass the income statement (dirty surplus items) are referred to as **other comprehensive income** (the relationship is Comprehensive income = Net income + Other comprehensive income). Strictly speaking, residual income models involve all items of income and expense (income under clean surplus accounting). If an analyst can reliably estimate material differences from clean surplus accounting expected in the future, an adjustment to net income may be appropriate. We explore violations of the clean surplus accounting assumption in more detail later.

2.2 Fundamental Determinants of Residual Income

In general, the residual income model makes no assumptions about future earnings and dividend growth. If constant earnings and dividend growth are assumed, a version of the residual income model that usefully illustrates the fundamental drivers of residual income can be derived. The following expression is used for justified P/B based on forecasted fundamentals, assuming the Gordon (constant growth) DDM and the sustainable growth rate equation, $g = b \times \text{ROE}$:

$$\frac{P_0}{B_0} = \frac{\text{ROE} - g}{r - g}$$

which is mathematically equivalent to

$$\frac{P_0}{B_0} = 1 + \frac{\text{ROE} - r}{r - g}$$

The justified price is the stock's intrinsic value ($P_0 = V_0$). Therefore, using the previous equation and remembering that residual income is earnings less the cost of equity, or (ROE × B_0) − (r × B_0), a stock's intrinsic value under the residual income model, assuming constant growth, can be expressed as:

$$V_0 = B_0 + \frac{\text{ROE} - r}{r - g} B_0 \quad (5)$$

Under this model, the estimated value of a share is the book value per share (B_0) plus the present value [(ROE − r)B_0/(r − g)] of the expected stream of residual income. In the case of a company for which ROE exactly equals the cost of equity, the intrinsic value is equal to the book value per share. Equation 5 is considered a single-stage (or constant-growth) residual income model.

In an idealized world, where the book value of equity represents the fair value of net assets and clean surplus accounting prevails, the term B_0 reflects the value of assets owned by the company less its liabilities. The second term, (ROE − r)B_0/(r − g), represents additional value expected because of the company's ability to generate returns in excess of its cost of equity; the second term is the present value of the company's expected economic profits. However, both IFRS and US GAAP allow companies to exclude some liabilities from their balance sheets, and neither set of rules reflects the fair value of many corporate assets. Internationally, however, a move toward fair value accounting is occurring, particularly for financial assets. Further, controversies, such as the failure of Enron Corporation in the United States, have highlighted the importance of identifying off-balance-sheet financing techniques.

The residual income model is most closely related to the P/B. A stock's justified P/B is directly related to expected future residual income. Another closely related concept is **Tobin's *q***, the ratio of the market value of debt and equity to the replacement cost of total assets:

$$\text{Tobin's } q = \frac{\text{Market value of debt and equity}}{\text{Replacement cost of total assets}}$$

Although similar to P/B, Tobin's q also has some obvious differences. The numerator includes the market value of total capital (debt as well as equity). The denominator uses total assets rather than equity. Further, assets are valued at replacement cost rather than at historical accounting cost; replacement costs take into account the effects of inflation. All else equal, Tobin's q is expected to be higher the greater the productivity of a company's assets (note that Tobin theorized that q would average to 1 for all companies because the economic rents or profits earned by assets would average to zero). One difficulty in computing Tobin's q is the lack of information on the replacement cost of assets. If available, market values of assets or replacement costs can be more useful in a valuation than historical costs.

SINGLE-STAGE RESIDUAL INCOME VALUATION AND MULTISTAGE RESIDUAL INCOME VALUATION 3

f calculate and interpret the intrinsic value of a common stock using single-stage (constant-growth) and multistage residual income models;

g calculate the implied growth rate in residual income, given the market price-to-book ratio and an estimate of the required rate of return on equity;

h explain continuing residual income and justify an estimate of continuing residual income at the forecast horizon, given company and industry prospects;

i compare residual income models to dividend discount and free cash flow models;

j explain strengths and weaknesses of residual income models and justify the selection of a residual income model to value a company's common stock;

The single-stage (constant-growth) residual income model assumes that a company has a constant return on equity and constant earnings growth rate through time. This model was given in Equation 5:

$$V_0 = B_0 + \frac{\text{ROE} - r}{r - g} B_0$$

EXAMPLE 6

Single-Stage Residual Income Model (1)

Joseph Yoh is evaluating a purchase of Koninklijke Philips N.V. Current book value per share is €13.22, and the current price per share is €35.40. Yoh expects the long-term ROE to be 12% and long-term growth to be 6.75%. Assuming a cost of equity of 8.5%, what is the intrinsic value of Canon stock calculated using a single-stage residual income model?

Solution:

Using Equation 5:

$$\boldsymbol{V}_0 = 13.22 + \frac{0.12 - 0.085}{0.085 - 0.675} \times 13.22$$

$$\boldsymbol{V}_0 = €39.66$$

Similar to the Gordon growth DDM, the single-stage RI model can be used to assess the market expectations of residual income growth—that is, an implied growth rate—by inputting the current price into the model and solving for g.

EXAMPLE 7

Single-Stage Residual Income Model (2)

Joseph Yoh is curious about the market-perceived growth rate, given that he is comfortable with his other inputs. By using the current price per share of €35.40 for Philips, Yoh solves the following equation for g:

$$35.40 = 13.22 + \frac{0.12 - 0.085}{0.085 - \boldsymbol{g}} \times 13.22$$

He finds an implied growth rate of 6.41%.

In Examples 6 and 7, the company was valued at almost 2.7× its book value because its ROE exceeded its cost of equity. If ROE was equal to the cost of equity, the company would be valued at book value. If ROE was lower than the cost of equity, the company would have negative residual income and be valued at less than book value. (When a company has no prospect of being able to cover its cost of capital, a liquidation of the company and redeployment of assets may be appropriate.)

In many applications, a drawback to the single-stage model is that it assumes the excess ROE above the cost of equity will persist indefinitely. More likely, a company's ROE will revert to a mean value of ROE over time, and at some point, the company's residual income will be zero. If a company or industry has an abnormally high ROE, other companies will enter the marketplace, thus increasing competition and lowering returns for all companies. Similarly, if an industry has a low ROE, companies will exit the industry (through bankruptcy or otherwise) and ROE will tend to rise over time. As with the single-stage DDM, the single-stage residual income model also assumes a constant growth rate through time. In light of these considerations, the residual income model has been adapted in practice to handle declining residual income. For example, Lee and Swaminathan (1999) and Lee, Myers, and Swaminathan (1999) used a residual income model to value the Dow 30 by assuming that ROE fades (reverts) to the industry mean over time. Lee and Swaminathan found that the residual income model had more ability than traditional price multiples to predict future returns. Fortunately, other models are available that enable analysts to relax the assumption of indefinite persistence of excess returns. The following section describes a multistage residual income model.

3.1 Multistage Residual Income Valuation

As with other valuation approaches, such as DDM and free cash flow, a multistage residual income approach can be used to forecast residual income for a certain time horizon and then estimate a terminal value based on continuing residual income at the end of that time horizon. **Continuing residual income** is residual income after the forecast horizon. As with other valuation models, the forecast horizon for the initial stage should be based on the ability to explicitly forecast inputs in the model. Because ROE has been found to revert to mean levels over time and may decline to the cost of equity in a competitive environment, residual income approaches often model ROE fading toward the cost of equity. As ROE approaches the cost of equity, residual income approaches zero. An ROE equal to the cost of equity would result in residual income of zero.

In residual income valuation, the current book value often captures a large portion of total value and the terminal value may not be a large component of total value because book value is larger than the periodic residual income and because ROE may fade over time toward the cost of equity. This contrasts with other multistage approaches (DDM and DCF), in which the present value of the terminal value is frequently a significant portion of total value.

Analysts make a variety of assumptions concerning continuing residual income. Frequently, one of the following assumptions is made:

- residual income continues indefinitely at a positive level;
- residual income is zero from the terminal year forward;
- residual income declines to zero as ROE reverts to the cost of equity through time; or
- residual income reflects the reversion of ROE to some mean level.

The following examples illustrate several of these assumptions.

One finite-horizon model of residual income valuation assumes that at the end of time horizon T, a certain premium over book value ($P_T - B_T$) exists for the company, in which case, current value equals the following (Bauman, 1999):

$$V_0 = B_0 + \sum_{t=1}^{T} \frac{(E_t - rB_{t-1})}{(1+r)^t} + \frac{P_T - B_T}{(1+r)^T} \tag{6}$$

Alternatively,

$$V_0 = B_0 + \sum_{t=1}^{T} \frac{(\text{ROE}_t - r)B_{t-1}}{(1+r)^t} + \frac{P_T - B_T}{(1+r)^T} \tag{7}$$

The last component in both specifications represents the premium over book value at the end of the forecast horizon. The longer the forecast period, the greater the chance that the company's residual income will converge to zero. For long forecast periods, this last term may be treated as zero. For shorter forecast periods, a forecast of the premium should be calculated.

EXAMPLE 8

Multistage Residual Income Model (1)

Diana Rosato, CFA, is considering an investment in Zenlandia Chemical Company, a fictitious manufacturer of specialty chemicals. Rosato obtained the following facts and estimates as of August 2020:

- Current price equals ZL$95.6.
- Cost of equity equals 12%.
- Zenlandia Chemical's ROE has ranged from 18% to 22.9% during the period 2015–2019. The only time ROE was below 20% during that period was in 2016.
- In 2019, the company paid a cash dividend of ZL$2.9995.
- Book value per share was ZL$28.8517 at the end of 2019.
- Rosato's forecasts of EPS are ZL$7.162 for 2020 and ZL$8.356 for 2021. She expects dividends of ZL$2.9995 for 2020 and ZL$3.2995 for 2021.
- Rosato expects Zenlandia Chemical's ROE to be 25% from 2022 through 2026 and then decline to 20% through 2039.
- For the period after 2021, Rosato assumes an earnings retention ratio of 60%.
- Rosato assumes that after 2039, ROE will be 12% and residual income will be zero; therefore, the terminal value would be zero. Rosato's residual income model is shown in Exhibit 5.

Exhibit 5 Zenlandia Chemical

Year	Book Value (ZL$)	Projected Income (ZL$)	Dividend per Share (ZL$)	Forecasted ROE (Beg. Equity, %)	COE (%)	COE (ZL$)	Residual Income (ZL$)	Present Value of Residual Income (ZL$)
2019	28.8517							28.85
2020	33.0142	7.1620	2.9995	24.82	12.00	3.4622	3.6998	3.30
2021	38.0707	8.3560	3.2995	25.31	12.00	3.9617	4.3943	3.50
2022	43.7813	9.5177	3.8071	25.00	12.00	4.5685	4.9492	3.52

Exhibit 5 (Continued)

Year	Book Value (ZL$)	Projected Income (ZL$)	Dividend per Share (ZL$)	Forecasted ROE (Beg. Equity, %)	COE (%)	COE (ZL$)	Residual Income (ZL$)	Present Value of Residual Income (ZL$)
2023	50.3485	10.9453	4.3781	25.00	12.00	5.2538	5.6916	3.62
2024	57.9008	12.5871	5.0349	25.00	12.00	6.0418	6.5453	3.71
2025	66.5859	14.4752	5.7901	25.00	12.00	6.9481	7.5271	3.81
2026	76.5738	16.6465	6.6586	25.00	12.00	7.9903	8.6562	3.92
2027	85.7626	15.3148	6.1259	20.00	12.00	9.1889	6.1259	2.47
2028	96.0541	17.1525	6.8610	20.00	12.00	10.2915	6.8610	2.47
2029	107.5806	19.2108	7.6843	20.00	12.00	11.5265	7.6843	2.47
2030	120.4903	21.5161	8.6065	20.00	12.00	12.9097	8.6065	2.47
2031	134.9492	24.0981	9.6392	20.00	12.00	14.4588	9.6392	2.47
2032	151.1431	26.9898	10.7959	20.00	12.00	16.1939	10.7959	2.47
2033	169.2802	30.2286	12.0914	20.00	12.00	18.1372	12.0914	2.47
2034	189.5938	33.8560	13.5424	20.00	12.00	20.3136	13.5424	2.47
2035	212.3451	37.9188	15.1675	20.00	12.00	22.7513	15.1675	2.47
2036	237.8265	42.4690	16.9876	20.00	12.00	25.4814	16.9876	2.47
2037	266.3657	47.5653	19.0261	20.00	12.00	28.5392	19.0261	2.47
2038	298.3296	53.2731	21.3093	20.00	12.00	31.9639	21.3093	2.47
2039	334.1291	59.6659	23.8664	20.00	12.00	35.7996	23.8664	2.47
							Present value ZL$	86.41

Terminal Premium = 0.00

The market price of ZL$95.6 exceeds the estimated value of ZL$86.41. The market price reflects higher forecasts of residual income during the period to 2039, a higher terminal premium than Rosato forecasts, and/or a lower cost of equity. If Rosato is confident in her forecasts she may conclude that the company is overvalued in the current marketplace.

Lee and Swaminathan (1999) and Lee, Myers, and Swaminathan (1999) have presented a residual income model based on explicit forecasts of residual income for three years. Thereafter, ROE is forecast to fade to the industry mean value of ROE. The terminal value at the end of the forecast horizon (T) is estimated as the terminal-year residual income discounted in perpetuity. Lee and Swaminathan stated that this assumes any growth in earnings after T is value neutral. Exhibit 6 presents sector ROE data from CSIMarket. In forecasting a fading ROE, the analyst should also consider any trends in industry ROE.

Exhibit 6 US Sector ROEs

Sectors	ROE (%)
Basic Materials	11.14
Consumer Goods	19.96
Consumer Non-cyclicals	26.59

(continued)

Exhibit 6 (Continued)

Sectors	ROE (%)
Energy	8.81
Financial	12.76
Healthcare	19.95
Industrial Goods	23.16
Retail	23.37
Technology	28.97
Transportation	21.49
Utilities	8.18

Source: Based on data from CSIMarket on 5 August 2019.

EXAMPLE 9

Multistage Residual Income Model (2)

Rosato's supervisor questions her assumption that Zenlandia Chemical will have no premium at the end of her forecast period. Rosato assesses the effect of a terminal value based on a perpetuity of Year 2039 residual income. She computes the following terminal value:

TV = ZL\$23.8664/0.12 = ZL\$198.8867

The present value of this terminal value is as follows:

PV = ZL\$198.8867/(1.12)^{20} = ZL\$20.6179

Adding ZL\$20.6179 to the previous value of ZL\$86.41 (for which the terminal value was zero) yields a total value of ZL\$107.03. Because the current market price of ZL\$95.6 is less than ZL\$107.03, market participants expect a continuing residual income that is lower than her new assumptions and/or are forecasting a lower interim ROE. If Rosato agrees with her supervisor and is confident in her new forecasts, she may now conclude that the company is undervalued.

Another multistage model assumes that ROE fades over time to the cost of equity. In this approach, ROE can be explicitly forecast each period until reaching the cost of equity. The forecast would then end and the terminal value would be zero.

Dechow, Hutton, and Sloan (1999) presented an analysis of a residual income model in which residual income fades over time:

$$V_0 = B_0 + \sum_{t=1}^{T-1} \frac{\left(E_t - rB_{t-1}\right)}{\left(1 + r\right)^t} + \frac{E_T - rB_{T-1}}{\left(1 + r - \omega\right)\left(1 + r\right)^{T-1}} \quad (8)$$

This model adds a persistence factor, ω, which is between zero and one. A persistence factor of one implies that residual income will not fade at all; rather it will continue at the same level indefinitely (i.e., in perpetuity). A persistence factor of zero implies that residual income will not continue after the initial forecast horizon. The higher the value of the persistence factor, the higher the stream of residual income in the final stage, and the higher the valuation, all else being equal. Dechow et al. found that in a large sample of company data from 1976 to 1995, the persistence factor equaled 0.62, which was interpreted by Bauman (1999) as equivalent to residual income decaying at an average rate of 38% a year. The persistence factor considers the long-run mean-reverting nature of ROE, assuming that in time ROE regresses toward r and that resulting residual income fades toward zero. Clearly, the persistence factor varies from company to company. For example, a company with a strong market leadership position would have a lower expected rate of decay (Bauman 1999). Dechow et al. provided insight into some characteristics, listed in Exhibit 7, that can indicate a lower or higher level of persistence.

Exhibit 7 Final-Stage Residual Income Persistence

Lower Residual Income Persistence	Higher Residual Income Persistence
Extreme accounting rates of return (ROE)	Low dividend payout
Extreme levels of special items (e.g., nonrecurring items)	High historical persistence in the industry
Extreme levels of accounting accruals	

Example 10 illustrates the assumption that continuing residual income will decline to zero as ROE approaches the required rate of return on equity.

EXAMPLE 10

Multistage Residual Income Model (3)

Rosato extends her analysis to consider the possibility that ROE will slowly decay toward r in 2040 and beyond, rather than using a perpetuity of Year 2037 residual income. Rosato estimates a persistence parameter of 0.60. The present value of the terminal value is determined as

$$\frac{E_T - rB_{T-1}}{(1 + r - \omega)(1 + r)^{T-1}}$$

with T equal to 20 and 2037 residual income equal to 23.8664, in which the 1.12 growth factor reflects a 12% growth rate calculated as the retention ratio multiplied by ROE, or (0.60)(20%) = 0.12.

$$\frac{23.8664}{(1 + 0.12 - 0.60)(1.12)^{19}} = 5.33$$

Total value is ZL\$86.26, calculated by adding the present value of the terminal value, ZL\$5.33, to ZL\$83.93 (the sum of the PV of residual income in the first 19 years). Rosato concludes that if Zenlandia Chemical's residual income does not persist at a stable level past 2039 and deteriorates through time, the shares are modestly overvalued at a price of ZL\$95.6.

In the previous example, the company's terminal residual value was estimated based on the residual income in the final year of stage 1 and on future growth or decay functions. As shown in Equations 6 and 7, the terminal residual value of the firm is $P_T - B_T$, the terminal price minus the terminal book value. The terminal price could be based on any valuation model, such as a DDM, a price–earnings multiple, or a price–book multiple. Example 11 uses a two-stage residual income model in which the terminal price per share is based on a P/B.

EXAMPLE 11

Two-Stage Residual Income Model

Andreea Popescu is using the two-stage residual income model to value the shares of URS Holdings. For her analysis, she assumes the following:

- Beginning book value per share is €15.00.
- EPS will be 25% of beginning book value for the next six years.
- Cash dividends will be 30% of EPS each year.
- At the end of six years, market price per share will be 1.80× book value per share.

1 Calculate per-share book value and residual income for the next three years.

2 Estimate the stock's value using the residual income model given in Equation 6:

$$V_0 = B_0 + \sum_{t=1}^{T} \frac{(E_t - rB_{t-1})}{(1+r)^t} + \frac{P_T - B_T}{(1+r)^T}$$

3 Confirm your valuation estimate in Part 2 using the discounted dividend approach (i.e., estimating the value of a share as the present value of expected future dividends and terminal price).

Solution to 1:

Exhibit 8 shows the book values, net income, dividends, and residual income.

Exhibit 8: Residual Income for URS Holdings

Year	Beginning Book Value	Net Income	Dividends	Ending Book Value	Residual Income	Present Value of Residual Income
1	15.000	3.750	1.125	17.625	2.558	2.369
2	17.625	4.406	1.322	20.709	3.005	2.579
3	20.709	5.177	1.553	24.334	3.531	2.807
4	24.334	6.083	1.825	28.592	4.149	3.055
5	28.592	7.148	2.144	33.595	4.875	3.325
6	33.595	8.399	2.520	39.475	5.728	3.620
			Sum of PV of Residual Income			17.755

Each year, net income is 25% of beginning book value, dividends are 30% of net income, ending book value is beginning book value plus net income minus dividends, and residual income is net income minus 7.95% of beginning book value.

Solution to 2:

In Exhibit 8, the present values of residual income are found by discounting at the 7.95% cost of equity. Using the logic in Equation 6, the value per share is:

Current book value per share		15.000
Present value of 6 years' residual income		17.755
Terminal value [$P_T - B_T = (1.8 \times B_T) - B_T$]	31.580	
Present value of terminal value (at 7.95%)		18,856
Value per share		€52.711

Solution to 3:

The value using a discounted dividend approach is

$$V_0 = \sum_{t=1}^{T} \frac{D_t}{(1+r)^t} + \frac{P_T}{(1+r)^T}$$

Exhibit 9 DDM Valuation of URS Holdings

Year	Dividends	PV of Dividends
1	1.125	1.042
2	1.322	1.134
3	1.553	1.235
4	1.825	1.344
5	2.144	1.463
6	2.520	1.592
Sum of PVs of six years' dividends		7.810
Terminal price = 1.8 × B_T	71.054	
PV of terminal price (@7.95%)		44.901
Value per share using DDM		€52.711

RESIDUAL INCOME VALUATION IN RELATION TO OTHER APPROACHES

4

i. compare residual income models to dividend discount and free cash flow models;

j. explain strengths and weaknesses of residual income models and justify the selection of a residual income model to value a company's common stock;

Before addressing accounting issues in using the residual income model, we briefly summarize the relationship of the residual income model to other valuation models.

Valuation models based on discounting dividends or on discounting free cash flows are as theoretically sound as the residual income model. Unlike the residual income model, however, the discounted dividend and free cash flow models forecast future cash flows and find the value of stock by discounting them back to the present using the required return. Recall that the required return is the cost of equity for both the DDM and the free cash flows to equity (FCFE) model. For the free cash flow to the firm (FCFF) model, the required return is the overall weighted average cost of capital.

The RI model approaches this process differently. It starts with a value based on the balance sheet, the book value of equity, and adjusts this value by adding the present values of expected future residual income. Thus, in theory, the recognition of value is different, but the total present value, whether using expected dividends, expected free cash flow, or book value plus expected residual income, should be consistent (Shrieves and Wachowicz, 2001).

Example 12 again illustrates the important point that the recognition of value in residual income models typically occurs earlier than in dividend discount models. In other words, residual income models tend to assign a relatively small portion of a security's total present value to the earnings that occur in later years. Note also that this example makes use of the fact that the present value of a perpetuity in the amount of X can be calculated as X/r.

EXAMPLE 12

Valuing a Perpetuity with the Residual Income Model

Assume the following data:

- A company will earn \$1.00 per share forever.
- The company pays out all earnings as dividends.
- Book value per share is \$6.00.
- The required rate of return on equity (or the percent cost of equity) is 10%.

1. Calculate the value of this stock using the DDM.
2. Calculate the level amount of per-share residual income that will be earned each year.
3. Calculate the value of the stock using a RI model.
4. Create a table summarizing the year-by-year valuation using the DDM and the RI model.

Solution to 1:

Because the dividend, D, is a perpetuity, the present value of D can be calculated as D/r.

$$V_0 = D/r = \$1.00/0.10 = \$10.00 \text{ per share}$$

Solution to 2:

Because each year all net income is paid out as dividends, book value per share will be constant at \$6.00. Therefore, with a required rate of return on equity of 10%, for all future years, per-share residual income will be as follows:

$$\text{RI}_t = E_t - rB_{t-1} = \$1.00 - 0.10(\$6.00) = \$1.00 - \$0.60 = \$0.40$$

Solution to 3:

Using a residual income model, the estimated value equals the current book value per share plus the present value of future expected residual income (which in this example can be valued as a perpetuity):

$$\begin{aligned} V_0 &= \text{Book value + PV of expected future per-share residual income} \\ &= \$6.00 + \$0.40/0.10 \\ &= \$6.00 + \$4.00 \\ &= \$10.00 \end{aligned}$$

Solution to 4:

Exhibit 10 summarizes the year-by-year valuation using the DDM and the RI models.

Exhibit 10 Value Recognition in the DDM and the RI Model

Dividend Discount Model			Residual Income Model	
Year	D_t	**PV of** D_t	B_0 **or** RI_t	**PV of** B_0 **or** RI_t
0			$6.00	$6.000
1	$1.00	$0.909	0.40	0.364
2	1.00	0.826	0.40	0.331
3	1.00	0.751	0.40	0.301
4	1.00	0.683	0.40	0.273
5	1.00	0.621	0.40	0.248
6	1.00	0.564	0.40	0.226
7	1.00	0.513	0.40	0.205
8	1.00	0.467	0.40	0.187
⋮	⋮	⋮	⋮	⋮
Total		$10.00		$10.00

In the RI model, most of the stock's total value is attributed to the earlier periods. Specifically, the current book value of $6.00 represents 60% of the stock's total present value of $10.

In contrast, in the DDM, value is derived from the receipt of dividends, and typically, a smaller proportion of value is attributed to the earlier periods. Less than $1.00 of the total $10 derives from the first year's dividend, and collectively, the first five years' dividends ($0.909 + $0.826 + $0.751 + $0.683 + $0.621 = $3.79) contribute only about 38% of the total present value of $10.

As shown earlier and illustrated again in Example 11, the dividend discount and residual income models are in theory mutually consistent. Because of the real-world uncertainty in forecasting distant cash flows, however, the earlier recognition of value in a residual income approach relative to other present value approaches is a practical advantage. In the dividend discount and free cash flow models, a stock's value is often modeled as the sum of the present value of individually forecasted dividends or free cash flows up to some terminal point plus the present value of the expected terminal value of the stock. In practice, a large fraction of a stock's total present value, in either the discounted dividend or free cash flow models, is represented by the present value of the expected terminal value. Substantial uncertainty, however, often surrounds the terminal value. In contrast, residual income valuations typically are less sensitive to terminal value estimates. (In some residual income valuation contexts, the terminal value may actually be set equal to zero.) The derivation of value from the earlier portion of a forecast horizon is one reason residual income valuation can be a useful analytical tool.

4.1 Strengths and Weaknesses of the Residual Income Model

Now that the implementation of the residual income model has been illustrated with several examples, a summary of the strengths and weaknesses of the residual income approach follows:

The strengths of residual income models include the following:

- Terminal values do not make up a large portion of the total present value, relative to other models.
- RI models use readily available accounting data.
- The models can be readily applied to companies that do not pay dividends or to companies that do not have positive expected near-term free cash flows.
- The models can be used when cash flows are unpredictable.
- The models have an appealing focus on economic profitability.

The potential weaknesses of residual income models include the following:

- The models are based on accounting data that can be subject to manipulation by management.
- Accounting data used as inputs may require significant adjustments.
- The models require either that the clean surplus relation (explained later) holds or that the analyst makes appropriate adjustments when the clean surplus relation does not hold.
- The residual income model's use of accounting income assumes that the cost of debt capital is reflected appropriately by interest expense.

4.2 Broad Guidelines for Using a Residual Income Model

The above list of potential weaknesses helps explain the following section's focus on accounting considerations. In light of its strengths and weaknesses, the following are broad guidelines for using a residual income model in common stock valuation.

A residual income model is most appropriate when:

- a company does not pay dividends, or its dividends are not predictable;
- a company's expected free cash flows are negative within the analyst's comfortable forecast horizon; or
- great uncertainty exists in forecasting terminal values using an alternative present value approach.

Residual income models are least appropriate when:

- significant departures from clean surplus accounting exist, or
- significant determinants of residual income, such as book value and ROE, are not predictable.

Because various valuation models can be derived from the same underlying theoretical model, when fully consistent assumptions are used to forecast earnings, cash flow, dividends, book value, and residual income through a full set of pro forma (projected) financial statements, and the same required rate of return on equity is used as the discount rate, the same estimate of value should result when using each model. Practically speaking, however, it may not be possible to forecast each of these items with the same degree of certainty. For example, if a company has near-term negative free cash flow and forecasts for the terminal value are uncertain, a residual income model may be more appropriate. But a company with positive, predictable cash flow that does not pay a dividend would be well suited for a discounted free cash flow valuation (Penman and Sougiannis 1998; Penman 2001; Lundholm and O'Keefe 2001a; and Lundholm and O'Keefe 2001b).

Residual income models, just like the discounted dividend and free cash flow models, can also be used to establish justified market multiples, such as P/E or P/B. For example, the value can be determined by using a residual income model and dividing by earnings to arrive at a justified P/E.

A residual income model can also be used in conjunction with other models to assess the consistency of results. If a wide variation of estimated value is found and each model appears appropriate, the inconsistency may lie with the assumptions used in the models. The analyst would need to perform additional work to determine whether the assumptions are mutually consistent and which model is most appropriate for the subject company.

5 ACCOUNTING AND INTERNATIONAL CONSIDERATIONS AND VIOLATIONS OF THE CLEAN SURPLUS RELATIONSHIP

k describe accounting issues in applying residual income models;

To most accurately apply the residual income model in practice, the analyst may need to adjust book value of common equity for off-balance-sheet items and adjust reported net income to obtain **comprehensive income** (all changes in equity other than contributions by, and distributions to, owners). In this section, we will discuss issues relating to these tasks.

Bauman (1999) has noted that the strength of the residual income model is that the two components (book value and future earnings) of the model have a balancing effect on each other, provided that the clean surplus relationship is followed:

> All other things held constant, companies making aggressive (conservative) accounting choices will report higher (lower) book values and lower (higher) future earnings. In the model, the present value of differences in future income is exactly offset by the initial differences in book value. (Bauman 1999, page 31)

Unfortunately, this argument has several problems in practice because the clean surplus relationship does not prevail, and analysts often use past earnings to predict future earnings. IFRS and US GAAP permit a variety of items to bypass the income statement and be reported directly in stockholders' equity. Further, off-balance-sheet liabilities or nonoperating and non-recurring items of income may obscure a company's financial performance. The analyst must thus be aware of such items when evaluating the book value of equity and return on equity to be used as inputs into a residual income model.

With regard to the possibility that aggressive accounting choices will lead to lower reported future earnings, consider an example in which a company chooses to capitalize an expenditure in the current year rather than expense it. Doing so overstates current-year earnings as well as current book value. If an analyst uses current earnings (or ROE) naively in predicting future residual earnings, the RI model will overestimate the company's value. Take, for example, a company with $1,000,000 of book value and $200,000 of earnings before taxes, after expensing an expenditure of $50,000. Ignoring taxes, this company has a ROE of 20%. If the company capitalized the expenditure rather than expensing it immediately, it would have a ROE of 23.81%

($250,000/$1,050,000). Although at some time in the future this capitalized item will likely be amortized or written off, thus reducing realized future earnings, analysts' expectations often rely on historical data. If capitalization of expenditures persists over time for a company whose size is stable, ROE can decline because net income will normalize over the long term, but book value will be overstated. For a growing company, for which the expenditure in question is increasing, ROE can continue at high levels over time. In practice, because the RI model uses primarily accounting data as inputs, the model can be sensitive to accounting choices, and aggressive accounting methods (e.g., accelerating revenues or deferring expenses) can result in valuation errors. The analyst must, therefore, be particularly careful in analyzing a company's reported data for use in a residual income model.

Two principal drivers of residual earnings are ROE and book value. Analysts must understand how to use historical reported accounting data for these items to the extent they use historical data in forecasting future ROE and book value. Elsewhere we have explained the DuPont analysis of ROE, which can be used as a tool in forecasting, and discussed the calculation of book value. We extend these discussions below with specific application to residual income valuation, particularly in addressing the following accounting considerations:

- violations of the clean surplus relationship;
- balance sheet adjustments for fair value;
- intangible assets;
- non-recurring items;
- aggressive accounting practices; and
- international considerations.

In any valuation, close attention must be paid to the accounting practices of the company being valued. The following sections address the aforementioned issues with respect to how they specifically affect residual income valuation.

5.1 Violations of the Clean Surplus Relationship

One potential accounting issue in applying a residual income model is a violation of the clean surplus accounting assumption. Violations of this assumption occur when accounting standards permit charges directly to stockholders' equity, bypassing the income statement. An example is the case of changes in the market value of "available-for-sale" investments under US GAAP and "equity instruments measured at fair value through other comprehensive income" under IFRS. Under both IFRS (IFRS 9 Financial Instruments, paragraph 5.7.5) and US GAAP (ASC 320-10-35-1), these categories of investments are shown on the balance sheet at market value. Any unrealized change in their market value, however, is reflected in other comprehensive income rather than as income on the income statement.

As stated earlier, comprehensive income is defined as all changes in equity during a period other than contributions by, and distributions to, owners. Comprehensive income includes net income reported on the income statement and *other comprehensive income*, which is the result of other events and transactions that result in a change to equity but are not reported on the income statement. Items that commonly bypass the income statement include

- unrealized changes in the fair value of some financial instruments, as already discussed;
- foreign currency translation adjustments;
- certain pension adjustments;

- a portion of gains and losses on certain hedging instruments;
- changes in revaluation surplus related to property, plant, and equipment or intangible assets (applicable under IFRS but not under US GAAP); and
- for certain categories of liabilities, a change in fair value attributable to changes in the liability's credit risk (applicable under IFRS but not under US GAAP).

Under both international and US standards, such items as fair value changes for some financial instruments and foreign currency translation adjustments bypass the income statement. In addition, under IFRS, which unlike US GAAP permits revaluation of fixed assets (IAS 16, paragraph 39–42), some changes in the fair value of fixed assets also bypass the income statement and directly affect equity.

In all of these cases in which items bypass the income statement, the book value of equity is stated accurately because it includes "accumulated other comprehensive income," but net income is not stated properly from the perspective of residual income valuation. The analyst should be most concerned with the effect of these items on forecasts of net income and ROE, which has net income in the numerator, and hence residual income. Note that for best results, historical ROE should be calculated at the aggregate level (e.g., as net income divided by shareholders' equity, rather than as earnings per share divided by book value per share), because such actions as share issuance and share repurchases can distort ROE calculated on a per-share basis. Because some items (including those listed earlier) bypass the income statement, they are excluded from historical ROE data. As noted by Frankel and Lee (1999), bias will be introduced into the valuation only if the present expected value of the clean surplus violations does not net to zero. In other words, reductions in income from some periods may be offset by increases from other periods. The analyst must examine the equity section of the balance sheet and the related statements of shareholders' equity and comprehensive income carefully for items that have bypassed the income statement. The analyst can then assess whether amounts are likely to be offsetting and can assess the effect on future ROE.

EXAMPLE 13

Evaluating Clean Surplus Violations

Excerpts from two companies' statements of changes in stockholders' equity are shown in Exhibits 11 and 12. The first statement, prepared under IFRS as of 31 December 2018, is for Nokia Corporation, a provider of network equipment, software, and services to telecom network companies. The second statement, prepared under US GAAP as of 31 December 2018, is for SAP AG, which is headquartered in Germany and is a worldwide provider of enterprise application software, including enterprise resource planning, customer relationship management, and supply chain management software.

Exhibit 12 SAP AG and Subsidiaries Statement of Changes in Shareholders' Equity (€ millions)

	Equity Attributable to Owners of Parent							
	Issued Capital	Share Premium	Retained Earnings	Other Components of Equity	Treasury Shares	Total	Non-controlling interests	Total Equity
1 January 2018	1,229	570	24,987	347	–1,591	25,542	31	25,573
Profit after tax			4,083			4,083	6	4,088

(continued)

PART OF EXAMPLE 13

Exhibit 11 Nokia Corporation Statement of Changes in Shareholders' Equity (€ millions except number of shares)

	Number of Shares Outstanding	Share Capital	Share Issue Premium	Treasury Shares	Translation Differences	Fair Value and Other Reserves	Reserve for Invested Unrestricted Equity	(Accumulated Deficit)/Retained Earnings	Attributable to Equity Holders of the Parent	Non-controlling Interests	Total Equity
As of 1 January 2018	**5,579,517**	**246**	**447**	**–1,480**	**–932**	**842**	**15,616**	**1,345**	**16,084**	**80**	**16,164**
Re-measurements and defined benefit pension plans, net of tax						293			293		293
Translation differences					402				402		402
Net investment hedges, net of tax					–61	3			–58		–58
Cash flow hedges, net of tax						–43			–43		–43
Financial assets at fair value through other comprehensive income, net of tax						–38			–38		–38
Other increase, net						6			5	1	6
Loss for the year									–340	5	–335
Total comprehensive income for the year					**341**	**221**			**221**	**6**	**227**
Share-based payment			68						68		68
Excess tax benefit on share-based payment	13,221		6						6		6

PART OF EXAMPLE 13

Exhibit 11 (Continued)

	Number of Shares Outstanding	Share Capital	Share Issue Premium	Treasury Shares	Translation Differences	Fair Value and Other Reserves	Reserve for Invested Unrestricted Equity	(Accumulated Deficit)/Retained Earnings	Attributable to Equity Holders of the Parent	Non-controlling Interests	Total Equity
Settlement of performance and restricted shares			–85	72			–11		–24		–24
Cancellation of treasury shares	424			1,000				–1,000			
Stock options exercised							1		1		1
Dividends								–1,063	–1,063	–5	–1,068
Acquisitions of non-controlling interests								–1	–1	1	0
Other movements					–1			–2	–3		–3
Total other equity movements		**0**	**–11**	**1,072**	**–1**	**0**	**–10**	**–2,066**	**–1,016**	**-4**	**-1020**
As of December 31, 2018	**5,593,162**	**246**	**436**	**–408**	**–592**	**1,063**	**15,606**	**–1,062**	**15,289**	**82**	**15371**

Source: www.nokia.com.

(continued)

Exhibit 12 (Continued)

	Issued Capital	Share Premium	Retained Earnings	Other Components of Equity	Treasury Shares	Equity Attributable to Owners of Parent: Total	Non-controlling interests	Total Equity
Other comprehensive income			11	887		898		898
Comprehensive income			**4,093**	**887**	**0**	**4,980**	**6**	**4,986**
Share-based payments		−40				−40		−40
Dividends			−1,671			−1,671	−13	−1,684
Reissuance of treasury shares under share-based payments		13			11	24		24
Shares to be issued			7			7		7
Hyperinflation			−8			−8		−8
Changes in non-controlling interests						0	19	19
Other changes			−2			−2	3	1
12/31/2018	**1,229**	**543**	**27,407**	**1,234**	**−1,580**	**28,832**	**45**	**28,877**

Source: www.sap.com.

For Nokia, items that have bypassed the income statement in 2018 are those in the columns labeled “Share issue premium,” “Translation differences,” “Fair value and other reserves,” and “Reserve for invested unrestricted equity.” For SAP, the amounts that bypassed the income statement in 2018 are “Share premium” and “Other components of equity.”

To illustrate the issues in interpreting these items, consider the columns “Translation differences” (Nokia) and “Other components of equity” (SAP). The amounts in these columns reflect currency translation adjustments to equity that have bypassed the income statement. For Nokia, the adjustment for the year 2018 was €341 million. Because this is a positive adjustment to stockholders’ equity, this item would have increased income if it had been reported on the income statement. For SAP, the “Other components of equity” adjustment (which includes translation adjustment for the year 2018) was €887 million. Again, because this is a positive adjustment to stockholders’ equity, this item would have increased income if it had been reported on the income statement. If the analyst expects this trend of positive translation adjustments to continue and has used historical data as the basis for initial estimates of ROE to be used in residual income valuation, an upward adjustment in that estimated future ROE might be warranted. It is possible, however, that future exchange rate movements will reverse this trend.

The examples we have explored used the actual beginning equity and a forecasted level of ROE (return on beginning equity) to compute the forecasted net income. Because equity includes accumulated other comprehensive income (AOCI), the assumptions about future other comprehensive income (OCI) will affect forecasted net income and thus residual income. To illustrate, Exhibit 13 shows a hypothetical company’s financials for a single previous year, labeled year $t - 1$, followed by three different forecasts for the following two years. In year $t - 1$, the company reports net income of $120, which is a 12% return on beginning equity of $1,000. The company paid no dividends, so ending retained earnings equal $120. In year $t - 1$, the company also reports OCI of –$100, a loss, so the ending amount shown in AOCI is –$100. (Companies typically label this line item "accumulated other comprehensive income (loss)," indicating that the amount is an accumulated loss when given in parentheses.)

All three forecasts in Exhibit 13 assume that ROE will be 12% and use this assumption to forecast net income for year t and $t + 1$ by using the expression 0.12 × Beginning book value. Each forecast, however, incorporates different assumptions about future OCI. Forecast A assumes that the company will have no OCI in year t or year $t + 1$, so the amount of AOCI does not change. Forecast B assumes that the company will continue to have the same amount of OCI in year t and year $t + 1$ as it had in the prior year, so the amount of AOCI becomes more negative each year. Forecast C assumes that the company's OCI will reverse in year t, so at the end of year t, AOCI will be zero. As shown, because the forecasts use the assumed ROE to compute forecasted net income, the forecasts for net income and residual income in year $t + 1$ vary significantly.

Because this example assumes all earnings are retained, a forecast of 12% ROE also implies that net income and residual income will grow at 12%. Only the year t to year $t + 1$ under Forecast A, which assumes no future OCI, correctly reflects that relationship. Specifically, in Forecast A, both net income and residual income increase by 12% from year t to year $t + 1$. Net income grows from $122.40 to $137.09, an increase of 12% [($137.09/$122.40) − 1]; and residual income grows from $20.40 to $22.85, an increase of 12% [($22.85/$20.40) − 1]. In contrast to Forecast A, neither Forecast B nor Forecast C correctly reflects the relationship between ROE and growth in income (net and residual). Growth in residual income from year t to year $t + 1$ was 2.2% under Forecast B and 21.8% under Forecast C.

If, alternatively, the forecasts of future ROE and the residual income computation had incorporated total comprehensive income (net income plus OCI), the results of the residual income computation would have differed significantly. For example, suppose that in Forecast B, which assumes the company will continue to have the same amount of OCI, the estimated future ROE was 2.0%, using total comprehensive income [($120 – $100)/$1,000 = $20/$1,000]. If the residual income computation had then also used forecasted total comprehensive income at time t, the amount of residual income would be negative. Specifically, for time t, forecast comprehensive income would be $22.40 (net income plus other comprehensive income), the equity charge would be $102 (required return of 10% multiplied by beginning equity of $1,020), and residual income would be –$79.60 (comprehensive income of $22.40 minus equity charge of $102). Clearly, residual income on this basis significantly falls short of the positive $20.40 when the violation of clean surplus is ignored. As this example demonstrates, using an ROE forecast or a net income forecast that ignores violations of clean surplus accounting will distort estimates of residual income. Unless the present value of such distortions net to zero, using those forecasts will also distort valuations.

What are the implications for implementing a residual-income-based valuation? If future OCI is expected to be significant relative to net income and if the year-to-year amounts of OCI are not expected to net to zero, the analyst should attempt to incorporate these items so that residual income forecasts are closer to what they would be if the clean surplus relation held. Specifically, when possible, the analyst should incorporate explicit assumptions about future amounts of OCI.

Example 14 illustrates, by reference to the DDM value, the error that results when OCI is omitted from residual income calculations (assuming an analyst has a basis for forecasting future amounts of OCI). The example also shows that the growth rate in residual income generally does not equal the growth rate of net income or dividends.

Exhibit 13 Hypothetical Company Alternative Forecasts with Different Assumptions about Comprehensive Income

	Actual	Forecast A		Forecast B		Forecast C	
Year	***t* − 1**	***t***	***t* + 1**	***t***	***t* + 1**	***t***	***t* + 1**
Beginning Balance Sheet							
Assets	$1,000.00	$1,020.00	$1,142.40	$1,020.00	$1,042.40	$1,020.00	$1,242.40
Liabilities	—	—	—	—	—	—	—
Common stock	1,000.00	1,000.00	1,000.00	1,000.00	1,000.00	1,000.00	1,000.00
Retained earnings	—	120.00	242.40	120.00	242.40	120.00	242.40
AOCI	—	(100.00)	(100.00)	(100.00)	(200.00)	(100.00)	—
Total equity	1,000.00	1,020.00	1,142.40	1,020.00	1,042.40	1,020.00	1,242.40
Total liabilities and total equity	$1,000.00	$1,020.00	$1,142.40	$1,020.00	$1,042.40	$1,020.00	$1,242.40
Net income	120.00	122.40	137.09	122.40	125.09	122.40	149.09
Dividends	—	—	—	—	—	—	—
Other comprehensive income	(100.00)	—	—	(100.00)	(100.00)	100.00	—
Ending Balance Sheet							
Assets	$1,020.00	$1,142.40	$1,279.49	$1,042.40	$1,067.49	$1,242.40	$1,391.49
Liabilities	—	—	—	—	—	—	—
Common stock	1,000.00	1,000.00	1,000.00	1,000.00	1,000.00	1,000.00	1,000.00
Retained earnings	120.00	242.40	379.49	242.40	367.49	242.40	391.49
AOCI	(100.00)	(100.00)	(100.00)	(200.00)	(300.00)	—	—
Total equity	$1,020.00	$1,142.40	$1,279.49	$1,042.40	$1,067.49	$1,242.40	$1,391.49
Total liabilities and total equity	$1,020.00	$1,142.40	$1,279.49	$1,042.40	$1,067.49	$1,242.40	$1,391.49
Residual income calculation based on beginning total equity							
Net income	120.00	122.40	137.09	122.40	125.09	122.40	149.09
Equity charge at 10%	100.00	102.00	114.24	102.00	104.24	102.00	124.24
Residual income	$20.00	$20.40	$22.85	$20.40	$20.85	$20.40	$24.85

EXAMPLE 14

Incorporating Adjustments in the Residual Income Model

Exhibit 14 gives per-share forecasts for Mannistore, Inc., a hypothetical company operating a chain of retail stores. The company's cost of equity capital is 10%.

Exhibit 14 Forecasts for Mannistore, Inc.

	Year				
Variable	1	2	3	4	5
Shareholders' equity$_{t-1}$	\$8.58	\$10.32	\$11.51	\$14.68	\$17.86
Plus net income	2.00	2.48	3.46	3.47	4.56
Less dividends	0.26	0.29	0.29	0.29	0.38
Less other comprehensive income	0.00	1.00	0.00	0.00	0.00
Equals shareholders' equity$_t$	\$10.32	\$11.51	\$14.68	\$17.86	\$22.04

1 Assuming the forecasted terminal price of Mannistore's shares at the end of Year 5 (time t = 5) is \$68.40, estimate the value per share of Mannistore using the DDM.

2 Given that the forecast terminal price of Mannistore's shares at the end of Year 5 (time t = 5) is \$68.40, estimate the value of a share of Mannistore using the RI model and calculate residual income based on:

  A net income without adjustment, and

  B net income plus other comprehensive income.

3 Interpret your answers to Parts 2A and 2B.

4 Assume that a forecast of the terminal price of Mannistore's shares at the end of Year 5 (time t = 5) is not available. Instead, an estimate of terminal price based on the Gordon growth model is appropriate. You estimate that the growth in net income and dividends from t = 5 to t = 6 will be 8%. Predict residual income for Year 6, and based on that 8% growth estimate, determine the growth rate in forecasted residual income from t = 5 to t = 6.

Solution to 1:

The estimated value using the DDM is

$$V_0 = \frac{\$0.26}{(1.10)^1} + \frac{\$0.29}{(1.10)^2} + \frac{\$0.29}{(1.10)^3} + \frac{\$0.29}{(1.10)^4} + \frac{\$0.38}{(1.10)^5} + \frac{\$68.40}{(1.10)^5} = \$43.59$$

Solution to 2:

A Calculating residual income as net income (NI) minus the equity charge, which is beginning shareholders' equity (SE) multiplied by the cost of equity capital (r), gives the following for years 1 through 5:

	Year				
	1	2	3	4	5
RI = NI − ($SE_{t-1} \times r$)	1.14	1.45	2.30	2.00	2.77

So, the estimated value using the RI model (using Equation 6), with residual income calculated based on net income, is

$$V_0 = \$8.58 + \frac{\$1.14}{(1.10)^1} + \frac{\$1.45}{(1.10)^2} + \frac{\$2.30}{(1.10)^3} + \frac{\$2.00}{(1.10)^4} + \frac{\$2.77}{(1.10)^5} + \frac{\$68.40 - \$22.04}{(1.10)^5}$$

$$V_0 = \$8.58 + 35.84 = \$44.42$$

B Calculating residual income as net income adjusted for OCI (NI + OCI) minus the equity charge, which equals beginning shareholders' equity (SE) multiplied by the cost of equity capital (r), gives the following for years 1 through 5:

	Year				
	1	2	3	4	5
RI = (NI + OCI) − ($SE_{t-1} \times r$)	\$1.14	\$0.45	\$2.30	\$2.00	\$2.77

So, the estimated value using the RI model, with residual income based on net income adjusted for OCI, is

$$V_0 = \$8.58 + \frac{\$1.14}{(1.10)^1} + \frac{\$.45}{(1.10)^2} + \frac{\$2.30}{(1.10)^3} + \frac{\$2.00}{(1.10)^4} + \frac{\$2.77}{(1.10)^5} + \frac{\$68.40 - \$22.04}{(1.10)^5}$$

$$V_0 = \$8.58 + 35.01 = \$43.59$$

Solution to 3:

The first calculation (2A) incorrectly omits an adjustment for a violation of the clean surplus relation. The second calculation (2B) includes an adjustment and yields the correct value estimate, which is consistent with the DDM estimate.

Solution to 4:

Given the estimated 8% growth in net income and dividends in Year 6, the estimated Year 6 net income is \$4.92 (\$4.56 × 1.08), and the estimated amount of Year 6 dividends is \$0.42 (\$0.38 × 1.08).

Residual income will then equal \$2.72 (which is net income of \$4.92 minus the equity charge of beginning book value of \$22.04 multiplied by the cost of capital of 10%). So, the growth rate in residual income is negative at approximately −2% (\$2.72/\$2.77 − 1).

Lacking a basis for explicit assumptions about future amounts of OCI, the analyst should nonetheless be aware of the potential effect of OCI on residual income and adjust ROE accordingly. Finally, as noted earlier, the analyst may decide that an alternative valuation model is more appropriate.

ACCOUNTING CONSIDERATIONS: OTHER

6

k describe accounting issues in applying residual income models;

To have a reliable measure of book value of equity, an analyst should identify and scrutinize significant off-balance-sheet assets and liabilities. Additionally, reported assets and liabilities should be adjusted to fair value when possible. Off-balance-sheet assets and liabilities may become apparent through an examination of the financial statement footnotes. Probably the most common example is the use of operating leases. Operating leases do not affect the amount of equity (because leases involve off-balance-sheet assets that offset the off-balance-sheet liabilities) but can affect an assessment of future earnings for the residual income component of value. Other assets and liabilities may be stated at values other than fair value. For example, inventory may be stated at LIFO and require adjustment to restate to current value. (LIFO is not permitted under IFRS.) The following are some common items to review for balance sheet adjustments. Note, however, that this list is not comprehensive:

- inventory;
- deferred tax assets and liabilities;
- operating leases;
- reserves and allowances (for example, bad debts); and
- intangible assets.

Additionally, the analyst should examine the financial statements and footnotes for items unique to the subject company.

6.1 Intangible Assets

Intangible assets can have a significant effect on book value. In the case of specifically identifiable intangibles that can be separated from the entity (e.g., sold), it is generally appropriate to include these in determining book value of equity. If these assets have a finite useful life, they will be amortized over time as an expense. Intangible assets, however, require special consideration because they are often not recognized as an asset unless they are obtained in an acquisition. For example, advertising expenditures can create a highly valuable brand, which is clearly an intangible asset. Advertising expenditures, however, are shown as an expense, and the value of a brand would not appear as an asset on the financial statements unless the company owning the brand was acquired.

To demonstrate this, consider a simplified example involving two companies, Alpha and Beta, with the following summary financial information (all amounts in thousands, except per-share data):

	Alpha (€)	Beta (€)
Cash	1,600	100
Property, plant, and equipment	3,400	900
Total assets	5,000	1,000
Equity	5,000	1,000
Net income	600	150

Each company pays out all net income as dividends (no growth), and the clean surplus relation holds. Alpha has a 12% ROE and Beta has a 15% ROE, both expected to continue indefinitely. Each has a 10% required rate of return. The fair market value of each company's property, plant, and equipment is the same as its book value. What is the value of each company in a residual income framework?

Using total book value rather than per-share data, the value of Alpha would be €6,000, determined as follows (note that result would be the same if calculated on a per-share basis):

$$V_0 = B_0 + \frac{\text{ROE} - r}{r - g} B_0 = 5{,}000 + \frac{0.12 - 0.10}{0.10 - 0.00} 5{,}000 = 6{,}000$$

Similarly, the value of Beta would be €1,500:

$$V_0 = B_0 + \frac{\text{ROE} - r}{r - g} B_0 = 1{,}000 + \frac{0.15 - 0.10}{0.10 - 0.00} 1{,}000 = 1{,}500$$

The value of the companies on a combined basis would be €7,500. Note that both companies are valued more highly than the book value of equity because they have ROE in excess of the required rate of return. Absent an acquisition transaction, the financial statements of Alpha and Beta do not reflect this value. If either is acquired, however, an acquirer would allocate the purchase price to the acquired assets, with any excess of the purchase price above the acquired assets shown as goodwill.

Suppose Alpha acquires Beta by paying Beta's former shareholders €1,500 in cash. Alpha has just paid €500 in excess of the value of Beta's total reported assets of €1,000. Assume that Beta's property, plant and equipment is already shown at its fair market value of €1,000, and that the €500 is considered to be the fair value of a license owned by Beta, say an exclusive right to provide a service. Assume further that the original cost of obtaining the license was an immaterial application fee, which does not appear on Beta's balance sheet, and that the license covers a period of 10 years. Because the entire purchase price of €1,500 is allocated to identifiable assets, no goodwill is recognized. Alpha's balance sheet immediately after the acquisition would be as follows:

	Alpha (€)
Cash	200
Property, plant, and equipment	4,300
License	500
Total assets	5,000
Equity	5,000

Note that the total book value of Alpha's equity did not change, because the acquisition was made for cash and thus did not require Alpha to issue any new shares. Also note that, for example, cash of €200 is calculated as €1,600 (cash of Alpha) + €100 (cash of Beta) – €1,500 (purchase price of Beta).

Under the assumption that the license is amortized over a 10-year period, the combined company's expected net income would be €700 (€600 + €150 – €50 amortization). If this net income number is used to derive expected ROE, the expected ROE would be 14%. Under a residual income model, with no adjustment for amortization, the value of the combined company would be

$$V_0 = B_0 + \frac{\text{ROE} - r}{r - g} B_0 = 5{,}000 + \frac{0.14 - 0.10}{0.10 - 0.00} 5{,}000 = 7{,}000$$

Why would the combined company be worth less than the two separate companies? If the assumption is made that a fair price was paid to Beta's former shareholders, the combined value should not be lower. The lower value using the residual income model results from a reduction in ROE as a result of the amortization of the intangible license asset. If this asset were not amortized (or if the amortization expense were added back before computing ROE), net income would be €750 and ROE would be 15%. The value of the combined entity would be

$$V_0 = B_0 + \frac{\text{ROE} - r}{r - g}B_0 = 5{,}000 + \frac{0.15 - 0.10}{0.10 - 0.00}5{,}000 = 7{,}500$$

This amount, €7,500, is the same as the sum of the values of the companies on a separate basis.

Would the answer be different if the acquiring company used newly issued stock rather than cash in the acquisition? The form of currency used to pay for the transaction should not affect the total value. If Alpha used €1,500 of newly issued stock to acquire Beta, its balance sheet would be as follows:

	Alpha (€)
Cash	1,700
Property, plant, and equipment	4,300
License	500
Total assets	6,500
Equity	6,500

Projected earnings, excluding the amortization of the license, would be €750, and projected ROE would be 11.538%. Value under the residual income model would be

$$V_0 = B_0 + \frac{\text{ROE} - r}{r - g}B_0 = 6{,}500 + \frac{0.11538 - 0.10}{0.10 - 0.00}6{,}500 = 7{,}500$$

The overall value remains unchanged. The book value of equity is higher but offset by the effect on ROE. Once again, this example assumes that the buyer paid a fair value for the acquisition. If an acquirer overpays for an acquisition, the overpayment should become evident in a reduction in future residual income.

Research and development (R&D) costs provide another example of an intangible asset that must be given careful consideration. Under US GAAP, R&D is generally expensed to the income statement directly (except in certain cases such as ASC 985-20-25, which permits the capitalization of R&D expenses related to software development after product feasibility has been established). Also, under IFRS, some R&D costs can be capitalized and amortized over time. R&D expenditures are reflected in a company's ROE, and hence residual income, over the long term. If a company engages in unproductive R&D expenditures, these will lower residual income through the expenditures made. If a company engages in productive R&D expenditures, these should result in higher revenues to offset the expenditures over time. In summary, on a continuing basis for a mature company, ROE should reflect the productivity of R&D expenditures without requiring an adjustment.

As explained in Lundholm and Sloan (2007), including and subsequently amortizing an asset that was omitted from a company's reported assets has no effect on valuation under a residual income model. Such an adjustment would increase the estimated equity value by adding the asset to book value at time zero but decrease the estimated value by an equivalent amount, which would include a) the present value of the asset when amortized in the future and b) the present value of a periodic capital charge based on the amount of the asset multiplied by the cost of equity. Expensing R&D, however, results in an immediately lower ROE vis-à-vis capitalizing R&D. But expensing R&D will result in a slightly higher ROE relative to capitalizing R&D in future years because this capitalized R&D is amortized. Because ROE is used in a number of expressions derived from the residual income model and may also be used in forecasting net income, the analyst should carefully consider a company's R&D expenditures and their effect on long-term ROE.

6.2 Non-recurring Items

In applying a residual income model, it is important to develop a forecast of future residual income based on recurring items. Companies often report non-recurring charges as part of earnings, which can lead to overestimates and underestimates of future residual earnings if no adjustments are made. No adjustments to book value are necessary for these items, however, because non-recurring gains and losses are reflected in the value of assets in place. Hirst and Hopkins (2000) noted that non-recurring items sometimes result from accounting rules and at other times result from "strategic" management decisions. Regardless, they highlighted the importance of examining the financial statement notes and other sources for items that may warrant adjustment in determining recurring earnings, such as

- unusual items;
- extraordinary items (applicable under US GAAP but not under IFRS);
- restructuring charges;
- discontinued operations; and
- accounting changes.

In some cases, management may record restructuring or unusual charges in every period. In these cases, the item may be considered an ordinary operating expense and may not require adjustment.

Companies sometimes inappropriately classify non-operating gains as a reduction in operating expenses (such as selling, general, and administrative expenses). If material, this inappropriate classification can usually be uncovered by a careful reading of financial statement footnotes and press releases. Analysts should consider whether these items are likely to continue and contribute to residual income in time. More likely, they should be removed from operating earnings when forecasting residual income.

6.3 Other Aggressive Accounting Practices

Companies may engage in accounting practices that result in the overstatement of assets (book value) and/or overstatement of earnings. We discussed some of these practices in the preceding sections. Other activities that a company may engage in include accelerating revenues to the current period or deferring expenses to a later period (Schilit and Perler 2010). Both activities simultaneously increase earnings and book value. For example, a company might ship unordered goods to customers at year-end, recording revenues and a receivable. As another example, a company could capitalize rather than expense a cash payment, resulting in lower expenses and an increase in assets.

Conversely, companies have also been criticized for the use of "cookie jar" reserves (reserves saved for future use), in which excess losses or expenses are recorded in an *earlier* period (for example, in conjunction with an acquisition or restructuring) and then used to reduce expenses and increase income in future periods. The analyst should carefully examine the use of reserves when assessing residual earnings. Overall, the analyst must evaluate a company's accounting policies carefully and consider the integrity of management when assessing the inputs in a residual income model.

6.4 International Considerations

Accounting standards differ internationally. These differences result in different measures of book value and earnings internationally and suggest that valuation models based on accrual accounting data might not perform as well as other present value models in international contexts. It is interesting to note, however, that Frankel and Lee (1999) found that the residual income model works well in valuing companies on an international basis. Using a simple residual income model without any of the adjustments discussed here, they found that their residual income valuation model accounted for 70% of the cross-sectional variation of stock prices among 20 countries. Frankel and Lee concluded that there are three primary considerations in applying a residual income model internationally:

- the availability of reliable earnings forecasts;
- systematic violations of the clean surplus assumption; and
- "poor quality" accounting rules that result in delayed recognition of value changes.

Analysts should expect the model to work best in situations in which earnings forecasts are available, clean surplus violations are limited, and accounting rules do not result in delayed recognition. Because Frankel and Lee found good explanatory power for a residual income model using unadjusted accounting data, one expects that if adjustments are made to the reported data to correct for clean surplus and other violations, international comparisons should result in comparable valuations. For circumstances in which clean surplus violations exist, accounting choices result in delayed recognition, or accounting disclosures do not permit adjustment, the residual income model would not be appropriate and the analyst should consider a model less dependent on accounting data, such as a FCFE model.

It should be noted, however, that IFRS is increasingly becoming widely used. As of 2019, according to AICPA (an accociation representing the accounting profession), approximately 120 nations and reporting jurisdictions permit or require IFRS for domestic listed companies, although approximately 90 countries have fully conformed with IFRS as promulgated by the IASB and include a statement acknowledging such conformity in audit reports. Furthermore, standard setters in numerous countries continue to work toward convergence between IFRS and home-country GAAP. In time, concerns about the use of different accounting standards should become less severe. Nonetheless, even within a single set of accounting standards, companies make choices and estimates that can affect valuation.

SUMMARY

We have discussed the use of residual income models in valuation. Residual income is an appealing economic concept because it attempts to measure economic profit, which are profits after accounting for all opportunity costs of capital.

- Residual income is calculated as net income minus a deduction for the cost of equity capital. The deduction, called the equity charge, is equal to equity capital multiplied by the required rate of return on equity (the cost of equity capital in percent).
- Economic value added (EVA) is a commercial implementation of the residual income concept. EVA = NOPAT − (C% × TC), where NOPAT is net operating profit after taxes, C% is the percent cost of capital, and TC is total capital.

- Residual income models (including commercial implementations) are used not only for equity valuation but also to measure internal corporate performance and for determining executive compensation.
- We can forecast per-share residual income as forecasted earnings per share minus the required rate of return on equity multiplied by beginning book value per share. Alternatively, per-share residual income can be forecasted as beginning book value per share multiplied by the difference between forecasted ROE and the required rate of return on equity.
- In the residual income model, the intrinsic value of a share of common stock is the sum of book value per share and the present value of expected future per-share residual income. In the residual income model, the equivalent mathematical expressions for intrinsic value of a common stock are

$$V_0 = B_0 + \sum_{t=1}^{\infty} \frac{\text{RI}_t}{(1+r)^t} = B_0 + \sum_{t=1}^{\infty} \frac{E_t - rB_{t-1}}{(1+r)^t}$$

$$= B_0 + \sum_{t=1}^{\infty} \frac{(\text{ROE}_t - r)B_{t-1}}{(1+r)^t}$$

where

V_0 = value of a share of stock today (t = 0)
B_0 = current per-share book value of equity
B_t = expected per-share book value of equity at any time t
r = required rate of return on equity (cost of equity)
E_t = expected earnings per share for period t
RI_t = expected per-share residual income, equal to $E_t - rB_{t-1}$ or to (ROE − r) × B_{t-1}
ROE_T = return on equity

- In the two-stage model with continuing residual income in stage two, the intrinsic value of a share of stock is

$$V_0 = B_0 + \sum_{t=1}^{T} \frac{\text{RI}_t}{(1+r)^t} + \frac{P_T - B_T}{(1+r)^T} = B_0 + \sum_{t=1}^{T} \frac{(E_t - rB_{t-1})}{(1+r)^t} + \frac{P_T - B_T}{(1+r)^T}$$

$$V_0 = B_0 + \sum_{t=1}^{T} \frac{(\text{ROE}_t - r)B_{t-1}}{(1+r)^t} + \frac{P_T - B_T}{(1+r)^T}$$

where

P_T = expected per share price at terminal time T
B_T = expected per share book value at terminal time T

- In most cases, value is recognized earlier in the residual income model compared with other present value models of stock value, such as the dividend discount model.
- Strengths of the residual income model include the following:
  - Terminal values do not make up a large portion of the value relative to other models.
  - The models use readily available accounting data.
  - The models can be used in the absence of dividends and near-term positive free cash flows.

- The models can be used when cash flows are unpredictable.

- Weaknesses of the residual income model include the following:
  - The models are based on accounting data that can be subject to manipulation by management.
  - Accounting data used as inputs may require significant adjustments.
  - The models require that the clean surplus relation holds, or that the analyst makes appropriate adjustments when the clean surplus relation does not hold.
- The residual income model is most appropriate in the following cases:
  - A company is not paying dividends or if it exhibits an unpredictable dividend pattern.
  - A company has negative free cash flow many years out but is expected to generate positive cash flow at some point in the future.
  - A great deal of uncertainty exists in forecasting terminal values.
- The fundamental determinants or drivers of residual income are book value of equity and return on equity.
- Residual income valuation is most closely related to P/B. When the present value of expected future residual income is positive (negative), the justified P/B based on fundamentals is greater than (less than) one.
- When fully consistent assumptions are used to forecast earnings, cash flow, dividends, book value, and residual income through a full set of pro forma (projected) financial statements, and the same required rate of return on equity is used as the discount rate, the same estimate of value should result from a residual income, dividend discount, or free cash flow valuation. In practice, however, analysts may find one model easier to apply and possibly arrive at different valuations using the different models.
- Continuing residual income is residual income after the forecast horizon. Frequently, one of the following assumptions concerning continuing residual income is made:
  - Residual income continues indefinitely at a positive level. (One variation of this assumption is that residual income continues indefinitely at the rate of inflation, meaning it is constant in real terms.)
  - Residual income is zero from the terminal year forward.
  - Residual income declines to zero as ROE reverts to the cost of equity over time.
  - Residual income declines to some mean level.
- The residual income model assumes the clean surplus relation of $B_t = B_{t-1} + E_t - D_t$. In other terms, the ending book value of equity equals the beginning book value plus earnings minus dividends, apart from ownership transactions.
- In practice, to apply the residual income model most accurately, the analyst may need to do the following:
  - adjust book value of common equity for:
    - off-balance-sheet items;
    - discrepancies from fair value; or
    - the amortization of certain intangible assets.
  - adjust reported net income to reflect clean surplus accounting.
  - adjust reported net income for non-recurring items misclassified as recurring items.

REFERENCES

Bauman, Mark P. 1999. "Importance of Reported Book Value in Equity Valuation." *Journal of Financial Statement Analysis*, vol. 4, no. 2:31–40.

Bernstein, Richard, and Carmen Pigler. 1997. "An Analysis of EVA®." *Quantitative Viewpoint*. Merrill Lynch. 19 December.

Bernstein, Richard, Kari Bayer, and Carmen Pigler. 1998. "An Analysis of EVA® Part II." *Quantitative Viewpoint*. Merrill Lynch. 3 February.

Dechow, Patricia M., Amy P. Hutton, and Richard G. Sloan. 1999. "An Empirical Assessment of the Residual Income Valuation Model." *Journal of Accounting and Economics*, vol. 26, no. 1-3:1–34.

Feltham, Gerald A., and James A. Ohlson. 1995. "Valuation and Clean Surplus Accounting for Operating and Financial Activities." *Contemporary Accounting Research*, vol. 11, no. 4:689–731.

Fleck, Shelby A., Scott D. Craig, Michael Bodenstab, Trevor Harris, and Elmer Huh. 2001. *Technology: Electronics Manufacturing Services*. Industry Overview; Morgan Stanley Dean Witter. 28 March.

Frankel, Richard M., and Charles M.C. Lee. 1999. "Accounting Diversity and International Valuation." Working Paper, May.

Hirst, D. Eric, and Patrick E. Hopkins. 2000. *Earnings: Measurement, Disclosure, and the Impact on Equity Valuation*. Charlottesville, VA: Research Foundation of AIMR and Blackwell Series in Finance.

Hitchner, James R. 2017. *Financial Valuation: Applications and Models*, 4th edition. Hoboken, NJ: John Wiley & Sons.

Lee, Charles M.C., and Bhaskaran Swaminathan. 1999. "Valuing the Dow: A Bottom-Up Approach." *Financial Analysts Journal*, vol. 55, no. 5:4–23.

Lee, Charles M.C., James Myers, and Bhaskaran Swaminathan. 1999. "What is the Intrinsic Value of the Dow?" *Journal of Finance*, vol. 54, no. 5:1693–1741.

Lo, Kin, and Thomas Lys. 2000. "The Ohlson Model: Contributions to Valuation Theory, Limitations, and Empirical Applications." *Journal of Accounting, Auditing & Finance*, vol. 15, no. 3:337–367.

Lundholm, Russell J., and Terrence B. O'Keefe. 2001a. "Reconciling Value Estimates from the Discounted Cash Flow Model and the Residual Income Model." *Contemporary Accounting Research*, vol. 18, no. 2:311–335.

Lundholm, Russell J., and Terrence B. O'Keefe. 2001b. "On Comparing Residual Income and Discounted Cash Flow Models of Equity Valuation: A Response to Penman 2001." *Contemporary Accounting Research*, vol. 18, no. 4:693–696.

Lundholm, Russell J., and Richard G. Sloan. 2007. *Equity Valuation and Analysis with eVal*, 2nd edition. McGraw-Hill Irwin. New York.

Marshall, Alfred"Income, Capital," in Principles of Economics, Book Two: Some Fundamental Notions, Chapter 4. London: Macmillan and Co., Ltd.).

Ohlson, James A. 1995. "Earnings, Book Values, and Dividends in Equity Valuation." *Contemporary Accounting Research*, vol. 11, no. 4:661–687.

Penman, Stephen H. 2001. "On Comparing Cash Flow and Accrual Accounting Models for Use in Equity Valuation: A Response to Lundholm and O'Keefe." *Contemporary Accounting Research*, vol. 18, no. 4:681–692.

Penman, Stephen H., and Theodore Sougiannis. 1998. "A Comparison of Dividend, Cash Flow and Earnings Approaches to Equity Valuation." *Contemporary Accounting Research*, vol. 15, no. 3:343–383.

Peterson, Pamela P., and David R. Peterson. 1996. *Company Performance and Measures of Value Added*. Charlottesville, VA: The Research Foundation of the ICFA.

Schilit and Perler. 2010. *Financial Shenanigans*. 3rd ed., McGraw-Hill.

Shrieves, Ronald E., and John M. Wachowicz, Jr. 2001. "Free Cash Flow (FCF), Economic Value Added (EVA™), and Net Present Value (NPV): A Reconciliation of Variations of Discounted-Cash-Flow (DCF) Valuation." *Engineering Economist*, vol. 46, no. 1:33–52.

Stewart, G. Bennett III. 1991. *The Quest for Value*. New York: HarperCollins.

Tobin, James. 1969. "A General Equilibrium Approach to Monetary Theory." *Journal of Money, Credit and Banking*, vol. 1, no. 1:15–29.

Young, S. David. 1999. "Some Reflections on Accounting Adjustments and Economic Value Added." *Journal of Financial Statement Analysis*, vol. 4, no. 2:7–19.

PRACTICE PROBLEMS

1 Based on the following information, determine whether Vertically Integrated Manufacturing (VIM) earned any residual income for its shareholders:

- VIM had total assets of $3,000,000, financed with twice as much debt capital as equity capital.
- VIM's pretax cost of debt is 6% and cost of equity capital is 10%.
- VIM had EBIT of $300,000 and was taxed at a rate of 40%.

Calculate residual income by using the method based on deducting an equity charge.

2 Use the following information to estimate the intrinsic value of VIM's common stock using the residual income model:

- VIM had total assets of $3,000,000, financed with twice as much debt capital as equity capital.
- VIM's pretax cost of debt is 6% and cost of equity capital is 10%.
- VIM had EBIT of $300,000 and was taxed at a rate of 40%. EBIT is expected to continue at $300,000 indefinitely.
- VIM's book value per share is $20.
- VIM has 50,000 shares of common stock outstanding.

3 Palmetto Steel, Inc. (PSI) maintains a dividend payout ratio of 80% because of its limited opportunities for expansion. Its return on equity is 15%. The required rate of return on PSI equity is 12%, and its long-term growth rate is 3%. Compute the justified P/B based on forecasted fundamentals, consistent with the residual income model and a constant growth rate assumption.

4 Because New Market Products (NMP) markets consumer staples, it is able to make use of considerable debt in its capital structure; specifically, 90% of the company's total assets of $450,000,000 are financed with debt capital. Its cost of debt is 8% before taxes, and its cost of equity capital is 12%. NMP achieved a pretax income of $5.1 million in 2006 and had a tax rate of 40%. What was NMP's residual income?

5 In 2020, Smithson–Williams Industries (SWI) achieved an operating profit after taxes of €10 million on total assets of €100 million. Half of its assets were financed with debt with a pretax cost of 9%. Its cost of equity capital is 12%, and its tax rate is 40%. Did SWI achieve a positive residual income?

6 Calculate the economic value added or residual income, as requested, for each of the following:

A NOPAT = $100

Beginning book value of debt = $200

Beginning book value of equity = $300

Weighted average cost of capital (WACC) = 11%

Calculate EVA.

B Net income = €5.00

Dividends = €1.00

Beginning book value of equity = €30.00

Required rate of return on equity = 11%

© 2017 CFA Institute. All rights reserved.

Calculate residual income.

C Return on equity = 18%

Required rate of return on equity = 12%

Beginning book value of equity = €30.00

Calculate residual income.

7 Jim Martin is using economic value added and market value added to measure the performance of Sundanci. Martin uses the fiscal year 2020 information below for his analysis.

- Adjusted net operating profit after taxes is $100 million.
- Total capital is $700 million (no debt).
- Closing stock price is $26.
- Total shares outstanding is 84 million.
- The cost of equity is 14%.

Calculate the following for Sundanci. Show your work.

A EVA for fiscal year 2020.

B MVA as of fiscal year-end 2020.

8 Protected Steel Corporation (PSC) has a book value of $6 per share. PSC is expected to earn $0.60 per share forever and pays out all of its earnings as dividends. The required rate of return on PSC's equity is 12%. Calculate the value of the stock using the following:

A Dividend discount model.

B Residual income model.

9 Notable Books (NB) is a family controlled company that dominates the retail book market. NB has book value of $10 per share, is expected to earn $2.00 per share forever, and pays out all of its earnings as dividends. Its required return on equity is 12.5%. Value the stock of NB using the following:

A Dividend discount model.

B Residual income model.

10 Simonson Investment Trust International (SITI) is expected to earn $4.00, $5.00, and $8.00 per share for the next three years. SITI will pay annual dividends of $2.00, $2.50, and $20.50 in each of these years. The last dividend includes a liquidating payment to shareholders at the end of Year 3 when the trust terminates. SITI's book value is $8 per share and its required return on equity is 10%.

A What is the current value per share of SITI according to the dividend discount model?

B Calculate per-share book value and residual income for SITI for each of the next three years and use those results to find the stock's value using the residual income model.

C Calculate return on equity and use it as an input to the residual income model to calculate SITI's value.

11 Foodsco Incorporated (FI), a leading distributor of food products and materials to restaurants and other institutions, has a remarkably steady track record in terms of both return on equity and growth. At year-end 2017, FI had a book value of $30 per share. For the foreseeable future, the company is expected to achieve a ROE of 15% (on trailing book value) and to pay out one-third of its earnings in dividends. The required return is 12%. Forecast FI's residual income for the year ending 31 December 2022.

12 Lendex Electronics (LE) had a great deal of turnover of top management for several years and was not followed by analysts during this period of turmoil. Because the company's performance has been improving steadily for the past three years, technology analyst Stephanie Kent recently reinitiated coverage of LE. A meeting with management confirms Kent's positive impression of LE's operations and strategic plan. Kent decides LE merits further analysis.

Careful examination of LE's financial statements revealed that the company had negative other comprehensive income from changes in the value of available-for-sale securities in each of the past five years. How, if at all, should this observation about LE's other comprehensive income affect the figures that Kent uses for the company's ROE and book value for those years?

13 Retail fund manager Seymour Simms is considering the purchase of shares in upstart retailer Hottest Topic Stores (HTR). The current book value of HTS is \$20 per share, and its market price is \$35. Simms expects long-term ROE to be 18%, long-term growth to be 10%, and cost of equity to be 14%. What conclusion would you expect Simms to arrive at if he uses a single-stage residual income model to value these shares?

14 Dayton Manufactured Homes (DMH) builds prefabricated homes and mobile homes. Favorable demographics and the likelihood of slow, steady increases in market share should enable DMH to maintain its ROE of 15% and growth rate of 10% through time. DMH has a book value of \$30 per share and the required rate of return on its equity is 12%. Compute the value of its equity using the single-stage residual income model.

15 Use the following inputs and the finite horizon form of the residual income model to compute the value of Southern Trust Bank (STB) shares as of 31 December 2020:

- ROE will continue at 15% for the next five years (and 10% thereafter) with all earnings reinvested (no dividends paid).
- Cost of equity equals 10%.
- B_0 = \$10 per share (at year-end 2020).
- Premium over book value at the end of five years will be 20%.

16 Shunichi Kobayashi is valuing Procter & Gamble Company (NYSE: PG). Kobayashi has made the following assumptions:

- Book value per share is estimated at \$21.30 on 31 March 2019.
- EPS will be 18% of the beginning book value per share for the next eight years.
- Cash dividends paid will be 70% of EPS.
- At the end of the eight-year period, the market price per share will be four times the book value per share.
- The beta for PG is 0.50, the risk-free rate is 2.0%, and the equity risk premium is 6.2%.

The current market price of PG is \$107.50, which indicates a current P/B of 5.05.

A Prepare a table that shows the beginning and ending book values, net income, and cash dividends annually for the eight-year period.

B Estimate the residual income and the present value of residual income for the eight years.

C Estimate the value per share of PG stock using the residual income model.

D Estimate the value per share of PG stock using the dividend discount model. How does this value compare with the estimate from the residual income model?

17 Thales S.A. (Paris: HO.PA) has a current stock price of €98.73. It also has book value per share of €26.83. and a P/B of 3.68. Assume that the single-stage growth model is appropriate for valuing the company. Thales S.A.'s adjusted beta is 0.68, the risk-free rate is 4.46%, and the equity risk premium is 5.50%.

A If the growth rate is 5.50% and the ROE is 20%, what is the justified P/B for Thales?

B If the growth rate is 5.50%, what ROE is required to yield Thales S.A.'s current P/B?

C If the ROE is 20%, what growth rate is required for Thales to have its current P/B?

18 Consider the following information about Industrias Gómez.

- Current book value per share is €20.00.
- Expected earnings per share for the next five years are €1.50, €2.50, €3.50, €4.50, and €5.50.
- Dividends per share are projected to be €1.00 for the first three years and €2.00 for the last two years.
- The terminal share price (at the end of Year 5) is expected to be 14× trailing earnings.
- The required rate of return on equity is 9%.
- Estimate the residual income each year, the terminal residual value, and the value per share of Industrias Gómez shares using the residual income model.
- Estimate the value per share of Industrias Gómez shares using the dividend discount model.

The following information relates to Questions 19–26

Mangoba Nkomo, CFA, a senior equity analyst with Robertson-Butler Investments, South Africa, has been assigned a recent graduate, Manga Mahlangu, to assist in valuations. Mahlangu is interested in pursuing a career in equity analysis. In their first meeting, Nkomo and Mahlangu discuss the concept of residual income and its commercial applications. Nkomo asks Mahlangu to determine the market value added for a hypothetical South African firm using the data provided in Exhibit 1.

Exhibit 1 Hypothetical Firm Data (amounts in South African rand)

Current share price	R25.43
Book value per share	R20.00
Total shares outstanding	30 million
Cost of equity	13%
Market value of debt	R55 million
Accounting book value of total capital	R650 million
Intrinsic share value of equity derived from residual income model	R22.00

Nkomo also shares his valuation report of the hypothetical firm with Mahlangu. Nkomo's report concludes that the intrinsic value of the hypothetical firm, based on the residual income model, is R22.00 per share. To assess Mahlangu's knowledge of residual income valuation, Nkomo asks Mahlangu two questions about the hypothetical firm:

Question 1 What conclusion can we make about future residual earnings given the current book value per share and my estimate of intrinsic value per share?

Question 2 Suppose you estimated the intrinsic value of a firm's shares using a constant growth residual income model, and you found that your estimate of intrinsic value equaled the book value per share. What would that finding imply about that firm's return on equity?

Satisfied with Mahlangu's response, Nkomo requests that Mahlangu use the single-stage residual income model to determine the intrinsic value of the equity of Jackson Breweries, a brewery and bottling company, using data provided in Exhibit 2.

Exhibit 2 Jackson Breweries Data (amounts in South African rand)

Constant long-term growth rate	9.5%
Constant long-term ROE	13%
Current market price per share	R150.70
Book value per share	R55.81
Cost of equity	11%

Nkomo also wants to update an earlier valuation of Amersheen, a food retailer. The valuation report, completed at the end of 2020, concluded an intrinsic value per share of R11.00 for Amersheen. The share price at that time was R8.25. Nkomo points out to Mahlangu that in late 2020, Amersheen announced a significant restructuring charge, estimated at R2 million, that would be reported as part of operating earnings in Amersheen's 2020 annual income statement. Nkomo asks Mahlangu the following question about the restructuring charge:

Question 3 What was the correct way to treat the estimated R2 million restructuring charge in my 2020 valuation report?

Satisfied with Mahlangu's response, Nkomo mentions to Mahlangu that Amersheen recently (near the end of 2021) completed the acquisition of a chain of convenience stores. Nkomo requests that Mahlangu complete, as of the beginning of 2022, an updated valuation of Amersheen under two scenarios:

Scenario 1 Estimate the value of Amersheen shares using a multistage residual income model with the data provided in Exhibit 3. Under Scenario 1, expected ROE in 2025 is 26%, but it is assumed that the firm's ROE will slowly decline towards the cost of equity thereafter.

Scenario 2 Estimate the value of Amersheen shares using a multistage residual income model with the data provided in Exhibit 3, but assume that at the end of 2024, share price is expected to equal book value per share.

Exhibit 3 Amersheen Data (amounts in South African rand)

Long-term growth rate starting in 2025	9.0%
Expected ROE in 2025	26%
Current market price per share	R16.55
Book value per share, beginning of 2022	R7.60
Cost of equity	10%
Persistence factor	0.70

	2022	2023	2024
Expected earnings per share	R3.28	R3.15	R2.90
Expected dividend per share	R2.46	R2.36	R2.06

19 Based on the information in Exhibit 1, the market value added of the hypothetical firm is *closest* to:

A R65 million.

B R113 million.

C R168 million.

20 The *most* appropriate response to Nkomo's Question 1 would be that the present value of future residual earnings is expected to be:

A zero.

B positive.

C negative.

21 The *most* appropriate response to Nkomo's Question 2 would be that the firm's return on equity is:

A equal to the firm's cost of equity.

B lower than the firm's cost of equity.

C higher than the firm's cost of equity.

22 Based on the information in Exhibit 2, the intrinsic value per share of the equity of Jackson Breweries is *closest* to:

A R97.67.

B R130.22.

C R186.03.

23 If Nkomo's 2020 year-end estimate of Amersheen shares' intrinsic value was accurate, then Amersheen's shares were *most likely*:

A overvalued.

B undervalued.

C fairly valued.

24 The *most* appropriate treatment of the estimated restructuring charge, in response to Nkomo's Question 3, would be:

A an upward adjustment to book value.

B an upward adjustment to the cost of equity.

C to exclude it from the estimate of net income.

25 Under Scenario 1, the intrinsic value per share of the equity of Amersheen is *closest* to:

A R13.29.

B R15.57.

C R16.31.

26 Under Scenario 2, the intrinsic value per share of the equity of Amersheen is *closest* to:

A R13.29.

B R15.57.

C R16.31.

The following information relates to Questions 27–36

Elena Castovan is a junior analyst with Contralith Capital, a long-only equity investment manager. She has been asked to value three stocks on Contralith's watch list: Portous, Inc. (PTU), SSX Financial (SSX), and Tantechi Ltd. (TTCI).

During their weekly meeting, Castovan and her supervisor, Ariana Beckworth, discuss characteristics of residual income models. Castovan tells Beckworth the following.

Statement 1 The present value of the terminal value in RI models is often a larger portion of the total intrinsic value than it is in other DCF valuation models.

Statement 2 The RI model's use of accounting income assumes that the cost of debt capital is appropriately reflected by interest expense.

Statement 3 RI models cannot be readily applied to companies that do not have positive expected near-term free cash flows.

Beckworth asks Castovan why an RI model may be more appropriate for valuing PTU than the dividend discount model or a free cash flow model. Castovan tells Beckworth that, over her five-year forecast horizon, she expects PTU to perform the following actions.

Reason 1 Pay dividends that are unpredictable

Reason 2 Generate positive and fairly predictable free cash flows

Reason 3 Report significant amounts of other comprehensive income

At the conclusion of their meeting, Beckworth asks Castovan to value SSX using RI models. Selected financial information on SSX is presented in Exhibit 1.

Exhibit 1 SSX Financial (SSX) Selected Financial Data

Total assets (millions)	€4,000.00
Capital structure	60% debt/40% equity
EBIT (millions)	€700.00
Tax rate	35.00%
Return on equity (ROE)	23.37%
Pretax cost of debt[a]	5.20%

(continued)

Exhibit 1 (Continued)	
Cost of equity	15.00%
Market price per share	€48.80
Price-to-book ratio	2.10

[a] Interest expense is tax-deductible.

Castovan's final assignment is to determine the intrinsic value of TTCI using both a single-stage and a multistage RI model. Selected data and assumptions for TTCI are presented in Exhibit 2.

Exhibit 2 Tantechi Ltd. (TTCI) Selected Financial Data and Assumptions	
Book value per share	€45.25
Market price per share	€126.05
Constant long-term ROE	12.00%
Constant long-term earnings growth rate	4.50%
Cost of equity	8.70%

For the multistage model, Castovan forecasts TTCI's ROE to be higher than its long-term ROE for the first three years. Forecasted earnings per share and dividends per share for TTCI are presented in Exhibit 3. Starting in Year 4, Castovan forecasts TTCI's ROE to revert to the constant long-term ROE of 12% annually. The terminal value is based on an assumption that residual income per share will be constant from Year 3 into perpetuity.

Exhibit 3 Tantechi Ltd. (TTCI) Forecasts of Earnings and Dividends

	Year 1	Year 2	Year 3
Earnings per share (€)	7.82	8.17	8.54
Dividends per share (€)	1.46	1.53	1.59

Beckworth questions Castovan's assumption regarding the implied persistence factor used in the multistage RI valuation. She tells Castovan that she believes that a persistence factor of 0.10 is appropriate for TTCI.

27 Which of Castovan's statements regarding residual income models is correct?

A Statement 1

B Statement 2

C Statement 3

28 Which of Castovan's reasons *best* justifies the use of a residual income model to value PTU?

A Reason 1

B Reason 2

C Reason 3

29 The forecasted item described in Reason 3 will *most likely* affect:

A earnings per share.

B dividends per share.

C book value per share.

30 Based on Exhibit 1, residual income for SSX is *closest* to:

A €40.9 million.

B €90.2 million.

C €133.9 million.

31 Based on Exhibit 1 and the single-stage residual income model, the implied growth rate of earnings for SSX is *closest* to:

A 5.8%.

B 7.4%.

C 11.0%.

32 Based on the single-stage RI model and Exhibit 2, Castovan should conclude that TTCI is:

A undervalued.

B fairly valued.

C overvalued.

33 Based on Exhibit 2, the justified price-to-book ratio for TTCI is *closest* to:

A 1.79.

B 2.27.

C 2.79.

34 Based on Exhibits 2 and 3 and the multistage RI model, Castovan should estimate the intrinsic value of TTCI to be *closest* to:

A €54.88.

B €83.01.

C €85.71.

35 The persistence factor suggested by Beckworth will lead to a multistage value estimate of TTCI's shares that is:

A less than Castovan's multistage value estimate.

B equal to Castovan's multistage value estimate.

C greater than Castovan's multistage value estimate.

36 The *best* justification for Castovan to use Beckworth's suggested persistence factor is that TTCI has:

A a low dividend payout.

B extreme accounting rates of return.

C a strong market leadership position.

SOLUTIONS

1 Yes, VIM earned a positive residual income of \$8,000.

EBIT	\$300,000	
Interest	120,000	(\$2,000,000 × 6%)
Pretax income	\$180,000	
Tax expense	72,000	
Net income	\$108,000	

$$\begin{aligned}\text{Equity charge} &= \text{Equity capital} \times \text{Required return on equity}\\ &= (1/3)(\$3{,}000{,}000) \times 0.10\\ &= \$1{,}000{,}000 \times 0.10 = \$100{,}000\end{aligned}$$

$$\begin{aligned}\text{Residual income} &= \text{Net income} - \text{Equity charge}\\ &= \$108{,}000 - \$100{,}000 = \$8{,}000\end{aligned}$$

2 According to the residual income model, the intrinsic value of a share of common stock equals book value per share plus the present value of expected future per-share residual income. Book value per share was given as \$20. If we note that debt is \$2,000,000 [(2/3)(\$3,000,000)] so that interest is \$120,000 (\$2,000,000 × 6%), VIM's residual income is \$8,000, which is calculated (as in Problem 1) as follows:

$$\begin{aligned}\text{Residual income} &= \text{Net income} - \text{Equity charge}\\ &= [(\text{EBIT} - \text{Interest})(1 - \text{Tax rate})] - [(\text{Equity capital})\\ &\quad (\text{Required return on equity})]\\ &= [(\$300{,}000 - \$120{,}000)(1 - 0.40)] - [(\$1{,}000{,}000)(0.10)]\\ &= \$108{,}000 - \$100{,}000\\ &= \$8{,}000\end{aligned}$$

Therefore, residual income per share is \$0.16 per share (\$8,000/50,000 shares). Because EBIT is expected to continue at the current level indefinitely, the expected per-share residual income of \$0.16 is treated as a perpetuity. The present value of \$0.16 is discounted at the required return on equity of 10%, so the present value of the residual income is \$1.60 (\$0.16/0.10).

$$\begin{aligned}\text{Intrinsic value} &= \text{Book value per share} +\\ &\quad \text{PV of expected future income per-share residual income}\\ &= \$20 + \$1.60 = \$21.60\end{aligned}$$

3 With $g = b \times \text{ROE} = (1 - 0.80)\ (0.15) = (0.20)\ (0.15) = 0.03$,

$$\begin{aligned}\text{P/B} &= (\text{ROE} - g)/(r - g)\\ &= (0.15 - 0.03)/(0.12 - 0.03)\\ &= 0.12/0.09 = 1.33\end{aligned}$$

or

$$\begin{aligned}\text{P/B} &= 1 + (\text{ROE} - r)/(r - g)\\ &= 1 + (0.15 - 0.12)/(0.12 - 0.03)\\ &= 1.33\end{aligned}$$

4 In this problem (unlike Problems 1 and 2), interest expense has already been deducted in arriving at NMP's pretax income of \$5.1 million.

Therefore,

$$\begin{aligned}\text{Net income} &= \text{Pretax income} \times (1 - \text{Tax rate})\\ &= \$5.1 \text{ million} \times (1 - 0.4)\\ &= \$5.1 \times 0.6 = \$3.06 \text{ million}\\ \text{Equity charge} &= \text{Total equity} \times \text{Cost of equity capital}\\ &= (0.1 \times \$450 \text{ million}) \times 12\%\\ &= \$45 \text{ million} \times 0.12 = \$5{,}400{,}000\\ \text{Residual income} &= \text{Net income} - \text{Equity charge}\\ &= \$3{,}0600{,}000 - \$5{,}400{,}000 = -\$2{,}340{,}000\end{aligned}$$

NMP had negative residual income of –$2,340,000.

5 To achieve a positive residual income, a company's net operating profit after taxes as a percentage of its total assets can be compared with its weighted average cost of capital. For SWI,

$$\begin{aligned}\text{NOPAT/Assets} &= €10 \text{ million}/€100 \text{ million} = 10\%\\ \text{WACC} &= \text{Percent of debt} \times \text{After-tax cost of debt} +\\ &\quad\ \text{Percent of equity} \times \text{Cost of equity}\\ &= (0.5)(0.09)(0.6) + (0.5)(0.12)\\ &= (0.5)(0.054) + (0.5)(0.12) = 0.027 + 0.06 = 0.087\\ &= 8.7\%\end{aligned}$$

Therefore, SWI's residual income was positive. Specifically, residual income equals €1.3 million [(0.10 – 0.087) × €100 million].

6 A $\text{EVA} = \text{NOPAT} - \text{WACC} \times \text{Beginning book value of assets}$

$$= \$100 - (11\%) \times (\$200 + \$300) = \$100 - (11\%)(\$500) = \$45$$

B $\text{RI}_t = E_t - rB_{t-1}$

$$= €5.00 - (11\%)(€30.00) = €5.00 - €3.30 = €1.70$$

C $\text{RI}_t = (\text{ROE}_t - r) \times B_{t-1}$

$$= (18\% - 12\%) \times (€30) = €1.80$$

7 A Economic value added = Net operating profit after taxes – (Cost of capital × Total capital) = $100 million – (14% × $700 million) = $2 million. In the absence of information that would be required to calculate the weighted average cost of debt and equity, and given that Sundanci has no long-term debt, the only capital cost used is the required rate of return on equity of 14%.

B Market value added = Market value of capital – Total capital = $26 stock price × 84 million shares – $700 million = $1.48 billion

8 A Because the dividend is a perpetuity, the no-growth form of the DDM is applied as follows:

$$\begin{aligned}V_0 &= D/r\\ &= \$0.60/0.12 = \$5 \text{ per share}\end{aligned}$$

B According to the residual income model, V_0 = Book value per share + Present value of expected future per-share residual income.

Residual income is calculated as:

$$\begin{aligned}\text{RI}_t &= E - rB_{t-1}\\ &= \$0.60 - (0.12)(\$6) = -\$0.12\end{aligned}$$

Present value of perpetual stream of residual income is calculated as:

$$\text{RI}_t/r = -\$0.12/0.12 = -\$1.00$$

The value is calculated as:

$$V_0 = \$6.00 - \$1.00 = \$5.00 \text{ per share}$$

9 **A** According to the DDM, $V_0 = D/r$ for a no-growth company.

$$V_0 = \$2.00/0.125 = \$16 \text{ per share}$$

B Under the residual income model, $V_0 = B_0$ + Present value of expected future per-share residual income.

Residual income is calculated as:

$$\begin{aligned} RI_t &= E - rB_{t-1} \\ &= \$2 - (0.125)(\$10) = \$0.75 \end{aligned}$$

Present value of stream of residual income is calculated as:

$$RI_t/r = 0.75/0.125 = \$6$$

The value is calculated as:

$$V_0 = \$10 + \$6 = \$16 \text{ per share}$$

10 **A** V_0 = Present value of the future dividends

$$\begin{aligned} &= \$2/1.10 + \$2.50\big/(1.1)^2 + \$20.50\big/(1.1)^3 \\ &= \$1.818 + \$2.066 + \$15.402 = \$19.286 \end{aligned}$$

B The book values and residual incomes for the next three years are as follows:

Year	1	2	3
Beginning book value	\$ 8.00	\$10.00	\$12.50
Retained earnings (Net income – Dividends)	2.00	2.50	(12.50)
Ending book value	\$10.00	\$12.50	\$ 0.00
Net income	\$ 4.00	\$ 5.00	\$ 8.00
Less equity charge (r × Book value)	0.80	1.00	1.25
Residual income	\$ 3.20	\$ 4.00	\$ 6.75

Under the residual income model,

$$\begin{aligned} V_0 &= B_0 + \text{Present value of expected future per-share residual income} \\ V_0 &= \$8.00 + \$3.20/1.1 + \$4.00/(1.1)^2 + \$6.75/(1.1)^3 \\ V_0 &= 8.00 + \$2.909 + \$3.306 + \$5.071 = \$19.286 \end{aligned}$$

C

Year	1	2	3
Net income (NI)	\$4.00	\$5.00	\$8.00
Beginning book value (BV)	8.00	10.00	12.50
Return on equity (ROE) = NI/BV	50%	50%	64%
ROE – r	40%	40%	54%
Residual income (ROE – r) × BV	\$3.20	\$4.00	\$6.75

Under the residual income model,

$$\begin{aligned} V_0 &= B_0 + \text{Present value of expected future per-share residual income} \\ V_0 &= \$8.00 + \$3.20/1.1 + \$4.00/(1.1)^2 + \$6.75/(1.1)^3 \\ V_0 &= 8.00 + \$2.909 + \$3.306 + \$5.071 = \$19.286 \end{aligned}$$

Note: Because the residual incomes for each year are necessarily the same in Parts B and C, the results for stock valuation are identical.

11

Year	2018	2019	2022
Beginning book value	$30.00	$33.00	$43.92
Net income = ROE × Book value	4.50	4.95	6.59
Dividends = payout × Net income	1.50	1.65	2.20
Equity charge (r × Book value)	3.60	3.96	5.27
Residual income = Net income − Equity charge	0.90	0.99	1.32
Ending book value	$33.00	$36.30	$48.32

The table shows that residual income in Year 2018 is $0.90, which equals Beginning book value × (ROE − r) = $30 × (0.15 − 0.12). The Year 2019 column shows that residual income grew by 10% to $0.99, which follows from the fact that growth in residual income relates directly to the growth in net income as this example is configured. When both net income and dividends are a function of book value and return on equity is constant, then growth, g, can be predicted from (ROE)(1 − Dividend payout ratio). In this case, g = 0.15 × (1 − 0.333) = 0.10 or 10%. Net income and residual income will grow by 10% annually.

Therefore, residual income in Year 2022 = (Residual income in Year 2018) × $(1.1)^4$ = 0.90 × 1.4641 = $1.32.

12 When such items as changes in the value of available-for-sale securities bypass the income statement, they are generally assumed to be nonoperating items that will fluctuate from year to year, although averaging to zero in a period of years. The evidence suggests, however, that changes in the value of available-for-sale securities are not averaging to zero but are persistently negative. Furthermore, these losses are bypassing the income statement. It appears that the company is either making an inaccurate assumption or misleading investors in one way or another. Accordingly, Kent might adjust LE's income downward by the amount of loss for other comprehensive income for each of those years. ROE would then decline commensurately. LE's book value would *not* be misstated because the decline in the value of these securities was already recognized and appears in the shareholders' equity account "Accumulated Other Comprehensive Income."

13
$$\begin{aligned} V_0 &= B_0 + (\text{ROE} - r)B_0/(r - g) \\ &= \$20 + (0.18 - 0.14)(\$20)/(0.14 - 0.10) \\ &= \$20 + \$20 = \$40 \end{aligned}$$

Given that the current market price is $35 and the estimated value is $40, Simms will probably conclude that the shares are somewhat undervalued.

14
$$\begin{aligned} V_0 &= B_0 + (\text{ROE} - r)B_0/(r - g) \\ &= \$30 + (0.15 - 0.12)(\$30)/(0.12 - 0.10) \\ &= \$30 + \$45 = \$75 \text{ per share} \end{aligned}$$

15

Year	Net Income (Projected)	Ending Book Value	ROE (%)	Equity Charge (in Currency)	Residual Income	PV of RI
2020		$10.00				
2021	$1.50	11.50	15	$1.00	$0.50	$0.45
2022	1.73	13.23	15	1.15	0.58	0.48

(continued)

Year	Net Income (Projected)	Ending Book Value	ROE (%)	Equity Charge (in Currency)	Residual Income	PV of RI
2023	1.99	15.22	15	1.32	0.67	0.50
2024	2.29	17.51	15	1.52	0.77	0.53
2025	2.63	20.14	15	1.75	0.88	0.55
						$2.51

Using the finite horizon form of residual income valuation,

$$\begin{aligned} V_0 &= B_0 + \text{Sum of discounted RIs} + \text{Premium (also discounted to present)} \\ &= \$10 + \$2.51 + (0.20)(20.14)/(1.10)^5 \\ &= \$10 + \$2.51 + \$2.50 = \$15.01 \end{aligned}$$

16 A Columns (a) through (d) in the table show calculations for beginning book value, net income, dividends, and ending book value.

	(a)	(b)	(c)	(d)	(e)	(f)
Year	Beginning Book Value	Net Income	Dividends	Ending Book Value	Residual Income	PV of RI
1	$21.300	$3.834	$2.684	$22.450	$2.748	$2.614
2	22.450	4.041	2.829	23.663	2.896	2.622
3	23.663	4.259	2.981	24.940	3.052	2.629
4	24.940	4.489	3.142	26.287	3.217	2.637
5	26.287	4.732	3.312	27.707	3.391	2.644
6	27.707	4.987	3.491	29.203	3.574	2.652
7	29.203	5.256	3.680	30.780	3.767	2.659
8	30.780	5.540	3.878	32.442	3.971	2.667
Total						$21.125

For each year, net income is 18% of beginning book value. Dividends are 70% of net income. The ending book value equals the beginning book value plus net income minus dividends.

B Column (e) of the table in Part A shows Residual income, which equals Net income – Cost of equity (%) × Beginning book value.

To find the cost of equity, use the CAPM:

$$r = R_F + \beta_i[E(R_M) - R_F] = 2\% + (0.50)(6.2\%) = 5.1\%$$

For Year 1 in the table,

$$\begin{aligned} \text{Residual income} = \text{RI}_t &= E - rB_{t-1} \\ &= 3.834 - (5.1\%)(21.30) \\ &= 3.834 - 1.086 = \$2.748 \end{aligned}$$

This same calculation is repeated for Years 2 through 8.

Column (f) of the table gives the present value of the calculated residual income, discounted at 5.1%.

C To find the stock value with the residual income method, use this equation:

$$V_0 = B_0 + \sum_{t=1}^{T} \frac{\left(E_t - rB_{t-1}\right)}{\left(1 + r\right)^t} + \frac{P_T - B_T}{\left(1 + r\right)^T}$$

- In this equation, B_0 is the current book value per share of $21.30.
- The second term, the sum of the present values of the eight years' residual income is shown in the table, $21.125.
- To estimate the final term, the present value of the excess of the terminal stock price over the terminal book value, use the assumption that the terminal stock price is assumed to be 4.0× the terminal book value. So, by assumption, the terminal stock price is $129.767 [$P_T$ = 4.0(32.442)]. $P_T - B_T$ is $97.325 (129.767 – 32.442), and the present value of this amount discounted at 5.1% for eight years is $65.374.
- Summing the relevant terms gives a stock price of $107.799 ($V_0$ = 21.30 + 21.125 + 65.374).

D The appropriate DDM expression expresses the value of the stock as the sum of the present value of the dividends plus the present value of the terminal value:

$$V_0 = \sum_{t=1}^{T} \frac{D_t}{\left(1 + r\right)^t} + \frac{P_T}{\left(1 + r\right)^T}$$

Discounting the dividends from the table shown in the solution to Part A above at 8.30% gives:

Year	Dividend	PV of Dividend
1	$2.684	2.554
2	2.829	2.561
3	2.981	2.568
4	3.142	2.575
5	3.312	2.583
6	3.491	2.590
7	3.680	2.598
8	3.878	2.605
All		$20.634

- The present value of the eight dividends is $20.634. The estimated terminal stock price, calculated in the solution to Part C above is $125.767, which equals $87.165 discounted at 5.1% for eight years.
- The value for the stock, the present value of the dividends plus the present value of the terminal stock price, is V_0 = 20.634 + 87.165 = $107.799.
- The stock values estimated with the residual income model and the dividend discount model are identical. Because they are based on similar financial assumptions, this equivalency is expected. Even though the two models differ in their timing of the recognition of value, their final results are the same.

17 A The justified P/B can be found with the following formula:

$$\frac{P_0}{B_0} = 1 + \frac{\text{ROE} - r}{r - g}$$

ROE is 20%, g is 5.5%, and r is 8.2% [$R_F + \beta_i[E(R_M) - RF]$ = 4.46% + (0.68)(5.5%)]. Substituting in the values gives a justified P/B of

$$\frac{P_0}{B_0} = 1 + \frac{0.20 - 0.082}{0.082 - 0.055} = 4.37$$

The assumed parameters give a justified P/B of 4.37, slightly above the current P/B of 3.68.

B To find the ROE that would result in a P/B of 3.68, we substitute 3.68, r, and g into the following equation:

$$\frac{P_0}{B_0} = 1 + \frac{\text{ROE} - r}{r - g}$$

This yields

$$3.68 = 1 + \frac{ROE - 0.082}{0.082 - 0.055}$$

Solving for ROE requires several steps to finally derive a ROE of 0.15435 or 15.4%. This value of ROE is consistent with a P/B of 3.68.

C To find the growth rate that would result with a P/B of 3.68, use the expression given in Part B, but solve for g instead of ROE:

$$\frac{P_0}{B_0} = 1 + \frac{\text{ROE} - r}{r - g}$$

Substituting in the values gives:

$$3.68 = 1 + \frac{0.20 - 0.082}{0.082 - g}$$

The growth rate g is 0.03797, or 3.8%. If we assume that the single-stage growth model is applicable to Thales, the current P/B and current market price can be justified with values for ROE or g that are quite a bit lower than the starting values of 20% and 5.5%, respectively.

18

A The value found with the residual income model is:

Year	Beginning BV	Net Income	Dividends	Ending BV	Residual Income	PV of Residual Income
1	20.00	1.50	1.00	20.50	–0.300	–0.275
2	20.50	2.50	1.00	22.00	0.655	0.551
3	22.00	3.50	1.00	24.50	1.520	1.174
4	24.50	4.50	2.00	27.00	2.295	1.626
5	27.00	5.50	2.00	30.50	3.070	1.995
			Sum PVRI			5.071
			Terminal $P_T - B_T$		46.500	
			PV of $P_T - B_T$			30.222
			B_0			20.000
			Total value:			€55.293

Residual income each year is Net income – 0.09 × (Beginning BV). The PV of residual income is found by discounting at 9%. The terminal price is 14 × EPS in Year 5, or 14 × 5.50 = €77.00. The terminal residual value is $P_T - B_T$ = 77.00 – 30.50 = €46.50. Discounted at 9%, the PV of €46.50 is €30.222. The value per share is B_0 + PV of residual income + PV of terminal residual value, which is €55.293.

B The value found with the dividend discount model is as follows:

Year	Dividend or Price	PV of Dividend or Price
1	1.00	0.917
2	1.00	0.842
3	1.00	0.772
4	2.00	1.417
5	2.00	1.300
5	77.00	50.045
	Total PV	€55.293

The values per share found with the DDM and the residual income model are an identical €55.293.

19 C is correct. Market value added equals the market value of firm minus total accounting book value of total capital.

Market value added = Market value of company – Accounting book value of total capital
Market value of firm = Market value of debt + Market value of equity
Market value of firm = R55 million + (30,000,000 × R25.43)
Market value of firm = R55 million + R762.9 million = R817.9 million
Market value added = R817.9 million – R650 million = R167.9 million, or approximately R168 million.

20 B is correct. The intrinsic value of R22.00 is greater than the current book value of R20.00. The residual income model states that the intrinsic value of a stock is its book value per share plus the present value of expected (future) per share residual income. The higher intrinsic value per share, relative to book value per share, indicates that the present value of expected per share residual income is positive.

21 A is correct because the intrinsic value is the book value per share, B_0, plus the expected residual income stream, or $B_0 + [(\text{ROE} - r)B_0/(r - g)]$. If ROE equals the cost of equity (r), then $V_0 = B_0$. This implies that ROE is equal to the cost of the equity, and therefore there is no residual income contribution to the intrinsic value. As a result, intrinsic value would be equal to book value.

22 B is correct. With a single-stage residual income (RI) model, the intrinsic value, $V_{0,}$ is calculated assuming a constant return on equity (ROE) and a constant earnings growth (g).

$$V_0 = B_0 + B_0\frac{(\text{ROE} - r)}{(r - g)}$$

$$V_0 = \text{R}55.81 + \text{R}55.81\frac{(0.13 - 0.11)}{(0.11 - 0.095)}$$

$$V_0 = \text{R}130.22$$

23 B is correct. The share price of R8.25 was lower than the intrinsic value of R11.00. Shares are considered undervalued when the current share price is less than intrinsic value per share.

24 C is correct. The restructuring charge is a non-recurring item and not indicative of future earnings. In applying a residual income model, it is important to develop a forecast of future residual income based on recurring items. Using the net income reported in Amersheen's 2020 net income statement to model subsequent future earnings, without adjustment for the restructuring charge, would understate the firm's future earnings. By upward adjusting the firm's net income, by adding back the R2 million restructuring charge to reflect the fact that the charge is non-recurring, future earnings will be more accurately forecasted.

25 C is correct. The multistage residual income model results in an intrinsic value of R16.31.

This variation of the multistage residual income model, in which residual income fades over time, is:

$$V_0 = B_0 + \sum_{t=1}^{T-1} \frac{(E_t - rB_{t-1})}{(1+r)^t} + \frac{(E_T - rB_{T-1})}{(1+r-\omega)(1+r)^{T-1}}$$

where ω is the persistence factor.

The first step is to calculate residual income per share for years 2022–2025:

	2022	2023	2024	2025
Beginning book value per share	R7.60 (given)	R7.60 + R3.28 – R2.46 = R8.42	R8.42 + R3.15 – R2.36 = R9.21	R9.21 + 2.90 – R2.06 = R10.05
ROE	R3.28/R7.60 = 0.4316	R3.15/R8.42 = 0.3741	R2.90/R9.21 = 0.3149	26% (given)
Retention rate	1 – (R2.46/R3.28) = 0.25	1 – (R2.36/R3.15) = 0.2508	1 – (R2.06/R2.90) = 0.2897	N/A
Growth rate	0.4316 × 0.25 =0.1079	0.3741 × 0.2508 = 0.0938	0.3149 × 0.2897 = 0.0912	9% (given)
Equity charge per share	R7.60 × 0.10 = R0.76	R8.42 × 0.10 = R0.842	R9.21 × 0.10 = R0.921	R10.05 × 0.10 = R1.005
Residual income per share	R3.28 – R0.76 = R2.52	R3.15 – R0.842 = R2.31	R2.90 – 0.921 = R1.98	[0.26 × R10.05] – R1.005 = R1.608

ROE = Earnings/Book value

Growth rate = ROE × Retention rate

Retention rate = 1 – (Dividends/Earnings)

Book value_t = Book value_{t-1} + Earnings_{t-1} – Dividends_{t-1}

Residual income per share = EPS – Equity charge per share

Equity charge per share = Book value per share_t × Cost of equity

Using the residual income per share for 2015 of R1.608, the second step is to calculate the present value of the terminal value:

$$\text{PV of Terminal Value} = \frac{\text{R}1.608}{(1+0.10-0.70)(1.10)^3} = \text{R}3.0203$$

Then, intrinsic value per share is:

$$V_0 = \text{R}7.60 + \frac{\text{R}2.52}{(1.10)} + \frac{\text{R}2.31}{(1.10)^2} + \frac{\text{R}1.98}{(1.10)^3} + \text{R}3.0203 = \text{R}16.31$$

26 A is correct. The multistage residual income model results in an intrinsic value of R13.29. The multistage residual income model, is:

$$V_0 = B_0 + \sum_{t=1}^{T} \frac{(E_t - rB_{t-1})}{(1+r)^t} + \frac{(P_T - B_T)}{(1+r)^T}$$

The first step is to calculate residual income per share for years 2022–2024:

	2022	2023	2024
Beginning book value per share	R7.60 (given)	R7.60 + R3.28 – R2.46 = R8.42	R8.42 + R3.15 – R2.36 = R9.21
ROE	R3.28/R7.60 = 0.4316	R3.15/R8.42 = 0.3741	R2.90/R9.21 = 0.3149
Retention rate	1 – (R2.46/R3.28) = 0.25	1 – (R2.36/R3.15) = 0.2508	1 – (R2.06/R2.90) = 0.2897
Growth rate	0.4316 × 0.25=0.1079	0.3741 × 0.2508 = 0.0938	0.3149 × 0.2897= 0.0912
Equity charge per share	R7.60 × 0.10 = R0.76	R8.42 × 0.10 = R0.842	R9.21 × 0.10 = R0.921
Residual income per share	R3.28 – R0.76 = R2.52	R3.15 – R0.842 = R2.31	R2.90 – 0.921= R1.98

ROE = Earnings/Book value

Growth rate = ROE × Retention rate

Retention rate = 1 – (Dividends/Earnings)

Book value$_t$ = Book value$_{t-1}$ + Earnings$_{t-1}$ – Dividends$_{t-1}$

Residual income per share = EPS – Equity charge per share

Equity charge per share = Book value per share$_t$ × Cost of equity

Under Scenario 2, at the end of 2024, it is assumed that share price will be equal to book value per share. This results in the second term in the equation above, the present value of the terminal value, being equal to zero.

Then, intrinsic value per share is:

$$V_0 = \text{R}7.60 + \frac{\text{R}2.52}{(1.10)} + \frac{\text{R}2.31}{(1.10)^2} + \frac{\text{R}1.98}{(1.10)^3} = \text{R}13.29$$

27 B is correct. The residual income model's use of accounting income assumes that the cost of debt capital is reflected appropriately by interest expense.

28 A is correct. Dividend payments are forecasted to be unpredictable over Castovan's five-year forecast horizon. A residual income model is appropriate when a company does not pay dividends or when its dividends are not predictable, which is the case for PTU.

29 C is correct. Other comprehensive income bypasses the income statement and goes directly to the statement of stockholders' equity (which is a violation of the clean surplus relationship). Therefore, book value per share for PTU will be affected by forecasted OCI.

30 C is correct. The residual income can be calculated using net income and the equity charge or using net operating profit after taxes and the total capital charge.

Residual income = Net income – Equity charge

Calculation of Net Income (values in millions):

EBIT	€700.0	
Less Interest expense	€124.8	(= €4,000 × 0.60 × 0.052)
Pretax income	€575.2	
Less Income tax expense	€201.3	(= €575.20 × 0.35)
Net income	€373.9	

Equity charge = Total assets × Equity weighting × Cost of equity

Equity charge = €4,000 million × 0.40 × 0.15 = €240 million

Therefore, residual income = €373.9 million – €240 million = €133.9 million.

Alternatively, residual income can be calculated from NOPAT as follows.

Residual income = NOPAT – Total capital charge

NOPAT = EBIT × (1 – Tax rate)

NOPAT = €700 million × (1 – 0.35) = €455 million

The total capital charge is as follows.

$$\begin{aligned}
\text{Equity charge} &= \text{Total assets} \times \text{Equity weighting} \times \text{Cost of equity} \\
&= €4{,}000 \text{ million} \times 0.40 \times 0.15 \\
&= €240 \text{ million} \\
\text{Debt charge} &= \text{Total assets} \times \text{Debt weighting} \times \text{Pretax cost of} \\
&\quad\ \text{debt} \times (1 - \text{Tax rate}) \\
&= €4{,}000 \text{ million} \times 0.60 \times 0.052(1 - 0.35) \\
&= €81.1 \text{ million} \\
\text{Total capital charge} &= €240 \text{ million} + €81.1 \text{ million} \\
&= €321.1 \text{ million}
\end{aligned}$$

Therefore, residual income = €455 million – €321.1 million = €133.9 million.

31 B is correct. The implied growth rate of earnings from the single-stage RI model is calculated by solving for g in the following equation:

$$V_0 = B_0 + \left(\frac{\text{ROE} - r}{r - g}\right)B_0$$

Book value per share can be calculated using the given price-to-book ratio and market price per share as follows.

$$\begin{aligned}
\text{Book value per share } (B_0) &= \text{Market price per share/Price-to-book} \\
&\quad\ \text{ratio} \\
&= €48.80/2.10 = €23.24
\end{aligned}$$

Then, solve for the implied growth rate.

$$€48.80 = €23.24 + \left(\frac{0.2337 - 0.15}{0.15 - g}\right)€23.24$$

g = 7.4%

32 C is correct. Using the single-stage RI model, the intrinsic value of TTCI is calculated as

$$\begin{aligned} V_0 &= B_0 + \left(\frac{\text{ROE} - r}{r - g}\right)B_0 \\ &= €45.25 + \left(\frac{0.12 - 0.087}{0.087 - 0.045}\right)€45.25 \\ &= €80.80 \end{aligned}$$

The intrinsic value of €80.80 is less than the market price of €126.05, so Castovan should conclude that the stock is overvalued.

33 A is correct. The justified price-to-book ratio is calculated as

$$\begin{aligned} \frac{P}{B} &= 1 + \left(\frac{\text{ROE} - r}{r - g}\right) \\ &= 1 + \left(\frac{0.12 - 0.087}{0.087 - 0.045}\right) = 1.79 \end{aligned}$$

34 C is correct. Residual income per share for the next three years is calculated as follows.

	Year 1	Year 2	Year 3
Beginning book value per share	45.25	51.61	58.25
Earnings per share	7.82	8.17	8.54
Less dividends per share	1.46	1.53	1.59
Change in retained earnings	6.36	6.64	6.95
Ending book value per share	51.61	58.25	65.20
Earnings per share	7.82	8.17	8.54
Less per share equity charge*	3.94	4.49	5.07
Residual income	3.88	3.68	3.47

* Per share equity charge = Beginning book value per share × Cost of equity
Year 1 per share equity charge = 45.25 × 0.087 = 3.94
Year 2 per share equity charge = 51.61 × 0.087 = 4.49
Year 3 per share equity charge = 58.25 × 0.087 = 5.07

Because Castovan forecasts that residual income per share will be constant into perpetuity, equal to Year 3 residual income per share, the present value of the terminal value is calculated using a persistence factor of 1.

$$\begin{aligned} \text{Present value of terminal value} &= \frac{8.54 - (0.087 \times 58.25)}{(1 + 0.087 - 1)(1 + 0.087)^2} \\ &= \frac{3.47}{(0.087)(1.087)^2} \\ &= 33.78 \end{aligned}$$

So, the intrinsic value of TTCI is then calculated as follows.

$$V_0 = €45.25 + \frac{3.88}{1.087} + \frac{3.68}{1.087^2} + 33.78 = €85.71$$

35 A is correct. In Castovan's multistage valuation, she assumes that TTCI's residual income will remain constant in perpetuity after Year 3. This perpetuity assumption implies a persistence factor of 1 in the calculation of the terminal value. A persistence factor of 0.10 indicates that TTCI's residual income is forecasted to decline at an average rate of 90% per year. This assumption would lead to a lower valuation than Castovan's multistage value estimate, which assumes that residual income will remain constant in perpetuity after Year 3.

36 B is correct. Beckworth's suggested persistence factor for TTCI is 0.10, which is quite low. Companies with extreme accounting rates of return typically have low persistence factors. Companies with strong market leadership positions and low dividend payouts are likely to have high persistence factors.

READING

27

Private Company Valuation

by Raymond D. Rath, ASA, CEIV, CFA

Raymond D. Rath, ASA, CEIV, CFA, is at Globalview Advisors LLC (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** compare public and private company valuation;
☐	**b.** describe uses of private business valuation and explain applications of greatest concern to financial analysts;
☐	**c.** explain the income, market, and asset-based approaches to private company valuation and factors relevant to the selection of each approach;
☐	**d.** explain cash flow estimation issues related to private companies and adjustments required to estimate normalized earnings;
☐	**e.** calculate the value of a private company using free cash flow, capitalized cash flow, and/or excess earnings methods;
☐	**f.** explain factors that require adjustment when estimating the discount rate for private companies;
☐	**g.** compare models used to estimate the required rate of return to private company equity (for example, the CAPM, the expanded CAPM, and the build-up approach);
☐	**h.** calculate the value of a private company based on market approach methods and describe advantages and disadvantages of each method;
☐	**i.** describe the asset-based approach to private company valuation;
☐	**j.** explain and evaluate the effects on private company valuations of discounts and premiums based on control and marketability.

© 2020 CFA Institute. All rights reserved.

1 INTRODUCTION, THE SCOPE AND DEFINITIONS OF PRIVATE COMPANY VALUATION

a compare public and private company valuation;

b describe uses of private business valuation and explain applications of greatest concern to financial analysts;

The valuation of the equity of private companies is a major field of application for equity valuation. Private companies are those whose shares are not listed on public markets. Generalist investment practitioners need to be familiar with issues associated with valuations of such companies. We use the terms "valuation" and "appraisal" interchangeably in this chapter.

Many public companies have start-up or other operations that can best be valued as if they were private companies. Companies may grow through the acquisition of competitors, including private companies, and analysts must be prepared to evaluate the price paid in such transactions. Furthermore, acquisitions often result in significant balances of intangible assets, including goodwill, that are reported on the balance sheets of acquiring companies. Goodwill balances require impairment assessment or formal testing on an annual basis under International Financial Reporting Standards (IFRS) and US GAAP. Impairment testing and other financial reporting initiatives increasingly result in the use of fair value estimates in financial statements. The concepts and methods we will discuss play important roles in this aspect of financial reporting. In addition, issues relating to private company valuation arise in the types of investment held by venture capital and other types of private equity funds.

The following sections illustrate key elements associated with the valuation of private companies and provide background for understanding private company valuation, including typical contrasts between public and private companies and the major purposes for which private valuations are performed. Later sections discuss earnings normalization and cash flow estimation; introduce the three major approaches recognized in private company valuation, valuation discounts, and premiums; and explain business valuation standards and practices.

1.1 The Scope of Private Company Valuation

Private companies range from single-employee, unincorporated businesses to formerly public companies that have been taken private in management buyouts or other transactions. Numerous large, successful companies also exist that have remained private since inception, such as IKEA and Bosch in Europe, Cargill and Bechtel in the United States, Alibaba in China, and Toyota in Japan. The diverse characteristics of private companies and the absence of a universally recognized body providing guidance on valuation methods and assumptions have contributed to the development of diverse valuation practices.

1.1.1 *Private and Public Company Valuation: Similarities and Contrasts*

We can gain some insight into the challenges of private company valuation by examining company- and stock-specific factors that mark key differences between private and public companies.

1.1.1.1 Company-Specific Factors Company-specific factors characterize the company itself, including its life-cycle stage, size, markets, and the goals and characteristics of management.

- *Stage in life cycle*. Private companies include companies at the earliest stages of development, whereas public companies are typically further advanced in their life cycle. Private companies may have minimal capital, assets, or employees. Private companies, however, also include large, stable, going concerns and failed companies in the process of liquidation. The stage of life cycle influences the valuation process for a company.
- *Size*. Relative size—whether based on income statement, balance sheet, or other measures—frequently distinguishes public and private companies; private companies in a given line of business tend to be smaller than public ones. Size has implications for the level of risk and, hence, relative valuation. Small size typically increases risk levels, and risk premiums for small size have often been applied in estimating required rates of return for private companies. For some private companies, small size may reduce growth prospects by reducing access to capital to fund growth of operations. The public equity markets are generally the best source for such funding. Conversely, for small companies, the costs of operating as a public company including compliance costs may outweigh any financing benefits.
- *Overlap of shareholders and management*. For many private companies, and in contrast to most public companies, top management has a controlling ownership interest. Therefore, they may not face the same pressure from external investors as public companies. **Agency issues**, such as monitoring costs arising from potentially conflicting interests of owners (principals) and managers (agents), may also be mitigated in private companies. For that reason, private company management may be able to take a longer-term perspective in decisions than public company management.
- *Quality/depth of management*. A small private company, especially if it has limited growth potential, would be expected to be less attractive to management candidates and have less management depth than a typical public company. The smaller scale of operation might also lead to less management depth compared with a public company. To the extent these considerations apply, they may increase risk and reduce growth prospects for the private company.
- *Quality of financial and other information*. Compared with the levels of disclosure by public companies, the more limited availability of financial and other information for private companies results in an increased burden for the prospective investor considering an equity investment or loan. This type of information difference presumably leads to greater uncertainty and, hence, risk. All else equal, the higher risk should lead to a relatively lower valuation. Note, however, that in certain private company valuations, such as fairness opinions prepared in the context of an acquisition, the analyst usually has unlimited access to books, records, contracts, and other information that would be unavailable to the public stock analyst.
- *Pressure from short-term investors*. Earnings consistency and growth rates are often perceived as critical to the stock price performance of public companies. Continued management employment and levels of incentive compensation are often linked to stock price performance, but many investors' interests may be of a trading or short-term nature. As a result, management may be motivated

to try to support share price in the short term. According to some observers, private companies typically do not experience similar stock price performance pressure and such companies can take a longer-term investment focus.

- *Tax concerns*. Reduction of reported taxable income and corporate tax payments may be a more important goal for private companies than for public companies because of greater benefit to the owners.

1.1.1.2 Stock-Specific Factors In addition to company-specific factors, the characteristics of the stock of private companies frequently differ markedly from that of public companies.

- *Liquidity of equity interests in business*. Stock in private companies is generally much less liquid than otherwise similar interests in public companies. Private companies typically have fewer shareholders. Shares of a private company have not been registered for sale in the public stock markets. The limited number of existing and potential buyers reduces the value of the shares in private companies.
- *Concentration of control*. Control of private companies is often concentrated in one or in very few investors. This concentration of control may lead to actions by a corporation that benefit some shareholders at the expense of other shareholders. Transactions with entities related to a controlling group at above-market prices would transfer value away from the corporation's non-controlling shareholders. Above-market compensation to a controlling shareholder is a typical perquisite. Please note that the "concentration of control" factor can also be viewed as a "company-specific" factor.
- *Potential agreements restricting liquidity*. Private companies may have shareholder agreements in place that restrict the ability to sell shares. These agreements may reduce the marketability of equity interests.

Generally, stock-specific factors are a negative for private company valuation, whereas company-specific factors are potentially positive or negative. The range of differences observed in private companies is such that the spectrum of risk and, therefore, the spectrum of return requirements are typically wider than for public companies. Another consequence is that the range of valuation methods and assumptions applied to private companies is typically more varied.

1.1.2 *Reasons for Performing Valuations*

Valuations of private businesses or equity interests therein fall into three groups: transaction related, compliance related, and litigation related.

Transactions encompass events affecting the ownership or financing of a business and represent a primary area of private company valuation. A variety of transaction types exist.

- *Private financing*. Raising capital is critical to development-stage companies. To reduce risk and maintain influence, **venture capital investors** (as equity investors in such companies are known) typically invest through multiple rounds of financing tied to the achievement of key developments ("milestones"). A high level of uncertainty concerning expected future cash flows results in valuations that are often informal and based on negotiations between the company and investors.
- *Initial public offering (IPO)*. An IPO is one liquidity option for a private company. Investment banking firms prepare valuations as part of the IPO process. A key element of an IPO-related valuation is frequently the identification of any public companies that are similar to the one going public.

- *Acquisition*. Acquisition can be an attractive option for development-stage or mature companies. Acquisition related valuations may be performed (and negotiated) by management of the target and/or buyer. Smaller companies may be sold with the assistance of a business broker. The sale of many larger companies is handled by investment banking firms.
- *Bankruptcy*. For companies operating under bankruptcy protection, valuations of the business and its underlying assets may help assess whether a company is more valuable as a going concern or in liquidation. For viable going concerns operating in bankruptcy, insights from valuation may be critical to the restructuring of an overleveraged capital structure.
- *Share-based payment (compensation)*. Share-based payments can be viewed as transactions between a company and its employees. These transactions often have accounting and tax implications to the issuer and the employee. Share-based payments can include stock option grants, restricted stock grants, and transactions involving an employee stock ownership plan (ESOP) in the United States and equivalent structures in other countries. Providing an incentive for improved employee performance is an important goal of such compensation mechanisms.

Compliance encompasses actions required by law or regulation. Compliance valuations are a second key area of valuation practice. Financial reporting and tax reporting are the two primary focuses of this type of valuation.

- *Financial reporting*. Financial reporting valuations are increasing in importance. Goodwill impairment is one of the most frequent financial reporting valuations that a securities analyst might observe. Goodwill impairment tests require a business valuation for a **cash-generating unit** (defined in IFRS as "the smallest identifiable group of assets that generates cash inflows that are largely independent of the cash inflows of other assets or groups of assets") of an entity or a **reporting unit** (defined in US GAAP as an operating segment or one level below an operating segment, called a component, that constitutes a business for which discrete financial information is available and for which segment management regularly reviews the operating results of that component). Essentially, components of public companies are valued using private company valuation techniques. For private companies, stock option grants will frequently require valuations.
- *Tax reporting*. Tax reporting is a longstanding area that requires valuations of private companies. Tax-related reasons for valuations include corporate and individual tax reporting. A variety of corporate activities, such as corporate restructurings, transfer pricing, and property tax matters, may require valuations. An individual's tax requirements, such as those arising from estate and gift taxation in some jurisdictions, may generate a need for private company valuations.

Litigation—legal proceedings including those related to damages, lost profits, shareholder disputes, and divorce—often requires valuations. Litigation may affect public or private companies or may be between shareholders with no effect at the corporate level.

As the foregoing descriptions make clear, each of the three major practice areas requires specialized knowledge and skills. This fact has led many valuation professionals to focus their efforts in one of these areas. Transactions, for example, often involve investment bankers. Compliance valuations are best performed by valuation professionals with knowledge of the relevant accounting or tax regulations. Litigation-related valuations require effective presentations in a legal setting.

Having provided an overview of the field of private company valuation, we can proceed to discussing how valuations are done. Logically, before developing an estimate of value, the valuator must understand the context of the valuation and its requirements. An important element in that process is knowing the definition(s) of value that the valuation must address, which we cover in the next section.

2 PRIVATE COMPANY VALUATION APPROACHES, EARNINGS NORMALIZATION AND CASH FLOW ESTIMATION ISSUES

d explain cash flow estimation issues related to private companies and adjustments required to estimate normalized earnings;

Private company valuation experts distinguish three major approaches to valuation.

- The **income approach** values an asset as the present discounted value of the income expected from it. The income approach has several variations depending on the assumptions the valuator makes.
- The **market approach** values an asset based on pricing multiples from sales of assets viewed as similar to the subject asset. ("Pricing multiples" may refer to multiples based on share price or multiples based on a measure of total company value.)
- The **asset-based approach** values a private company based on the values of its underlying assets less the value of any related liabilities.

Valuation approaches for private companies are conceptually similar to those used for public companies, although the labels used for them by experts in each field and the details of application may differ. The income approach corresponds to what public equity analysts call discounted cash flow models or present value models. Along with asset-based models, discounted cash flow models are classified as absolute valuation models. By contrast, analysts use a relative valuation model when they apply a market-based approach in evaluating price and enterprise multiples relative to the value of a comparable.

Analysts select approach(es) depending on specific factors. The nature of operations and stage in life cycle are important considerations. For a development-stage company with the potential to operate as a successful large public company, the valuation methods may change over time. At the earliest stages of development, the company may best be valued using an asset-based approach because the going-concern **premise of value** may be uncertain and/or future cash flows may be extremely difficult to predict. With progress to a development-stage company in a high-growth mode, the company might be valued using a free cash flow method, which in private business appraisal is known as an income approach. A stable, mature company might be best valued on the basis of the market approach. Specific facts and circumstances may suggest different valuation methods.

Size is an important criterion in assessing valuation approaches and valuation methods. Multiples from public companies may be inappropriate for a small, relatively mature private company with very limited growth prospects. Comparisons to public companies are not a good basis of valuation for a private company if risk and growth prospects differ materially.

Public and private companies may consist of a variety of operating and non-operating assets. Non-operating assets are defined as assets not necessary to the ongoing operations of the business enterprise. Excess cash and investment balances are typical examples of non-operating assets. In principle, the value of a company is the sum of the value of operating assets and the value of non-operating assets. Thus, non-operating assets should be included in the valuation of an enterprise regardless of the valuation approach or method being used.

Before we illustrate the application of the three approaches to valuation, we need to address certain typical issues relating to valuation model inputs that arise when valuing private companies.

2.1 Earnings Normalization and Cash Flow Estimation Issues

The next two sections cover earnings normalization and cash flow estimation in the context of private company valuation. Potential acquirers of private companies may find that current earnings reflect inefficiencies or redundancies that detract from their relevance as a baseline for forecasting future earnings under new ownership. In such cases, the earnings should be adjusted or "normalized" to a basis that is relevant for forecasting future results, given that the company is acquired. Essentially, the valuator is seeking to understand accurately the earnings and cash flow capacity of the business enterprise if it is acquired and run efficiently.

2.1.1 *Earnings Normalization Issues for Private Companies*

Private company valuations may require significant adjustments to estimate the company's normalized earnings. As defined in the International Glossary of Business Valuation Terms (IGBVT, developed through joint efforts of several North American accountancy and appraisal bodies to explain frequently used terms), **normalized earnings** are "economic benefits adjusted for non-recurring, non-economic, or other unusual items to eliminate anomalies and/or facilitate comparisons." (Note that the term "normalized earnings" can also refer to earnings adjusted for the effects of a business cycle.) As a result of the concentration of control in many private businesses, reported earnings may reflect discretionary expenses or expenses that are not at arm's-length amounts. Tax and other motivations may also result in reporting earnings that may differ from the normalized earnings of a private company. The smaller size of many private companies potentially increases the relative impact of discretionary expenses on company value.

When comparing the reported earnings of private companies with those of public companies, a key area of difference is the possible effect of transactions between the company and owners working in the business or with entities controlled by controlling shareholders. Many adjustments required to normalize earnings involve items that reduce the reported earnings of a profitable, private company. The controlling or sole shareholder is often active in the business and controls the board of directors as well as all policy and operating decisions. Above-market compensation or other expenses will reduce taxable income and income tax expense at the corporate level and subsequent taxes upon the payment of dividends to the controlling shareholder and other shareholders. Above-market expenses can also result in the controlling shareholder receiving a disproportionately high return in relation to other shareholders.

Compensation expense is a key area requiring possible adjustment. Profitable, private companies may report compensation expense to owner/employees above amounts that would be paid to a non-owner employee. Family members may also be included as employees and paid amounts above the market value of their services. For private companies with limited profits or reported losses, expenses may actually be understated with the reported income of the entity overstated. Owners active in the business may not take compensation commensurate with market levels required by an employee for similar activities.

A number of other areas exist for consideration for possible adjustments. Personal expenses may be included as expenses of the private company. Personal-use assets and excess entertainment expenses are areas for consideration. Personal residences, aircraft, and luxury or excessive use of corporate vehicles for personal use may require an adjustment. Life insurance and loans to shareholders merit review.

Real estate used by the private company is also an area for consideration. When a private company owns real estate, some analysts separate the real estate from the operating company. This separation consists of removing any revenues and expenses associated with the real estate from the income statement. If the company is using owned property in its business operations, adding a market rental charge for the use of the real estate to the expenses of the company would produce a more accurate estimate of the earnings of the business operations. Adjusting reported earnings to include a provision for third-party real estate costs would produce a value of the business operations excluding the owned real estate. Because the real estate is still owned by the entity, its value would represent a non-operating asset of the entity. These adjustments for the financial impact of owned real estate can be appropriate because the business operations and real estate have different risk levels and growth expectations.

Without these adjustments to eliminate the effect of owned real estate on reported financial performance, the private company may be incorrectly valued. Rent charges for the use of real estate include return "of" and "on" investment components. Depreciation reflects return "of" investment. If real property is owned, depreciation expense would reflect the historical acquisition cost rather than current replacement cost. For owned real estate, the return "on" component of the rental charge would not be included at a market level charge. Applying a capitalization rate for the business operations to an earnings figure that includes some of the benefit from the owned real estate may misvalue the private company. The business operations and real estate may have different levels of risk and expected future growth that require separate valuation. If real estate is leased to the private company by a related entity, the level of expense may require an adjustment to a market rental rate. If real estate is leased from an unrelated party but the rental charge is not at a market level, an adjustment to normalize this expense may also be appropriate.

Example 1 illustrates a case in which a prospective buyer of a private business would need to make adjustments to reported financial results for a more accurate picture of the company's normalized earnings and value under new ownership.

EXAMPLE 1

Able Manufacturing: Normalized Earnings Adjustments

John Smith is the sole shareholder and CEO of Able Manufacturing, Inc. Smith has put Able up for sale in advance of his retirement. James Duvall, a manager in the corporate venturing unit of a public company, is evaluating the purchase of Able. Duvall notes the following facts affecting the most recent fiscal year's reported results:

- Smith's compensation for the year was $1.5 million. Duvall's executive compensation consultant believes a normalized compensation expense of $500,000 for a CEO of a company like Able is appropriate. Compensation is included in selling, general, and administrative expenses (SG&A).

- Certain corporate assets including ranch property and a condominium are in Duvall's judgment not required for the company's core operations. Fiscal year expenses associated with the ranch and condominium were $400,000, including $300,000 of such operating expenses as property upkeep, property taxes, and insurance reflected in SG&A expenses, and depreciation expense of $100,000. All other asset balances (including cash) are believed to be at normal levels required to support current operations.
- Able's debt balance of $2,000,000 (interest rate of 7.5%) was lower than the optimal level of debt expected for the company. As reported interest expense did not reflect an optimal charge, Duvall believes the use of an earnings figure that excludes interest expense altogether, specifically operating income after taxes, will facilitate the assessment of Able.

Duvall uses the reported income statement to show the derivation of reported operating income after taxes, as given in the following table.

Able Manufacturing, Inc. Operating Income after Taxes

As of 31 December 2020	As Reported
Revenues	$50,000,000
Cost of goods sold	30,000,000
Gross profit	20,000,000
Selling, general, and admin. expenses	5,000,000
EBITDA	15,000,000
Depreciation and amortization	1,000,000
Earnings before interest and taxes	14,000,000
Pro forma taxes (at 30.0%)	4,200,000
Operating income after taxes	$9,800,000

Based only on the information given, address the following:

1 Identify the adjustments that Duvall would make to reported financials to estimate normalized operating income after taxes—that is, what the operating income after taxes would have been under ownership by Duvall's unit.

2 Based on your answer to Part 1, construct a pro forma statement of normalized operating income after taxes for Able.

Solution to 1:

First, SG&A should be reduced by $1,500,000 – $500,000 = $1,000,000 to reflect the expected level of salary expense under professional management at a market rate of compensation. Second, the ranch and condominium are non-operating assets—they are not needed to generate revenues—so expense items should be adjusted to reflect their removal (e.g., through a sale). Two income statement lines are affected: SG&A expenses should be reduced by $300,000, and depreciation and amortization reduced by $100,000.

Solution to 2:

The pro forma statement of normalized operating income after taxes would be as follows:

Able Manufacturing, Inc. Pro Forma Normalized Operating Income after Taxes

As of 31 December 2020	Pro Forma
Revenues	$50,000,000
Cost of goods sold	30,000,000
Gross profit	20,000,000
Selling, general, and admin. expenses	3,700,000
EBITDA	16,300,000
Depreciation and amortization	900,000
Earnings before interest and taxes	15,400,000
Pro forma taxes (at 30.0%)	4,620,000
Operating income after taxes	$10,780,000

In addition to the various adjustments noted, a variety of other areas exist for possible adjustment that are similar for valuing both public and private companies (e.g., adjustments related to inventory accounting methods, depreciation assumptions, and capitalization versus expensing of various costs). Private companies may have their financial statements reviewed rather than audited. **Reviewed financial statements** provide an opinion letter with representations and assurances by the reviewing accountant that are less than those in audited financial statements. The preparation of reviewed rather than audited financial statements as well as other factors suggest a potentially greater need for analyst adjustments to the reported financials of some private companies. **Compiled financial statements** (that are not accompanied by an auditor's opinion letter) suggest an even greater need for analytical adjustments.

2.1.2 *Cash Flow Estimation Issues for Private Companies*

In addition to earnings normalization, cash flow estimation is an important element of the valuation process. Free cash flow (FCF) is the relevant concept of cash flow in this context. Free cash flow to the firm (FCFF) represents free cash flow at the business enterprise level and is used to value the firm or, indirectly, the firm's equity. Alternatively, free cash flow to equity (FCFE) can be used to value equity directly.

Cash flow estimation for private companies raises some important challenges, including those related to the nature of the interest being valued, potentially acute uncertainties regarding future operations, and managerial involvement in forecasting.

The nature of assumptions in cash flow estimates depends on a variety of factors. The equity interest appraised and the intended use of the appraisal are key in determining the appropriate definition of value for a specific valuation. The assumptions included in cash flow estimates may differ if a small minority equity interest is appraised rather than the total equity of a business. For example, an investment value standard may lead to different cash flow estimates than a fair value standard related to a financial reporting valuation assignment.

In assessing future cash flow estimates, uncertainty regarding a potentially wide range of future cash flow possibilities also creates challenges for valuation using FCF. Many development-stage companies and some mature companies are subject to significant uncertainties regarding future operations and cash flows. One possible solution involves projecting the different possible future scenarios. For a privately held development-stage company, the possible scenarios could include initial public offering, acquisition, continued operation as a private company, or bankruptcy. For a larger, mature company, the scenarios might be chosen to cover the range of possible levels of growth and profitability.

In valuing an individual scenario, the discount rate chosen should reflect the risk of achieving the projected cash flows in that scenario. The analyst must also estimate the probability of each scenario occurring. The overall value estimate for a company is then a probability-weighted average of the company's estimated scenario values. Alternatively, the expected future cash flows based on the scenarios could be discounted using a conventional, single discount rate to obtain an overall value estimate. Although the trend is generally to more robust models, in current practice private company valuation more frequently reflects an average or most likely scenario than an explicit multiple scenario analysis.

Managers of private companies generally command much more information about their business than outside analysts. Management may develop cash flow forecasts to be used in a valuation with appraiser input, or appraisers may develop their own forecasts consulting management as needed. The appraiser should be aware of potential managerial biases, such as to possibly overstate values in the case of goodwill impairment testing or understate values in the case of incentive stock option grants. Appraisers should also pay attention to whether projections adequately capture capital needs.

The actual process for estimating FCFF and FCFE is similar for private and public companies.

EXAMPLE 2

Able Manufacturing: Pro Forma Free Cash Flow to the Firm

Duvall, the manager of the corporate venturing unit introduced in Example 1, has decided to make a bid for Able Manufacturing. Duvall has decided to use an income approach to value Able. As stated in Example 1, Able's debt is $2,000,000. Considering the nature of Able's business, its size, and the financial leverage used by competitors, Duvall has concluded that Able has a low level of debt relative to its capacity and that it will be optimal to increase its debt if Duvall's unit succeeds in purchasing Able. Because of that anticipated change in leverage, Duvall has decided to use a FCFF approach rather than FCFE to value Able.

Based on available information, Duvall makes the following assumptions:

- Long-term growth of revenues and after-tax operating income is 3% annually.
- The gross profit margin will remain at 40%.
- Depreciation will remain at 1.8% of sales.
- SG&A expenses can be maintained at the prior year's level of $3,700,000 at least for two years.
- Working capital equal to 10% of revenues is required (e.g., if the increase in revenues is $X from the prior year, additional working capital of 0.10 × $X would be needed).
- Capital expenditures are expected to equal projected depreciation expense (to support current operations) plus 3% of incremental revenues (to support future growth).

1 Should Duvall use reported earnings or normalized earnings in estimating FCFF for Able? Explain.

2 Forecast FCFF for Able for the upcoming year (from the perspective of a knowledgeable buyer).

Solution to 1:

For the valuation of Able in a purchase transaction, the normalized earnings of Able should be used to estimate FCFF. Normalized earnings would more accurately reflect the income expected by a willing buyer of Able than reported earnings.

Solution to 2:

Duvall assumed long-term growth of 3% into the foreseeable future. With the \$50 million revenue base from the prior year and the 3% annual revenue growth, a \$1.5 million increase in revenues is forecast when moving from the last historical year to the year ahead. Given depreciation of \$927,000 (1.8% of \$51.5 million) and incremental sales of \$1,500,000, forecast capital expenditure sums to \$927,000 + 0.03(\$1,500,000) = \$927,000 + \$45,000 = \$972,000. A requirement for incremental working capital of 10% of the increase in revenues equates to a \$150,000 deduction in calculating free cash flow. Based on these assumptions, free cash flow to the firm of \$10,986,100 was calculated as follows.

Able Manufacturing, Inc. Calculation of Next Year's Projected Free Cash Flow to Firm	
Revenues (\$50,000,000 × 1.03 =)	\$51,500,000
Cost of goods sold (0.60 × Revenues =)	30,900,000
Gross profit (Revenue – Cost of goods sold =)	20,600,000
SG&A expenses (maintained at 2020 level)	3,700,000
Pro forma EBITDA	16,900,000
Deprec. and amort. (0.018 × \$51,500,000 =)	927,000
Pro forma earnings before interest and taxes	15,973,000
Pro forma taxes on EBIT (at 30.0%)	4,791,900
Operating income after tax	11,181,100
Plus: Depreciation and amortization	927,000
Less: Capital expenditures (\$927,000 + 0.035 × \$1,500,000)	972,000
Less: Increase in working capital (0.10 × (\$51,500,000 – \$50,000,000).	150,000
Free cash flow to firm	\$10,986,100

INCOME APPROACH METHODS AND REQUIRED RATE OF RETURN: MODELS AND ESTIMATION ISSUES 3

c explain the income, market, and asset-based approaches to private company valuation and factors relevant to the selection of each approach;

f explain factors that require adjustment when estimating the discount rate for private companies;

g compare models used to estimate the required rate of return to private company equity (for example, the CAPM, the expanded CAPM, and the build-up approach);

The income approach obtains its conceptual support from the assumption that value is based on expectations of future income and cash flows. The income approach converts future economic benefits into a present value equivalent. For IFRS and US GAAP, assets are defined as probable future economic benefits. This definition provides strong support for the application of the income approach to valuation of an interest in a public or private company. The income approach takes three main forms.

- **Free cash flow method**. This method is often referred to as the **discounted cash flow method** in the appraisal community. It values an asset based on estimates of future cash flows that are discounted to present value by using a discount rate reflective of the risks associated with the cash flows. For a going concern, this method frequently includes a series of discrete cash flow projections followed by an estimate of the value of the business enterprise as a going concern at the end of the projection period.
- **Capitalized cash flow method**. Also referred to as the **capitalized income method** or **capitalization of earnings method**, this approach values a private company by using a single representative estimate of economic benefits and dividing that estimate by an appropriate capitalization rate to derive an indication of value.
- **Residual income method**. Frequently referred to as the **excess earnings method** in the valuation community, this method is sometimes categorized under the asset approach because it involves marking the tangible assets to market and estimating the value of intangible assets. For valuing a business enterprise, the excess earnings method consists of estimating the value of all of the company's intangible assets by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets. The value of the intangible assets is added to the values of working capital and fixed assets to arrive at the value of the business enterprise.

Whichever income approach method is used, an appropriate required rate of return estimate is needed for discounting expected future cash flows.

3.1 Required Rate of Return: Models and Estimation Issues

A variety of factors make estimating a required rate of return for a private company challenging.

- *Application of size premiums*. In assessing private company valuations, size premiums are frequently used in developing equity return requirements by private company appraisers. This practice seems to be less prevalent in the valuation of public companies. Furthermore, size premium estimates based on public company data for the smallest market cap segments can capture premiums for financial and/or operating distress that may be irrelevant to the company being valued.

- *Use of the CAPM*. Some parties have questioned whether the CAPM is appropriate for developing discount rate estimates for small private company valuations. Small companies that have little prospect of going public or being acquired by a public company may be viewed as not comparable to the public companies for which market-data-based beta estimates are available.
- *Expanded CAPM*. The **expanded CAPM** is an adaptation of the CAPM that adds to the CAPM premiums for small size and company-specific risk (Pratt and Grabowski, 2014). Estimation of company-specific risk has been a very subjective element of the valuation process. Several valuation professionals have presented methodologies to develop quantitative estimates of company-specific risk. These tools are being vetted in the valuation community.
- *Elements of the build-up approach*. The build-up approach, introduced earlier as part of return concept topics, relies on building up the required rate of return as a set of premia added to the risk-free rate. The added premia are typically based on factors such as size and company risk. When **guideline public companies** (public-company comparables for the company being valued) are unavailable or of questionable comparability, appraisers may rely on a build-up method rather than the CAPM or other models. The build-up method, unlike the expanded CAPM, excludes the application of beta to the equity risk premium. In the build-up model, in which beta is implicitly assumed equal to 1, an argument exists to include an industry risk adjustment (premium or discount), although there are challenges in measuring industry risk adjustments. For the baseline implementation of the build-up model, we take the model with an industry risk adjustment.
- *Relative debt availability and cost of debt*. Another valuation challenge involves correctly estimating a private company's debt capacity. In calculating a weighted average cost of capital (WACC) for a valuation based on FCFF, analysts should note that a private company may have less access to debt financing than a similar public company. This lesser access means the private company may need to rely more on equity financing, which would tend to increase its WACC. Furthermore, a private company's typically smaller size could lead to greater operating risk and a higher cost of debt.
- *Discount rates in an acquisition context*. In evaluating an acquisition, finance theory indicates that the cost of capital used should be based on the target company's capital structure and the riskiness of the target company's cash flows—the buyer's cost of capital is irrelevant. In the context of acquisitions made by larger, more mature companies of smaller, riskier target companies, the buyer would be expected to have a lower cost of capital than the target. Both of these practices in general incorrectly transfer value from the buyer to the seller because the buyer would be paying the seller for possible value it brings to a transaction (Damodaran, 2012).
- *Discount rate adjustment for projection risk*. Any lesser amount of information concerning a private company's operations or business model compared with a similar public company introduces greater uncertainty into projections that may lead to a higher required rate of return. As a second area of concern, management of a private company (on whom analysts may need to rely for forecasts) may have less experience forecasting future financial performance. Projections

may reflect excessive optimism or pessimism. Any adjustments to a discount rate to account for projection risk or managerial inexperience in forecasting, however, would typically be highly judgmental.

EXAMPLE 3

Developing a Discount Rate for a Private Company

Duvall and his advisers have decided to use an income approach to value Able Manufacturing.

Because of its years of operating successfully and its owner's conservative nature, Able operated with little debt. Smith explored various sources of debt financing to operate Able with a lower overall cost of capital. Analysis of public companies in Able's industry indicated several guideline public companies for possible use in estimating a discount rate for Able. Duvall and his advisers agreed on the following estimates:

- Risk-free rate: Estimated at 3.8%.
- Equity risk premium: The parties agreed that a 5% equity risk premium was appropriate.
- Beta: A beta of 1.1 was estimated based on publicly traded companies in the same industry.
- Small stock premium: The smaller size and less diversified operations suggest greater risk for Able relative to public companies. A small stock premium of 3% was included in the equity return calculation for these expected risks.
- Company-specific risk premium: Assessment of Able indicated that beyond Smith's key role at the company, no other unusual elements created additional risk. A 1% company-specific risk adjustment was included.
- Industry risk premium (build-up method only): The industry risk premium was 0% because no industry-related factors were viewed as materially affecting the overall required return on equity estimate.
- Pre-tax cost of debt: Estimated at 7.5%.
- Ratio of debt to total capital for public companies in the same industry: Estimated at 20%.
- Optimal ratio of debt to total capital: The ratio was estimated at 10% based on discussions with various sources of financing. Able would be unable to achieve the industry capital structure based on its smaller size compared with public companies and the greater risk of its operations as a standalone company.
- Actual ratio of debt to total capital: For Able, the actual ratio was 2%.
- Combined corporate tax rate: Estimated at 30%.

Based only on the information given, address the following:

1. Calculate the required return on equity for Able using the CAPM.
2. Calculate the required return on equity for Able using the expanded CAPM.
3. Calculate the required return on equity for Able using the build-up method.
4. Discuss the selection of the capital structure weights to use in determining the weighted average cost of capital for Able.

5 Calculate the WACC for Able using the current capital structure and a 13% cost of equity.

6 Calculate the WACC for Able based on the optimal capital structure for Able and a 13% cost of equity.

Solution to 1:

According to the CAPM, Required return on share i = Current expected risk-free return + β_i (Equity risk premium) = 3.8% + 1.1(5%) = 9.30%.

Solution to 2:

The required rate of return is 13.3%, which is shown in the following tabular format.

Able Manufacturing, Inc. Expanded CAPM: Required Rate of Return on Equity	
Risk-free rate	3.8%
Plus: Equity risk premium adjusted for beta[a]	5.5
Plus: Small stock premium	3.0
Plus: Company-specific risk adjustment	1.0
Indicated required return on equity	13.3%

[a] 1.1 beta × 5% equity risk premium = 5.5%.

Solution to 3:

The required rate of return is 12.8%. Note the absence of a beta adjustment. Note also that the fact that beta (1.1) is close to 1.0 possibly suggests any industry risk adjustment that could be made would be small in magnitude.

Able Manufacturing, Inc. Build-Up Method: Required Rate of Return on Equity	
Risk-free rate	3.8%
Plus: Equity risk premium	5.0
Plus: Small stock premium	3.0
Plus: Industry risk premium	0.0
Plus: Company-specific risk adjustment	1.0
Indicated return on equity	12.8%

Solution to 4:

For valuation concerning the possible sale of Able, it is appropriate to assume the weights in the optimal capital structure in calculating WACC because an acquirer would be able and motivated to establish the optimum. Able's current capital structure involves less debt than the optimal one; thus the company's WACC is currently higher than it needs to be. Note, however, that the weight on debt of similar large public companies may be higher than what is optimal for Able. Large public companies would be expected to have better access to public debt markets. Also, Able's small size increases its risk relative to larger public companies. These two factors tend to increase Able's cost of debt relative to a large public comparable and lead to a lower optimal weight of debt compared with such a public company.

Solution to 5:

The cost of capital for Able based on the existing capital structure was calculated as follows:

Able Manufacturing, Inc. Calculation of Weighted Average Cost of Capital Current Capital Structure		
Pre-tax cost of debt	7.5%	
Tax rate complement (1 – Tax rate)	0.70	
After-tax cost of debt	5.3%	
Weight	×0.02	
Weighted cost of debt		0.1%
Cost of equity	13.0%	
Weight	×0.98	
Weighted cost of equity		12.7%
Weighted average cost of capital		12.8%

Solution to 6:

The overall cost of capital using the optimal capital structure for Able reflected a higher level of debt financing. The WACC was calculated as follows:

Able Manufacturing, Inc. Calculation of Weighted Average Cost of Capital Optimal Capital Structure		
Pre-tax cost of debt	7.5%	
Tax rate complement (1 – Tax rate)	0.70	
After-tax cost of debt	5.3%	
Weight	0.10	
Weighted cost of debt		0.5%
Cost of equity	13.0%	
Weight	0.90	
Weighted cost of equity		11.7%
Weighted average cost of capital		12.2%

Note: Rounded figures are used.

For early stage development companies, discount rate estimation concerns are magnified. Very high levels of company-specific risk, for example, may make using the CAPM problematic. Several life cycle stages exist with perceived broad ranges of absolute rate of return requirements for companies operating in each stage. Further, there can be uncertainty in classifying a company in a specific life cycle stage.

AICPA practice aids

The American Institute of Certified Public Accountants (AICPA) released practice aids ("Valuation of Privately-Held-Company Equity Securities Issued as Compensation" and "Assets Acquired in a Business Combination to Be Used in Research and Development Activities: A Focus on Software, Electronic Devices, and Pharmaceutical Industries") to provide technical guidance for stock valuation in the context of stock option grants and other share-based payments. Paragraph 119 of "Valuation of Privately-Held-Company Equity Securities Issued as Compensation" notes that "One of the objectives and benefits of becoming a public enterprise is the ability to access the public capital markets, with the associated benefits of a lower cost of both equity and debt capital." The practice aids also provide descriptive information on various stages in the early life cycle of development-stage companies and estimated return requirements.

4 FREE CASH FLOW, CAPITALIZED CASH FLOW AND EXCESS EARNINGS METHODS

c explain the income, market, and asset-based approaches to private company valuation and factors relevant to the selection of each approach;

e calculate the value of a private company using free cash flow, capitalized cash flow, and/or excess earnings methods;

Free cash flow valuation for private and public companies is substantially similar. For example, in the case of Able Manufacturing, a FCF valuation might involve projecting individual free cash flows for a number of years, finding the present value of those projected free cash flows, followed by finding the present value of a terminal value estimate that captures the business enterprise value at the end of the initial projection period. In principle, discrete free cash flow forecasts should be made until cash flows are expected to stabilize at a constant growth rate.

To value the business enterprise at the end of the initial projection period, the capitalized cash flow method incorporating a sustainable long-term growth rate is a theoretically preferred method. Some appraisers, however, will calculate the terminal value using pricing multiples developed in the market approach. For a company in a high-growth industry, market multiples would be expected to capture rapid growth in the near future and "normal" growth into the indefinite future. If we use these multiples to estimate terminal value, the residual enterprise value may be inappropriate because rapid growth was incorporated twice: once in the cash flow projections over the projection period and also in the market multiple used in calculating the residual enterprise value.

4.1 Capitalized Cash Flow Method

The capitalized cash flow method (CCM) estimates value based on the expression for the value of a growing perpetuity and is essentially a stable growth (single-stage) free cash flow model. Although rarely used for the valuation of public companies, larger private companies, or in the context of acquisitions or financial reporting, the CCM may be appropriate in valuing a private company for which no projections are available and an expectation of stable future operations exists. If market pricing evidence from public companies or transactions is limited, a CCM valuation may also be a feasible alternative.

For companies that are not expected to grow at a constant rate, FCF valuation using a series of discrete cash flow projections is theoretically preferable to the CCM. The CCM could provide assistance in assessing the discount rate or growth assumptions embedded in value indications from the market approach.

At the firm level, the formula for the capitalized cash flow to the firm is

$$V_f = \text{FCFF}_1/(\text{WACC} - g_f) \quad (1)$$

where

V_f = value of the firm
FCFF_1 = free cash flow to the firm for next 12 months
WACC = weighted average cost of capital
g_f = sustainable growth rate of free cash flow to the firm

The value of equity is found as the value of the company less the market value of debt, or V_f – (Market value of debt). An implicit assumption in using WACC for discounting FCFF in Equation 1 is that a constant capital structure at market values in the future exists.

To value equity directly, the inputs for free cash flow would reflect FCFE and the equity return requirement would be substituted for the WACC:

$$V = \text{FCFE}_1/(r - g) \quad (2)$$

where r is the required return on equity and g is the sustainable growth rate of free cash flow to equity. In Equations 1 and 2, the denominator is known as the **capitalization rate**. Thus, the estimate of value in each is calculated as the forecasted Year 1 FCF divided by the capitalization rate. Example 4 illustrates the application of the CCM.

EXAMPLE 4

Valuation Using the Capitalized Cash Flow Method

Duvall and his team are comfortable with the normalized earnings, growth, and discount rate estimated for Able. The company's management has not developed detailed projections for Able. Suppose that free cash flow to the firm is expected to grow at 3% annually from the level of $10,986,100 forecast earlier.

1 Explain the rationale for using the CCM in this case.
2 Calculate the value of the equity of Able using the CCM and a WACC of 12.2% based on Able's optimal capital structure.
3 Calculate the value of Able's equity using the WACC of 12.8% based on the existing capital structure.
4 Discuss factors leading to the difference in the computed values.

Solution to 1:

The CCM is appropriate given the assumption that free cash flow to the firm grows at a constant rate (here 3%) is accurate. Otherwise, at best it provides a rough value estimate.

Solution to 2:

Using the estimated free cash flow to the firm, a capitalization rate of 9.2% (12.2% – 3%) was applied to derive a valuation indication for the business enterprise. Able's debt balance was subtracted to arrive at an equity value calculated as follows.

Able Manufacturing, Inc. Capitalized Cash Flow Method—Optimal Capital Structure

Free cash flow to firm		$10,986,100
Weighted average cost of capital	12.2%	
Long-term growth rate	3.0%	
Capitalization rate		9.2%
Indicated value of invested capital		119,414,130
Less: Debt capital (actual, assumed to equal market value)		2,000,000
Indicated value of equity		$117,414,130

Solution to 3:

This calculation is similar to the one in the Solution to 2 except for the use of a capitalization rate of 9.8% (12.8% – 3%).

Able Manufacturing, Inc. Capitalized Cash Flow Method—Existing Capital Structure

Free cash flow to firm		$10,986,100
Weighted average cost of capital	12.8%	
Long-term growth rate	3.0%	
Capitalization rate		9.8%
Indicated value of invested capital		112,103,061
Less: Debt capital		2,000,000
Indicated value of equity		$110,103,0615

Solution to 4:

The low level of debt in the existing capital structure results in a higher WACC and a lower valuation conclusion for Able relative to the optimal capital structure.

4.2 Excess Earnings Method

In a business valuation context, the excess earnings method (EEM) involves estimating the earnings remaining after deducting amounts that reflect the required returns to working capital and fixed assets (i.e., the tangible assets). This residual amount of earnings (i.e., "excess earnings") is capitalized by using the growing perpetuity formula from the CCM to obtain an estimate of the value of intangible assets. Generally, the

EEM has been used to value intangible assets and very small businesses when other such market approach methods are not feasible. For valuing the entire business, the values of working capital and fixed assets are added to the capitalized value of intangibles.

Applying the EEM to value a business enterprise involves the following steps:

1. Estimate values of working capital and fixed assets (typically, fair value estimates are used). Suppose these are €200,000 and €800,000, respectively.
2. Determine the normalized earnings of the business enterprise. Suppose normalized earnings are €120,000 for the year just ended.
3. Develop discount rates for working capital and fixed assets. Working capital is viewed as the lowest-risk and most liquid asset with the lowest required rate of return. Fixed assets require a somewhat greater rate of return. Intangible assets, given their limited liquidity and high risk, often require the highest return. Suppose the required returns on working capital and fixed assets are 5% and 11%, respectively.
4. Calculate required returns associated with working capital and fixed assets and subtract the required returns on working capital and fixed assets from the normalized earnings of the business enterprise to estimate the residual income. This residual income, if any, must reflect the value associated with intangible assets. In this case, residual income is €120,000 – 0.05(€200,000) – 0.11(€800,000) = €22,000. Assume that residual income grows at 3% annually.
5. Estimate discount rate and capitalization rate required for valuing the intangible assets. This estimate typically represents all intangible assets (including customer relationships, technology, trade names, and the assembled work force, among others). The details of such a calculation are outside the scope of our coverage. Significant judgement is associated with many of these estimates (for further discussion, see The Appraisal Foundation's best practice guide, "The Identification of Contributory Assets and Calculation of Economic Rents"). For the purpose of this illustration, assume the discount rate is 12%.
6. Value the enterprise's intangible assets using the formula for a growing perpetuity. The total value of intangible assets is (1.03)(€22,000)/(0.12 – 0.03) ≈ €251,778. (Because €22,000 is associated with normalized income for the most recent year, it is increased by its assumed 3% growth rate to obtain a forecast of the year-ahead residual income.)
7. Total of working capital, fixed assets, and intangibles equals the value of the business. The EEM estimate is €200,000 + €800,000 + €251,778 = €1,251,778.

As mentioned, the EEM is used only rarely in pricing entire private businesses, and then only for small ones. Some have viewed the specific return requirements for working capital, tangible assets, and the residual income associated with intangible assets as not readily measurable (again, for further detail, see The Appraisal Foundation's best practice guide, "The Identification of Contributory Assets and Calculation of Economic Rents").

For financial reporting, the concept of residual income is an important element of intangible asset valuations and has wide acceptance. Residual income is the subject of significant discussion among appraisers who perform purchase price allocation valuations of intangible assets pursuant to IFRS 3 or Accounting Standards Codification (ASC) 805. An analyst considering intangible asset amortization and goodwill impairment issues would benefit from an understanding of residual income concepts.

5 MARKET APPROACH METHODS AND THE GUIDELINE PUBLIC COMPANY METHOD

c explain the income, market, and asset-based approaches to private company valuation and factors relevant to the selection of each approach;

h calculate the value of a private company based on market approach methods and describe advantages and disadvantages of each method;

The market approach uses direct comparisons to public companies and acquired enterprises to estimate the fair value of an equity interest in a private company. Three major variations of the market approach exist:

- The **guideline public company method** (GPCM) establishes a value estimate based on the observed multiples from trading activity in the shares of public companies viewed as reasonably comparable to the subject private company. The multiples from the public companies are adjusted to reflect differences in the relative risk and growth prospects of the subject private company compared with the guideline public companies.
- The **guideline transactions method** (GTM) establishes a value estimate based on pricing multiples derived from the acquisition of control of entire public or private companies that were acquired. Whereas GPCM uses a multiple that could be associated with trades of any size, GTM uses a multiple that specifically relates to sales of entire companies.
- The **prior transaction method** (PTM) considers actual transactions in the stock of the subject private company. The actual price paid for shares or the pricing multiples implied by past transactions in the stock can be used for this method.

Because the market approach relies on data generated in actual market transactions, some consider it conceptually preferable to the income- and asset-based approaches for private company valuation. In the United States, tax courts assessing private company valuations have generally stated a preference for valuation based on market transactions, although they often accept valuations based on the income approach. IFRS 13 and ASC 820 also present a fair value hierarchy that gives the highest priority to market-based evidence. Specifically, IFRS 13 states that "fair value hierarchy gives the highest priority to quoted prices (unadjusted) in active markets for identical assets or liabilities (Level 1) and the lowest priority to unobservable inputs (Level 3)." The primary assumption of the market approach is that transactions providing pricing evidence are reasonably comparable to the subject company.

A primary challenge in using the market approach is finding comparable companies and accurately assessing their pricing. All of the company-specific factors noted previously may lead to different levels of expected risk and growth for a private company relative to a public one. Market multiples reflect both expected risk and growth. Risk and growth assumptions should be extracted and multiples adjusted to reflect any differences of the subject company vis-à-vis the chosen comparable(s). The stock-specific factors associated with private companies may create additional uncertainties regarding levels of risk and growth.

The pricing of shares in public companies reflects stock price volatility as a result of, in part, their ready marketability. Rapid movements in public companies' stock prices can lead to changes in pricing multiples that often serve as a basis for private company valuations. Interests in private companies have much more limited marketability and often require longer periods to completely sell. The extended period to sell such an interest in full and the likely movement of pricing multiples over the sale period create uncertainty in the determination of a pricing multiple and thus in the final value conclusion.

Factors for identifying guideline companies are similar for public and private companies. Key factors include industry membership, form of operations, trends, and current operating status, among others. As previously noted, life cycle and size differences may create significant challenges in applying the market approach.

Public and private company analysis may differ in the financial metrics used in the valuation process. Price-to-earnings methods are frequently cited in the valuation of public companies, with other multiples considered as well. For larger, mature private companies, pricing multiples are frequently based on EBITDA and/or EBIT. EBITDA is best compared with the **market value of invested capital** (MVIC), defined as the market value of debt and equity, in forming the valuation metric. In addition to MVIC, other similar terms include enterprise value (EV), business enterprise value (BEV), and firm value. Definitions for enterprise value vary but most frequently start with MVIC and subtract any cash and cash equivalents. BEV is typically synonymous with EV.

With a calculation of MVIC for a private company, the value of debt can be subtracted to produce an estimate of equity value. Because current transaction market values for debt are unavailable in many cases, some estimate of the market value of debt is needed. The use of the face value of debt as an estimate may be acceptable in many situations in which debt represents a small fraction of overall financing and operations are stable. For companies with highly leveraged financial conditions and/or significant volatility expected in future financial performance, the valuation of equity as the residual obtained by subtracting the face value of debt from the value of the business enterprise is frequently not appropriate. Consider highly leveraged companies or companies with significant volatility of financial performance whose debt may be valued at significant discounts from face value. In these cases, option pricing theory can be used to value each debt and equity instrument as a separate call option on the company's BEV. An alternative in such cases is to estimate market value based on debt characteristics, known as matrix prices. For many very small private companies with limited asset bases, net income–based multiples may be more commonly used than EBITDA multiples. For extremely small companies, multiples of revenue may even be commonly applied. This convention considers the likely absence of meaningful financial data and the greater impact and subjectivity associated with such items as owner compensation.

Non-financial metrics may be an appropriate means of valuation for certain industries. These metrics would probably best be used in addition to financial metrics. Significant reliance on these metrics would be appropriate only if the non-financial measure is generally accepted within the industry. Examples of non-financial metrics include price per subscriber in cable and price per bed for hospital and skilled nursing and other healthcare facilities.

5.1 Guideline Public Company Method

In private company valuation, as has been noted, valuation based on multiples of similar public companies is often referred to as the guideline public company method. The valuation process is essentially similar for a public or a private company. A group of public companies is identified, the relevant pricing multiples for the guideline companies are derived, and adjustments to the multiples are made that reflect the risk and growth prospects of the subject company relative to the publicly traded companies. For a private company, this method would lead to a conclusion of value. For a public company, applying this method helps assess over- or undervaluation of a company relative to similar companies at a specific point in time.

The primary advantage of this method is the potentially large pool of guideline companies and the significant descriptive, financial, and trading information available to the analyst/appraiser. Disadvantages include possible issues regarding comparability and subjectivity in the risk and growth adjustments to the pricing multiple.

Control premiums may be used in valuing a controlling interest in a company. Defined in the IGBVT, a **control premium** is an amount or percentage by which the pro rata value of a controlling interest exceeds the pro rata value of a non-controlling interest in a business enterprise, to reflect the power of control. For the valuation of a controlling interest, a control premium has often been added if the value is derived from the GPCM. The trading of interests in public companies typically reflects small blocks without control of the entity. Given this information, many but not all believe the resulting pricing multiples do not reflect control of the entity.

Control Premium Recommendation

The Appraisal Foundation's 2017 guide *The Measurement and Application of Market Participant Acquisition Premiums* proposes major changes to improve consistency among appraisers' practice in this area. In its draft form, the document recommends that any control premium be justified based on an analysis of projected cash flows after an acquisition and, when justified, that control premiums be calculated at the MVIC rather than the equity level.

A control premium adjustment may be appropriate depending on the specific facts. Historically, control premiums have been estimated based on transactions in which public companies were acquired. Several factors require careful consideration in estimating a control premium.

- *Type of transaction*. Some transaction databases classify acquisitions as either financial or strategic transactions. A **strategic transaction** involves a buyer that would benefit from certain synergies associated with owning the target firm. These synergies could include enhanced revenues, cost savings, or other possible benefits. A **financial transaction** involves a buyer having essentially no material synergies with the target. As an example, the purchase of a private company by a company in an unrelated industry would typically be a financial transaction. Compared with financial transactions, control premiums for an acquisition by a strategic buyer are typically larger because of the expected synergies.
- *Industry factors*. Industry sectors with acquisition activity are considered to be "in play" at a valuation date; that is, pricing of public companies in the sector may reflect some part of a possible control premium in the share prices. Control premiums measured at a date significantly before a valuation date might reflect a different industry environment from that of the valuation date.
- *Form of consideration*. Transactions involving the exchange of significant amounts of stock (as opposed to all-cash transactions) might be less relevant as a basis of measuring a control premium because of the possibility that acquiring companies time such transactions during periods when their management perceives that shares of their company are overvalued in the marketplace.

The multiple resulting from applying a control premium to pricing multiples from publicly traded companies should be assessed for reasonableness. Suppose that a public company, which is viewed as comparable to a private company being appraised, was acquired at an 8× pricing multiple. A control premium of 30% control is paid based on the stock's price prior to the acquisition. Pricing multiples for guideline public companies, however, are 10× at the valuation date. Applying a 30% control premium would suggest a 13× pricing multiple. The dramatically different value indications resulting from applying a 8× transaction multiple and a 13× multiple suggest the need for further investigation before accepting the 13× multiple. Comparability issues or dramatic pricing changes may be factors leading to this material difference.

EXAMPLE 5

Valuation Using Guideline Public Company Method

Duvall decides to use the GPCM to develop a value indication for Able that is independent of the FCF indication he is also pursuing. He believes that many acquirors apply a multiple of market value of invested capital to EBITDA to value companies in Able's industry. A search for comparable public companies indicated several companies that might serve as guidelines or benchmarks for valuing Able; however, all of these are much larger than Able. Duvall's research on guideline public companies indicates the following:

- The MVIC to EBITDA multiples of such public companies cluster near 7.0.
- A combined downward adjustment of 15% for relative risk and growth characteristics of Able compared with the guideline public companies suggests an adjusted MVIC to EBITDA multiple of 5.95, rounded to 6, for Able.
- A control premium of 20% was reported in a single strategic acquisition from several years ago. The transaction involved an exchange of stock with no cash consideration paid.
- Duvall is unaware of any strategic buyers that might incorporate synergies into their valuation of Able.
- Normalized EBITDA is $16,900,000.
- Market value of debt capital is $2,000,000.

1 Explain the elements included in the calculation of a pricing multiple for Able.
2 Calculate the pricing multiple appropriate for Able, including a control premium adjustment.
3 Calculate the value of Able using the guideline public company method.

Solution to 1:

Able's value in relation to a possible acquisition is desired. Pricing multiples from guideline public companies provide a starting point to develop a pricing multiple. The pricing multiples for the guideline public companies must be adjusted to reflect any differences in risk and growth expectations for Able compared with the guideline public companies. As a final element, the pricing multiple should consider the inclusion of a control premium given the possible sale of Able.

Solution to 2:

Considering the absence of any strategic buyers, in the present instance a control premium of 0% is a reasonable baseline. There was a single strategic transaction for the acquisition of a public company several years prior to the acquisition. The age of the transaction, however, creates concern regarding the relevance of the indicated control premium.

Based on the information provided, the MVIC to EBITDA multiple for Able can be taken to approximately 6, reflecting no control premium adjustment.

Able Manufacturing, Inc. Development for Pricing Multiple for Guideline Public Company Method

Initial MVIC to EBITDA from public companies		7.0
Relative risk and growth adjustment for Able	−15%	(1.05)
Multiple before control adjustment		5.95
Control Premium adjustment*	0%	0
Multiple after control adjustment		5.95
Rounded to		6.0

* Control premiums are measured based on the value of the equity or the MVIC of public companies before and after an acquisition. When an equity control premium has been estimated, a valuation on an MVIC basis (as is often the case in a transaction setting) would require an adjustment to the equity control premium. In the example, no control premium was concluded to be appropriate. Assuming an equity control premium of 30% was deemed appropriate based on different facts, a normalized capital structure of one-third debt and two-thirds equity would suggest a 20% control premium (two-thirds of 30%) if applied to an MVIC-multiple-based value from guideline public companies. Control premium data vary markedly, and divergence in practice exists in this area of valuation.

Solution to 3:

Able Manufacturing, Inc. Valuation Using Guideline Public Company Method

Normalized EBITDA	$16,900,000
Pricing multiple	6.0
Indicated value of invested capital	101,400,000
Less: Debt capital	2,000,000
Indicated value of equity	$99,400,000

6 GUIDELINE TRANSACTIONS AND PRIOR TRANSACTION METHODS

h calculate the value of a private company based on market approach methods and describe advantages and disadvantages of each method;

The guideline transactions method is conceptually similar to the guideline public company method. Unlike the GPCM, the GTM uses pricing multiples derived from acquisitions of public or private companies. Transaction data available on publicly reported acquisitions are compiled from public filings made by parties to the transaction with the regulatory bodies, such as the Financial Conduct Authority in the United Kingdom or the Securities and Exchange Commission (SEC) in the United

States. Data on transactions not subject to public disclosure may be available from certain transaction databases. Because information may be limited and is generally not readily confirmed, many appraisers challenge the reliability of this data. All other things equal, transaction multiples would be the most relevant evidence for valuation of a controlling interest in a private company.

A number of factors need to be considered in assessing transaction-based pricing multiples.

- *Synergies*. The pricing of strategic acquisitions may include payment for anticipated synergies. The relevance of payments for synergies to the case at hand merits consideration.
- *Contingent consideration*. **Contingent consideration** represents potential future payments to the seller that are contingent on the achievement of certain agreed-on occurrences. Obtaining some form of regulatory approval or achieving a targeted level of EBITDA are two types of contingencies. Contingent consideration may be included in the structure of acquisition. The inclusion of contingent consideration in the purchase price paid for an enterprise often reflects uncertainty regarding the entity's future financial performance. IFRS 3 and ASC 805 changed the requirements for measuring and reporting contingent consideration in the context of a business combination (further details in "Valuation of Contingent Consideration" by The Appraisal Foundation, February 2019).
- *Non-cash consideration*. Acquisitions may include stock in the consideration. The cash equivalent value of a large block of stock may create uncertainty regarding the transaction price. For example, the 2001 merger of America Online (AOL) and Time Warner Corporation was a stock swap that occurred at a time when AOL stock was trading based on expectations of significant future growth. In 2002, the combined company reported two charges for goodwill impairment expense totaling $99 billion. The level of this impairment expense raises questions regarding whether the initial transaction price reflected temporary overvaluation of AOL stock relative to its intrinsic value.
- *Availability of transactions*. Meaningful transactions for a specific private company may be limited. The relevance of pricing indications from a transaction that occurred a significant period prior to a valuation date can be challenged—especially if evidence indicates changes in the subject company, industry, or economy between the transaction date and the valuation date.
- *Changes between transaction date and valuation date*. Unlike the guideline public company method, which develops pricing multiples based on stock prices at or very near the valuation date, the guideline transactions method relies on pricing evidence from acquisitions of control of firms at different points in the past. In many industries, transactions are limited and transactions several months or more from a valuation date may be the only transaction evidence available. Changes in the marketplace could result in differing risk and growth expectations requiring an adjustment to the pricing multiple.

EXAMPLE 6

Valuation Using Guideline Transactions Method

In addition to the income approach and the guideline public company method, the guideline transactions method was considered and applied. Duvall and his advisers noted the following:

- Pricing multiples from several recent acquisitions of private companies in the industry indicated a MVIC to EBITDA multiple of 6.0.
- Several of the acquisitions studied were viewed as similar to Able because of similar revenue bases and limited diversification. The overall risk and growth characteristics of the acquired companies and Able were viewed as similar.

1 Discuss differences between pricing multiples from the guideline transactions and guideline public company methods.
2 Explain the calculation of a pricing multiple using the guideline transactions method.
3 Calculate the pricing multiple appropriate for Able.
4 Calculate the value of Able using the guideline transactions method.

Solution to 1:

The guideline transactions method considers market transactions involving the acquisition of the total equity of companies. As such, the pricing multiple, compared to other methods, more accurately reflects the value of total companies. Pricing multiples from guideline public companies typically reflect public trading in small blocks of stock. The multiples may not reflect the value of the total equity of the public companies.

Solution to 2:

The pricing multiples from acquisitions are the basis for the pricing multiple. The risk and growth prospects of the acquired companies and the subject private company are assessed and an adjustment factor is applied. Because the multiples reflect acquisitions of total equity, they reflect the value of total equity. No control premium adjustment is necessary.

Solution to 3:

Calculation of the initial pricing multiple follows:

Able Manufacturing, Inc. Development of Pricing Multiple for Guideline Transactions Method

Initial MVIC to EBITDA from transactions		6.0
Relative risk and growth adjustment for Able	0%	0.0
Indicated multiple		6.0
Rounded to		6.0

Solution to 4:

Valuation using the guideline transactions method is similar to that from the guideline public company method, except any control premium is already incorporated in the transaction multiple.

Able Manufacturing, Inc. Guideline Transactions Method	
EBITDA	$16,900,000
Pricing multiple	6.0
Indicated value of invested capital	101,400,000
Less: Debt capital	2,000,000
Indicated value of equity	$99,400,000

6.1 Prior Transaction Method

The prior transaction method considers actual transactions in the stock of the subject company. Valuation can be based on either the actual price paid or the multiples implied from the transaction. The PTM is generally most relevant when considering the value of a minority equity interest in a company. For many private companies, there are no or very limited transactions in the stock.

If actual transactions in the stock of the subject company took place, were done at arm's length, and information is available, the PTM would be expected to provide the most meaningful evidence of value. The PTM provides less reliable valuation evidence if transactions are infrequent. Also, uncertainly regarding the parties' motivations, or special circumstances surrounding a prior transaction, can create uncertainty regarding PTM data reliability. Transactions at different points in time may require significant adjustment. As an example, an early stage venture capital–funded company experiences rapid value increases resulting from successful execution of its development plans. A transaction conducted prior to the company achieving a significant value event might not provide meaningful value insights at a subsequent date.

The PTM can provide insights on the value of development-stage entities when revenues and cash flows are highly speculative. Many development-stage companies fund development activities through several rounds of equity financing. As such, there may be a series of prior transactions providing valuation evidence. The equity financing often involves the sale of preferred stock with liquidation preferences and rights to convert to common stock. Because development-stage entities often have complex capital structures with different classes of equity securities that have differing rights, significant adjustments are required. This process is complex and requires significant judgment. The AICPA guide, "Valuation of Privately-Held-Company Equity Securities Issued as Compensation" (2013) provides further insights.

7 ASSET BASED APPROACH

c explain the income, market, and asset-based approaches to private company valuation and factors relevant to the selection of each approach;

i describe the asset-based approach to private company valuation;

The principle underlying the asset-based approach is that the value of ownership of an enterprise is equivalent to the fair value of its assets less the fair value of its liabilities. Of the three approaches to valuation, the asset-based approach (also referred to as the **cost approach** by many in the valuation profession) is generally considered to be the weakest from a conceptual standpoint for valuing an ongoing business enterprise.

The asset-based approach is rarely used for the valuation of going concerns. Reasons include the limited market data available to directly value intangible assets, difficulties in valuing certain tangible assets (such as special use plant and equipment), and the more readily available information to value operating companies as an integrated whole rather than on an asset-by-asset basis.

An operating company with nominal profits relative to the values of assets used and without prospects for doing better in the future might best be valued using an asset-based approach assuming the winding up of operations. In this case, its value as a going concern might be less than its value in liquidation (the value that could be realized through the liquidation of its assets) because the assets might be redeployed by buyers to higher-valued uses. Resource and financial companies might also be valued based on an asset-based approach. Banks and finance companies largely consist of loan and securities portfolios that can be priced based on market variables. In such cases, a summation of individual asset value estimates may give a lower-bound-type estimate of the company's overall value. The asset-based approach may be appropriate for the valuation of holding (investment) companies, such as real estate investment trusts (REITs) and closed end investment companies (CEICs). For these entities, the underlying assets typically consist of real estate or securities that were valued using the market and/or income approaches. An asset-based approach may also be appropriate for very small businesses with limited intangible value or for early stage companies.

For the valuation of an interest in a pooled investment vehicle, certain factors may suggest a value different from the net asset value per share. Management and incentive fees may lead to an expectation of proceeds available to an investor and a value estimate that is less than the net asset value per share. The relative growth and profit as a result of management expertise may also merit an upward or downward adjustment to the net asset value. Other factors, such as the possible effect of tax attributes (tax basis in the assets held by the entity) and diversification, and professional management benefits may also affect value.

Exhibit 1 illustrates four definitions of values that a private business appraiser used to value the financial services subsidiary of a public company.

Exhibit 1

Valuation of a Financial Services Company

In a valuation of a financial services company, a business appraiser estimated four values for the company using four different approaches, which she characterized as follows:

1 *Discounted cash flow approach*. The appraiser estimated value as the present value of projected FCFE for the next 10 years plus the present value of the capitalized value of the 11th-year cash flow.

Exhibit 1 (Continued)

2 *Market approach*. The appraiser used the GPCM with price-to-cash flow, price-to-book (mainly in the case of financial services companies), and price-to-earnings multiples, and she made adjustments to reflect differences in risk and growth, applying the resulting multiples to the company's cash flow, book value, and earnings, respectively.

3 *Adjusted book value approach, going-concern basis*. The appraiser adjusted the book values of assets and liabilities to better reflect market values and obtained the adjusted book value of equity, which was the estimate of value based on this approach. The definition of market value used was: "Market value is…the most probable price that an asset should bring in a competitive and open market under all conditions requisite to a fair sale, the buyer and seller each acting prudently and knowledgeably, and assuming the price is not affected by undue stimulus." This approach is relevant mostly for financial services companies.

4 *Adjusted book value approach, orderly liquidation basis*. The appraiser adjusted the book values of assets and liabilities to better reflect orderly liquidation values and obtained the liquidation book value of equity, which was the estimate of value based on this approach. The definition of orderly liquidation value used was: "Orderly liquidation value [is] the price [the asset] would bring if exposed for sale on the open market, with a reasonable time allowed to find a purchaser, both buyer and seller having knowledge of the uses and purposes to which the asset is adapted and for which it is capable of being used, the seller being compelled to sell and the buyer being willing, but not compelled, to buy."

State and explain which of the foregoing methods would be expected to produce the lowest value estimate.

Methods 1, 2, and 3 recognize a going-concern value for the company; Method 4 does not, so the value estimates under 4 should be the lowest unless the entity is unable to generate sufficient income to justify continued operations. In general, using individual assets in a coordinated way in the operation of a business as implicitly assumed in 1 and 2 should increase value. Between Methods 3 and 4, the element of the seller being compelled to sell should result in 4 being the lowest estimate. Method 3 may also produce a low value because of difficulties in identifying and valuing intangible assets. Because goodwill includes a variety of items, including potential future customers and other intangibles that may be created, its direct valuation is extremely difficult.

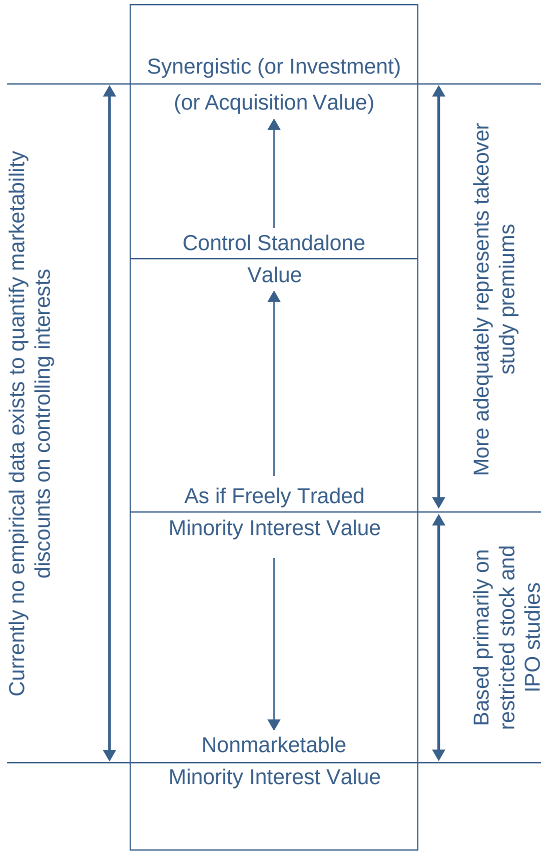
VALUATION DISCOUNTS AND PREMIUMS

8

j explain and evaluate the effects on private company valuations of discounts and premiums based on control and marketability.

Control and/or marketability adjustments are often included in valuations of interests in private companies. This area is one of the primary differences in the valuation of interests in private companies compared with public companies. The chart in Exhibit 2 is adapted from Hitchner (2017) and presents the relationship of these concepts and other concepts that we discuss. As the chart indicates, the inclusion of discounts depends, in part, on the starting point of a valuation.

Exhibit 2 Valuation of Private Companies



Starting at the top of the chart, the highest possible value indication for an entity would be its **investment value** to the optimal synergistic buyer. This value reflects a controlling interest assumption, which also increases value. Below the control value of the enterprise to a strategic buyer is the value of the enterprise to a standalone (financial) buyer. In this case, specific synergies to the buyer are unavailable. The "As If Freely Traded/Minority Interest Value" represents the value of a non-controlling equity interest that is readily marketable. This value would be equivalent to the price at which most publicly traded companies trade in the market. The lowest level of value is the "Nonmarketable/Minority Interest Value." This value reflects the reduction to value associated with the lack of control and ready marketability associated with small equity interests in private companies.

The application of both control premiums and lack of control and marketability discounts is fact-specific, and estimates may vary dramatically. Variations in estimated discounts and premiums may relate to the challenging comparability of the data used to quantify discounts. Discounts may also vary based on interpretation of the importance of the size of shareholding and distribution of shares, the relationship of parties, state law affecting minority shareholder rights, and other factors.

The timing of a potential liquidity event is one key consideration. An interest in a private company that is pursuing either an IPO or a strategic sale might be valued with relatively modest valuation discounts. An equity interest in a private company that has not paid dividends and has no prospect for a liquidity event would likely require much higher valuation discounts.

8.1 Lack of Control Discounts

A **discount for lack of control** (DLOC), as defined in the IGBVT, is an amount or percentage deducted from the pro rata share of 100% of the value of an equity interest in a business to reflect the absence of some or all of the powers of control.

Lack of control discounts may be necessary for valuing non-controlling equity interests in private companies if the value of total equity was developed on a controlling interests basis. The lack of control may be disadvantageous to an investor because of the inability to select directors, officers, and management that control an entity's operations. Without control, an investor is unable to distribute cash or other property, to buy and sell assets, to obtain financing, and to bring about other actions, all of which could affect the value of the investment, the timing of distributions, and the ultimate return to the investor.

Although an interest may lack control, the effect of this lack of control on value is uncertain. The US SEC suggests that evidence of "disproportionate returns" is important in supporting the application of lack of control discounts. Disproportionate returns would result when control shareholders increase their returns through above-market compensation and other actions that reduce the returns available to minority shareholders. For private companies seeking a liquidity event through an IPO or strategic sale of the entity, the likelihood decreases that a controlling group will take actions that reduce an entity's earnings.

Data available for estimating a lack of control discount are limited and interpretations can vary markedly. For interests in operating companies, control premium data from acquisitions of public companies have been used frequently in the past. The factors cited earlier on calculating a control premium should also be considered for estimating a lack of control discount. Noting the uncertainties in demonstrating the adverse financial impact of an interest's lack of control and finding appropriate data to measure that lack of control, the equation used frequently in the calculation of a lack of control discount is

$$\text{DLOC} = 1 - [1/(1 + \text{Control premium})].$$

For example, if a 15% control premium is assumed, the associated DLOC is 1 – (1/1.15) = 0.130, or 13.0%.

The following sets forth the typical application of DLOC based on the different methods of valuation.

Method	Basis of Valuation	DLOC Expected?
GTM	Control	Yes
GPCM	Typically minority	No
CCM/FCF	Control or minority	Depends on cash flows

Valuation indications from the CCM and FCF methods of the income approach are generally agreed to be a controlling interest value if cash flows and the discount rate are estimated on a controlling interest basis. If control cash flows are not used and/or the discount rate does not reflect an optimal capital structure, the resulting value is generally believed to reflect a lack of control basis.

Some analysts believe trading in REITs and CEICs may provide a basis for estimating lack of control discounts as well. Because individual REITs and CEICs may trade at premiums, discounts, or near their net asset value at different points in time, the use of this data to quantify the lack of control is challenging and outside the scope of our coverage.

8.2 Lack of Marketability Discounts

A **discount for lack of marketability** (DLOM), as defined in the IGBVT, is an amount or percentage deducted from the value of an ownership interest to reflect the relative absence (compared with publicly traded companies) of a ready market for a company's shares.

Lack of marketability discounts are frequently applied in the valuation of non-controlling equity interests in private companies. Although a DLOM is different from a DLOC, the two discounts are often linked; that is, if a valuation is on a non-controlling interest basis, a lack of marketability discount is typically appropriate. Key variables affecting a marketability discount include prospects for liquidity, contractual arrangements affecting marketability (such as lock-up agreements), restrictions on transferability, pool of potential buyers, risk or volatility, size and timing of distributions (duration of asset), uncertainty of value, and concentration of ownership (AICPA "Valuation of Privately-Held-Company Equity Securities", 2013). At a minimum, an interest that lacks marketability involves a potential opportunity cost associated with the inability to redeploy investment funds.

Restricted stock transactions and IPOs are two types of data used to quantify lack of marketability discounts. A variety of option pricing models are being used to develop marketability discount estimates as well. Although valuation professionals generally agree that these sources offer the best available data to support discounts, all of these approaches are subject to significant differences in their interpretation.

In the United States, SEC Rule 144 provides certain restrictions on the resale of unregistered stock in public companies. Shares acquired prior to an IPO are an example of shares that might be subject to Rule 144 restrictions. These restrictions prevent resale of shares subject to the requirements of Rule 144 in an attempt to maintain an orderly trading market for the publicly traded shares. Restricted stock is essentially identical to freely traded stock of a public company except for the trading restrictions. Unlike interests in private companies, restricted stock transactions typically involve shares that will enjoy ready marketability in the near future. Note that it is sometimes observed that the sale of blocks of restricted stock that significantly exceed public trading activity in the stock may be the most comparable data for quantifying a lack of marketability discount. If the block size significantly exceeds trading volumes, large blocks of restricted shares may still be illiquid when Rule 144 restrictions terminate. A private sale of such a block may reflect a valuation discount related to the price risk associated with the holding.

The relationship of stock sales prior to IPOs is another source of marketability discounts. In many companies (especially early stage or high-growth companies) approaching an IPO, value may be increasing as levels of risk and uncertainty decline because the company is progressing in its development. Reduction in risk associated with realization of the predicted cash flows or a narrowing of the ranges of possible future cash flows would lead to a reduction in the implied marketability discount. Some studies have attempted to adjust for this factor. According to AICPA ("Valuation of Privately-Held-Company Equity Securities", 2013, known as "Stock Practice Aid"), "The cost of equity capital for a private enterprise prior to its IPO generally ranges from 20% to 35%," (paragraph 117); and "By contrast, the cost of equity capital for a newly public enterprise generally ranges from 15% to 25%" (paragraph 119).

A variety of models involving put options have also been used to quantify lack of marketability discounts. As the first step of this process, an at-the-money put option is priced. The value of the put option as a percentage of the value of the stock before any DLOM provides an estimate of the DLOM as a percentage. DLOM based on put options are often used for equity interests in development stage companies. For these companies, liquidity in the short to intermediate term is frequently a key objective of investors.

The key assumptions are the expected term until a liquidity event and the level of volatility associated with the company. One advantage of the put option analysis is the ability to directly address perceived risk of the private company through the volatility estimate. The volatility estimate may better capture the risks of the stock compared with restricted stock or IPO transactions in which volatility may be one of many variables influencing the level of discount. An estimate of volatility can be developed at the valuation date based on either historical volatilities of public companies or the volatility estimates embedded in the prices of publicly traded options. Put options provide only price protection (the protection lasts for the life of the option). They do not, however, provide liquidity for the asset holding, raising a concern on the use of this form of estimate of the DLOM. Put options also allow the holder of the underlying security to benefit from potential price increases in the security's value and thus do not exactly model lack of marketability.

In addition to control and marketability discounts, a variety of other potential valuation discounts exist that may require consideration. These include key person discounts, portfolio discounts (discount for non-homogeneous assets), and possible discounts for non-voting shares.

If both lack of control and lack of marketability discounts are appropriate, these discounts are applied in sequence and are essentially multiplicative rather than additive. The discounts are multiplicative because the valuation process involves discrete steps—first moving from a controlling to a non-controlling basis and then moving from a marketable to a non-marketable basis. For an equity interest in which a 10% lack of control discount and a 20% lack of marketability discount are believed to be appropriate, the total discount is 28% [1 – (1 – 10%)(1 – 20%)] rather than 30% (10% + 20%).

EXAMPLE 7

Application of Valuation Discounts

Suppose that Jane Doe owns 10% of the stock of Able, and the remaining 90% is held by CEO John Smith. Smith is interested in selling Able to a third party. Smith advises Doe that if Able is not sold, he has no reason to purchase Doe's 10% interest. Assume the following:

- Valuation discounts assuming imminent transaction:
  - Lack of control discount = 0%.
  - Lack of marketability discount = 5%.
- Valuation discounts assuming continued operation as a private company:
  - Lack of control discount: incorporated through use of reported earnings rather than normalized earnings.
  - Lack of marketability discount = 25%.
- Indicated value of equity in operations:
  - $96,000,000 in sale scenario.
  - $80,000,000 in "stay private scenario."
- Discuss the relevance of valuation discounts assuming an imminent sale of Able.
- Explain which estimate of equity value should be used and calculate the value of Doe's equity interest in Able assuming a sale is likely.
- Discuss the relevance of valuation discounts assuming Able continues as a private company.

- Explain which estimate of equity value should be used and calculate the value of Doe's equity interest assuming Able continues as a private company.
- Contrast the valuation conclusions and discuss factors that contribute to the difference in the concluded values.

Solution to 1:

The sale of Able can be completed only with Smith's concurrence, given his 90% equity interest. If a sale of Able seems imminent, valuation discounts associated with Doe's 10% equity interest would be modest. The controlling shareholder, Smith, would maximize the sales proceeds to himself and any other shareholder(s). Hence, the lack of control associated with a small minority equity interest would not be a factor. Note that when the controlling stockholder sells, he is not always obligated to offer the minority shareholders the same price. The analyst should investigate this fact and consider factors including 1) intent of the controlling stockholder, 2) articles of incorporation, and 3) legal statutes on corporate governance and shareholder rights. The pending transaction being driven by the controlling shareholder reduces the adverse impact of the limited marketability of an interest in a private company.

Solution to 2:

If a sale is viewed as highly likely, the $96,000,000 equity value would be appropriate. This equity value uses normalized earnings and a discount rate based on an optimal capital structure in calculating the capitalization rate applied to earnings.

Able Manufacturing, Inc. Valuation of Doe's 10% Equity Interest Sale of Company Viewed as Highly Likely

Indicated value of equity in operations	$96,000,000
Interest appraised	10%
Pro rata value of 10% equity interest	9,600,000
Less: Lack of control discount of 0%	0
Value assuming ready marketability	9,600,000
Less: Lack of marketability discount of 5%	480,000
Indicated value of Doe's 10% equity interest	$9,120,000

Solution to 3:

If Smith has no intent to sell the company, the above-market expenses may continue. With the above-market expenses, the reported earnings would be lower than the normalized earnings. Use of reported earnings rather than normalized earnings is one possible means of capturing the adverse impact associated with the lack of control of a small minority equity interest.

Given the absence of any potential liquidity event and the above-market expenses, little market for the stock exists. A higher lack of marketability discount would be appropriate for the interest in this situation.

Solution to 4:

If continuing as a private company is viewed as highly likely, the $80,000,000 equity value would be appropriate. This equity value uses reported earnings and a discount rate based on the actual (not optimal) capital structure in calculating the capitalization rate applied to earnings.

Able Manufacturing, Inc. Valuation of Doe's 10% Equity Interest Continued Operation as a Private Company Likely

Indicated value of equity in operations	$80,000,000
Interest appraised	10%
Pro rata value of 10% equity interest	8,000,000
Less: Lack of control discount*	0
Value assuming ready marketability	8,000,000
Less: Lack of marketability discount of 25%	2,000,000
Indicated value of Doe's 10% equity interest	$6,000,000

* As noted in the example, the impact on the value of the 10% equity interest was assumed to be captured in the use of reported rather than normalized earnings. Also, the actual capital structure was used rather than the optimal capital structure. A wide range of practice exists in the treatment of the lack of control for a minority equity interest in a private firm.

Solution to 5:

The value of Doe's 10% minority equity interest differs markedly in the two scenarios. The imminent sale scenario results in a higher value indication for Doe's equity interest as a result of the company's higher value and the lower valuation discounts. The company's value would be higher because of the use of normalized earnings rather than reported earnings. A lower pricing multiple might also be warranted. The discount rate might be lower in the event an optimal capital structure is used rather than the existing structure. The lack of control becomes less important in the event of an imminent liquidity event such as a sale. The lack of marketability of a small equity interest also becomes less important in this instance.

Note that the treatment of non-operating assets varies when a minority interest in the stock is appraised. In the event of a sale, many buyers would not be interested in the non-operating assets—and those assets could be distributed to the shareholders prior to the sale of the stock to a buyer. Alternatively, Able could sell the operating assets and liabilities to a buyer, resulting in Able holding the real estate assets and cash from sale of the business operations. Specific circumstances would determine whether or not these assets and cash should be included in equity valuation.

We have seen that in private company valuation, as in most types of valuation beyond the simplest, a range of approaches and estimates can be argued even apart from differences resulting from different forecasts or business assumptions. The investment community also has the perception that valuation practices and estimates of value diverge from each other, as well as that valuation standards could benefit consumers of valuations.

SUMMARY

We have provided an overview of key elements of private company valuation and contrasted public and private company valuations.

- Company- and stock-specific factors may influence the selection of appropriate valuation methods and assumptions for private company valuations. Stock-specific factors may result in a lower value for an equity interest in a private company relative to a public company.
- Company-specific factors in which private companies differ from public companies include:
  - stage in life cycle;
  - size;
  - overlap of shareholders and management;
  - quality/depth of management;
  - quality of financial and other information;
  - pressure from short-term investors; and
  - tax concerns.
- Stock-specific factors that frequently affect the value of private companies include
  - liquidity of equity interests in business;
  - concentration of control; and
  - potential agreements restricting liquidity.
- Private company valuations are typically performed for three different reasons: transactions, compliance (financial or tax reporting), or litigation. Acquisition-related valuation issues and financial reporting valuation issues are of greatest importance in assessing public companies.
- Different definitions (standards) of value exist. The use of a valuation and key elements pertaining to the appraised company will help determine the appropriate definition. Key definitions of value include
  - fair market value;
  - market value;
  - fair value for financial reporting;
  - fair value in a litigation context;
  - investment value; and
  - intrinsic value.
- Private company valuations may require adjustments to the income statement to develop estimates of the company's normalized earnings. Adjustments may be required for non-recurring, non-economic, or other unusual items to eliminate anomalies and/or facilitate comparisons.
- Within the income approach, the FCF method is frequently used to value larger, mature private companies. For smaller companies or in special situations, the capitalized cash flow method and residual income method may also be used.
- Within the market approach, three methods are regularly used: the guideline public company method, guideline transactions method, and prior transactions method.

- An asset-based approach is infrequently used in valuing private companies. This approach may be appropriate for companies that are worth more in liquidation than as going concerns. This approach is also applied for asset holding companies, very small companies, or companies formed recently that have limited operating histories.
- Control and marketability issues are important and challenging elements in the valuation of private companies and equity interests therein.
- If publicly traded companies are used as the basis for pricing multiple(s), control premiums may be appropriate in measuring the total equity value of a private company. Control premiums have also been used to estimate lack of control discounts.
- Discounts for lack of control are used to convert a controlling interest value into a non-controlling equity interest value. Evidence of the adverse impact of the lack of control is an important consideration in assessing this discount.
- Discounts for lack of marketability are often used in valuing non-controlling equity interests in private companies. A DLOM may be inappropriate if the company has a high likelihood of a liquidity event in the immediate future.
- Quantification of DLOMs can be challenging because of limited data, differences in the interpretation of available data, and different interpretations of the lack of marketability's effect on a private company.
- DLOM can be estimated based on 1) private sales of restricted stock in public companies relative to their freely traded share price, 2) private sales of stock in companies prior to a subsequent IPO, and 3) the pricing of put options.

REFERENCES

AICPA."*Valuation of Privately-Held-Company Equity Securities Issued as Compensation – Accounting and Valuation Guide*" (2013)

AICPA."*Valuation of Portfolio Company Investments of Venture Capital and Private Equity Funds and Other Investment Companies – Accounting and Valuation Guide*" (2019).

Damodaran, Aswath. 2012. *Investment Valuation Tools and Techniques for Determining the Value of Any Asset*. Hoboken, NJ: John Wiley & Sons.

Hitchner, James R. 2017. *Financial Valuation: Applications and Models*. 4th ed. Hoboken, NJ: John Wiley & Sons.

Pratt, Shannon P., and Roger J. Grabowski. 2014. *Cost of Capital: Applications and Examples*. 5th ed. Hoboken, NJ: John Wiley & Sons.

"Valuation of Contingent Consideration" by The Appraisal Foundation, February 2019.

PRACTICE PROBLEMS

1 Two companies are considering the acquisition of Target Company. Buyer A is a strategic buyer and Buyer B is a financial buyer. The following information pertains to Target Company:

Sales = £28,000,000

Reported EBITDA = £4,500,000

Reported executive compensation = £1,000,000

Normalized executive compensation = £500,000

Reduced SG&A from eliminating duplicate general and administrative functions = £600,000

Calculate the pro forma EBITDA estimates that the strategic and financial buyers would each develop in an acquisitions analysis of Target Company.

2 Using the build-up method and assuming that no adjustment for industry risk is required, calculate an equity discount rate for a small company, given the following information:

Equity risk premium = 5.0%

Midcap equity risk premium = 3.5%

Small stock risk premium = 4.2%

Income return on long-term bonds = 5.1%

Total return on intermediate-term bonds = 5.3%

Company-specific risk premium = 3.0%

20-Year Treasury bond yield as of the valuation date = 4.5%

3 Using the capitalized cash flow method (CCM), calculate the fair market value of 100% of the equity of a hypothetical company, given the following information:

Current year's reported free cash flow to equity = $1,400,000

Current year's normalized free cash flow to equity = $1,800,000

Long-term interest-bearing debt = $2,000,000

Weighted average cost of capital = 15%

Equity discount rate = 18%

Long-term growth rate of FCFE = 5.5%

4 You have been asked to value Pacific Corporation, Inc., using an excess earnings method, given the following information:

Working capital balance = $2,000,000

Fair value of fixed assets = $5,500,000

Book value of fixed assets = $4,000,000

© 2021 CFA Institute. All rights reserved.

Normalized earnings of firm = \$1,000,000

Required return on working capital = 5.0%

Required return on fixed assets = 8.0%

Required return on intangible assets = 15.0%

Weighted average cost of capital = 10.0%

Long-term growth rate of residual income = 5.0%

Based on this information:

A What is the value of Pacific's intangible assets?

B What is the market value of invested capital?

5 An appraiser has been asked to determine the combined level of valuation discounts for a small equity interest in a private company. The appraiser concluded that an appropriate control premium is 15%. A discount for lack of marketability was estimated at 25%. Given these factors, what is the combined discount?

The following information relates to Questions 6–10

Alan Chin, CEO of Thunder Corporation, has asked his chief financial officer, Constance Ebinosa, to prepare a valuation of Thunder for the purpose of selling the company to a private investment partnership. Thunder is a profitable US-domiciled manufacturer of generic household products with \$200 million in annual sales. Customers consist of several grocery store chains in the United States. Competitors include large companies such as Procter & Gamble, The Clorox Company, and Unilever. Thunder has been in business for 15 years and is privately owned by the original shareholders, none of whom are employed by the company. Thunder's senior management has been in charge of the company's operations for most of the past 15 years and expects to remain in that capacity after any sale.

The partnership has expectations about Thunder similar to the current shareholders and management of Thunder. These investors expect to hold Thunder for an intermediate period and then bring the company public when market conditions are more favorable than currently.

Chin is concerned about what definition of value to use in analyzing Thunder. He notes that the stock market has been very volatile recently. He also wonders whether fair market value can be realistically estimated when the most similar recent private market transactions may not have been at arm's length.

Chin asks Ebinosa whether there will be differences in the process of valuing a private company like Thunder compared with a public company. Ebinosa replies that differences do exist and mentions several factors an analyst must consider.

Ebinosa also explains that several approaches are available for valuing private companies. She mentions that one possibility is to use an asset-based approach because Thunder has a relatively large and efficient factory and warehouse for its products. A real estate appraiser can readily determine the value of these facilities. A second method would be the market approach and using an average of the price-to-earnings multiples for Procter & Gamble and Clorox. A third possibility is a discounted free cash flow approach. The latter would focus on a continuation of Thunder's trend of slow profitable growth during the past 10 years.

The private investment partnership has mentioned that it is likely to use an income approach as one of its valuation methods. Ebinosa decides to validate the estimates they make. She assumes that for the next 12 months, Thunder's revenues will increase by the long-term annual growth rate of 3%. She also makes the following assumptions to calculate the free cash flow to the firm for the next 12 months:

- Gross profit margin is 45%.
- Depreciation is 2% of revenues.
- Selling, general, and administrative expenses are 24% of revenues.
- Capital expenditures equal 125% of depreciation to support the current level of revenues.
- Additional capital expenditures of 15% of incremental revenues are needed to fund future growth.
- Working capital investment equals 8% of incremental revenues.
- Marginal tax rate on EBIT is 35%.

Chin knows that if an income approach is used, the choice of discount rate may have a large influence on the estimated value. He makes two statements regarding discount rate estimates:

1 If the CAPM method is used to estimate the discount rate with a beta estimate based on public companies with operations and revenues similar to Thunder, then a small stock premium should be added to the estimate.
2 The weighted average cost of capital of the private investment partnership should be used to value Thunder.

Ebinosa decides to calculate a value of Thunder's equity using the capitalized cash flow method and decides to use the build-up method to estimate Thunder's required return on equity. She makes the following assumptions:

- Growth of FCFE is at a constant annual rate of 3%.
- Free cash flow to equity for the year ahead is $2.5 million.
- Risk-free rate is 4.5%.
- Equity risk premium is 5.0%.
- Size premium is 2.0%.

6 The *least likely* factor that would be a source of differences in valuing Thunder compared with valuing a publicly traded company is:
 A access to public debt markets.
 B agency problems.
 C the size of the company.
7 Ebinosa can *best* value Thunder using the:
 A excess earnings approach.
 B asset-based approach.
 C discounted free cash flow approach.
8 The free cash flow to the firm is *closest* to:
 A $23,031,000.
 B $25,441,000.
 C $36,091,000.
9 Regarding the two statements about discount rate estimates, Chin is:

A correct with respect to adding the small stock premium and correct with respect to the weighted average cost of capital.

B correct with respect to adding the small stock premium and incorrect with respect to the weighted average cost of capital.

C incorrect with respect to adding the small stock premium and incorrect with respect to the weighted average cost of capital.

10 The indicated value of Thunder's equity using the build-up method and the capitalized cash flow method (CCM) based on free cash flow to equity is *closest* to:

A $29.41 million.

B $38.46 million.

C $125.00 million.

The following information relates to Questions 11–16

The senior vice president of acquisitions for Northland Industries, Angela Lanton, and her head analyst, Michael Powell, are evaluating several potential investments. Northland is a diversified holding company for numerous businesses. One of Northland's divisions is a manufacturer of fine papers, and that division has alerted Lanton about Oakstar Timber, a supplier that may be available for purchase. Oakstar's sole owner, Felix Tanteromo, has expressed interest in exchanging his ownership of Oakstar for a combination of cash and Northland Industries securities.

Oakstar's main asset is 10,000 hectares of timberland in western Canada. The land is a combination of new and old growth Douglas fir trees. The value of this timberland has been steadily increasing since Oakstar acquired it. Oakstar manages the land on a sustained yield basis (i.e., so it continues to produce timber indefinitely) and contracts with outside forestry companies to evaluate, harvest, and sell the timber. Oakstar's income is in the form of royalties (fees paid to Oakstar based on the number of cubic meters harvested). Oakstar's balance sheet as of 31 December 20X0, in Canadian dollars, is as follows.

Oakstar Timber Balance Sheet
Year Ended 31 December 20X0

Assets	
Cash	C$500,000
Inventory	25,000
Accounts receivable	50,000
Plant and equipment (cost less depreciation)	750,000
Land	10,000,000
Total assets	C$11,325,000

Liabilities and Equity	
Accounts payables	C$75,000
Long-term bank loan	1,500,000
Common stock	9,750,000
Total liabilities and equity	C$11,325,000

In addition to the balance sheet, Powell is gathering other data to assist in valuing Oakstar and has found information on recent sales of timberland in western Canada. Douglas fir properties have averaged C$6,178 per hectare for tracts that are not contiguous and do not have a developed road system for harvesting the timber. For tracts with these features, as possessed by Oakstar, the average price is C$8,750 per hectare. Properties near urban areas and having potential for residential and recreational second home development command up to C$20,000 per hectare. Oakstar's land lacks this potential. Lanton believes these values would form the basis of an asset-based valuation for Oakstar, with the additional assumption that other assets and liabilities on the balance sheet are assumed to be worth their stated values.

The second company under evaluation, FAMCO, Inc., is a family-owned electronic manufacturing company with annual sales of US$120 million. The family wants to monetize the value of its ownership in FAMCO with a view to later investing part of the proceeds in a diversified stock portfolio. Lanton has asked Powell to obtain data for both an income-based and market-based valuation. Powell has obtained the recent annual income statement and additional data needed to calculate normalized earnings as follows.

FAMCO, Inc.
Income Statement
Year Ending 31 December 20XO

Revenues		US$120,000,000
Gross profit		85,000,000
Selling, general, and administrative expenses		23,000,000
Pro forma EBITDA		US$62,000,000
Depreciation and amortization		3,500,000
Pro forma earnings before interest and taxes		US$58,500,000
Less: Interest		1,000,000
Earnings before taxes (EBT)		US$57,500,000
Pro forma taxes on EBT	40%	23,000,000
Operating income after tax		US$34,500,000

Additional data for FAMCO is provided in the following table. Included are estimates by Powell of the compensation paid to family members and the smaller amount of salary expense for replacement employees if Northland acquires the company (reflecting perceived above-market compensation of the family group executives). He believes the current debt of FAMCO can be replaced with a more optimal level of debt at a lower interest rate. The additional data will be reflected in a normalized income statement.

FAMCO, Inc.	
Current debt level	US$10,000,000
Current interest rate	10%
Salaries of employed family members	US$7,000,000
Salaries of replacement employees	US$5,400,000
New debt level	US$25,000,000
New interest rate	8%

Powell also recognizes that a value needs to be assigned to FAMCO's intangibles consisting of patents and other intangible assets. Powell prepares an additional estimate of excess earnings and intangibles value using the capitalized cash flow method. He projects the following data for 20X1:

FAMCO, Inc.—Intangibles Valuation Data	
Working capital balance	US$10,000,000
Fair value of fixed assets	US$45,000,000
Normalized income to the company	US$35,000,000
Required return on working capital	8%
Required return on fixed assets	12%
Required return on intangible assets	20%
Weighted average cost of capital	14.5%
Future growth rate	6%

Lanton asks Powell to also use the market approach to valuation with a focus on the guideline transactions method. Powell prepares a table showing relevant information regarding three recent guideline transactions and market conditions at the time of the transactions. Powell's assumptions about FAMCO include its expected fast growth and moderate level of risk.

Target Firm	Target's Risk	Target's Growth	Consideration	Market Conditions
Firm 1	High	Slow	Cash	Normal, rising trend
Firm 2	Moderate	Fast	Stock	Prices near peak
Firm 3	Moderate	Fast	Cash	Normal, rising trend

Although Northland is interested in acquiring all of the stock of FAMCO, the acquisition of a 15% equity interest in FAMCO is also an option. Lanton asks Powell about the valuation of small equity interests in private entities and notes that control and marketability are important factors that lead to adjustments in value estimates for small equity interests. Powell mentions that the control premium paid for the most similar guideline firm used in the analysis suggests a discount for lack of control of 20%. The discount for lack of marketability was estimated at 15%.

11 Which of the following statements concerning asset-based valuation as applied to Oakstar is *most* accurate? The approach is applicable:

A only when a guideline public company is unavailable for the valuation.

B because natural resources with determinable market values constitute the majority of Oakstar's total value.

C because as a passive collector of royalties, Oakstar has no meaningful capital expenditures and free cash flow is irrelevant.

12 Using an asset-based approach, the value (net of debt) of Oakstar is *closest* to:

A C$62,250,000.

B C$87,250,000.

C C$199,750,000.

13 The normalized earnings after tax for FAMCO is *closest* to:

A US$32,940,000.

B US$34,260,000.

C US$34,860,000.

14 Using the excess earnings method, the value of FAMCO's intangibles is *closest* to:

A US$144.0 million.

B US$205.7 million.

C US$338.8 million.

15 The guideline transaction that is *most likely* applicable to FAMCO is:

A Firm 1.

B Firm 2.

C Firm 3.

16 The total discount for both control and marketability is *closest* to:

A 15%.

B 32%.

C 35%.

The following information relates to Questions 17–23

Donald Schmidt is a portfolio manager at FutureTech, an information technology company that is expanding into the food industry. Schmidt meets with Tim Beckett, a FutureTech analyst, to discuss several privately owned companies being investigated for possible acquisition.

Schmidt and Beckett first analyze Dairy Foody, a large, privately owned online food retailer.–Beckett observes that the free cash flow growth of Dairy Foody has been quite unstable over time. Schmidt and Beckett discuss the valuation method that is most appropriate to value 100% of Dairy Foody. Schmidt prefers the valuation to reflect market pricing from recent acquisition activity in the food industry, which has been high in recent months.

Schmidt and Beckett then discuss FastDelivery, a family-owned food delivery service. Pro forma financial data for FastDelivery, representing performance mid-economic cycle, are presented in Exhibit 1.

Exhibit 1 FastDelivery: Pro Forma Operating Income after Taxes

Revenues	$50,000,000
COGS	30,000,000
Gross Profit	20,000,000
SG&A	10,000,000
EBITDA	10,000,000
Depreciation and amortization	3,000,000
EBIT	7,000,000
Pro forma taxes (at 25%)	1,750,000
Operating income after taxes	$5,250,000

Beckett notices two items in FastDelivery's financials that could detract from the reliability of forecasting normalized earnings. Specifically, Beckett notes that in Exhibit 1, SG&A includes the following:

- Management salaries of $5,000,000
- The family's personal expenses of $550,000

Beckett estimates the actual market rate of the management salaries to be $3,500,000.

Schmidt also asks Beckett to calculate the required rate of return on equity for FastDelivery using the build-up method. Beckett collects the following information:

- The risk-free rate is 1%.
- The equity risk premium is 6%.
- The industry risk premium is 2%.
- The small-stock risk premium is 3%.

Schmidt transitions the conversation to Spice World, a privately owned grocery retailer. Schmidt estimates a required rate of return from a selection of guideline public companies. He tells Beckett that the required rate of return needs additional adjustments so that it can be applied to Spice World. First, Schmidt points out that Spice World faces operational and customer concentration risks that do not affect the selected guideline public companies. Second, Schmidt notes that Spice World has fewer employees and a lower level of total assets relative to the guideline public companies.

Having estimated a required rate of return on equity for Spice World, Beckett calculates the company's weighted average cost of capital (WACC) using its existing capital structure, as Schmidt believes that the current debt level is optimal. Beckett utilizes the following information to calculate the WACC for Spice World:

- The cost of equity is 18%.
- The corporate tax rate is 25%.
- The pre-tax cost of debt is 9%.
- The ratio of debt to total capital is 20%.

Schmidt also requests that Beckett value a 15% interest in Food Garden, another grocery retailer. Based on a recent transaction involving a change in control, Beckett estimates the total equity value of Food Garden to be $100,000,000. In arriving at a value for the 15% interest, Beckett concludes that a discount for lack of control (DLOC) of 20% and a discount for lack of marketability (DLOM) of 35% are applicable.

Finally, Beckett and Schmidt discuss LiveLong Foods, a privately owned food preservation company. Schmidt asks Beckett to employ the guideline public company method to value 100% of LiveLong Foods. Beckett observes that a group of comparable public companies has a median market value of invested capital (MVIC) to EBITDA multiple of 10. Beckett notes the following:

- LiveLong Foods' EBITDA is $12,000,000.
- The market value of LiveLong Foods' debt is $6,000,000.
- A combined downward adjustment of 12% for relative risk and growth characteristics should be applied to the median multiple.
- An 18% control premium should also be applied to the median multiple.

17 Given Schmidt's preference, which of the following valuation methods is *most appropriate* to value Dairy Foody?

A Excess earnings method

B Capitalized cash flow method

C Guideline transactions method

18 The normalized operating income for FastDelivery is:

A $5,662,500.

B $6,375,000.

C $6,787,500.

19 Using the build-up method, the required rate of return on equity for FastDelivery is:

A 10%.

B 11%.

C 12%.

20 Given the observations about Spice World noted by Schmidt, the required rate of return for Spice World should be adjusted to reflect:

A only a size premium.

B only a company-specific risk premium.

C both a size premium and a company-specific risk premium.

21 The WACC for Spice World is *closest* to:

A 12.15%.

B 15.75%.

C 16.20%.

22 Beckett should estimate the value of the 15% equity interest in Food Garden to be:

A $6,750,000.

B $7,800,000.

C $12,750,000.

23 Using the guideline public company method, Beckett should estimate the value of 100% of LiveLong Foods to be:

A $118,608,000.

B $121,200,000.

C $124,608,000.

SOLUTIONS

1 A strategic buyer seeks to eliminate unnecessary expenses. The strategic buyer would adjust the reported EBITDA by the amount of the officers' excess compensation. A strategic buyer could also eliminate redundant manufacturing costs estimated at £600,000. The pro forma EBITDA a strategic buyer might use in its acquisition analysis is the reported EBITDA of £4,500,000 plus the non-market compensation expense of £500,000 plus the operating synergies (cost savings) of £600,000. The adjusted EBITDA for the strategic buyer is £4,500,000 + £500,000 + £600,000 = £5,600,000. The financial buyer would also make the adjustment to normalize officers' compensation but would be unable to eliminate redundant manufacturing expenses. Thus, adjusted EBITDA for the financial buyer would be £4,500,000 + £500,000 = £5,000,000.

2 The build-up method is substantially similar to the extended CAPM except that beta is excluded from the calculation. The equity return requirement is calculated as risk-free rate plus equity risk premium for large-capitalization stocks plus small stock risk premium plus company-specific risk premium: 4.5 + 5.0 + 4.2 + 3.0 = 16.7%. Although practice may vary, in this case, there was no adjustment for industry risk.

3 There are FCFF and FCFE variations of the CCM. In this problem, the data permit the application of only the FCFE variation. According to that variation, the estimated value of equity equals the normalized free cash flow to equity estimate for next period divided by the capitalization rate for equity. The capitalization rate is the required rate of return for equity less the long-term growth rate in free cash flow to equity. Using the current $1.8 million of free cash flow to equity, the 18% equity discount rate, and the long-term growth rate of 5.5% yields a value indication of [($1.8 million)(1.055)]/(0.18 – 0.055) = $1.899 million/0.125 = $15.19 million.

4 The excess earnings consist of any remaining income after returns to working capital and fixed assets are considered. Fair value estimates and rate of return requirements for working capital and fixed assets are provided. The return required for working capital is $2,000,000 × 5.0% = $100,000, and the return required for fixed assets is $5,500,000 × 8.0% = $440,000, or $540,000 in total.

   A The residual income for intangible assets is $460,000 (the normalized earnings of $1,000,000 less the $540,000 required return for working capital and fixed assets). The value of intangible assets can then be calculated using the capitalized cash flow method. The intangibles value is $4,830,000 based on $483,000 of year-ahead residual income available to the intangibles capitalized at 10.0% (15.0% discount rate for intangibles less 5.0% long-term growth rate of residual income).

   B The market value of invested capital is the total of the values of working capital, fixed assets, and intangible assets. This value is $2,000,000 + $5,500,000 + $4,830,000 = $12,330,000.

5 The valuation of a small equity interest in a private company would typically be calculated on a basis that reflects the lack of control and lack of marketability of the interest. The control premium of 15% must first be used to provide an indication of a discount for lack of control (DLOC). A lack of control discount can be calculated using the formula Lack of control discount = 1 – [1/ (1 + Control premium)]. In this case, a lack of control discount of approximately 13% is calculated as 1 – [1/(1 + 15%)]. The discount for lack of marketability (DLOM) was specified. Valuation discounts are applied sequentially and are not added.

The formula is (Pro rata control value) × (1 – DLOC) × (1 – DLOM). A combined discount of approximately 35% is calculated as 1 – (1 – 13%) × (1 – 25%) = 0.348, or 34.8%.

6 B is correct. Thunder's size and its probable lack of access to public debt markets are potential factors affecting its valuation compared with a public company. Given that the separation of ownership and control at Thunder is similar to that at public companies, however, agency problems are not a distinguishing factor in its valuation.

7 C is correct. The excess earnings method would rarely be applied to value a company's equity, particularly when it is not needed to value intangibles. The asset-based approach is less appropriate because it is infrequently used to estimate the business enterprise value of operating companies. By contrast, the free cash flow method is broadly applicable and readily applied in this case.

8 A is correct. Using Ebinosa's assumptions:

Revenues ($200,000,000 × 1.03 =)		$206,000,000
Gross profit	45%[a]	92,700,000
Selling, general, and administrative expenses	24%[a]	49,440,000
Pro forma EBITDA		43,260,000
Depreciation	2%[a]	4,120,000
Pro forma EBIT		39,140,000
Pro forma taxes on EBIT	35%[b]	13,699,000
Operating income after tax		25,441,000
Plus: Depreciation		4,120,000
Less: Capital expenditures on current sales	125%[c]	5,150,000
Less: Capital expenditures to support future sales	15%[d]	900,000
Less: Working capital requirement	8%[d]	480,000
Free cash flow to the firm		$23,031,000

[a] Percentage of revenues
[b] Percentage of EBIT
[c] Percentage of depreciation
[d] Percentage of incremental revenues

9 C is correct. Both statements by Chin are incorrect. If the CAPM is used with public companies with similar operations and similar revenue size, as stated, then the calculation likely captures the small stock premium and should not be added to the estimate. Small stock premiums are associated with build-up models and the expanded CAPM, rather than the CAPM per se. The correct weighted average cost of capital should reflect the risk of Thunder's cash flows, not the risk of the acquirer's cash flows.

10 A is correct. The return on equity is the sum of the risk-free rate, equity risk premium, and the size premium for a total of 4.5 + 5.0 + 2.0 = 11.5%. The value of the firm using the CCM is $V = FCFE_1/(r - g) = 2.5/(0.115 - 0.03) =$ \$29.41 million.

11 B is correct. Oakstar's primary asset is timberland, the market value of which can be determined from comparable land sales.

12 B is correct. In the absence of market value data for assets and liabilities, the analyst usually must use book value data (the assumption is explicitly made that book values accurately reflect market values as well). Except for timberland,

market values for assets are unavailable. Thus, all other assets are assumed to be valued by their book values, which sum to C$500,000 + C$25,000 + C$50,000 + C$750,000 = C$1,325,000. The land's value is determined by the value of C$8,750 per hectare for properties comparable to Oakstar's. Thus, the value of Oakstar's land is C$8,750 × 10,000 = C$87,500,000. Liabilities are assumed to be worth the sum of their book value, or C$1,575,000. Thus, Estimated value = Total assets – Liabilities = C$1,325,000 + C$87,500,000 – C$1,575,000 = C$87,250,000.

13 C is correct. The new interest level is US$2,000,000 instead of US$1,000,000. SG&A expenses are reduced by US$1,600,000 (= US$5,400,000 – US$7,000,000) to US$21,400,000 by salary expense savings. Other than a calculation of a revised provision for taxes, no other changes to the income statement results in normalized earnings before tax of US$58,100,000 and normalized earnings after tax of US$34,860,000.

14 B is correct:

Return on working capital = 0.08 × US$10,000,000 = US$800,000

Return on fixed assets = 0.12 × US$45,000,000 = US$5,400,000

Return on intangibles = US$35,000,000 – US$800,000 – US$5,400,000 = US$28,800,000

Value of intangibles using CCM = US$28,800,000/(0.20 – 0.06) = US$205.71 million.

15 C is correct. Firm 3 matches FAMCO in both risk and growth. Firm 1 fails on these factors. In addition, Firm 3 is a better match to FAMCO than Firm 2 because the offer for Firm 3 was a cash offer in normal market conditions, whereas Firm 2 was a stock offer in a boom market and the value does not reflect risk and growth in the immediate future.

16 B is correct. Both discounts apply, and they are multiplicative rather than additive:

$$1 - (1 - 0.20)(1 - 0.15) = 1 - 0.68$$

17 C is correct. Schmidt and Beckett are valuing 100% of Dairy Foody as part of their acquisition analysis. The guideline transactions method (GTM) uses a multiple that specifically relates to the sale of entire companies and establishes the value estimate of a business based on pricing multiples derived from the acquisition of control of entire public or private companies. The GTM is appropriate given that Schmidt prefers the valuation to reflect market pricing from recent acquisition activity in the food industry, which has been high in recent months.

A is incorrect because the excess earnings method (EEM) would not likely be used to value 100% of Dairy Foody. Generally, the EEM is used to value intangible assets and very small businesses when other such market approach methods are not feasible, and the EEM is rarely used in pricing entire private businesses. Additionally, Schmidt prefers the valuation to reflect market pricing from recent acquisition activity in the food industry, which has been high in recent months; therefore, the GTM would be more appropriate than the EEM.

B is incorrect because the capitalized cash flow method (CCM) would not likely be used to value 100% of Dairy Foody. The CCM is essentially a stable, growth-free cash flow model that is not a preferable valuation approach when a company is not expected to grow at a constant rate as the method employs a sustainable growth rate to estimate firm value. Dairy Foody has a growth rate that has been quite unstable over time, so the CCM would not be a preferable method to estimate its value. Additionally, the CCM is rarely used for the valuation of public companies, larger private companies, or in the context of acquisitions, but it may be appropriate if market pricing evidence from public companies or transactions is limited. Thus, the CCM would not likely be used to value Dairy Foody because Dairy Foody is a large, privately owned company that is being valued in the context of an acquisition. Finally, Schmidt prefers the valuation to reflect market pricing from recent acquisition activity in the food industry, which has been high in recent months; therefore, the GTM would be more appropriate than the CCM.

18 C is correct. In computing normalized operating income for FastDelivery, SG&A expenses are first reduced by $1,500,000 (from the current $5 million to the market rate of $3.5 million) to reflect the change of management from family members to professional management at the market rate. Second, SG&A should be reduced by an additional $550,000 to remove the family's personal expenses. Based on those adjustments, FastDelivery's pro forma income statement is:

FastDelivery: Pro Forma Operating Income after Taxes	
Revenues	$50,000,000
COGS	30,000,000
Gross profit	20,000,000
SG&A	7,950,000
EBITDA	12,050,000
Depreciation and amortization	3,000,000
EBIT	9,050,000
Pro forma taxes (at 25%)	2,262,500
Operating income after taxes	$6,787,500

A is incorrect because it fails to adjust SG&A for the change in management salaries from $5 million to the market rate of $3.5 million:

FastDelivery: Pro Forma Operating Income after Taxes	
Revenues	$50,000,000
COGS	30,000,000
Gross profit	20,000,000
SG&A	9,450,000
EBITDA	10,550,000
Depreciation and amortization	3,000,000
EBIT	7,550,000
Pro forma taxes (at 25%)	1,887,500
Operating income after taxes	$5,662,500

B is incorrect because it fails to remove the family's personal expenses from SG&A:

FastDelivery: Pro Forma Operating Income after Taxes	
Revenues	$50,000,000
COGS	30,000,000
Gross profit	20,000,000
SG&A	8,500,000
EBITDA	11,500,000
Depreciation and amortization	3,000,000
EBIT	8,500,000
Pro forma taxes (at 25%)	2,125,000
Operating income after taxes	$6,375,000

19 C is correct. The required rate of return on equity for FastDelivery is 12%:

Risk-free rate	1%
Plus: Small-stock risk premium	3%
Plus: Equity risk premium	6%
Plus: Industry risk premium	2%
Indicated return on equity	12%

A is incorrect because the industry risk premium is excluded from the calculation:

Risk-free rate	1%
Plus : Small-stock risk premium	3%
Plus: Equity risk premium	6%
Indicated return on equity	10%

B is incorrect because the risk-free rate is incorrectly excluded from the calculation:

Equity risk premium	6%
Plus: Small-stock risk premium	3%
Plus: Industry risk premium	2%
Indicated return on equity	11%

20 C is correct. The required return for Spice World should be adjusted to reflect both a size premium and a company-specific risk premium. Spice World is smaller than the guideline public companies, as it has fewer employees and a lower level of total assets. Therefore, the required rate of return should be adjusted to reflect a size premium so that it reflects the level of risk of Spice World. Further, Spice World faces operational and customer concentration risks that do not impact the selected guideline public companies—risks specific to Spice World. These company-specific risks increase the level of risk of Spice World relative to that of the guideline public companies. Consequently, the required return for Spice World should also reflect a company-specific risk premium.

A is incorrect because the required return for Spice World should be adjusted to reflect both a size premium and a company-specific risk premium (not just a size premium). Spice World faces operational and customer concentration risks that do not impact the selected guideline public companies—risks specific to Spice World. These company-specific risks increase the level of risk of Spice World relative to that of the guideline public companies. Consequently, the required return for Spice World should also reflect a company-specific risk premium.

B is incorrect because the required rate of return for Spice World should be adjusted to reflect both a size premium and a company-specific risk premium (not just a company-specific risk premium). Spice World is smaller than the guideline public companies, as it has fewer employees and a lower level of total assets. Therefore, the required return should be adjusted to reflect a size premium so that it reflects the level of risk of Spice World.

21 B is correct. Spice World's WACC is calculated as:

Pre-tax cost of debt	9.00%	
Tax rate complement (1 – tax rate)	× 0.75	
After-tax cost of debt	6.75%	
Weight	× 0.20	
Weighted cost of debt		1.35%
Cost of equity	18.00%	
Weight	× 0.80	
Weighted cost of equity		14.40%
Weighted average cost of capital		15.75%

A is incorrect because Spice World's cost of equity is incorrectly adjusted for taxes:

Pre-tax cost of debt	9%	
Tax rate complement (1 – tax rate)	× 0.75	
After-tax cost of debt	6.75%	
Weight	× 0.20	
Weighted cost of debt		1.35%
Cost of equity	18.0%	
Tax rate complement (1 – tax rate)	× 0.75	
After-tax cost of equity	13.50	
Weight	× 0.80	
Weighted cost of equity		10.80%
Weighted average cost of capital		12.15%

C is incorrect because Spice World's pre-tax cost of debt, instead of its after-tax cost of debt, is incorrectly used to calculate its weighted cost of debt:

Pre-tax cost of debt	9%	
Weight	× 0.20	
Weighted cost of debt		1.8%
Cost of equity	18.00%	
Weight	× 0.80	
Weighted cost of equity		14.40%
Weighted average cost of capital		16.20%

22 B is correct. The lack of control and lack of marketability discounts are applied in sequence. The discounts are multiplicative because the valuation process involves discrete steps—first moving from a controlling to a non-controlling basis, and then moving from a marketable to a non-marketable basis. Therefore, the 15% equity interest in Food Garden is calculated as:

Indicated value of equity in operations	$100,000,000
Interest appraised	×15%
Pro rata value of 15% equity interest	15,000,000
Less: Lack of control discount of 20%	3,000,000
Value assuming ready marketability	12,000,000
Less: Lack of marketability discount of 35%	4,200,000
Indicated value of 15% equity interest	$7,800,000

A is incorrect because the lack of control and marketability discounts are incorrectly added together. Valuation discounts are multiplicative and applied sequentially, not added:

Indicated value of equity in operations	$100,000,000
Interest appraised	×15%
Pro rata value of 15% equity interest	15,000,000
Less: 55%, the sum of the DLOC of 20% and the DLOM of 35%	8,250,000
Indicated value of 15% equity interest	$6,750,000

C is incorrect because the DLOC is incorrectly subtracted from the lack of marketability discount. Valuation discounts are multiplicative and applied sequentially, not subtracted from one another:

Indicated value of equity in operations	$100,000,000
Interest appraised	×15%
Pro rata value of 15% equity interest	15,000,000
Less: 15%, the difference between the DLOM of 35% and the DLOC of 20%	2,250,000
Indicated value of 15% equity interest	$12,750,000

23 A is correct. Using the guideline public company method, the value of LiveLong Foods is calculated in two steps:

First, calculate the pricing multiple for LiveLong Foods:

Initial MVIC to EBITDA from public companies		10.0
Relative risk and growth adjustment for LiveLong Foods	–12%	(1.2)
Multiple before control adjustment		8.8
Control premium adjustment	18%	1.584
Multiple after control adjustment		10.384

Second, calculate the valuation of LiveLong Foods:

Normalized EBITDA	$12,000,000
Pricing multiple	× 10.384
Indicated value of invested capital	$124,608,000
Less: Debt capital	$6,000,000
Indicated value of equity	$118,608,000

B is incorrect because the downward adjustment to LiveLong Foods' pricing multiple and control premium adjustment were incorrectly added together. The valuation adjustments should be applied sequentially, not added:

Initial MVIC to EBITDA from public companies		10.0
6%, the sum of the control premium adjustment and the relative risk and growth adjustment for LiveLong Foods	18% – 12%	0.6
Multiple after adjustment		10.6

Normalized EBITDA	$12,000,000
Pricing multiple	× 10.6
Indicated value of invested capital	$127,200,000
Less: Debt capital	$6,000,000
Indicated value of equity	$121,200,000

C is incorrect because the market value of LiveLong Foods' debt is incorrectly not deducted from the indicated value of invested capital in arriving at the value of LiveLong Foods' equity:

Normalized EBITDA	$12,000,000
Pricing multiple	× 10.384
Indicated value of invested capital	$124,608,000

Fixed Income

STUDY SESSIONS

Study Session 11	Fixed Income (1)
Study Session 12	Fixed Income (2)

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to estimate the risks and expected returns for fixed-income instruments, analyze the term structure of interest rates and yield spreads, and evaluate fixed-income instruments with embedded options and unique features.

Understanding interest rate dynamics including changes in the yield curve is critical for investment activities such as economic and capital market forecasting, asset allocation, and active fixed-income management. Active fixed-income managers, for instance, must identify and exploit perceived investment opportunities, manage interest rate and yield curve exposure, and report on benchmark relative performance.

Many fixed-income securities contain embedded options. Issuers use bonds with call provisions to manage interest rate exposure and interest payments. Investors may prefer bonds granting early redemption or equity conversion rights. Given their widespread use and inherent complexity, investors and issuers should understand when option exercise might occur and how to value these bonds.

Evaluating bonds for credit risk is very important. As demonstrated by the 2008 global financial crisis, systemic mispricing of risk can have wide ranging and severe consequences that extend far beyond any individual position or portfolio.

© 2021 CFA Institute. All rights reserved.

FIXED INCOME

STUDY SESSION

11

Fixed Income (1)

This study session introduces the yield curve and key relationships underlying its composition. Traditional and modern theories and models explaining the shape of the yield curve are presented. An arbitrage-free framework using observed market prices is introduced for valuing option-free bonds. This approach also holds for more complex valuation of bonds with embedded options and other bond types.

READING ASSIGNMENTS

Reading 28	The Term Structure and Interest Rate Dynamics by Thomas S.Y. Ho, PhD, Sang Bin Lee, PhD, and Stephen E. Wilcox, PhD, CFA
Reading 29	The Arbitrage-Free Valuation Framework by Steven V. Mann, PhD

© 2021 CFA Institute. All rights reserved.

READING

28

The Term Structure and Interest Rate Dynamics

by Thomas S.Y. Ho, PhD, Sang Bin Lee, PhD, and Stephen E. Wilcox, PhD, CFA

Thomas S.Y. Ho, PhD, is at Thomas Ho Company Ltd (USA). Sang Bin Lee, PhD, is at Hanyang University (South Korea). Stephen E. Wilcox, PhD, CFA, is at Minnesota State University, Mankato (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe relationships among spot rates, forward rates, yield to maturity, expected and realized returns on bonds, and the shape of the yield curve;
☐	**b.** describe how zero-coupon rates (spot rates) may be obtained from the par curve by bootstrapping;
☐	**c.** describe the assumptions concerning the evolution of spot rates in relation to forward rates implicit in active bond portfolio management;
☐	**d.** describe the strategy of rolling down the yield curve;
☐	**e.** explain the swap rate curve and why and how market participants use it in valuation;
☐	**f.** calculate and interpret the swap spread for a given maturity
☐	**g.** describe short-term interest rate spreads used to gauge economy-wide credit risk and liquidity risk;
☐	**h.** explain traditional theories of the term structure of interest rates and describe the implications of each theory for forward rates and the shape of the yield curve;
☐	**i.** explain how a bond's exposure to each of the factors driving the yield curve can be measured and how these exposures can be used to manage yield curve risks;
☐	**j.** explain the maturity structure of yield volatilities and their effect on price volatility;
☐	**k.** explain how key economic factors are used to establish a view on benchmark rates, spreads, and yield curve changes.

© 2021 CFA Institute. All rights reserved.

1 SPOT RATES, FORWARD RATES, AND THE FORWARD RATE MODEL

a describe relationships among spot rates, forward rates, yield-to-maturity, expected and realized returns on bonds, and the shape of the yield curve

b describe how zero-coupon rates (spot rates) may be obtained from the par curve by bootstrapping

Interest rates are both a barometer of the economy and an instrument for its control. The term structure of interest rates—market interest rates at various maturities—is a vital input into the valuation of many financial products. The quantification of interest rate risk is of critical importance to risk managers. Understanding the determinants of interest rates, and thus the drivers of bond returns, is imperative for fixed-income market participants. Here, we explore the tools necessary to understand the term structure and interest rate dynamics—that is, the process by which bond yields and prices evolve over time.

Section 1 explains how spot (or current) rates and forward rates, which are set today for a period starting in the future, are related, as well as how their relationship influences yield curve shape. Section 2 builds upon this foundation to show how forward rates impact the yield-to-maturity and expected bond returns. Section 3 explains how these concepts are put into practice by active fixed-income portfolio managers.

The swap curve is the term structure of interest rates derived from a periodic exchange of payments based on fixed rates versus short-term market reference rates rather than default-risk-free government bonds. Sections 4 and 5 describe the swap curve and its relationship to government yields, known as the swap spread, and explains their use in valuation.

Section 6 describes traditional theories of the term structure of interest rates. These theories outline several qualitative perspectives on economic forces that may affect the shape of the term structure.

Section 7 describes yield curve factor models. The focus is a popular three-factor term structure model in which the yield curve changes are described in terms of three independent movements: level, steepness, and curvature. These factors can be extracted from the variance–covariance matrix of historical interest rate movements.

Section 8 builds on the factor model and describes how to manage the risk of changing rates over different maturities. Section 9 concludes with a discussion of key variables known to influence interest rates, the development of interest rate views based on forecasts of those variables, and common trades tailored to capitalize on an interest rate view. A summary of key points concludes the reading.

1.1 Spot Rates and Forward Rates

We first explain the relationships among spot rates, forward rates, yield-to-maturity, expected and realized returns on bonds, and the shape of the yield curve. We then discuss the assumptions made about forward rates in active bond portfolio management.

The price of a risk-free single-unit payment (e.g., \$1, €1, or £1) after N periods is called the **discount factor** with maturity N, denoted by PV_N. The yield-to-maturity of the payment is called a **spot rate**, denoted by Z_N. That is,

$$DF_N = \frac{1}{(1 + Z_N)^N} \tag{1}$$

The N-period discount factor, DF_N, and the N-period spot rate, Z_N, for a range of maturities in years $N > 0$ are called the **discount function** and the **spot yield curve** (or, more simply, **spot curve**), respectively. This spot curve represents the term structure of interest rates. Note that the discount function completely identifies the spot curve and vice versa, because both contain the same set of information about the time value of money.

The spot curve shows, for various maturities, the annualized return on an option-free and default-risk-free **zero-coupon bond** (**zero** for short) with a single payment at maturity. For this reason, spot rates are also referred to as zero-coupon yields or zero rates. The spot rate as a yield concept avoids the need for a reinvestment rate assumption for coupon-paying securities.

As Equation 1 suggests, the spot curve is a benchmark for the time value of money received on a future date as determined by the market supply and demand for funds. It is viewed as the most basic term structure of interest rates because no reinvestment risk is involved; the stated yield equals the actual realized return if the zero is held to maturity. Thus, the yield on a zero-coupon bond maturing in year T is regarded as the most accurate representation of the T-year interest rate.

A **forward rate** is an interest rate determined today for a loan that will be initiated in a future period. The set of forward rates for loans of different maturities with the same future start date is called the **forward curve**. Forward rates and forward curves can be mathematically derived from the current spot curve.

Denote the forward rate of a loan initiated A periods from today with tenor (further maturity) of B periods by $f_{A,B-A}$. Consider a forward contract in which one party, the buyer, commits to pay another party, the seller, a forward contract price $f_{A,B-A}$ at time A for a zero-coupon bond with maturity $B - A$ and unit principal. Because this is an agreement to do something in the future, no money is exchanged at contract initiation. At A, the buyer will pay the seller the contracted forward price and will receive from the seller at time B a payment defined here as a single currency unit.

The **forward pricing model** describes the valuation of forward contracts. The no-arbitrage principle, which simply states that tradable securities with identical cash flow payments must have the same price, may be used to derive the model as shown in Equation 2:

$$DF_B = DF_A \times F_{A,B-A} \tag{2}$$

The discount factors DF_A and DF_B represent the respective prices for period A and a longer period B needed to derive the forward price, $F_{A,B-A}$, a contract which starts in the future at time A and ends at time B.To understand the reasoning behind Equation 2, consider two alternative investments: (1) buying a two-year zero-coupon bond at a cost of $DF_2 = 0.93$ and (2) entering into a one-year forward contract to purchase a one-year zero-coupon bond for $DF_1 = 0.95$. Because the payoffs in two years are the same and the initial costs of the investments must be equal, the no-arbitrage forward price $F_{1,1}$ must equal 0.93/0.95, or 0.9789. Otherwise, any trader could sell the overvalued investment and buy the undervalued investment with the proceeds to generate risk-free profits with zero net investment.

Example 1 should help confirm your understanding of discount factors and forward prices. Please note that the solutions in the examples that follow may be rounded to two or four decimal places.

EXAMPLE 1

Spot and Forward Prices and Rates (1)

Consider a two-year loan beginning in one year (A = 1, B = 3). The one-year spot rate is $z_1 = z_A$ = 7% = 0.07. The three-year spot rate is $z_3 = z_B$ = 9% = 0.09.

1 Calculate the one-year discount factor: $DF_A = DF_1$.
2 Calculate the three-year discount factor: $DF_B = DF_3$.
3 Calculate the forward price of a two-year bond to be issued in one year: $F_{A,B-A} = F_{1,3}$.
4 Interpret your answer to Problem 3.

Solution to 1:

Using Equation 1,

$$DF_1 = \frac{1}{(1+0.07)^1} = 0.9346$$

Solution to 2:

$$DF_3 = \frac{1}{(1+0.09)^3} = 0.7722$$

Solution to 3:

Using Equation 2,

$0.7722 = 0.9346 \times F_{1,3}$.

$F_{1,3} = 0.7722 \div 0.9346 = 0.8262$.

Solution to 4:

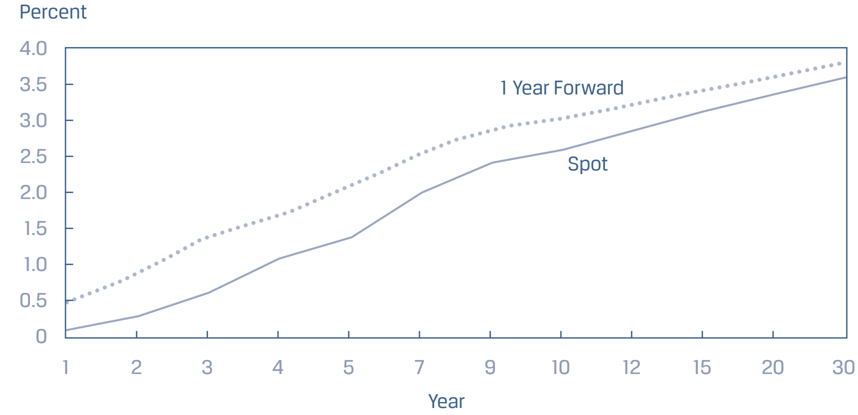
The forward contract price of $DF_{1,3}$ = 0.8262 is the price agreed on today, to be paid one year from today for a bond with a two-year maturity and a risk-free unit-principal payment (e.g., \$1, €1, or £1) at maturity in three years. As shown in the solution to 3, it is calculated as the three-year discount factor, DF_3 = 0.7722, divided by the one-year discount factor, DF_1 = 0.9346.

1.1.1 *The Forward Rate Model*

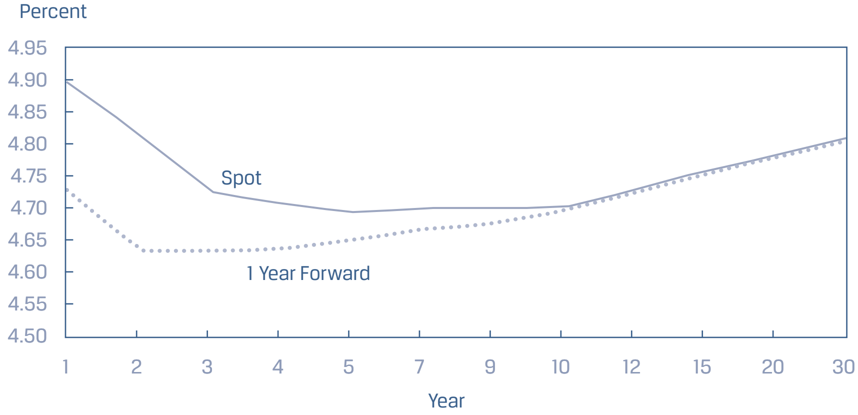
This section uses the forward rate model to establish that forward rates are above spot rates when the spot curve is upward sloping and below spot rates when the spot curve slopes downward. Exhibit 1 shows these spot versus forward relationships for the US Treasury yield curve in July 2013 versus December 2006, respectively. As we illustrate later, the relationship between spot and forward rates is important for future rate expectations as well as valuing fixed-income instruments.

Exhibit 1 Spot and Forward Curves

A. Spot vs. Forward US Treasury Yields, July 2013



B. Spot vs. Forward US Treasury Yields, December 2006



In contrast to the forward price $F_{A,B-A}$, the forward *rate* $f_{A,B-A}$ is the discount rate for a risk-free unit-principal payment (e.g., \$1, €1, or £1) B periods from today, valued at time A, such that the present value equals the forward contract price, $DF_{A,B-A}$. Then, by definition,

$$DF_{A,B-A} = \frac{1}{\left(1 + F_{A,B-A}\right)^{B-A}} \tag{3}$$

By substituting Equations 1 and 3 into Equation 2, the forward pricing model can be expressed in terms of rates as noted by Equation 4, which is the **forward rate model**:

$$\left(1 + z_B\right)^B = \left(1 + z_A\right)^A \left(1 + f_{A,B-A}\right)^{B-A} \tag{4}$$

Thus, the spot rate for B periods, which is z_B, and the spot rate for A periods, which is z_A, imply a value for the $(B\text{-}A)$-period forward rate at A, $f_{A,B\text{-}A}$. Equation 4 is important because it shows how forward rates may be extrapolated from spot rates—that is, they are implicit in the spot rates at any given point in time.

Equation 4 suggests two ways to interpret forward rates. For example, suppose $f_{7,1}$, the rate agreed on today for a one-year loan to be made seven years from today, is 3%. Then 3% is the

- reinvestment rate that would make an investor indifferent between buying an eight-year zero-coupon bond or investing in a seven-year zero-coupon bond and at maturity reinvesting the proceeds for one year. In this sense, the forward rate can be viewed as a type of breakeven interest rate.
- one-year rate that can be locked in today by buying an eight-year zero-coupon bond rather than investing in a seven-year zero-coupon bond and, when it matures, reinvesting the proceeds in a zero-coupon instrument that matures in one year. In this sense, the forward rate can be viewed as a rate that can be locked in by extending maturity by one year.

Example 2 addresses forward rates and the relationship between spot and forward rates.

EXAMPLE 2

Spot and Forward Prices and Rates (2)

The spot rates for three hypothetical zero-coupon bonds (zeros) with maturities of one, two, and three years are given in the following table.

Maturity (T)	1	2	3
Spot rates	z_1 = 9%	z_2 = 10%	z_3 = 11%

1 Calculate the forward rate for a one-year zero issued one year from today, $f_{1,1}$.

2 Calculate the forward rate for a one-year zero issued two years from today, $f_{2,1}$.

3 Calculate the forward rate for a two-year zero issued one year from today, $f_{1,2}$.

4 Based on your answers to 1 and 2, describe the relationship between the spot rates and the implied one-year forward rates.

Solution to 1:

$f_{1,1}$ is calculated as follows (using Equation 4):

$$(1 + z_2)^2 = (1 + z_1)^1 (1 + f_{1,1})^1$$

$$(1 + 0.10)^2 = (1 + 0.09)^1 (1 + f_{1,1})^1$$

$$f_{1,1} = \frac{(1.10)^2}{1.09} - 1 = 11.01\%$$

Solution to 2:

$f(2,1)$ is calculated as follows:

$$(1 + z_3)^3 = (1 + z_2)^2 (1 + f_{2,1})^1$$

$$(1 + 0.11)^3 = (1 + 0.10)^2 (1 + f_{2,1})^1$$

$$f_{2,1} = \frac{(1.11)^3}{(1.10)^2} - 1 = 13.03\%$$

Solution to 3:

$f_{1,2}$ is calculated as follows:

$$(1+z_3)^3 = (1+z_1)^1(1+f_{1,2})^2$$

$$(1+0.11)^3 = (1+0.09)^1(1+f_{1,2})^2$$

$$f_{1,2} = \sqrt[2]{\frac{(1.11)^3}{(1.09)^1}} - 1 = 12.01\%$$

Solution to 4:

The upward-sloping zero-coupon yield curve is associated with an upward-sloping forward curve (a series of increasing one-year forward rates because 13.03% is greater than 11.01%). This dynamic is explained further in the following discussion.

The relationship between spot rates and one-period forward rates may be demonstrated using the forward rate model and successive substitution, resulting in Equations 5a and 5b:

$$(1+z_T)^T = (1+z_1)(1+f_{2,1})(1+f_{3,1})\ldots(1+f_{T-1,1}) \quad \text{(5a)}$$

$$z_T = \{(1+z_1)(1+f_{2,1})(1+f_{3,1})\ldots(1+f_{T-1,1})\}^{\frac{1}{T}} - 1 \quad \text{(5b)}$$

Equation 5b shows that the spot rate for a security with a maturity of $T > 1$ can be expressed as a geometric mean of the spot rate for a security with a maturity of $T = 1$ and a series of $T - 1$ forward rates.

Equation 5b is critical for active fixed-income portfolio managers. Although the question of whether forward rates are unbiased estimators of market consensus expectations remains open to debate, implied forward rates are generally the best available and most accessible proxy for market expectations of future spot rates. If an active trader can identify a series of short-term bonds whose actual returns exceed today's quoted forward rates, then the total return over her investment horizon would exceed the return on a maturity-matching, buy-and-hold strategy if the yield curve were to remain relatively stable. Later, we will apply this concept to dynamic hedging strategies and the local expectations theory.

Examples 3 and 4 explore the relationship between spot and forward rates.

EXAMPLE 3

Spot and Forward Prices and Rates (3)

Given the data and conclusions for z_1, $f_{1,1}$, and $f_{2,1}$ from Example 2:

$z_1 = 9\%$

$f_{1,1} = 11.01\%$

$f_{2,1} = 13.03\%$

Show that the two-year spot rate of $z_2 = 10\%$ and the three-year spot rate of $z_3 = 11\%$ are geometric averages of the one-year spot rate and the forward rates.

Solution:

Using Equation 5a,

$$(1 + z_2)^2 = (1 + z_1)(1 + f_{1,1})$$

$$z_2 = \sqrt[2]{(1 + 0.09)(1 + 0.1101)} - 1 \approx 10\%$$

$$(1 + z_3)^3 = (1 + z_1)(1 + f_{1,1})(1 + f_{2,1})$$

$$z_3 = \sqrt[3]{(1 + 0.09)(1 + 0.1101)(1 + 0.1303)} - 1 \approx 11\%$$

We can now consolidate our knowledge of spot and forward rates to explain important relationships between the spot and forward rate curves. The forward rate model (Equation 4) can also be expressed as Equation 6.

$$\left\{\frac{1 + z_B}{1 + z_A}\right\}^{\frac{A}{B-A}}(1 + z_B) = 1 + f_{A,B-A} \quad (6)$$

To illustrate, suppose $A = 1$, $B = 5$, $z_1 = 2\%$, and $z_5 = 3\%$; the left-hand side of Equation 6 is

$$\left(\frac{1.03}{1.02}\right)^{\frac{1}{4}}(1.03) = (1.0024)(1.03) = 1.0325$$

so $f_{1,4} = 3.25\%$. Given that the yield curve is upward sloping—so, $z_B > z_A$—Equation 6 implies that the forward rate from A to B is greater than the long-term spot rate: $f_{A,B-A} > z_B$. This is the case in our example, because 3.25% > 3.00%. Conversely, when the yield curve is downward sloping, then $z_B < z_A$ and the forward rate from A to B is lower than the long-term spot rate: $f_{A,B-A} < z_B$. Equation 6 also shows that if the spot curve is flat, all one-period forward rates equal the spot rate. For an upward-sloping yield curve— $z_B > z_A$ —the forward rate rises as time periods increase. For a downward-sloping yield curve— $z_B < z_A$ —the forward rate declines as time periods increase.

EXAMPLE 4

Spot and Forward Prices and Rates (4)

Given the spot rates $z_1 = 9\%$, $z_2 = 10\%$, and $z_3 = 11\%$, as in Examples 2 and 3:

1 Determine whether the forward rate $f_{1,2}$ is greater than or less than the long-term rate, z_3.

2 Determine whether forward rates rise or fall as the initiation date, A, for the forward rate is later.

Solution to 1:

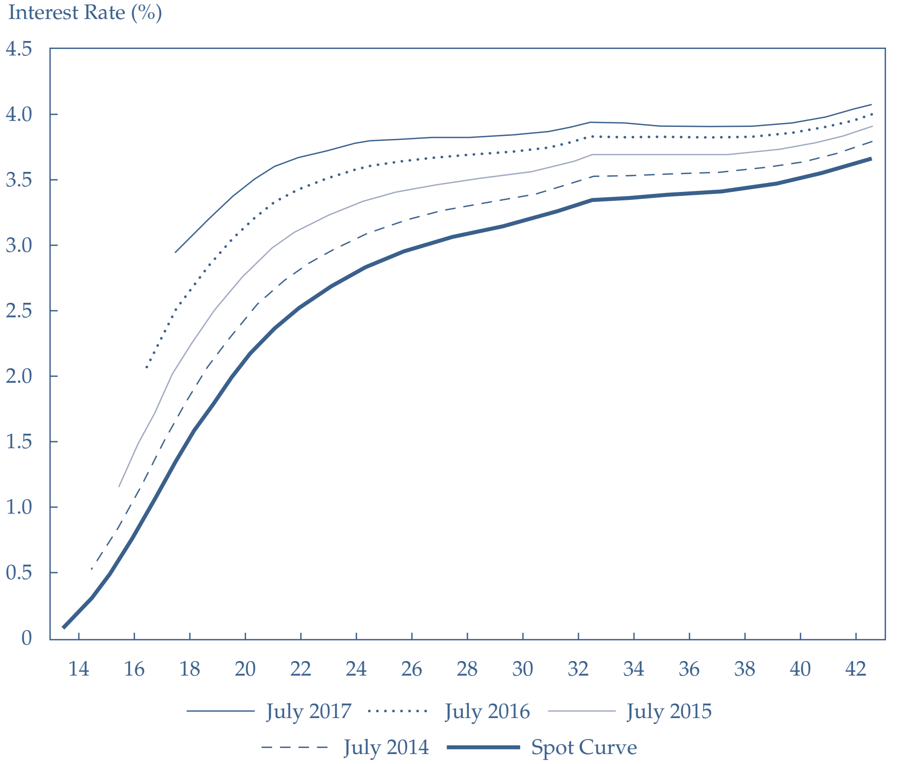
The spot rates imply an upward-sloping yield curve, $z_3 > z_2 > z_1$, or in general, $z_B > z_A$. Thus, the forward rate will be greater than the long-term rate, or $f_{A,B-A} > z_B$. Note from Example 2 that $f_{1,2} = 12.01\% > z_3 = 11\%$.

Solution to 2:

The spot rates imply an upward-sloping yield curve, $z_3 > z_2 > z_1$. Thus, the forward rates will rise with increasing A. This relationship was shown in Example 2, in which $f_{1,1} = 11.01\%$ and $f_{2,1} = 13.03\%$.

These relationships are illustrated in Exhibits 2 and 3 as an extension of Exhibit 1. The spot rates for US Treasuries as of 31 July 2013 constructed using interpolation are the lowest, as shown in the table following the exhibit. Note that the spot curve is upward sloping. The forward curves for the end of July 2014, 2015, 2016, and 2017 are also presented in Exhibit 2. Because the yield curve is upward sloping, these forward curves are all above the spot curve and become successively higher and steeper as the forward period increases, the highest of which is that for July 2017.

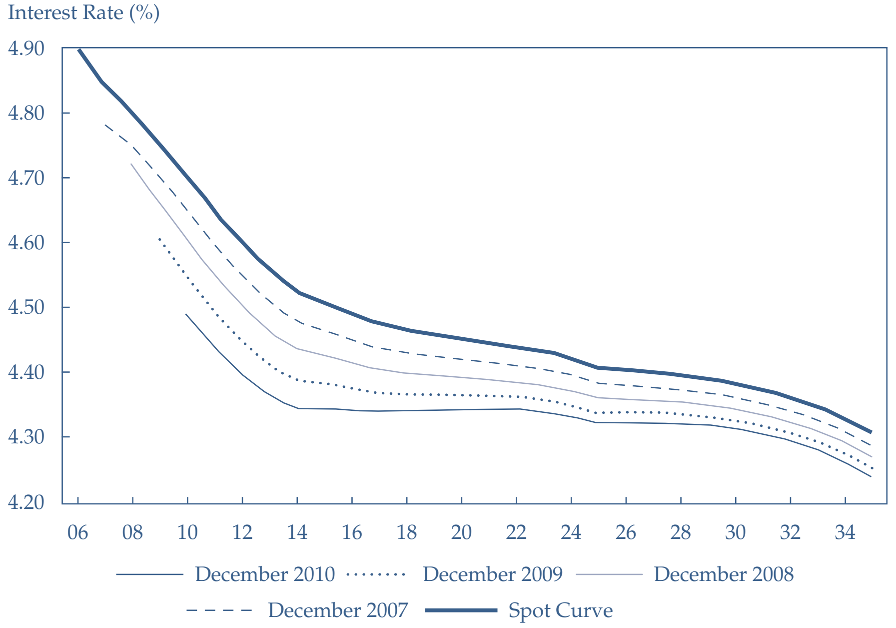
Exhibit 2 Historical Example: Upward-Sloping Spot Curve vs. Forward Curves, 31 July 2013



Maturity (years)	1	2	3	5	7	10	20	30
Spot rate (%)	0.11	0.33	0.61	1.37	2.00	2.61	3.35	3.66

Exhibit 3 shows the opposite case of a downward sloping spot curve based on US Treasury rates as of 31 December 2006. This data also uses interpolation and is somewhat modified to make the yield curve more downward sloping for illustrative purposes. The spot curve and forward curves for the end of December 2007, 2008, 2009, and 2010 are presented in Exhibit 3.

Exhibit 3 Historical Example: Downward-Sloping Spot Curve vs. Forward Curves, 31 December 2006 (modified for illustrative purposes)



Maturity (years)	1	2	3	5	7	10	20	30
Spot rate (%)	4.90	4.82	4.74	4.70	4.60	4.51	4.41	4.31

The highest curve is the spot yield curve, and it is downward sloping. The forward curves are below the spot curve, with longer forward periods associated with lower forward curves, the lowest of which is dated December 2010.

An important point that can be inferred from Exhibit 2 and Exhibit 3 is that forward rates do not extend beyond the longest maturity on today's yield curve. For example, if yields reach a 30-year maturity on today's yield curve, then a three-year forward model will extend just 27 years. Similarly, four years hence, the longest-maturity forward rate would be $f_{4,26}$.

In summary, when the spot curve slopes upward, the forward curve will lie above the spot curve. Conversely, when the spot curve slopes downward, the forward curve will lie below the spot curve. This dynamic reflects the basic mathematical truth that when an average is rising (falling), the marginal data point must be above (below) the average. In this case, the spot curve represents an average over an entire period and the forward rates represent the marginal changes between future periods.

We have thus far discussed the spot curve and the forward curve. Another curve important in practice is the government par curve. The **par curve** represents the yields to maturity on coupon-paying government bonds, priced at par, over a range of maturities. In practice, recently issued ("on the run") bonds are most often used to create the par curve, because these securities are most liquid and typically priced at or close to par.

The par curve is important for valuation in that it can be used to construct a zero-coupon yield curve. The process considers a coupon-paying bond as a portfolio of zero-coupon bonds. The zero-coupon rates are determined by using the par yields and solving for the zero-coupon rates one by one, from the shortest to longest maturities using a forward substitution process known as **bootstrapping**.

WHAT IS BOOTSTRAPPING?

Because the practical details of deriving the zero-coupon yield are beyond the scope of this reading, the concept of bootstrapping may be best shown using a numerical illustration. Suppose the following yields are observed for annual coupon sovereign debt:

Par Rates:

One-year par rate = 5%, two-year par rate = 5.97%, three-year par rate = 6.91%, four-year par rate = 7.81%. From these data, we can bootstrap zero-coupon rates.

Zero-Coupon Rates:

Given annual coupons, the one-year zero-coupon rate equals the one-year par rate because it has one cash flow, whereas two-year and longer maturity bonds have coupon payments prior to maturity.

The derivation of zero-coupon rates begins with the two-year maturity. The two-year zero-coupon rate is determined by using z_1 = 5% and solving for z_2 in the following equation for of one monetary unit of current market value:

$$1 = \frac{0.0597}{(1.05)} + \frac{1 + 0.0597}{(1 + z_2)^2}$$

In the equation, 0.0597 and 1.0597 represent payments from interest and principal and interest, respectively, per unit of principal value. The equation implies that z_2 = 6%. We have bootstrapped the two-year spot rate. Continuing with forward substitution, the three-year zero-coupon rate can be bootstrapped by solving for z_3 using the known values of the one-year and two-year spot rates of 5% and 6%:

$$1 = \frac{0.0691}{(1.05)} + \frac{0.0691}{(1.06)^2} + \frac{1 + 0.0691}{(1 + z_3)^3}$$

Thus, z_3 = 7%. Finally, we solve for the four-year zero-coupon rate, z_4:

$$1 = \frac{0.0781}{(1.05)} + \frac{0.0781}{(1.06)^2} + \frac{0.0781}{(1.07)^3} + \frac{1 + 0.0781}{(1 + z_4)^4}$$

In summary, z_1 = 5%, z_2 = 6%, z_3 = 7%, and z_4 = 8%.

In the preceding discussion, we considered an upward-sloping (spot) yield curve (Exhibit 2) and an inverted or downward-sloping (spot) yield curve (Exhibit 3). In developed markets, yield curves are most commonly upward sloping with diminishing marginal increases in yield for identical changes in maturity; that is, the yield curve "flattens" at longer maturities. Because nominal yields incorporate a premium for expected inflation, an upward-sloping yield curve is generally interpreted as reflecting a market expectation of rising or at least stable future inflation (associated with relatively strong economic growth). The existence of risk premiums (e.g., for the greater interest rate risk of longer-maturity bonds) also contributes to a positive slope.

An inverted yield curve (Exhibit 3) is less common. Such a term structure may reflect a market expectation of declining future inflation rates (because a nominal yield incorporates a premium for expected inflation) from a relatively high current level. Expectations of an economic slowdown may be one reason to anticipate a decline in inflation, and a downward-sloping yield curve is frequently observed before recessions. A flat yield curve typically occurs briefly in the transition from an upward-sloping to a downward-sloping yield curve, or vice versa. A humped yield curve, which is relatively rare, occurs when intermediate-term interest rates are higher than short- and long-term rates.

2 YIELD-TO-MATURITY IN RELATION TO SPOT AND FORWARD RATES

c describe the assumptions concerning the evolution of spot rates in relation to forward rates implicit in active bond portfolio management

Yield-to-maturity (YTM) is perhaps the most familiar pricing concept in bond markets. In this section, we clarify how it is related to spot rates and a bond's expected and realized returns.

How is the yield-to-maturity related to spot rates? In bond markets, most bonds outstanding have coupon payments and many have various options, such as a call provision. The YTM of these bonds with maturity T would not be the same as the spot rate at T but should be mathematically related to the spot curve. Because the principle of no arbitrage shows that a bond's value is the sum of the present values of payments discounted by their corresponding spot rates, the YTM of the bond should be some weighted average of spot rates used in the valuation of the bond.

Example 5 addresses the relationship between spot rates and YTM.

EXAMPLE 5

Spot Rate and Yield-to-Maturity

Recall from earlier examples the spot rates z_1= 9%, z_2 = 10%, and z_3 = 11%. Let y_T be the YTM.

1 Calculate the price of a two-year annual coupon bond using the spot rates. Assume the coupon rate is 6% and the face value is $1,000. Next, state the formula for determining the price of the bond in terms of its YTM. Is z_2 greater than or less than y_2? Why?

2 Calculate the price of a three-year annual coupon-paying bond using the spot rates. Assume the coupon rate is 5% and the face value is £100. Next, write a formula for determining the price of the bond using the YTM. Is z_3 greater or less than y_3? Why?

Solution to 1:

Using the spot rates,

$$\text{Price} = \frac{\$60}{(1+0.09)^1} + \frac{\$1{,}060}{(1+0.10)^2} = \$931.08$$

Using the YTM,

$$Price = \frac{\$60}{(1+y_2)} + \frac{\$1{,}060}{(1+y_2)^2} = \$931.08$$

Note that y_2 is used to discount both the first- and second-year cash flows. Because the bond can have only one price, it follows that $z_1 < y_2 < z_2$ because y_2 is a weighted average of z_1 and z_2 and the yield curve is upward sloping. Using a calculator, one can calculate the YTM as y_2 = 9.97%, which is less than z_2 = 10% and greater than z_1 = 9%, just as we would expect. Note that y_2 is much closer to z_2 than to z_1 because the bond's largest cash flow occurs in Year 2, thereby giving z_2 a greater weight than z_1 in the determination of y_2.

Solution to 2:

Using the spot rates,

$$\text{Price} = \frac{£5}{(1+0.09)^1} + \frac{£5}{(1+0.10)^2} + \frac{£105}{(1+0.11)^3} = £85.49$$

Using the yield-to-maturity,

$$\text{Price} = \frac{£5}{(1+y_3)} + \frac{£5}{(1+y_3)^2} + \frac{£105}{(1+y_3)^3} = £85.49.$$

Note that y_3 is used to discount all three cash flows. Because the bond can have only one price, y_3 must be a weighted average of z_1, z_2, and z_3. Given that the yield curve is upward sloping in this example, $y_3 < z_3$. Using a calculator to compute YTM, y_3 = 10.93%, which is less than z_3 = 11% and greater than z_1 = 9%—just as we would expect, because the weighted YTM must lie between the highest and lowest spot rates. Note that y_3 is much closer to z_3 than it is to z_2 or z_1 because the bond's largest cash flow occurs in Year 3, thereby giving z_3 a greater weight than z_1 and z_2 in the determination of y_3.

Investors can expect to earn the yield-to-maturity on a bond only under extremely restrictive assumptions. The YTM is the expected rate of return for a bond held to maturity, assuming that all promised coupon and principal payments are made in full when due and that coupons are reinvested at the original YTM. As interest rates change, the reinvestment of coupons at the original YTM is unlikely. The YTM can provide a poor estimate of expected return if (1) interest rates are volatile, (2) the yield curve is sloped either upward or downward, (3) there is significant risk of default, or (4) the bond has one or more embedded options (e.g., put, call, or conversion). If either (1) or (2) is the case, reinvestment of coupons would not be expected to be at the assumed rate (YTM). Case 3 implies that actual cash flows may differ from those assumed in the YTM calculation, and in Case 4, the exercise of an embedded option would result in a holding period shorter than the bond's original maturity.

The realized return is the actual bond return during an investor's holding period. It is based on actual reinvestment rates and the yield curve at the end of the holding period. If we had perfect foresight, the expected bond return would equal the realized bond return.

To illustrate these concepts, assume that z_1 = 5%, z_2 = 6%, z_3 = 7%, z_4 = 8%, and z_5 = 9%. Consider a five-year annual coupon bond with a coupon rate of 10%. The forward rates extrapolated from the spot rates are $f_{1,1}$ = 7.0%, $f_{2,1}$ = 9.0%, $f_{3,1}$ = 11.1%, and $f_{4,1}$ = 13.1%. The price, determined as a percentage of par, is 105.43.

The yield-to-maturity of 8.62% can be determined by solving

$$105.43 = \frac{10}{(1+y_5)} + \frac{10}{(1+y_5)^2} + \frac{10}{(1+y_5)^3} + \frac{10}{(1+y_5)^4} + \frac{110}{(1+y_5)^5}$$

The yield-to-maturity of 8.62% is the bond's expected return assuming no default, a holding period of five years, and a reinvestment rate of 8.62%. But what if the forward rates are assumed to be the future spot rates?

Using the forward rates as the expected reinvestment rates results in the following expected cash flow at the end of Year 5:

$$\begin{aligned} &10(1+0.07)(1+0.09)(1+0.111)(1+0.131) + 10(1+0.09)(1+0.011)(1+0.131) \\ &+ 10(1+0.111)(1+0.131) + 10(1+0.131) + 110 \approx 162.22 \end{aligned}$$

Therefore, the expected bond return is (162.22 – 105.43)/105.43 = 53.87% and the expected annualized rate of return is 9.00% [solve $(1 + x)^5 = 1 + 0.5387$].

From this example, we can see that the expected rate of return is not equal to the YTM even if we make the generally unrealistic assumption that the forward rates are the future spot rates. The YTM is generally a realistic estimate of expected return only if the yield curve is flat. Note that in the foregoing formula, all cash flows were discounted at 8.62% regardless of maturity.

Example 6 will reinforce your understanding of various yield and return concepts.

EXAMPLE 6

Yield and Return Concepts

1 When the spot curve is upward sloping, the forward curve:

- A lies above the spot curve.
- B lies below the spot curve.
- C is coincident with the spot curve.

2 Which of the following statements concerning the YTM of a default-risk-free bond is *most* accurate? The YTM of such a bond:

- A equals the expected return on the bond if the bond is held to maturity.
- B can be viewed as a weighted average of the spot rates applying to its cash flows.
- C will be closer to the realized return if the spot curve is upward sloping rather than flat through the life of the bond.

3 When the spot curve is downward sloping, a later initiation date results in a forward curve that is:

- A closer to the spot curve.
- B a greater distance above the spot curve.
- C a greater distance below the spot curve.

Solution to 1:

A is correct. Points on a spot curve can be viewed as an average of single-period rates over given maturities, whereas forward rates reflect the marginal changes between future periods.

Solution to 2:

B is correct. The YTM is the discount rate that, when applied to a bond's promised cash flows, equates those cash flows to the bond's market price and the fact that the market price should reflect discounting promised cash flows at appropriate spot rates.

Solution to 3:

C is correct. This answer follows from the forward rate model as expressed in Equation 6. If the spot curve is downward sloping (upward sloping), a later initiation date will result in a forward curve that is a greater distance below (above) the spot curve. See Exhibit 2 and Exhibit 3.

2.1 Yield Curve Movement and the Forward Curve

This section establishes several important results concerning forward prices and the spot yield curve to demonstrate the relevance of the forward curve to active bond investors.

The forward contract price remains unchanged as long as future spot rates evolve as predicted by today's forward curve. If a trader expects the future spot rate to be below what is predicted by the prevailing forward rate, the forward contract value is expected to increase and the trader would buy the forward contract. Conversely, if the trader expects the future spot rate to be above that predicted by the existing forward rate, then the forward contract value is expected to decrease and the trader would sell the forward contract.

Using the forward pricing model defined by Equation 2, we can determine the forward contract price that delivers a $(B - A)$-period-maturity bond at time A, $F_{A,B-A}$, using Equation 7 (which is Equation 2 solved for the forward price):

$$F_{A,B-A} = \frac{DF_B}{DF_A} \tag{7}$$

Now suppose that after t periods, the new discount function for some maturity time T period, denoted as DF_T^{new}, is the same as the forward discount function implied by today's discount function, as shown by Equation 8.

$$DF_T^{new} = \frac{DF_{t+T}}{DF_t} \tag{8}$$

Next, after a lapse of t periods, the time to expiration of the contract is $A - t$, and the forward contract price at time t is $F_{A-t,B-A}^{new}$. Equation 7 can be rewritten as Equation 9:

$$F_{A-t,B-A}^{new} = \frac{DF_{B-t}^{new}}{DF_{A-t}^{new}} \tag{9}$$

Substituting Equation 8 into Equation 9 and adjusting for the lapse of time t results in Equation 10:

$$F_{A-t,B-A}^{new} = \frac{DF_{B-t}^{new}}{DF_{A-t}^{new}} = \frac{\dfrac{DF_B}{DF_t}}{\dfrac{DF_A}{DF_t}} = \frac{DF_B}{DF_A} = F_{A,B-A} \tag{10}$$

Equation 10 shows that the forward contract price remains unchanged as long as future spot rates are equal to what is predicted by today's forward curve. Therefore, a change in the forward price is the result of a deviation of the spot curve from what is predicted by today's forward curve.

To make these observations concrete, consider a flat yield curve for which the interest rate is 4%. Using Equation 1, the discount factors for the one-year, two-year, and three-year terms are, to four decimal places, as follows:

$$DF_1 = \frac{1}{(1 + 0.04)} = 0.9615$$

$$DF_2 = \frac{1}{(1 + 0.04)^2} = 0.9246$$

$$DF_3 = \frac{1}{(1 + 0.04)^3} = 0.8890$$

Therefore, using Equation 7, the forward contract price that delivers a one-year bond at Year 2 is

$$F_{2,1} = \frac{DF_3}{DF_2} = \frac{0.8890}{0.9246} = 0.9615$$

Suppose the future discount function at Year 1 is the same as the forward discount function implied by the Year 0 spot curve. The lapse of time is $t = 1$. Using Equation 8, the discount factors for the one-year and two-year terms one year from today are as follows:

$$DF_1^{new} = \frac{DF_2}{DF_1} = \frac{0.9246}{0.9615} = 0.9616$$

$$DF_2^{new} = \frac{DF_3}{DF_1} = \frac{0.8890}{0.9615} = 0.9246$$

Using Equation 9, the price of the forward contract one year from today is

$$F_{2,1}^{new} = \frac{DF_2^{new}}{DF_1^{new}} = \frac{0.9246}{0.9615} = 0.9616$$

The price of the forward contract is nearly unchanged. This will be the case as long as future discount functions are the same as those based on today's forward curve.

From this numerical example, we can see that if the spot rate curve is unchanged, then each bond "rolls down" the curve and earns the current one-period spot rate and subsequent forward rates. Specifically, when one year passes, a three-year bond will return (0.9246 − 0.8890)/0.8890 = 4%, which is equal to the spot rate. Furthermore, if another year passes, the bond will return (0.9615 − 0.9246)/0.9246 = 4%, which is equal to the implied forward rate for a one-year security one year from today.

3 ACTIVE BOND PORTFOLIO MANAGEMENT

d describe the strategy of rolling down the yield curve

One way that active bond portfolio managers attempt to outperform the bond market's return is by anticipating changes in interest rates relative to the projected evolution of spot rates reflected in today's forward curves.

The forward rate model (Equation 4) provides insight into these issues. By rearranging terms in Equation 4 and setting the time horizon to one period, $A = 1$, we obtain

$$\frac{(1 + z_B)^B}{(1 + f_{A,B-A})^{B-A}} = (1 + z_A)^A \quad \textbf{(11)}$$

The numerator of the left-hand side of Equation 11 is for a bond with an initial maturity of B periods and a remaining maturity of $B - A$ periods after A periods pass. Suppose the prevailing spot yield curve after one period ($A = 1$) is the current forward curve; then, Equation 11 shows that the total return on the bond is the one-period risk-free rate. The following sidebar shows that returns on bonds of varying tenor over a one-year period always equal the one-year rate (the risk-free rate over the one-year period) if the spot rates evolve as implied by the current forward curve at the end of the first year.

WHEN SPOT RATES EVOLVE AS IMPLIED BY THE CURRENT FORWARD CURVE

As in earlier examples, assume the following:

$z_1 = 9\%$

$z_2 = 10\%$

$z_3 = 11\%$

$f_{1,1} = 11.01\%$

$f_{1,2} = 12.01\%$

If the spot curve one year from today reflects the current forward curve, the return on a zero-coupon bond for the one-year holding period is 9%, regardless of the bond's maturity. The following computations assume a par amount of 100 and represent the percentage change in price. Given the rounding of price and the forward rates to the nearest hundredth, the returns all approximate 9%. With no rounding, however, all answers would be precisely 9%.

The return of the one-year zero-coupon bond over the one-year holding period is 9%. The bond is purchased at a price of 91.74 and is worth the par amount of 100 at maturity.

$$\left(100 \div \frac{100}{1 + z_1}\right) - 1 = \left(100 \div \frac{100}{1 + 0.09}\right) - 1 = \frac{100}{91.74} - 1 = 9\%$$

The return of the two-year zero-coupon bond over the one-year holding period is 9%. The bond is purchased at a price of 82.64. One year from today, the two-year bond has a remaining maturity of one year. Its price one year from today is 90.08, determined as the par amount divided by 1 plus the forward rate for a one-year bond issued one year from today.

$$\left(\frac{100}{\left(1 + f_{1,1}\right)} \div \frac{100}{\left(1 + z_2\right)^2}\right) - 1 = \left(\frac{100}{\left(1 + 0.1101\right)} \div \frac{100}{\left(1 + 0.10\right)^2}\right) - 1 = \frac{90.08}{82.64} - 1 = 9\%$$

The return of the three-year zero-coupon bond over the one-year holding period is 9%. The bond is purchased at a price of 73.12. One year from today, the three-year bond has a remaining maturity of two years. Its price one year from today of 79.71 reflects the forward rate for a two-year bond issued one year from today.

$$\left(\frac{100}{\left(1 + f_{1,2}\right)^2} \div \frac{100}{\left(1 + z_3\right)^3}\right) - 1 = \left(\frac{100}{\left(1 + 0.1201\right)^2} \div \frac{100}{\left(1 + 0.11\right)^3}\right) - 1 = \frac{79.71}{73.12} - 1 \approx 9\%$$

This numerical example shows that the return of a bond over a one-year period is always the one-year rate (the risk-free rate over the one period) if the spot rates evolve as implied by the current forward curve.

But if the spot curve one year from today differs from today's forward curve, the returns on each bond for the one-year holding period will not all be 9%. To show that the returns on the two-year and three-year bonds over the one-year holding period are not 9%, we assume that the spot rate curve at Year 1 is flat with yields of 10% for all maturities.

The return on a one-year zero-coupon bond over the one-year holding period is

$$\left(100 \div \frac{100}{1 + 0.09}\right) - 1 = 9\%$$

The return on a two-year zero-coupon bond over the one-year holding period is

$$\left(\frac{100}{1 + 0.10} \div \frac{100}{\left(1 + 0.10\right)^2}\right) - 1 = 10\%$$

The return on a three-year zero-coupon bond over the one-year holding period is

$$\left(\frac{100}{\left(1 + 0.10\right)^2} \div \frac{100}{\left(1 + 0.11\right)^3}\right) - 1 = 13.03\%$$

The bond returns are 9%, 10%, and 13.03%. The returns on the two-year and three-year bonds differ from the one-year risk-free interest rate of 9%.

Equation 11 provides a total return investor with a means to evaluate the cheapness or expensiveness of a bond of a certain maturity. If any of the investor's expected future spot rates is below a quoted forward rate for the same maturity, then (all else being equal) the investor would perceive the bond to be undervalued, in the sense that the market is effectively discounting the bond's payments at a higher rate than the investor and the bond's market price is below the intrinsic value perceived by the investor.

Another example will reinforce the point that if a portfolio manager's projected spot curve is above (below) the forward curve and his expectation turns out to be true, the return will be less (more) than the one-period risk-free interest rate.

For the sake of simplicity, assume a flat yield curve of 8% and that a trader holds a three-year bond paying an 8% annual coupon. Assuming a par value of 100, the current market price is also 100. If today's forward curve turns out to be the spot curve one year from today, the trader will earn an 8% return.

If the trader projects that the spot curve one year from today is above today's forward curve—for example, a flat yield curve of 9%—the trader's expected rate of return is 6.24%, which is less than 8%:

$$\frac{8 + \dfrac{8}{1 + 0.09} + \dfrac{108}{(1 + 0.09)^2}}{100} - 1 = 6.24\%$$

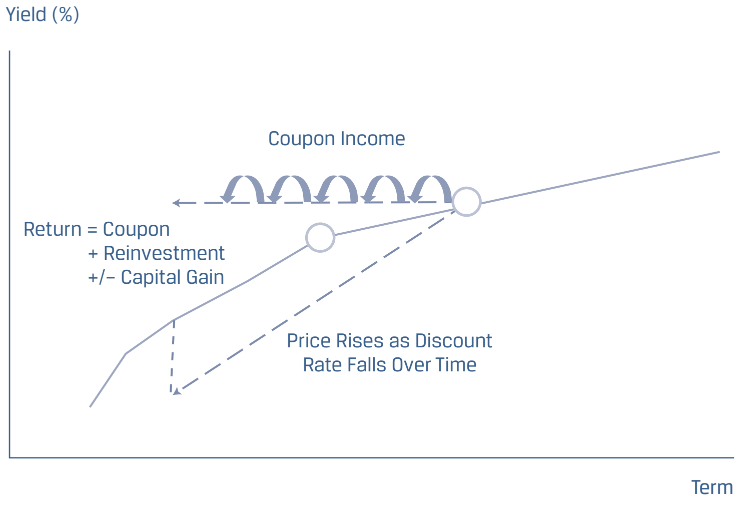
If the trader predicts a flat yield curve of 7%, the trader's expected return is 9.81%, which is greater than 8%:

$$\frac{8 + \dfrac{8}{1 + 0.07} + \dfrac{108}{(1 + 0.07)^2}}{100} - 1 = 9.81\%$$

As the gap between the projected future spot rate and the forward rate widens, so too will the difference between the trader's expected return and the original YTM of 8%.

This logic is the basis for a popular yield curve trade called **rolling down the yield curve**, also referred to as riding the yield curve. As we have noted, when a yield curve is upward sloping, the forward curve is always above the current spot curve. If the trader expects the yield curve to remain static over an investment horizon, then buying bonds with a maturity longer than the investment horizon would provide a total return greater than the return on a maturity-matching strategy. The bond's total return will depend on the spread between the forward rate and the spot rate as well as the maturity of the bond. The longer the bond's maturity, the more sensitive its total return is to the spread. This strategy is shown in Exhibit 4.

Exhibit 4 Rolling Down the Yield Curve



The return on a yield curve rolldown strategy may be demonstrated using a simple example. As stated earlier, the investment return on a fixed-rate (non-defaulted and non-callable) bond return may be defined as follows:

Bond return = Receipt of promised coupons (and principal)
+ Reinvestment of coupon payments (12)
+/– Capital gain/Loss on sale prior to maturity

Say we observe one-, three-, four-, five- and six-year spot rates on annual coupon bonds trading at par of 2%, 4%, 5%, 6%, and 7%, respectively. An investor with a five-year maturity target decides to forgo a matched-maturity 6% five-year bond in favor of the 7%, six-year bond given her expectation of an unchanged yield curve over the next two years. We can compare the annualized return over two years for both bonds, assuming unchanged yields, as follows.

The 6% five-year bond purchased for 100 returns 120.61 in two years [(6 × 1.02) + 6 + 108.49], which consists of the first year's coupon reinvested at the one-year rate, the second annual coupon, and the capital gain on the sale of the 6% bond with three years to maturity at an unchanged three-year yield of 4% [108.49 = $6/1.04 + 6/(1.04)^2 + 106/(1.04)^3$]. The annualized rate of return is 9.823% [solve for r, where $(120.61/100) = (1 + r)^2$].

The 7% six-year bond purchased at par returns 125.03 in two years [(7 × 1.02) + 7 + 110.89] with an annualized return of 11.817%. The excess return of nearly 2% results from both higher coupon income than the five-year matched maturity bond as well as a larger capital gain on the sale of the 7% bond with four years to maturity at an unchanged four-year yield of 5% [110.89 = $7/1.05 + 7/(1.05)^2 + + 7/(1.05)^3 + 107/(1.05)^4$].

In the years following the 2008 financial crisis, many central banks acted to keep short-term interest rates very low. As a result, yield curves subsequently had a steep upward slope (see Exhibit 2). For active fixed-income managers, this situation provided an incentive to access short-term funding and invest in long-term bonds. This is just one form of a carry trade, referred to as a maturity spread carry trade, and is subject to significant interest rate risk, such as an unexpected increase in future spot rates (e.g., as a result of a spike in inflation). The maturity spread carry trade, in which the trader borrows short term and lends long term in the same currency, is common in an upward-sloping yield curve environment.

In summary, when the yield curve slopes upward, as a bond approaches maturity or "rolls down the yield curve," it is valued at successively lower yields and higher prices. Using this strategy, a bond can be held for a period of time as it appreciates in price and then sold before maturity to realize a higher return. As long as interest rates remain stable and the yield curve retains an upward slope, this strategy can continuously add to the total return of a bond portfolio.

Example 7 addresses how the preceding analysis relates to active bond portfolio management.

EXAMPLE 7

Active Bond Portfolio Management

1 The "rolling down the yield curve" strategy is executed by buying bonds whose maturities are:
   - A equal to the investor's investment horizon.
   - B longer than the investor's investment horizon.
   - C shorter than the investor's investment horizon.
2 A bond will be overvalued if the expected spot rate is:
   - A equal to the current forward rate.
   - B lower than the current forward rate.
   - C higher than the current forward rate.
3 Assume a flat yield curve of 6%. A three-year £100 bond is issued at par paying an annual coupon of 6%. What is the portfolio manager's expected return if he predicts that the yield curve one year from today will be a flat 7%?
   - A 4.19%
   - B 6.00%
   - C 8.83%
4 A forward contract price will increase if:
   - A future spot rates evolve as predicted by current forward rates.
   - B future spot rates are lower than what is predicted by current forward rates.
   - C future spot rates are higher than what is predicted by current forward rates.

Solution to 1:

B is correct. A bond with a longer maturity than the investor's investment horizon is purchased but then sold prior to maturity at the end of the investment horizon. If the yield curve is upward sloping and yields do not change, the bond will be valued at successively lower yields and higher prices over time. The bond's total return will exceed that of a bond whose maturity is equal to the investment horizon.

Solution to 2:

C is correct. If the expected discount rate is higher than the forward rate, then the bond will be overvalued. The expected price of the bond is lower than the price obtained from discounting using the forward rate.

Solution to 3:

A is correct. Expected return will be less than the current YTM of 6% if yields increase to 7%. The expected return of 4.19% is computed as follows:

$$\frac{6 + \dfrac{6}{1 + 0.07} + \dfrac{106}{(1 + 0.07)^2}}{100} - 1 \approx 4.19\%$$

Solution to 4:

B is correct. The forward rate model can be used to show that a change in the forward contract price requires a deviation of the spot curve from that predicted by today's forward curve. If the future spot rate is lower than what is predicted by the prevailing forward rate, the forward contract price will increase because it is discounted at an interest rate that is lower than the originally anticipated rate.

THE SWAP RATE CURVE 4

e explain the swap rate curve and why and how market participants use it in valuation

Earlier, we described the spot rate curve of default-risk-free bonds as a measure of the time value of money. The swap rate curve, or swap curve for short, is another important representation of the time value of money used in fixed-income markets. Here we will discuss how the swap curve is used in valuation, where the spread of swap rates over government benchmark rates is a proxy for perceived credit risk relative to risk-free debt.

4.1 Swap Rate Curve

Interest rate swaps are an integral part of the fixed-income market. These derivative contracts usually involve the net exchange, or swap, of fixed-rate for floating-rate interest payments, and these contracts are an essential tool for investors who use them to hedge, speculate on, or otherwise modify risk. The fixed and floating payments are determined by multiplying the respective rate by a principal (or notional) amount for each interest period over the swap maturity. The rate for the fixed leg of an interest rate swap is known as the **swap rate**. The swap rate is analogous to the YTM on a government bond, which as we saw earlier may be derived from zero rates using bootstrapping. The key difference between the swap rate and the government bond rate is that the swap rate is derived using short-term lending rates rather than default-risk-free rates. Swap floating rates historically referenced short-term survey-based interest rates, such as three- or six-month US dollar Libor (London Interbank Offered Rate) and are slated to transition to transaction-based market reference rates (MRR) based on secured overnight funding transactions. The yield curve of swap rates is called the **swap rate curve** or, more simply, the **swap curve**. Because it is based on so-called **par swaps**, in which the fixed rate is set so that no money is exchanged at contract initiation—the present values of the fixed-rate and benchmark floating-rate legs being equal— the swap curve is a type of par curve. When we refer to the "par curve" here, however, the reference is to the government par yield curve.

The swap market is a highly liquid market for two reasons. First, unlike bonds, a swap does not have multiple borrowers or lenders, only counterparties who exchange cash flows. Such arrangements offer significant flexibility and customization in the swap contract's design. Second, swaps provide one of the most efficient ways to hedge interest rate risk. The Bank for International Settlements (BIS) estimates that the notional amount outstanding on interest rate swaps was nearly $350 trillion as of June 2020.

Many countries do not have a liquid government bond market with maturities longer than one year. The swap curve is a necessary market benchmark for interest rates in these countries. In countries where the private sector is much bigger than the public sector, the swap curve is a far more relevant measure of the time value of money than is the government's cost of borrowing.

Swaps are frequently used as a benchmark in Europe, whereas in Asia, the swap markets and the government bond markets have developed in parallel, and both are used in valuation in credit and loan markets.

4.2 Why Do Market Participants Use Swap Rates When Valuing Bonds?

Government spot curves and swap rate curves are the chief reference curves in fixed-income valuation. The choice between them can depend on multiple factors, including the relative liquidity of these two markets. In the United States, where there is both an active Treasury security market and a swap market, the choice of a benchmark for the time value of money often depends on the interest rate exposure profile of the institution using the benchmark. On one hand, wholesale banks frequently use the swap curve to value assets and liabilities because they hedge their balance sheet with swaps. On the other hand, retail banks with little exposure to the swap market are more likely to use the government spot curve as their benchmark.

Let us illustrate how a financial institution uses the swap market for its internal operations. Consider the case of a bank raising funds using a certificate of deposit (CD). Assume the bank can borrow $10 million in the form of a CD that bears interest of 1.5% for a two-year term. Another $10 million CD offers 1.70% for a three-year term. The bank can arrange two swaps: (1) The bank receives 1.50% fixed and pays MRR minus 10 bps with a two-year term and a notional amount of $10 million, and (2) the bank receives 1.70% fixed and pays MRR minus 15 bps with a three-year term and a notional amount of $10 million. After issuing the two CDs and committing to the two swaps, the bank has raised $20 million with an annual funding cost for the first two years of MRR minus 12.5 bps applied to the total notional amount of $20 million. The fixed interest payments received from the counterparty to the swap are paid to the CD investors; in effect, fixed-rate liabilities have been converted to floating-rate liabilities. The margins on the floating rates become the standard by which value is measured in assessing the bank's total funding cost.

By using the swap curve as a benchmark for the time value of money, the investor can adjust the swap spread so that the swap will be fairly priced given the spread. Conversely, given a swap spread, the investor can determine a fair price for the bond. We will use the swap spread in the following section to determine the value of a bond.

4.3 How Do Market Participants Use the Swap Curve in Valuation?

Although benchmark swap rates are quoted for specific maturities, swap contracts may be customized by two parties in the over-the-counter market. The fixed payment can be specified by an amortization schedule or involve a coupon with non-standard payment dates. In this section, we will focus on zero-coupon bonds. The yields on these bonds determine the swap curve, which, in turn, can be used to determine bond values.

Each forward date has an associated discount factor that represents the value today of a unit payment that one would hypothetically receive on the forward date expressed as a decimal fraction. For example, if we expect to receive ₩10,000 (10,000 South Korean won) in one year and the current price of the security is ₩9,259.30, then the discount factor for one year will be 0.92593 (= ₩9,259.30/₩10,000). Note that the rate associated with this discount factor is 1/0.92593 − 1 ≈ 8.00%.

To price a swap using current market rates, as mentioned we must solve for a constant fixed rate that sets the present value of fixed-leg payments equal to the present value of floating-leg payments over the life of the swap. Once established, the fixed cash flows are specified by the coupon rate set at the time of the original agreement. Pricing the floating leg is more complex than pricing the fixed leg because, by definition, its cash flows change with future changes in interest rates. The forward rate for each floating payment date is calculated by using the forward curves.

Let s_T stand for the T-period swap rate. Because the value of a swap at origination is set to zero, the swap rates must satisfy Equation 13. Note that the swap rates can be determined from the spot rates and the spot rates can be determined from the swap rates.

$$\sum_{t=1}^{T} \frac{s_T}{(1+z_t)^t} + \frac{1}{(1+z_T)^T} = 1 \tag{12}$$

The right-hand side of Equation 12 is the value of the floating leg, which is always 1 at origination. The swap rate is determined by equating the value of the fixed leg, on the left-hand side, to the value of the floating leg.

Example 8 addresses the relationship between the swap rate curve and spot curve.

EXAMPLE 8

Determining the Swap Rate Curve

Suppose a government spot curve implies the following discount factors:

$DF_1 = 0.9524$

$DF_2 = 0.8900$

$DF_3 = 0.8163$

$DF_4 = 0.7350$

Given this information, determine the swap rate curve.

Solution:

Recall from Equation 1 that $DF_N = \dfrac{1}{(1+Z_N)^N}$. Therefore,

$$z_N = \left(\frac{1}{DF_N}\right)^{1/N} - 1$$

$$z_1 = \left(\frac{1}{0.9524}\right)^{1/1} - 1 = 5.00\%$$

$$z_2 = \left(\frac{1}{0.8900}\right)^{1/2} - 1 = 6.00\%$$

$$z_3 = \left(\frac{1}{0.8163}\right)^{1/3} - 1 = 7.00\%$$

$$z_4 = \left(\frac{1}{0.7350}\right)^{1/4} - 1 = 8.00\%$$

Using Equation 12, for $N = 1$,

$$\frac{s_1}{(1 + z_1)} + \frac{1}{(1 + z_1)} = \frac{s_1 + 1}{(1 + 0.05)} = 1$$

Therefore, $s_1 = 5\%$.

For $T = 2$,

$$\frac{s_2}{(1 + z_1)} + \frac{s_2}{(1 + z_2)^2} + \frac{1}{(1 + z_2)^2} = \frac{s_2}{(1 + 0.05)} + \frac{s_2 + 1}{(1 + 0.06)^2} = 1$$

Therefore, $s_2 = 5.97\%$.

For $T = 3$,

$$\frac{s_3}{(1 + z_1)} + \frac{s_3}{(1 + z_2)^2} + \frac{s_3}{(1 + z_3)^3} + \frac{1}{(1 + z_3)^3}$$
$$= \frac{s_3}{(1 + 0.05)} + \frac{s_3}{(1 + 0.06)^2} + \frac{s_3}{(1 + 0.07)^3} + \frac{1}{(1 + 0.07)^3} = 1$$

Therefore, $s_3 = 6.91\%$.

For $T = 4$,

$$\frac{s_4}{(1 + z_1)} + \frac{s_4}{(1 + z_2)^2} + \frac{s_4}{(1 + z_3)^3} + \frac{s_4}{(1 + z_4)^4} + \frac{1}{(1 + z_4)^4}$$
$$= \frac{s_4}{(1 + 0.05)} + \frac{s_4}{(1 + 0.06)^2} + \frac{s_4}{(1 + 0.07)^3} + \frac{s_4}{(1 + 0.08)^4} + \frac{1}{(1 + 0.08)^4} = 1$$

Therefore, $s_4 = 7.81\%$.

Note that the swap rates, spot rates, and discount factors are all mathematically linked together. Having access to data for one of the series allows you to calculate the other two.

5 THE SWAP SPREAD AND SPREADS AS A PRICE QUOTATION CONVENTION

f calculate and interpret the swap spread for a given maturity

g describe short-term interest rate spreads used to gauge economy-wide credit risk and liquidity risk

The swap spread is a popular way to indicate credit spreads in a market. The **swap spread** is defined as the spread paid by the fixed-rate payer of an interest rate swap over the rate of the "on-the-run" (most recently issued) government security with the same maturity as the swap. The spread captures the yield premium required for credit relative to the benchmark government bond. Because swap rates are built from market rates for short-term risky debt, this spread is a barometer of the market's perceived credit risk relative to default-risk-free rates. This spread typically widens countercyclically, exhibiting greater values during recessions and lower values during economic expansions.

The term "swap spread" is sometimes also used as a reference to a bond's basis point spread over the interest rate swap curve and is a measure of the credit and/or liquidity risk of a bond. Here, a swap spread is an excess yield of swap rates over the yields on government bonds, and we use the terms I-spread, ISPRD, or interpolated spread to refer to bond yields net of the swap rates of the same maturities. In its simplest form, the I-spread can be measured as the difference between the yield-to-maturity of the bond and the swap rate given by a straight-line interpolation of the swap curve.

Often, fixed-income prices will be quoted as a swap rate plus (or minus) a spread, for which the yield is simply the yield on an equal-maturity government bond plus the swap spread. For example, if the fixed rate of a five-year fixed-for-float MRR swap is 2.00% and the five-year Treasury is yielding 1.70%, the swap spread is 2.00% − 1.70% = 0.30%, or 30 bps.

For euro-denominated swaps, the government yield used as a benchmark is most frequently Bunds (German government bonds) with the same maturity. Gilts (UK government bonds) are used as a benchmark in the United Kingdom.

Although the Libor swap curve is being phased out, it has historically been considered to reflect the default risk of A1/A+ rated commercial banks. The transition from Libor to MRR based on secured overnight funding rates will increase the influence of demand and supply conditions in government debt markets on swap rates. Another reason for the popularity of the swap market is that it is led by major financial institutions rather than controlled by governments, so swap rates are more comparable across different countries. The swap market also has more maturities with which to construct a yield curve than do government bond markets. Historically, cash or deposit rates such as Libor have been used for short-maturity yields; interest rate futures such as Eurodollar futures contracts have maturities of up to a year; and swap rates extend to maturities of up to 50 years in US dollars or euro. As the market transitions from Libor, the concept of this spread will be consistent with whichever market-based alternative to Libor emerges.

HISTORY OF THE US SWAP SPREAD SINCE 2008

The fact that governments generally pay less than private entities do in order to borrow suggests that swap spreads should always be positive. However, the 30-year Treasury swap spread turned negative following the collapse of Lehman Brothers Holdings Inc. in September 2008. Strong demand for duration combined with tighter liquidity and greater counterparty risk were widely cited as reasons for this phenomenon. For the period shown, the 30-year Treasury swap spread hit a record low (–62 bps intramonth) during November 2008. The 30-year Treasury swap spread was at or above zero for more than a year before becoming negative once again (see Exhibit 5). A recent study by the Federal Reserve Bank of New York (Boyarchenko, Gupta, Steele, and Yen, 2018) suggests that negative swap spreads have persisted because of increased regulatory capital requirements among swap dealers following the financial crisis.

Exhibit 5 US Swap Spread, January 2008–May 2020 (monthly data)

To illustrate the use of the swap spread in fixed-income pricing, consider a US$1 million investment in GE Capital (GECC) notes with a coupon rate of 1 5/8% (1.625%) that matures on 2 July 2024. Coupons are paid semiannually. The evaluation date is 12 July 2021, so the remaining maturity is 2.97 years [= 2 + (350/360)]. The Treasury rates for two-year and three-year maturities are 0.525% and 0.588%, respectively. By simple interpolation between these two rates, the US Treasury rate for 2.97 years is 0.586% [= 0.525% + (350/360)(0.588% − 0.525%)]. If the swap spread for the same maturity is 0.918%, then the yield-to-maturity on the bond is 1.504% (= 0.918% + 0.586%). Given the yield-to-maturity, the invoice price (price including accrued interest) for US$1 million face value is as follows:

$$\frac{1{,}000{,}000\left(\frac{0.01625}{2}\right)}{\left(1+\frac{0.01504}{2}\right)^{\left(1-\frac{10}{180}\right)}}+\frac{1{,}000{,}000\left(\frac{0.01625}{2}\right)}{\left(1+\frac{0.01504}{2}\right)^{\left(2-\frac{10}{180}\right)}}+\cdots+$$

$$\frac{1{,}000{,}000\left(\frac{0.01625}{2}\right)}{\left(1+\frac{0.01504}{2}\right)^{\left(6-\frac{10}{180}\right)}}+\frac{1{,}000{,}000}{\left(1+\frac{0.01504}{2}\right)^{\left(6-\frac{10}{180}\right)}}=\text{US\$}1{,}003{,}954.12$$

The left-hand side sums the present values of the semiannual coupon payments and the final principal payment of US$1,000,000. The accrued interest rate amount is US$451.39 [= 1,000,000 × (0.01625/2)(10/180)]. Therefore, the clean price (price not including accrued interest) is US$1,003,502.73 (= 1,003,954.12 – 451.39).

The swap spread helps an investor to identify the time value, credit, and liquidity components of a bond's YTM. If the bond is default free, then the swap spread could provide an indication of the bond's liquidity, or it could provide evidence of market mispricing. The higher the swap spread, the higher the return that investors require for credit and/or liquidity risks. Another approach introduced in an earlier reading is to calculate a constant yield spread over a government (or interest rate swap) spot curve instead. This spread is known as the zero volatility spread (Z-spread) of a bond over the benchmark rate.

5.1 Spreads as a Price Quotation Convention

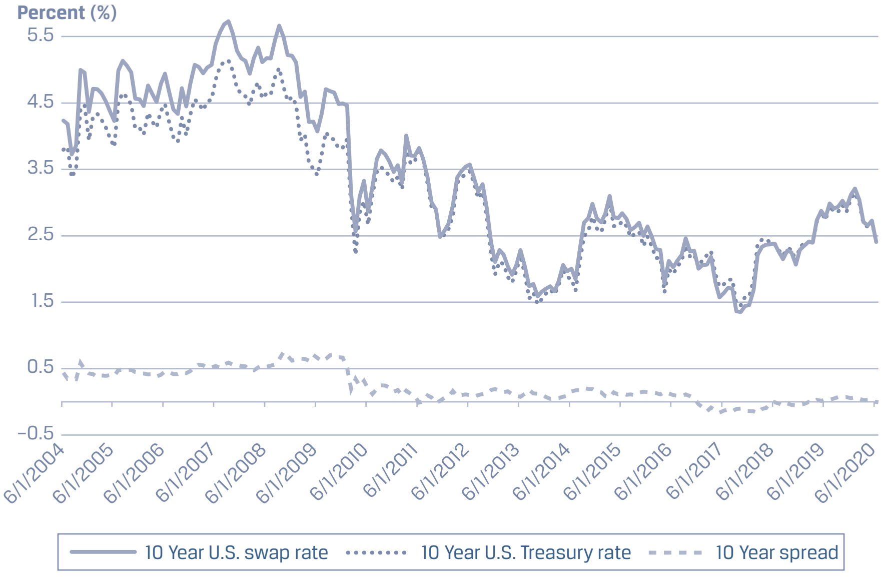
Treasury curves and swap curves represent different benchmarks for fixed-income valuation. It is therefore important to distinguish between a bond price quote that uses the bond yield net of a benchmark Treasury yield and one that uses a swap rate.

The Treasury rate can differ from the swap rate for the same term for several reasons. Unlike the cash flows from US Treasury bonds, the cash flows from swaps are subject to greater default risk. Market liquidity for specific maturities may differ. For example, some parts of the term structure of interest rates may be more actively traded with swaps than with Treasury bonds. Finally, arbitrage between these two markets cannot be perfectly executed.

Swap spreads to the Treasury rate (as opposed to **I-spreads**, which are bond rates net of the swap rates of the same maturities) are simply the differences between swap rates and government bond yields of a particular maturity. One problem in defining swap spreads is that, for example, a 10-year swap matures in exactly 10 years, whereas this condition is true for a 10-year government bond only at the time of issuance. By convention, therefore, the 10-year swap spread is defined as the difference between the 10-year swap rate and the 10-year on-the-run government bond. Swap spreads of other maturities are defined similarly.

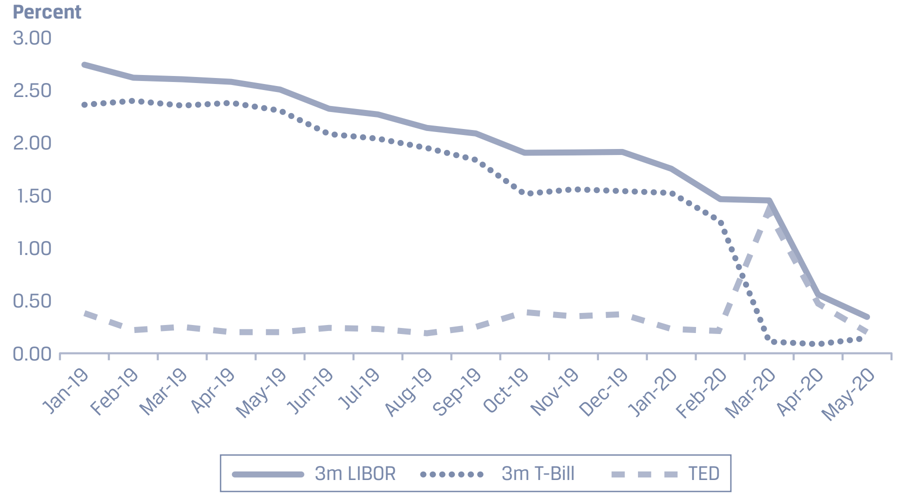
The curves in Exhibit 6 show the relationship between 10-year Treasury notes and 10-year swap rates. The 10-year swap spread is the 10-year swap rate less the 10-year Treasury note yield. Although positive swap spreads reflecting the difference between Libor-based rates and default-risk-free US government yields were historically the norm, these spreads have narrowed to zero or negative levels since the 2008 financial crisis because of higher swap dealer capital requirements and leverage constraints.

Exhibit 6 10-Year US Swap Rate vs. 10-Year US Treasury Rate



Market participants often use interest rate spreads between short-term government and risky rates as a barometer to evaluate relative credit and liquidity risk. For example, the difference between Libor and the yield on a Treasury bill of the same maturity, or **TED spread**, has historically been a key indicator of perceived credit and liquidity risk. TED is an acronym formed from an abbreviation for the US T-bill (T) and the ticker symbol for the Libor-based Eurodollar futures contract (ED). Exhibit 7 shows the historical TED spread. An increase in the TED spread signals greater perceived credit and liquidity risk, as occurred in early 2020 amid market turmoil related to the COVID-19 pandemic.

Exhibit 7 TED Spread, January 2019–May 2020 (end-of-month data)



Another popular measure of such risk is the **Libor–OIS spread**, which is the difference between Libor and the **overnight indexed swap (OIS) rate**. An OIS is an interest rate swap in which the periodic floating rate of the swap equals the geometric average of a daily unsecured overnight rate (or overnight index rate). The index rate is typically the rate for overnight unsecured lending between banks, such as the federal funds rate for US dollars or Eonia (Euro OverNight Index Average) for euros. As market participants transition away from survey-based Libor to alternative benchmarks based on actual transaction data, the **secured overnight financing rate (SOFR)**, or overnight cash borrowing rate collateralized by US Treasuries, has gained prominence and is expected to replace Libor in the future. A barometer of the US Treasury repurchase (or repo) market, SOFR is a daily volume-weighted index of all qualified repo market transactions and is influenced by supply and demand conditions in secured funding markets. The shift to overnight secured funding benchmarks extends globally—for example, the secured European Short-Term Rate (ESTR) has been recommended to replace Eonia, and the Canadian Overnight Repo Rate Average (CORRA) is proposed to replace the survey-based unsecured Canadian Dollar Offered Rate (CDOR).

TRADITIONAL THEORIES OF THE TERM STRUCTURE OF INTEREST RATES

6

h explain traditional theories of the term structure of interest rates and describe the implications of each theory for forward rates and the shape of the yield curve

This section presents four traditional theories of the underlying economic factors that affect the shape of the yield curve.

6.1 Expectations Theory

One branch of traditional term structure theory focuses on interpreting term structure shape in terms of investors' expectations. Historically, the first such theory is known as the **unbiased expectations theory**, also called **pure expectations theory**. It says that the forward rate is an unbiased predictor of the future spot rate; its broadest interpretation is that bonds of any maturity are perfect substitutes for one another. For example, buying a bond with a maturity of five years and holding it for three years has the same expected return as buying a three-year bond or buying a series of three one-year bonds.

The predictions of the unbiased expectations theory are consistent with the assumption of risk neutrality. In a risk-neutral world, investors are unaffected by uncertainty and risk premiums do not exist. Every security is risk free and yields the risk-free rate for that particular maturity. Although such an assumption leads to interesting results, it clearly is in conflict with the large body of evidence showing that investors are risk averse.

A theory that is similar but more rigorous than the unbiased expectations theory is the **local expectations theory**. Rather than asserting that every maturity strategy has the same expected return over a given investment horizon, this theory instead contends that the expected return for every bond over short periods is the risk-free rate. This conclusion results from an assumed no-arbitrage condition in which bond pricing does not allow for traders to earn arbitrage profits.

The primary way that the local expectations theory differs from the unbiased expectations theory is that it can be extended to a world characterized by risk. Although the theory requires that risk premiums be nonexistent for very short holding periods, no such restrictions are placed on longer-term investments. Thus, the theory is applicable to both risk-free as well as risky bonds.

Although the local expectations theory is economically appealing, it is often observed that short-holding-period returns on long-dated bonds in fact exceed those on short-dated bonds. The need for liquidity and the ability to hedge risk essentially ensure that the demand for short-term securities will exceed that for long-term securities. Thus, both the yields and the actual returns for short-dated securities are typically lower than those for long-dated securities.

6.2 Liquidity Preference Theory

Whereas expectations theories leave no room for risk aversion, liquidity preference theory attempts to account for it. **Liquidity preference theory** asserts that **liquidity premiums** exist to compensate investors for the added interest rate risk they face when lending long term and that these premiums increase with maturity. Thus, given an expectation of unchanging short-term spot rates, liquidity preference theory predicts an upward-sloping yield curve. The forward rate provides an estimate of the expected spot rate that is biased upward by the amount of the liquidity premium, which invalidates the unbiased expectations theory. The liquidity premium for each consecutive future period should be no smaller than that for the prior period.

For example, the US Treasury offers bonds that mature in 30 years. Most investors, however, have shorter investment horizons than 30 years. For investors to hold these bonds, they would demand a higher return for taking the risk that the yield curve changes and that they must sell the bond prior to maturity at an uncertain price. That incrementally higher return is the liquidity premium. Note that this premium is not to be confused with a yield premium for the lack of liquidity that thinly traded bonds may bear. Rather, it is a premium applying to all long-term bonds, including those with deep markets.

Liquidity preference theory fails to offer a complete explanation of the term structure. Rather, it simply argues for the existence of liquidity premiums. For example, a downward-sloping yield curve could still be consistent with the existence of liquidity premiums if one of the factors underlying the shape of the curve is an expectation of deflation (i.e., a negative rate of inflation resulting from monetary or fiscal policy actions). Expectations of sharply declining spot rates may also result in a downward-sloping yield curve if the expected decline in interest rates is severe enough to offset the effect of the liquidity premiums.

In summary, liquidity preference theory claims that lenders require a liquidity premium as an incentive to lend long term. Thus, forward rates derived from the current yield curve provide an upwardly biased estimate of expected future spot rates. Although downward-sloping or hump-shaped yield curves may sometimes occur, the existence of liquidity premiums implies that the yield curve will typically be upward sloping.

6.3 Segmented Markets Theory

Unlike expectations theory and liquidity preference theory, **segmented markets theory** allows for lender and borrower preferences to influence the shape of the yield curve. The result is that yields are not a reflection of expected spot rates or liquidity premiums. Rather, they are solely a function of the supply and demand for funds of a particular maturity. That is, each maturity sector can be thought of as a segmented market in which yield is determined independently from the yields that prevail in other maturity segments.

The theory is consistent with a world in which asset/liability management constraints exist, either regulatory or self-imposed. In such a world, investors might restrict their investment activity to a maturity sector that provides the best match for the maturity of their liabilities. Doing so avoids the risks associated with an asset/liability mismatch.

For example, because life insurers sell long-term liabilities against themselves in the form of life insurance contracts, they tend to be most active as buyers in the long end of the bond market. Similarly, because the liabilities of pension plans are long term, they typically invest in long-term securities. Why would they invest short term given that those returns might decline while the cost of their liabilities stays fixed? In contrast, money market funds would be limited to investing in debt with maturity of one year or less, in general.

In summary, the segmented markets theory assumes that market participants are either unwilling or unable to invest in anything other than securities of their preferred maturity. It follows that the yield of securities of a particular maturity is determined entirely by the supply and demand for funds of that particular maturity.

6.4 Preferred Habitat Theory

The **preferred habitat theory** is similar to the segmented markets theory in proposing that many borrowers and lenders have strong preferences for particular maturities, but it does not assert that yields at different maturities are determined independently of each other.

The theory contends, however, that if the expected additional returns to be gained become large enough, institutions will be willing to deviate from their preferred maturities or habitats. For example, if the expected returns on longer-term securities exceed those on short-term securities by a large enough margin, an intermediate-term bond fund might lengthen the maturities of their assets. And if the excess returns expected from buying short-term securities become large enough, life insurance companies might stop limiting themselves to long-term securities and place a larger part of their portfolios in shorter-term investments.

The preferred habitat theory is based on the realistic notion that agents and institutions will accept additional risk in return for additional expected returns. In accepting elements of both the segmented markets theory and the unbiased expectations theory, yet rejecting their extreme polar positions, the preferred habitat theory moves closer to explaining real-world phenomena. In this theory, both market expectations and the institutional factors emphasized in the segmented markets theory influence the term structure of interest rates.

PREFERRED HABITAT AND QE

The term "quantitative easing" (QE) refers to an unconventional monetary policy used by central banks to increase the supply of money in an economy when central bank and/or interbank interest rates are already close to zero. The first of several QE efforts by the US Federal Reserve began in late 2008, following the establishment of a near-zero target range for the federal funds rate. Since then, the Federal Reserve has greatly expanded its holdings of long-term securities via a series of asset purchase programs, with the goal of putting downward pressure on long-term interest rates and thereby making financial conditions even more accommodative. Exhibit 8 presents information regarding the securities held by the Federal Reserve on 20 September 2007 (when all securities held by the Fed were US Treasury issuance) and on 29 October 2014 (when the Federal Reserve ended its third round of QE).

Exhibit 8 Securities Held by the US Federal Reserve

(US$ billions)	20 Sep 2007	29 Oct 2014
Securities held outright	780	4,219
US Treasury	780	2,462
Bills	267	0
Notes and bonds, nominal	472	2,347
Notes and bonds, inflation indexed	36	115
Inflation compensation	5	16
Federal agency	0	40
Mortgage-backed securities	0	1,718

As Exhibit 8 shows, the Federal Reserve's security holdings on 20 September 2007 consisted entirely of US Treasury securities, and about 34% of those holdings were short term in the form of T-bills. On 29 October 2014, only about 58% of the Federal Reserve's security holdings were Treasury securities, and none were T-bills. Furthermore, the Federal Reserve held well over US$1.7 trillion of mortgage-backed securities (MBS), which accounted for 41% of all securities held.

Prior to the QE efforts, the yield on MBS was typically in the 5%–6% range. It declined to less than 2% by the end of 2012. Concepts related to preferred habitat theory could possibly help explain that drop in yield.

The purchase of MBS by the Federal Reserve reduced the supply of these securities that was available for private purchase. Assuming that many MBS investors are either unwilling or unable to withdraw from the MBS market because of their comparative experience and expertise in managing interest rate and repayment risks of MBS versus option-free bonds, MBS investing institutions would have a "preferred habitat" in the MBS market. If they were unable to meet investor demand without bidding more aggressively, these buyers would drive down yields on MBS.

The Federal Reserve's purchase of MBS also resulted in a reduction in MBS yields. If a homeowner prepays on a mortgage, the payment is sent to MBS investors on a pro rata basis. Although investors are uncertain about when such a prepayment will be received, prepayment is more likely in a declining interest rate environment.

Use Example 9 to test your understanding of traditional term structure theories.

EXAMPLE 9

Traditional Term Structure Theories

1 Many fixed-income portfolio managers are limited in or prohibited from high-yield bond investments. When a bond is downgraded from an investment-grade to a high-yield (junk) rating, it is referred to as a *fallen angel*. Because of restrictions, many pension funds sell fallen angels when they are downgraded from investment grade to high yield (junk). This coordinated selling action often results in depressed prices and attractive yields for the fallen angels. Which of the following reasons best explains why fallen angel yields often exceed otherwise identical bonds?

   A The preferred habitat theory

   B The segmented markets theory

   C The local expectations theory

2 The term structure theory in which investors can be induced by relatively attractive yields to hold debt securities whose maturities do not match their investment horizon is *best* described as the:

   A preferred habitat theory.

   B segmented markets theory.

   C unbiased expectations theory.

3 The unbiased expectations theory assumes investors are:

   A risk averse.

   B risk neutral.

   C risk seeking.

4 Market evidence shows that forward rates are:

   A unbiased predictors of future spot rates.

   B upwardly biased predictors of future spot rates.

   C downwardly biased predictors of future spot rates.

5 Market evidence shows that short holding-period returns on short-maturity bonds *most* often are:

A less than those on long-maturity bonds.

B about equal to those on long-maturity bonds.

C greater than those on long-maturity bonds.

Solution to 1:

B is correct. Market segmentation in this example results from the requirement that some fixed-income fund managers are prohibited or limited in their capacity to hold high-yield bonds. The segmentation results in selling pressure on fallen angels that depresses their prices.

Solution to 2:

A is correct. Preferred habitat theory asserts that investors are willing to deviate from their preferred maturities if yield differentials encourage the switch. Segmented markets theory is more rigid than preferred habitat in that asset/liability management constraints force investors to buy securities whose horizons match those of their liabilities. The unbiased expectations theory makes no assumptions about maturity preferences. Rather, it contends that forward rates are unbiased predictors of future spot rates.

Solution to 3:

B is correct. The unbiased expectations theory asserts that different maturity strategies, such as rollover, maturity matching, and riding the yield curve, have the same expected return. By definition, a risk-neutral party is indifferent about choices with equal expected payoffs, even if one choice is riskier. Thus, the predictions of the theory are consistent with the existence of risk-neutral investors.

Solution to 4:

B is correct. The existence of a liquidity premium ensures that the forward rate is an upwardly biased estimate of the future spot rate. Market evidence clearly shows that liquidity premiums exist, and this evidence effectively refutes the predictions of the unbiased expectations theory.

Solution to 5:

A is correct. Although the local expectations theory predicts that the short-run return for all bonds will equal the risk-free rate, most of the evidence refutes that claim. Returns from long-dated bonds are generally higher than those from short-dated bonds, even over relatively short investment horizons. This market evidence is consistent with the risk–expected return trade-off that is central to finance and the uncertainty surrounding future spot rates.

YIELD CURVE FACTOR MODELS

7

i. explain how a bond's exposure to each of the factors driving the yield curve can be measured and how these exposures can be used to manage yield curve risks

The effect of yield volatilities on price is an important consideration in fixed-income investment, particularly for risk management and portfolio evaluation. In this section, we describe measuring and managing the interest rate risk of bonds.

7.1 A Bond's Exposure to Yield Curve Movement

Shaping risk is defined as the sensitivity of a bond's price to the changing shape of the yield curve. The yield curve's shape changes continually, and yield curve shifts are rarely parallel. For active bond management, a bond investor may want to base trades on a forecasted yield curve shape or may want to hedge the yield curve risk on a bond portfolio using swaps. Shaping risk also affects the value of many options, which is very important because many fixed-income instruments have embedded options.

Exhibits 9 and 10 show historical yield curve movements for US and European swap rates from March 2006 until March 2020. The exhibits show the considerable swap yield curve changes over time. In both cases, the pre-financial-crisis March 2006 yield curves represent the highest swap yields and those from March 2020 (amid the COVID-19 pandemic-related market turmoil) the lowest. In the United States, however, the end of quantitative easing and tighter monetary policy resulted in a rebound in swap yields prior to 2020, whereas in Europe, yields remained low or negative because of continued accommodative monetary policy. Note that the vertical axis values of the three exhibits differ, and the horizontal axis is not to scale.

Exhibit 9 Historical US Swap Yield Curve Movements

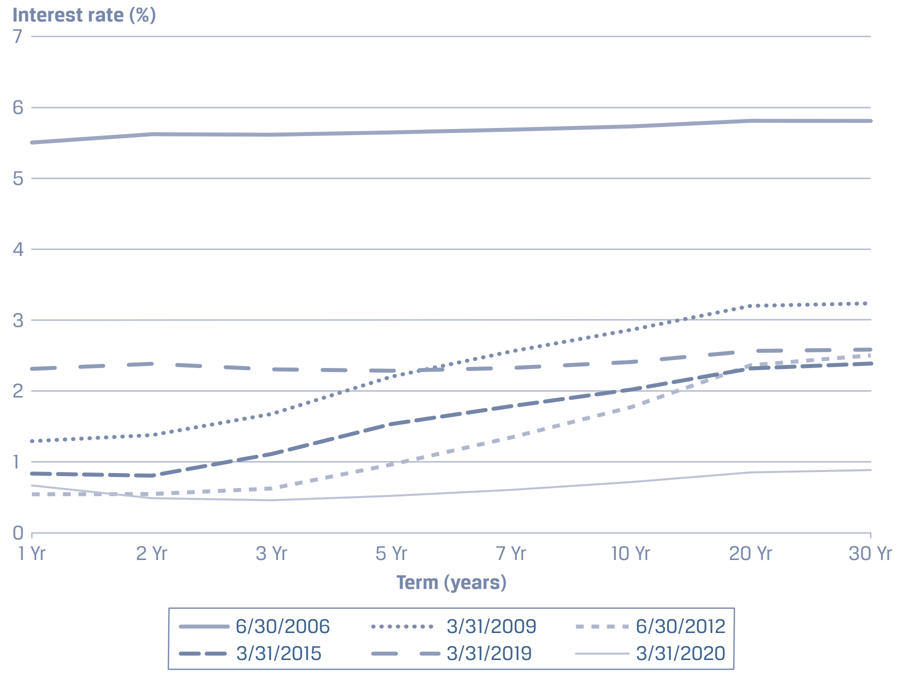
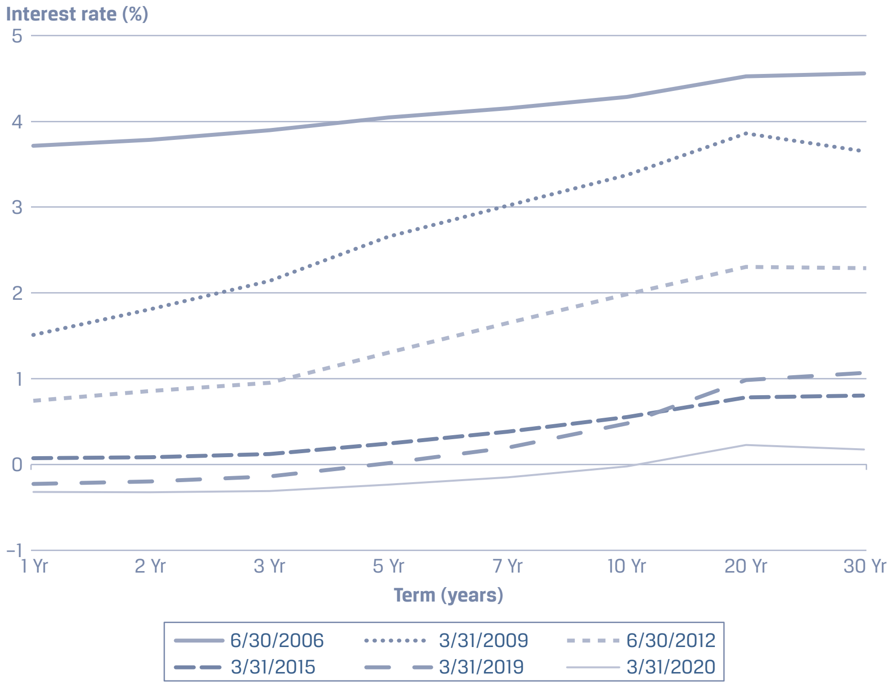


Exhibit 10 Historical European Swap Yield Curve Movements

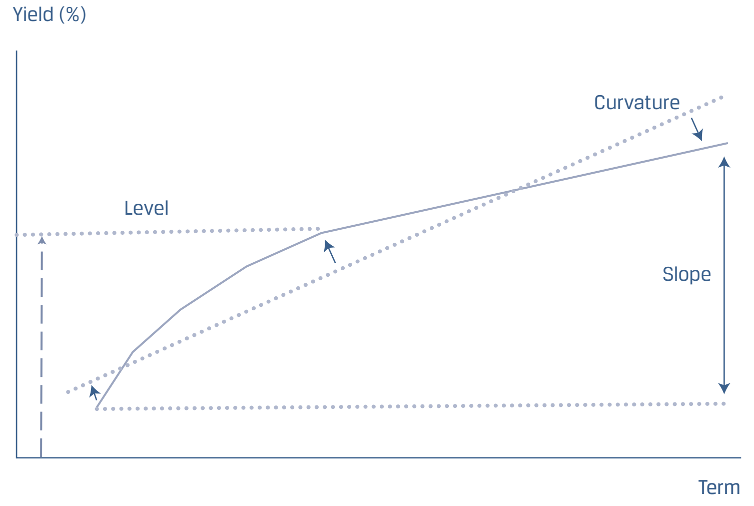


7.2 Factors Affecting the Shape of the Yield Curve

The previous section showed that the yield curve can take nearly any shape. The challenge for a fixed-income manager is to implement a process to manage the yield curve shape risk in her portfolio. One approach is to find a model that reduces most of the possible yield curve movements to a probabilistic combination of a few standardized yield curve movements. This section presents one of the best-known yield curve factor models.

A **yield curve factor model** is defined as a model or a description of yield curve movements that can be considered realistic when compared with historical data. Research has led to models that can describe these movements with some accuracy. One specific yield curve factor model is the three-factor model of Litterman and Scheinkman (1991), who found that yield curve movements are historically well described by a combination of three independent movements, which they interpreted as **level**, **steepness**, and **curvature**. The level movement refers to an upward or downward shift in the yield curve. The steepness movement refers to a non-parallel shift in the yield curve when either short-term rates change more than long-term rates or long-term rates change more than short-term rates. The curvature movement is a reference to movement in three segments of the yield curve: The short-term and long-term segments rise while the middle-term segment falls, or vice versa. Exhibit 11 illustrates these factors.

Exhibit 11 Primary Yield Curve Factors: Level, Slope, and Curvature



In practice, the level movement factor explains most of the total changes in swap and bond market yields. This factor may be interpreted as a reflection of parallel yield curve moves in which rates move in the same direction and by a similar order of magnitude. The steepness factor addresses the shape of the curve, with short-term yields typically moving more than long-term yields. These changes take place over time and therefore explain less of the total variance in rates than the level factor. Finally, the third factor, curvature, tends to have a negative impact on intermediate yields and a positive impact on short- and long-term yields. This variable explaining the "twist" in the yield curve has the smallest impact of the three.

8 THE MATURITY STRUCTURE OF YIELD CURVE VOLATILITIES AND MANAGING YIELD CURVE RISKS

j explain the maturity structure of yield volatilities and their effect on price volatility

8.1 Yield Volatility

Quantifying interest rate volatilities is important for fixed income managers for at least two reasons. First, most fixed-income instruments and derivatives have embedded options. Option values, and hence the values of the fixed-income instrument, crucially depend on the level of interest rate volatilities. Second, fixed-income interest rate risk management is clearly an important part of any management process, and such risk management includes controlling the impact of interest rate volatilities on the instrument's price volatility.

The term structure of interest rate volatilities is a representation of the yield volatility of a zero-coupon bond for every maturity of security. This volatility curve (or "vol") or volatility term structure measures yield curve risk.

Interest rate volatility is not the same for all interest rates along the yield curve. On the basis of the typical assumption of a lognormal model, the uncertainty of an interest rate is measured by the annualized standard deviation of the proportional change in a bond yield over a specified interval. For example, if the interval is a one-month period, then the specified interval equals 1/12 years. This measure, called interest rate volatility, is denoted $\sigma(t,T)$, which is the volatility of the rate for a security with maturity T at time t. The term structure of volatilities is given by Equation 13:

$$\sigma(t,T) = \frac{\sigma\left[\Delta r(t,T)/r(t,T)\right]}{\sqrt{\Delta t}} \tag{13}$$

In Exhibit 12, to illustrate a term structure of volatility, the data series is deliberately chosen to end before the 2008 financial crisis, which was associated with some unusual volatility magnitudes.

Exhibit 12 Historical Example: US Treasuries, August 2005–December 2007

Maturity (years)	0.25	0.50	1	2	3	5	7	10	20	30
$\sigma(t,T)$	0.3515	0.3173	0.2964	0.2713	0.2577	0.2154	0.1885	0.1621	0.1332	0.1169

For example, the 35.15% standard deviation for the three-month T-bill in Exhibit 12 is based on a monthly standard deviation of 0.1015 = 10.15%, which annualizes as

$$0.1015 \div \sqrt{\frac{1}{12}} = 0.3515 = 35.15\%$$

The volatility term structure typically shows that short-term rates are more volatile than long-term rates. That said, long-term bond *prices* tend to vary more than short-term bond prices given the impact of duration. Research indicates that short-term volatility is most strongly linked to uncertainty regarding monetary policy, whereas long-term volatility is most strongly linked to uncertainty regarding the real economy and inflation. Furthermore, most of the co-movement between short-term and long-term volatilities appears to depend on the ever-changing correlations among these three determinants (monetary policy, the real economy, and inflation). During the period of August 2005–December 2007, long-term volatility was lower than short-term volatility, falling from 35.15% for the 0.25-year rate to 11.69% for the 30-year rate.

8.2 Managing Yield Curve Risks Using Key Rate Duration

Yield curve risk—the risk to portfolio value arising from unanticipated changes in the yield curve—can be managed on the basis of several measures of sensitivity to yield curve movements. Management of yield curve risk involves changing the identified exposures to desired values by trades in security or derivative markets (the details fall under the rubric of fixed-income portfolio management and thus are outside the scope of this reading).

One available measure of yield curve sensitivity is effective duration, which measures the sensitivity of a bond's price to a small parallel shift in a benchmark yield curve. Another is based on **key rate duration**, which measures a bond's sensitivity to a small change in a benchmark yield curve at a specific maturity segment. A further measure can be developed on the basis of the factor model developed in Section 6.3. Using one of these last two measures allows identification and management of "shaping risk"—that is, sensitivity to changes in the shape of the benchmark yield curve—in addition to the risk associated with parallel yield curve changes, which is addressed adequately by effective duration.

To make the discussion more concrete, consider a portfolio of 1-year, 5-year, and 10-year zero-coupon bonds with \$100 value in each position; total portfolio value is therefore \$300. Also consider the hypothetical set of factor movements shown in the following table:

Year	1	5	10
Parallel	1	1	1
Steepness	–1	0	1
Curvature	1	0	1

In the table, a parallel movement or shift means that all the rates shift by an equal amount—in this case, by a unit of 1. A steepness movement means that the yield curve steepens with the long rate shifting up by one unit and the short rate shifting down by one unit. A curvature movement means that both the short rate and the long rate shift up by one unit, whereas the medium-term rate remains unchanged. These movements need to be defined, as they are here, such that none of the movements can be a linear combination of the other two movements. Next, we address the calculation of the various yield curve sensitivity measures.

Because the bonds are zero-coupon bonds, each bond's effective duration is the same as its maturity. The portfolio's effective duration is the weighted sum of the effective duration of each bond position; for this equally weighted portfolio, effective duration is 0.333(1 + 5 + 10) = 5.333.

To calculate key rate durations, consider various yield curve movements. First, suppose that the one-year rate changes by 100 bps while the other rates remain the same; the sensitivity of the portfolio to that shift is 1/[(300)(0.01)] = 0.3333. We conclude that the key rate duration of the portfolio ($KeyDur_{Full}$) to the one-year rate, denoted $KeyDur_1$, is 0.3333. Likewise, the key rate durations of the portfolio to the 5-year rate, $KeyDur_5$, and the 10-year rate, $KeyDur_{10}$, are 1.6667 and 3.3333, respectively. Note that the sum of the key rate durations is 5.333, which is the same as the effective duration of the portfolio. This fact can be explained intuitively. Key rate duration measures the portfolio risk exposure to each key rate. If all the key rates move by the same amount, then the yield curve has made a parallel shift, and as a result, the proportional change in value has to be consistent with effective duration. The related model for yield curve risk based on key rate durations ($KeyDur$) is as follows:

$$\begin{aligned} KeyDur_{Full} &= \%\Delta P = \left(\frac{\Delta P}{P}\right) \approx -KeyDur_1 \Delta z_1 - KeyDur_5 \Delta z_5 - KeyDur_{10} \Delta z_{10} \\ &= -0.3333\Delta z_1 - 1.6667\Delta z_5 - -3.3333\Delta z_{10} \end{aligned} \quad \textbf{(14)}$$

Next, we can calculate a measure based on the decomposition of yield curve movements into parallel, steepness, and curvature movements, as described earlier. Define D_L, D_S, and D_C as the sensitivities of portfolio value to small changes in the level, steepness, and curvature factors, respectively. Based on this factor model, Equation 15 shows the proportional change in portfolio value that would result from a small change in the level factor (Δx_L), the steepness factor (Δx_S), and the curvature factor (Δx_C).

$$KeyDur_{Full} = \%\Delta P = \left(\frac{\Delta P}{P}\right) \approx -KeyDur_L \Delta x_L - KeyDur_S \Delta z_S - KeyDur_C \Delta z_C \quad \textbf{(15)}$$

Because $KeyDur_L$ is by definition sensitivity to a parallel shift, the proportional change in the portfolio value per unit shift (the line for a parallel movement in the table) is 5.3333 = (1 + 5 + 10)/[(300)(0.01)]. The sensitivity for steepness movement can be calculated as follows (see the line for steepness movement in the table). When the steepness makes an upward shift of 100 bps, it would result in a downward shift of 100 bps for the 1-year rate, resulting in a gain of \$1, and an upward shift for the 10-year rate, resulting in a loss of \$10. The change in value is therefore (1 – 10). $KeyDur_S$ is the negative of the proportional change in price per unit change in this movement and in this case is 3.0 = −(1 − 10)/[(300)(0.01)]. Considering the line for curvature movement in the table, $KeyDur_C$ = 3.6667 = (1 + 10)/[(300)(0.01)]. Thus, for our hypothetical bond portfolio, we can analyze the portfolio's yield curve risk using the following equation:

$$KeyDur_{Full} = \%\Delta P = \left(\frac{\Delta P}{P}\right) \approx -5.3333\Delta x_L - 3.0\Delta z_S - 3.6667\Delta z_C$$

For example, if Δx_L = −0.0050, Δx_S = 0.002, and Δx_C = 0.001, the predicted change in portfolio value would be +1.7%. It can be shown that key rate durations are directly related to level, steepness, and curvature in this example and that one set of sensitivities can be derived from the other. One can use the numerical example to verify that relation by decomposing changes in the term structure into level, slope, and curvature factors:

$KeyDur_L = KeyDur_1 + KeyDur_5 + KeyDur_{10}$

$KeyDur_S = -\ KeyDur_1 + KeyDur_{10}$

$KeyDur_C = KeyDur_1 + KeyDur_{10}$

Example 10 reviews concepts from this section and the preceding sections.

EXAMPLE 10

Term Structure Dynamics

1 The most important factor in explaining changes in the yield curve has been found to be:

A level.

B curvature.

C steepness.

2 A movement of the yield curve in which the short rate decreases by 150 bps and the long rate decreases by 50 bps would *best* be described as a:

A flattening of the yield curve resulting from changes in level and steepness.

B steepening of the yield curve resulting from changes in level and steepness.

C steepening of the yield curve resulting from changes in steepness and curvature.

3 The yield curve starts off flat, and then intermediate-maturity yields decrease by 10 bps while short- and long-maturity yields remain constant. This movement is *best* described as involving a change in:

A level only.

B curvature only.

C level and curvature.

4 Typically, short-term interest rates:

A are less volatile than long-term interest rates.

B are more volatile than long-term interest rates.

C have about the same volatility as long-term interest rates.

5 Suppose for a given portfolio that key rate changes are considered to be changes in the yield on 1-year, 5-year, and 10-year securities. Estimated key rate durations are $KeyDur_1$ = 0.50, $KeyDur_2$ = 0.70, and $KeyDur_3$ = 0.90. What is the percentage change in the value of the portfolio if a parallel shift in the yield curve results in all yields declining by 50 bps?

A −1.05%.

B +1.05%.

C +2.10%.

Solution to 1:

A is correct. Research shows that upward and downward shifts in the yield curve explain more than 75% of the total change in the yield curve.

Solution to 2:

B is correct. Both the short-term and long-term rates have declined, indicating a change in the level of the yield curve. Short-term rates have declined more than long-term rates, indicating a change in the steepness of the yield curve.

Solution to 3:

B is correct. The curve starts off flat, with identical short, intermediate, and long rates. Both the short-term and long-term rates remained constant, indicating no change in the level of the yield curve. Intermediate rates decreased, however, resulting in curvature.

Solution to 4:

B is correct. A possible explanation is that expectations for long-term inflation and real economic activity affecting longer-term interest rates are slower to change than those related to shorter-term interest rates.

Solution to 5:

B is correct. A decline in interest rates would lead to an increase in bond portfolio value: −0.50(−0.005) − 0.70(−0.005) − 0.90(−0.005) = 0.0105 = 1.05%.

9 DEVELOPING INTEREST RATE VIEWS USING MACROECONOMIC VARIABLES

k explain how key economic factors are used to establish a view on benchmark rates, spreads, and yield curve changes

Interest rate dynamics such as changes in spot versus forward rates and the level, steepness, and curvature of the yield curve are influenced by key economic variables and market events. Implied forward rates serve as market-neutral reference points for fixed income traders. As we illustrated earlier, if today's forward rates are realized in the future, then bond values will simply roll down the yield curve. In practice, active fixed-income market participants establish their own views on future interest rate developments and then position their portfolios in order to capitalize on differences between their own rate view and the market consensus. If their forecast is accurate, the portfolio generates greater returns than it would have otherwise.

This section reviews the key drivers of interest rates before moving on to establishing views and positioning fixed-income portfolios to capitalize on a specific interest rate view.

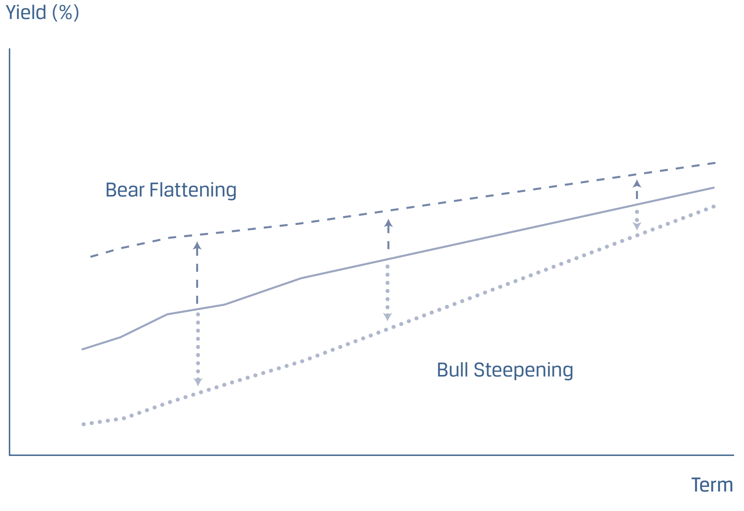
The term **bond risk premium** refers to the expected excess return of a default-free long-term bond less that of an equivalent short-term bond or the one-period risk-free rate. This premium is also referred to as the term (or duration) premium, and it is usually measured using government bonds to capture uncertainty of default-free rates, whereas credit, liquidity, and other risks may increase the overall risk premium for a specific bond. Unlike *ex post* observed historical returns, the bond risk premium is a forward-looking expectation and must be estimated.

Several macroeconomic factors influence bond pricing and required returns such as inflation, economic growth, and monetary policy, among others.

Research shows that although inflation, GDP, and monetary policy explain most of the variance of bond yields, short- and intermediate-term bond yields are driven mostly by inflation, whereas other factors such as monetary policy are key drivers of long-term yields. Inflation explains about two-thirds of short- and intermediate-term bond yield variation, with the remaining third roughly equally attributable to economic growth and factors including monetary policy. In contrast, monetary policy explains nearly two-thirds of long-term yield variation, and the remaining third is largely attributable to inflation.

Monetary policy impacts the bond risk premium. Central banks such as the European Central Bank control the money supply and influence interest rates through policy tools in order to achieve stable prices and sustainable economic growth. During economic expansions, monetary authorities raise benchmark rates to help control inflation. This action is often consistent with **bearish flattening**, or short-term bond yields rising more than long-term bond yields, resulting in a flatter yield curve. During economic recessions or anticipated recessions, the monetary authority cuts benchmark rates to help stimulate economic activity. The lowering of interest rates is associated with **bullish steepening**, in which short-term rates fall by more than long-term yields, resulting in a steeper term structure. These monetary policy actions lead to procyclical short-term interest rate changes. Exhibit 13 shows these two yield curve changes.

Exhibit 13 Examples of Yield Curve Flattening and Steepening



In recent years, central banks have increasingly used their balance sheets for large-sale asset purchases. For example, the Federal Reserve has bought large quantities of US Treasury bonds and mortgage-backed securities. The intended purpose is to stimulate economic activity by increasing the money supply through benchmark bond purchases and driving down the bond risk premium, encouraging capital allocation to incrementally higher-risk assets. Asset purchases impact the term structure by raising demand in a range of maturity segments.

Other factors that influence bond prices, yields, and the bond risk premium include fiscal policy, the maturity structure of debt, and investor demand.

Benchmark government bonds are the means by which nations fund their cumulative (current and past) budget deficits. Greater deficits require more borrowing, which influences both bond supply and required yield. Thus, fiscal supply-side effects affect bond prices and yields by increasing (decreasing) yields when budget deficits rise (fall). In the late 1990s, market participants believed the US government would run fiscal surpluses, leading to a reduction in government bond supply as the Treasury stopped issuing new 30-year bonds for four years. The expected reduction in supply drove long-maturity Treasury yields lower.

Longer government debt maturity structures predict greater excess bond returns. This is effectively a segmented market factor, wherein the greater supply of bonds of long-term maturity increases the yield in that market segment.

Domestic investor demand is a key driver of bond prices, especially among pension funds and insurance companies that use long-dated government bonds to match expected future liabilities. Greater domestic investor demand increases prices and reduces the bond risk premium.

Non-domestic investor demand influences government bond prices and may result either from holding reserves or from actions associated with currency exchange rate management. Non-domestic flows significantly influence bond prices because inflows (outflows) bid up (down) bond prices, lowering (raising) the bond risk premium.

During highly uncertain market periods, investors flock to government bonds in what is termed a **flight to quality**. This term refers to investors' selling off higher-risk asset classes such as stocks and commodities in favor of default-risk-free government bonds. A flight to quality is often associated with **bullish flattening**, in which the yield curve flattens as long-term rates fall by more than short-term rates.

Fixed-income trades based on interest rate forecasts can take a variety of forms, often using bond futures contracts to avoid significant portfolio turnover. Remember that any interest rate view must be evaluated relative to the current short rate and forward curve, because they reflect returns earned by investors rolling down the curve under the current set of implied forward rates.

Investors expecting interest rates to fall will generally extend portfolio duration relative to a benchmark to take advantage of bond price increases from falling rates, whereas investors expecting higher rates will shorten portfolio duration to reduce exposure to falling bond prices.

To capitalize on a steeper curve under which long-term rates rise relative to short-term rates, traders will short long-term bonds and purchase short-term bonds. If on the other hand a trader forecasts curve flattening, whereby short-term rates rise relative to long-term rates, she may capitalize on this trend by purchasing long-term bonds and selling short-term bonds short. In both the expected steepening and flattening trades, the position may be designed as duration neutral in order to insulate from changes in the level of the term structure. Fixed-income investors with long-only investment mandates may alternate between portfolios concentrated in a single maturity, known as a **bullet portfolio**, and those with similar duration that combine short and long maturities, known as a **barbell portfolio**. For example, an investor may seek to capitalize on an expected bullish flattening of the yield curve by shifting from a bullet to a barbell position.

EXAMPLE 11

Building a Rate View Based On Economic Forecasts and Monetary Policy

Morgan Salaz is a fixed income analyst responsible for advising fixed income clients about bond trading opportunities. In the current recessionary environment, the level of government bond yields is low and the term structure is nearly flat. Salaz's firm forecasts that after a brief recession, economic growth will return quickly during the coming 12 months.

1 Which of the following changes to the yield curve is consistent with Salaz's expectation of increasing economic growth over the coming year?

  A Decrease in the level

  B Decrease in the term spread of long-term rates over short-term rates

  C Increase in the term spread of long-term rates over short-term rates

Answer: C is correct. Economic growth forecasts impact long-term rates. The view that economic growth will return to robust levels is consistent with a shift to a positively sloped term structure.

2 Salaz also expects the Federal Reserve to decrease asset purchases of long-term bonds as the economic recovery continues. Which of the following scenarios is consistent with this view? The reduced asset purchases will likely:

  A amplify the effect of increased economic activity on the term spread.

  B dampen the effect of increased economic activity on the term spread.

  C have no effect on the term spread.

Answer: A. Reduced asset purchases constitute a negative shift in demand for longer-term bonds, which raises their yields. The reduced asset purchases of long-maturity bonds would add to the effect of greater economic activity, both of which will increase the term spread.

SUMMARY

- The spot rate for a given maturity can be expressed as a geometric average of the short-term rate and a series of forward rates.
- Forward rates are above (below) spot rates when the spot curve is upward (downward) sloping, whereas forward rates are equal to spot rates when the spot curve is flat.
- If forward rates are realized, then all bonds, regardless of maturity, will have the same one-period realized return, which is the first-period spot rate.
- If the spot rate curve is upward sloping and is unchanged, then each bond "rolls down" the curve and earns the forward rate that rolls out of its pricing (i.e., an N-period zero-coupon bond earns the N-period forward rate as it rolls down to be a $N – 1$ period security). This dynamic implies an expected return in excess of short-maturity bonds (i.e., a **term premium**) for longer-maturity bonds if the yield curve is upward sloping.

- Active bond portfolio management is consistent with the expectation that today's forward curve does not accurately reflect future spot rates.
- The swap curve provides another measure of the time value of money.
- Swaps are an essential tool frequently used by investors to hedge, take a position in, or otherwise modify interest rate risk.
- Bond quote conventions often use measures of spreads. Those quoted spreads can be used to determine a bond's price.
- Swap curves and Treasury curves can differ because of differences in their credit exposures, liquidity, and other supply/demand factors.
- Market participants often use interest rate spreads between short-term government and risky rates as a barometer to evaluate relative credit and liquidity risk.
- The local expectations theory, liquidity preference theory, segmented markets theory, and preferred habitat theory provide traditional explanations for the shape of the yield curve.
- Historical yield curve movements suggest that they can be explained by a linear combination of three principal movements: level, steepness, and curvature.
- The volatility term structure can be measured using historical data and depicts yield curve risk.
- The sensitivity of a bond value to yield curve changes may make use of effective duration, key rate durations, or sensitivities to parallel, steepness, and curvature movements. Using key rate durations or sensitivities to parallel, steepness, and curvature movements allows one to measure and manage shaping risk.
- The term bond risk premium refers to the expected excess return of a default-free long-term bond less that of an equivalent short-term bond or the one-period risk-free rate
- Several macroeconomic factors influence bond pricing and required returns such as inflation, economic growth, and monetary policy, among others.
- During highly uncertain market periods, investors flock to government bonds in a flight to quality that is often associated with bullish flattening, in which long-term rates fall by more than short-term rates.
- Investors expecting rates to fall will generally extend (shorten) portfolio duration to take advantage of expected bond price increases (decreases)
- When investors expect a steeper (flatter) curve under which long-term rates rise (fall) relative to short-term rates, they will sell (buy) long-term bonds and purchase (sell) short-term bonds.

REFERENCES

Boyarchenko, Nina, Pooja Gupta, Nick Steele, and Jacqueline Yen.October 2018. “Negative Swap Spreads.” *Federal Reserve Bank of New York Economic Policy Review* 24 (2): 1–14.

Haubrich, Joseph G. 2006. “Does the Yield Curve Signal Recession?” Federal Reserve Bank of Cleveland (15 April).

Litterman, Robert, and José Scheinkman. 1991. “Common Factors Affecting Bond Returns.” Journal of Fixed Income 1 (1): 54–61. doi:10.3905/jfi.1991.692347

PRACTICE PROBLEMS

1 Given spot rates for one-, two-, and three-year zero coupon bonds, how many forward rates can be calculated?

2 Give two interpretations for the following forward rate: The two-year forward rate one year from now is 2%.

3 Describe the relationship between forward rates and spot rates if the yield curve is flat.

4 A Define the yield-to-maturity for a coupon bond.

B Is it possible for a coupon bond to earn less than the yield-to-maturity if held to maturity?

5 If a bond trader believes that current forward rates overstate future spot rates, how might she profit from that conclusion?

6 Explain the strategy of rolling down the yield curve.

7 What are the advantages of using the swap curve as a benchmark of interest rates relative to a government bond yield curve?

8 What is the TED spread, and what type of risk does it measure?

9 What is the SOFR rate, and which market conditions does it reflect?

10 According to the local expectations theory, what would be the difference in the one-month total return if an investor purchased a five-year zero-coupon bond versus a two-year zero-coupon bond?

11 Compare the segmented market and the preferred habitat term structure theories.

12 A List the three factors that have empirically been observed to affect Treasury security returns and explain how each of these factors affects returns on Treasury securities.

B What has been observed to be the most important factor in affecting Treasury returns?

C Which measures of yield curve risk can measure shaping risk?

13 Which forward rate cannot be computed from the one-, two-, three-, and four-year spot rates? The rate for a:

A one-year loan beginning in two years

B two-year loan beginning in two years

C three-year loan beginning in two years

14 Consider spot rates for three zero-coupon bonds: z(1) = 3%, z(2) = 4%, and z(3) = 5%. Which statement is correct? The forward rate for a one-year loan beginning in one year will be:

A less than the forward rate for a one-year loan beginning in two years.

B greater than the forward rate for a two-year loan beginning in one year.

C greater than the forward rate for a one-year loan beginning in two years.

15 If one-period forward rates are decreasing with maturity, the yield curve is *most likely*:

A flat.

© 2020 CFA Institute. All rights reserved.

B upward sloping.

C downward sloping.

The following information relates to Questions 16–19

A one-year zero-coupon bond yields 4.0%. The two- and three-year zero-coupon bonds yield 5.0% and 6.0%, respectively.

16 The rate for a one-year loan beginning in one year is *closest* to:

A 4.5%.

B 5.0%.

C 6.0%.

17 The forward rate for a two-year loan beginning in one year is *closest* to:

A 5.0%.

B 6.0%.

C 7.0%.

18 The forward rate for a one-year loan beginning in two years is *closest* to:

A 6.0%.

B 7.0%.

C 8.0%.

19 The five-year spot rate is not provided here; however, the forward price for a two-year zero-coupon bond beginning in three years is known to be 0.8479. The price today of a five-year zero-coupon bond is *closest* to:

A 0.7119.

B 0.7835.

C 0.9524.

20 The one-year spot rate z_1 is 4%, the forward rate for a one-year loan beginning in one year is 6%, and the forward rate for a one-year loan beginning in two years is 8%. Which of the following rates is *closest* to the three-year spot rate?

A 4.0%

B 6.0%

C 8.0%

21 The one-year spot rate z_1 is 5%, and the forward price for a one-year zero-coupon bond beginning in one year is 0.9346. The spot price of a two-year zero-coupon bond is *closest* to:

A 0.87.

B 0.89.

C 0.93.

22 In a typical interest rate swap contract, the swap rate is *best* described as the interest rate for the:

A fixed-rate leg of the swap.

B floating-rate leg of the swap.

C difference between the fixed and floating legs of the swap.

23 A two-year fixed-for-floating MRR swap is 1.00%, and the two-year US Treasury bond is yielding 0.63%. The swap spread is *closest* to:

A 37 bps.

B 100 bps.

C 163 bps.

24 The swap spread is quoted as 50 bps. If the five-year US Treasury bond is yielding 2%, the rate paid by the fixed payer in a five-year interest rate swap is *closest* to:

A 0.50%.

B 1.50%.

C 2.50%.

25 If the three-month T-bill rate drops and Libor remains the same, the relevant TED spread:

A increases.

B decreases.

C does not change.

26 Given the yield curve for US Treasury zero-coupon bonds, which spread is *most* helpful pricing a corporate bond? The:

A Z-spread.

B TED spread.

C Libor–OIS spread.

The following information relates to Questions 27–33

Jane Nguyen is a senior bond trader for an investment bank, and Chris Alexander is a junior bond trader at the bank. Nguyen is responsible for her own trading activities and also for providing assignments to Alexander that will develop his skills and create profitable trade ideas. Exhibit 1 presents the current par and spot rates.

Exhibit 1 Current Par and Spot Rates

Maturity	Par Rate	Spot Rate
One year	2.50%	2.50%
Two years	2.99%	3.00%
Three years	3.48%	3.50%
Four years	3.95%	4.00%
Five years	4.37%	

Note: Par and spot rates are based on annual-coupon sovereign bonds.

Nguyen gives Alexander two assignments that involve researching various questions:

Assignment 1 What is the yield-to-maturity of the option-free, default-risk-free bond presented in Exhibit 2? Assume that the bond is held to maturity, and use the rates shown in Exhibit 1.

Exhibit 2 Selected Data for $1,000 Par Bond

Bond Name	Maturity (T)	Coupon
Bond Z	Three years	6.00%

Note: Terms are today for a T-year loan.

Assignment 2 Assuming that the projected spot curve two years from today will be below the current forward curve, is Bond Z fairly valued, undervalued, or overvalued?

After completing his assignments, Alexander asks about Nguyen's current trading activities. Nguyen states that she has a two-year investment horizon and will purchase Bond Z as part of a strategy to ride the yield curve. Exhibit 1 shows Nguyen's yield curve assumptions implied by the spot rates.

27 Based on Exhibit 1, the five-year spot rate is *closest to*:

A 4.40%.

B 4.45%.

C 4.50%.

28 Based on Exhibit 1, the market is *most likely* expecting:

A deflation.

B inflation.

C no risk premiums.

29 Based on Exhibit 1, the forward rate of a one-year loan beginning in three years is *closest to*:

A 4.17%.

B 4.50%.

C 5.51%.

30 Based on Exhibit 1, which of the following forward rates can be computed?

A A one-year loan beginning in five years

B A three-year loan beginning in three years

C A four-year loan beginning in one year

31 For Assignment 1, the yield-to-maturity for Bond Z is *closest* to the:

A one-year spot rate.

B two-year spot rate.

C three-year spot rate.

32 For Assignment 2, Alexander should conclude that Bond Z is currently:

A undervalued.

B fairly valued.

C overvalued.

33 By choosing to buy Bond Z, Nguyen is *most likely* making which of the following assumptions?

A Bond Z will be held to maturity.

B The three-year forward curve is above the spot curve.

C Future spot rates do not accurately reflect future inflation.

The following information relates to Questions 34–38

Laura Mathews recently hired Robert Smith, an investment adviser at Shire Gate Advisers, to assist her in investing. Mathews states that her investment time horizon is short, approximately two years or less. Smith gathers information on spot rates for on-the-run annual-coupon government securities and swap spreads, as presented in Exhibit 1. Shire Gate Advisers recently published a report for its clients stating its belief that, based on the weakness in the financial markets, interest rates will remain stable, the yield curve will not change its level or shape for the next two years, and swap spreads will also remain unchanged.

Exhibit 1 Government Spot Rates and Swap Spreads

	Maturity (years)			
	1	2	3	4
Government spot rate	2.25%	2.70%	3.30%	4.05%
Swap spread	0.25%	0.30%	0.45%	0.70%

Smith decides to examine the following three investment options for Mathews:

Investment 1: Buy a government security that would have an annualized return that is nearly risk free. Smith is considering two possible implementations: a two-year investment or a combination of two one-year investments.

Investment 2: Buy a four-year, zero-coupon corporate bond and then sell it after two years. Smith illustrates the returns from this strategy using the swap rate as a proxy for corporate yields.

Investment 3: Buy a lower-quality, two-year corporate bond with a coupon rate of 4.15% and a Z-spread of 65 bps.

When Smith meets with Mathews to present these choices, Mathews tells him that she is somewhat confused by the various spread measures. She is curious to know whether there is one spread measure that could be used as a good indicator of the risk and liquidity of money market securities during the recent past.

34 In his presentation of Investment 1, Smith could show that under the no-arbitrage principle, the forward price of a one-year government bond to be issued in one year is *closest* to:

A 0.9662.

B 0.9694.

C 0.9780.

35 In presenting Investment 1, using Shire Gate Advisers' interest rate outlook, Smith could show that riding the yield curve provides a total return that is *most likely*:

A lower than the return on a maturity-matching strategy.

B equal to the return on a maturity-matching strategy.

C higher than the return on a maturity-matching strategy.

36 In presenting Investment 2, Smith should show an annual return *closest* to:

A 4.31%.

B 5.42%.

C 6.53%.

37 The bond in Investment 3 is *most likely* trading at a price of:

A 100.97.

B 101.54.

C 104.09.

38 The *most* appropriate response to Mathews question regarding a spread measure is the:

A Z-spread.

B TED spread.

C Libor–OIS spread.

The following information relates to Questions 39–42

Rowan Madison is a junior analyst at Cardinal Capital. Sage Winter, a senior portfolio manager and Madison's supervisor, meets with Madison to discuss interest rates and review two bond positions in the firm's fixed-income portfolio.

Winter begins the meeting by asking Madison to state her views on the term structure of interest rates. Madison responds:

> "Yields are a reflection of expected spot rates and risk premiums. Investors demand risk premiums for holding long-term bonds, and these risk premiums increase with maturity."

Winter tells Madison that, based on recent changes in spreads, she is concerned about a perceived increase in counterparty risk in the economy and its effect on the portfolio. Madison asks Winter:

> "Which spread measure should we use to assess changes in counterparty risk in the economy?"

Winter is also worried about the effect of yield volatility on the portfolio. She asks Madison to identify the economic factors that affect short-term and long-term rate volatility. Madison responds:

> "Short-term rate volatility is mostly linked to uncertainty regarding monetary policy, whereas long-term rate volatility is mostly linked to uncertainty regarding the real economy and inflation."

Finally, Winter asks Madison to analyze the interest rate risk portfolio positions in a 5-year and a 20-year bond. Winter requests that the analysis be based on level, slope, and curvature as term structure factors. Madison presents her analysis in Exhibit 1.

Exhibit 1 Three-Factor Model of Term Structure

	Time to Maturity (years)	
Factor	**5**	**20**
Level	–0.4352%	–0.5128%
Steepness	–0.0515%	–0.3015%
Curvature	0.3963%	0.5227%

Note: Entries indicate how yields would change for a one standard deviation increase in a factor.

Winter asks Madison to perform two analyses:

Analysis 1: Calculate the expected change in yield on the 20-year bond resulting from a two-standard-deviation increase in the steepness factor.

Analysis 2: Calculate the expected change in yield on the five-year bond resulting from a one-standard-deviation decrease in the level factor and a one-standard-deviation decrease in the curvature factor.

39 Madison's views on the term structure of interest rates are *most* consistent with the:

A local expectations theory.

B segmented markets theory.

C liquidity preference theory.

40 Is Madison's response regarding the factors that affect short-term and long-term rate volatility correct?

A Yes

B No, she is incorrect regarding factors linked to long-term rate volatility

C No, she is incorrect regarding factors linked to short-term rate volatility

41 Based on Exhibit 1, the results of Analysis 1 should show the yield on the 20-year bond decreasing by:

A 0.3015%.

B 0.6030%.

C 0.8946%.

42 Based on Exhibit 1, the results of Analysis 2 should show the yield on the five-year bond:

A decreasing by 0.8315%.

B decreasing by 0.0389%.

C increasing by 0.0389%.

The following information relates to Questions 43–50

Liz Tyo is a fund manager for an actively managed global fixed-income fund that buys bonds issued in Countries A, B, and C. She and her assistant are preparing the quarterly markets update. Tyo begins the meeting by distributing the daily rates sheet, which includes the current government spot rates for Countries A, B, and C as shown in Exhibit 1.

Exhibit 1 Today's Government Spot Rates

Maturity	Country A	Country B	Country C
One year	0.40%	–0.22%	14.00%
Two years	0.70	–0.20	12.40
Three years	1.00	–0.12	11.80
Four years	1.30	–0.02	11.00
Five years	1.50	0.13	10.70

Tyo asks her assistant how these spot rates were obtained. The assistant replies, "Spot rates are determined through the process of bootstrapping. It entails backward substitution using par yields to solve for zero-coupon rates one by one, in order from latest to earliest maturities."

Tyo then provides a review of the fund's performance during the last year and comments, "The choice of an appropriate benchmark depends on the country's characteristics. For example, although Countries A and B have both an active government bond market and a swap market, Country C's private sector is much bigger than its public sector, and its government bond market lacks liquidity."

Tyo further points out, "The fund's results were mixed; returns did not benefit from taking on additional risk. We are especially monitoring the riskiness of the corporate bond holdings. For example, our largest holdings consist of three four-year corporate bonds (Bonds 1, 2, and 3) with identical maturities, coupon rates, and other contract terms. These bonds have Z-spreads of 0.55%, 1.52%, and 1.76%, respectively."

Tyo continues, "We also look at risk in terms of the swap spread. We considered historical three-year swap spreads for Country B, which reflect that market's credit and liquidity risks, at three different points in time." Tyo provides the information in Exhibit 2.

Exhibit 2 Selected Historical Three-Year Rates for Country B

Period	Government Bond Yield (%)	Fixed-for-Floating Libor Swap (%)
1 month ago	–0.10	0.16
6 months ago	–0.08	0.01
12 months ago	–0.07	0.71

Tyo then suggests that the firm was able to add return by riding the yield curve. The fund plans to continue to use this strategy but only in markets with an attractive yield curve for this strategy.

She moves on to present her market views on the respective yield curves for a five-year investment horizon.

> **Country A:** "The government yield curve has changed little in terms of its level and shape during the last few years, and I expect this trend to continue. We assume that future spot rates reflect the current forward curve for all maturities."
>
> **Country B:** "Because of recent economic trends, I expect a reversal in the slope of the current yield curve. We assume that future spot rates will be higher than current forward rates for all maturities."
>
> **Country C:** "To improve liquidity, Country C's central bank is expected to intervene, leading to a reversal in the slope of the existing yield curve. We assume that future spot rates will be lower than today's forward rates for all maturities."

Tyo's assistant asks, "Assuming investors require liquidity premiums, how can a yield curve slope downward? What does this imply about forward rates?"

Tyo answers, "Even if investors require compensation for holding longer-term bonds, the yield curve can slope downward—for example, if there is an expectation of severe deflation. Regarding forward rates, it can be helpful to understand yield curve dynamics by calculating implied forward rates. To see what I mean, we can use Exhibit 1 to calculate the forward rate for a two-year Country C loan beginning in three years."

43 Did Tyo's assistant accurately describe the process of bootstrapping?

- **A** Yes
- **B** No, with respect to par yields
- **C** No, with respect to backward substitution

44 The swap curve is a better benchmark than the government spot curve for:

- **A** Country A.
- **B** Country B.
- **C** Country C.

45 Based on Exhibit 2, the implied credit and liquidity risks as indicated by the historical three-year swap spreads for Country B were the lowest:

- **A** 1 month ago.
- **B** 6 months ago.
- **C** 12 months ago.

46 Based on Exhibit 1 and Tyo's expectations, which country's term structure is currently best for traders seeking to ride the yield curve?

- **A** Country A
- **B** Country B
- **C** Country C

47 Based on Exhibit 1 and assuming Tyo's market views on yield curve changes are realized, the forward curve of which country will lie below its spot curve?

- **A** Country A
- **B** Country B
- **C** Country C

48 Based on Exhibit 1 and Tyo's expectations for the yield curves, Tyo *most likely* perceives the bonds of which country to be fairly valued?

A Country A

B Country B

C Country C

49 With respect to their discussion of yield curves, Tyo and her assistant are *most likely* discussing which term structure theory?

A Pure expectations theory

B Local expectations theory

C Liquidity preference theory

50 Tyo's assistant should calculate a forward rate *closest* to:

A 9.07%.

B 9.58%.

C 9.97%.

51 During economic expansions, monetary authorities raise benchmark rates to help control inflation. This action is *most* often consistent with:

A bearish flattening.

B bullish steepening.

C bearish steepening.

52 When government budget deficits fall, fiscal supply-side effects are *most* likely to result in:

A higher bond yields.

B a steeper yield curve.

C lower bond yields.

53 A flight to quality is most often associated with:

A a general rise in the level of interest rates.

B bullish flattening.

C bearish flattening.

SOLUTIONS

1 Three forward rates can be calculated from the one-, two- and three-year spot rates. The rate on a one-year loan that begins at the end of Year 1 can be calculated using the one- and two-year spot rates; in the following equation, one would solve for $f_{1,1}$:

$$[1 + z_2]^2 = [1 + z_1]^1[1 + f_{1,1}]^1$$

The rate on a one-year loan that starts at the end of Year 2 can be calculated from the two- and three-year spot rates. In the following equation, one would solve for $f_{2,1}$:

$$[1 + z_3]^3 = [1 + z_2]^2[1 + f_{2,1}]^1$$

Additionally, the rate on a two-year loan that begins at the end of Year 1 can be computed from the one- and three-year spot rates. In the following equation, one would solve for $f_{1,2}$:

$$[1 + z_3]^3 = [1 + z_1]^1[1 + f_{1,2}]^2$$

2 For the two-year forward rate one year from now of 2%, the two interpretations are as follows:

- 2% is the rate that will make an investor indifferent between buying a three-year zero-coupon bond or investing in a one-year zero-coupon bond and, when it matures, reinvesting in a zero-coupon bond that matures in two years.
- 2% is the rate that can be locked in today by buying a three-year zero-coupon bond rather than investing in a one-year zero-coupon bond and, when it matures, reinvesting in a zero-coupon bond that matures in two years.

3 A flat yield curve implies that all spot interest rates are the same. When the spot rate is the same for every maturity, successive applications of the forward rate model will show that all the forward rates will also be the same and equal to the spot rate.

4 **A** The yield-to-maturity of a coupon bond is the expected rate of return on a bond if the bond is held to maturity, there is no default, and the bond and all coupons are reinvested at the original yield-to-maturity.

B Yes, it is possible. For example, if reinvestment rates for the future coupons are lower than the initial yield-to-maturity, a bondholder may experience lower realized returns.

5 If forward rates are higher than expected future spot rates, the market price of the bond will be lower than the intrinsic value. This dynamic occurs because, everything else held constant, the market is currently discounting the bonds cash flows at a higher rate than the investor's expected future spot rates. The investor can capitalize on this scenario by purchasing the undervalued bond. If expected future spot rates are realized, then bond prices should rise, thus generating gains for the investor.

6 The strategy of rolling down the yield curve is one in which a bond trader attempts to generate a total return over a given investment horizon that exceeds the return to bond with maturity matched to the horizon. The strategy involves buying a bond with maturity more distant than the investment horizon. Assuming an upward-sloping yield curve, if the yield curve does not change

level or shape, as the bond approaches maturity (or rolls down the yield curve) it will be priced at successively lower yields. So as long as the bond is held for a period less than maturity, it should generate higher returns because of price gains.

7 Some countries do not have active government bond markets with trading at all maturities. For those countries without a liquid government bond market but with an active swap market, there are typically more points available to construct a swap curve than a government bond yield curve. For those markets, the swap curve may be a superior benchmark.

8 The TED spread is the difference between Libor and the US T-bill rate of matching maturity. It is an indicator of perceived credit and liquidity risk. In particular, because sovereign debt instruments are typically the benchmark for the lowest default risk instruments in a given market, and loans between banks (often at Libor) have some counterparty risk, the TED spread is considered to at least in part reflect default (or counterparty) risk in the banking sector.

9 The secured overnight financing rate (SOFR), or overnight cash borrowing rate collateralized by US Treasuries, is a barometer of the US Treasury repurchase (or repo) market. SOFR is a volume-weighted index of all qualified repo market transactions on a given day and is influenced by supply and demand conditions in secured funding markets.

10 The local expectations theory asserts that the total return over a one-month horizon for a five-year zero-coupon bond would be the same as for a two-year zero-coupon bond.

11 Both theories attempt to explain the shape of any yield curve in terms of supply and demand for bonds. In segmented market theory, bond market participants are limited to purchase of maturities that match the timing of their liabilities. In the preferred habitat theory, participants have a preferred maturity for asset purchases, but they may deviate from it if they feel returns in other maturities offer sufficient compensation for leaving their preferred maturity segment.

12 **A** Studies have shown that three factors affect Treasury returns: (1) changes in the level of the yield curve, (2) changes in the slope of the yield curve, and (3) changes in the curvature of the yield curve. Changes in the level refer to upward or downward shifts in the yield curve. For example, an upward shift in the yield curve is likely to result in lower returns across all maturities. Changes in the slope of the yield curve relate to the steepness of the yield curve. Thus, if the yield curve steepens, higher returns for short-maturity bonds and lower returns for long-maturity bonds will likely occur. An example of a change in the curvature of the yield curve is a situation where rates fall at the short and long end of the yield curve while rising for intermediate maturities. In this situation, returns on short and long maturities are likely to rise while declining for intermediate-maturity bonds.

B Empirically, the most important factor is the change in the level of interest rates.

C Key rate durations and a measure based on sensitivities to level, slope, and curvature movements can address shaping risk, but effective duration cannot.

13 C is correct. There is no spot rate information to provide rates for a loan that terminates in five years. That is $f_{2,3}$ is calculated as follows:

$$f_{2,3} = \sqrt[3]{\frac{[1 + z_5]^5}{[1 + z_2]^2}} - 1$$

This equation indicates that in order to calculate the rate for a three-year loan beginning at the end of two years, one needs the five-year spot rate, z_5, and the two-year spot rate, z_2. However, z_5 is not provided.

14 A is correct. The forward rate for a one-year loan beginning in one year, $f_{1,1}$, is $1.04^2/1.03 - 1 = 5\%$. The rate for a one-year loan beginning in two years, $f_{2,1}$, is $1.05^3/1.04^2 - 1 = 7\%$. This confirms that an upward-sloping yield curve is consistent with an upward-sloping forward curve.

15 C is correct. If one-period forward rates are decreasing with maturity, then the forward curve is downward sloping. This turn implies a downward-sloping yield curve where longer-term spot rates z_{B-A} are less than shorter-term spot rates z_A.

16 C is correct. From the forward rate model, we have

$$[1 + z_2]^2 = [1 + z_1]^1[1 + f_{1,1}]^1$$

Using the one- and two-year spot rates, we have

$$(1 + 0.05)^2 = (1 + 0.04)^1[1 + f_{1,1}]^1\text{, so } \frac{(1 + 0.05)^2}{(1 + 0.04)^1} - 1 = f_{1,1} = 6.010\%.$$

17 C is correct. From the forward rate model,

$$[1 + z_3]^3 = [1 + z_1]^1[1 + f_{1,2}]^2$$

Using the one- and three-year spot rates, we find

$$(1 + 0.06)^3 = (1 + 0.04)^1[1 + f_{1,2}]^2\text{, so } \sqrt{\frac{(1 + 0.06)^3}{(1 + 0.04)^1}} - 1 = f_{1,2} = 7.014\%.$$

18 C is correct. From the forward rate model,

$$[1 + z_3]^3 = [1 + z_2]^2[1 + f_{2,1}]^1$$

Using the two- and three-year spot rates, we find

$$(1 + 0.06)^3 = (1 + 0.05)^2[1 + f_{2,1}]^1\text{, so } \frac{(1 + 0.06)^3}{(1 + 0.05)^2} - 1 = f_{2,1} = 8.029\%.$$

19 A is correct. We can convert spot rates to spot prices to find $DF_3 = \dfrac{1}{(1.06)^3} =$ 0.8396. The forward pricing model can be used to find the price of the five-year zero as $DF_B = DF_A \times F_{A,B-A}$, so $DF_5 = DF_3F_{3,2} = 0.8396 \times 0.8479 = 0.7119$.

20 B is correct. Applying the forward rate model, we find

$$[1 + z_3]^3 = [1 + z_1]^1[1 + f_{1,1}]^1[1 + f_{2,1}]^1$$

So $[1 + z_3]^3 = (1 + 0.04)^1(1 + 0.06)^1(1 + 0.08)^1$, $\sqrt[3]{1.1906} - 1 = z_3 = 5.987\%$.

21 B is correct. We can convert spot rates to spot prices and use the forward pricing model, so we have $DF_1 = \dfrac{1}{(1.05)^1} = 0.9524$. The forward pricing model is $DF_B = DF_A \times F_{A,B-A}$, so $DF_2 = DF_1F_{1,1} = 0.9524 \times 0.9346 = 0.8901$.

22 A is correct. The swap rate is the interest rate for the fixed-rate leg of an interest rate swap.

23 A is correct. The swap spread = 1.00% – 0.63% = 0.37%, or 37 bps.

24 C is correct. The fixed leg of the five-year fixed-for-floating swap will be equal to the five-year Treasury rate plus the swap spread: 2.0% + 0.5% = 2.5%.

25 A is correct. The TED spread is the difference between the three-month Libor and the three-month Treasury bill rate. If the T-bill rate falls and Libor does not change, the TED spread will increase.

26 A is correct. The Z spread is the single rate that, when added to the rates of the spot yield curve, will provide the correct discount rates to price a particular risky bond.

27 B is correct. The five-year spot rate is determined by using forward substitution and using the known values of the one-year, two-year, three-year, and four-year spot rates, as follows:

$$1 = \frac{0.0437}{1.025} + \frac{0.0437}{(1.03)^2} + \frac{0.0437}{(1.035)^3} + \frac{0.0437}{(1.04)^4} + \frac{1 + 0.0437}{(1 + z_5)^5}$$

$$z_5 = \sqrt[5]{\frac{1.0437}{0.8394}} - 1 = 4.453\%$$

28 B is correct. The spot rates imply an upward-sloping yield curve, $z_3 > z_2 > z_1$. Because nominal yields incorporate a premium for expected inflation, an upward-sloping yield curve is generally interpreted as reflecting a market expectation of increasing, or at least level, future inflation (associated with relatively strong economic growth).

29 C is correct. A one-year loan beginning in three years, or $f_{3,1}$, is calculated as follows:

$$\left[1 + z_{3+1}\right]^{3+1} = \left[1 + z_3\right]^3 \left[1 + f_{3,1}\right]^1$$

$$\left[1.04\right]^4 = \left[1.035\right]^3 \left[1 + f_{3,1}\right]$$

$$f_{3,1} = \frac{(1.04)^4}{(1.035)^3} - 1 = 5.514\%$$

30 C is correct. Exhibit 1 provides five years of par rates, from which the spot rates for z_1, z_2, z_3, z_4, and z_5 can be derived. Thus the forward rate $f_{1,4}$ can be calculated as follows:

$$f_{1,4} = \sqrt[4]{\frac{\left[1 + z_5\right]^5}{\left[1 + z_1\right]^1} - 1}$$

31 C is correct. The yield-to-maturity, y_3, of Bond Z should be a weighted average of the spot rates used in the valuation of the bond. Because the bond's largest cash flow occurs in Year 3, z_3 will have a greater weight than z_1 and z_2 in determining y_3.

Using the spot rates:

$$\text{Price} = \frac{\$60}{(1.025)^1} + \frac{\$60}{(1.030)^2} + \frac{\$1{,}060}{(1.035)^3} = \$1{,}071.16$$

Using the yield-to-maturity:

$$\text{Price} = \frac{\$60}{\left[1 + y(3)\right]^1} + \frac{\$60}{\left[1 + y(3)\right]^2} + \frac{\$1{,}060}{\left[1 + y(3)\right]^3} = \$1{,}071.16$$

The computed result is y_3 = 3.46%, which is closest to the three-year spot rate of 3.50%.

32 A is correct. Alexander projects that the spot curve two years from today will be below the current forward curve, which implies that her expected future spot rates beyond two years will be lower than the quoted forward rates. Alexander would perceive Bond Z to be undervalued in the sense that the market is effectively discounting the bond's payments at a higher rate than she would, and the bond's market price is below her estimate of intrinsic value.

33 B is correct. Nguyen's strategy is to ride the yield curve, which is appropriate when the yield curve is upward sloping. The yield curve implied by Exhibit 1 is upward sloping, which implies that the three-year forward curve is above the current spot curve. When the yield curve slopes upward, as a bond approaches maturity or "rolls down the yield curve," the bond is valued at successively lower yields and higher prices.

34 B is correct. The forward pricing model is based on the no-arbitrage principle and is used to calculate a bond's forward price based on the spot yield curve. The spot curve is constructed by using annualized rates from option-free and default-risk-free zero-coupon bonds.

Equation 2: $DF_B = DF_A \times F_{A,B-A}$; we need to solve for $F_{1,1}$.

$DF_1 = 1/(1 + 0.0225)^1$ and $DF_2 = 1/(1 + 0.0270)^2$,

$F_{1,1} = DF_2/DF_1 = 0.9481/0.9780 = 0.9694$.

35 C is correct. When the spot curve is upward sloping and its level and shape are expected to remain constant over an investment horizon (Shire Gate Advisers' view), buying bonds with a maturity longer than the investment horizon (i.e., riding the yield curve) will provide a total return greater than the return on a maturity-matching strategy.

36 C is correct. The swap spread is a common way to indicate credit spreads in a market. The four-year swap rate (fixed leg of an interest rate swap) can be used as an indication of the four-year corporate yield. Riding the yield curve by purchasing a four-year zero-coupon bond with a yield of 4.75% {i.e., 4.05% + 0.70%, [$P_4 = 100/(1 + 0.0475)^4 = 83.058$]} and then selling it when it becomes a two-year zero-coupon bond with a yield of 3.00% {i.e., 2.70% + 0.30%, [$P_2 = 100/(1 + 0.0300)^2 = 94.260$]} produces an annual return of 6.53%: $(94.260/83.058)^{0.5} - 1.0 = 0.0653$.

37 B is correct. The Z-spread is the constant basis point spread that is added to the default-free spot curve to price a risky bond. A Z-spread of 65 bps for a particular bond would imply adding a fixed spread of 65 bps to maturities along the spot curve to correctly price the bond. Therefore, for the two-year bond, z_1 = 2.90% (i.e., 2.25% + 0.65%), z_2 = 3.35% (i.e., 2.70% + 0.65%), and the price of the bond with an annual coupon of 4.15% is as follows:

$P = 4.15/(1 +0.\ 029)^1 + 4.15/(1 + 0.0335)^2 + 100/(1 + 0.0335)^2$,

$P = 101.54$.

38 C is correct. The Libor–OIS spread is considered an indicator of the risk and liquidity of money market securities. This spread measures the difference between Libor and the OIS rate.

39 C is correct. Liquidity preference theory asserts that investors demand a risk premium, in the form of a liquidity premium, to compensate them for the added interest rate risk they face when buying long-maturity bonds. The theory also states that the liquidity premium increases with maturity.

40 A is correct. Madison's response is correct; research indicates that short-term rate volatility is mostly linked to uncertainty regarding monetary policy, whereas long-term rate volatility is mostly linked to uncertainty regarding the real economy and inflation.

41 B is correct. Because the factors in Exhibit 1 have been standardized to have unit standard deviations, a two-standard-deviation increase in the steepness factor will lead to the yield on the 20-year bond decreasing by 0.6030%, calculated as follows:

Change in 20-year bond yield = –0.3015% × 2 = –0.6030%

42 C is correct. Because the factors in Exhibit 1 have been standardized to have unit standard deviations, a one-standard-deviation decrease in both the level factor and the curvature factor will lead to the yield on the five-year bond increasing by 0.0389%, calculated as follows:

Change in five-year bond yield = 0.4352% – 0.3963% = 0.0389%

43 C is correct. The assistant states that bootstrapping entails *backward* substitution using par yields to solve for zero-coupon rates one by one, in order from latest to earliest maturities. Bootstrapping entails *forward* substitution, however, using par yields to solve for zero-coupon rates one by one, in order from earliest to latest maturities.

44 C is correct. Country C's private sector is much bigger than the public sector, and the government bond market in Country C currently lacks liquidity. Under such circumstances, the swap curve is a more relevant benchmark for interest rates.

45 B is correct. The historical three-year swap spread for Country B was the lowest six months ago. Swap spread is defined as the spread paid by the fixed-rate payer of an interest rate swap over the rate of the "on the run" (most recently issued) government bond security with the same maturity as the swap. The lower (higher) the swap spread, the lower (higher) the return that investors require for credit and/or liquidity risks.

The fixed rate of the three-year fixed-for-floating Libor swap was 0.01% six months ago, and the three-year government bond yield was –0.08% six months ago. Thus the swap spread six months ago was 0.01% – (–0.08%) = 0.09%.

One month ago, the fixed rate of the three-year fixed-for-floating Libor swap was 0.16%, and the three-year government bond yield was –0.10%. Thus the swap spread one month ago was 0.16% – (–0.10%) = 0.26%.

Twelve months ago, the fixed rate of the three-year fixed-for-floating Libor swap was 0.71%, and the three-year government bond yield was –0.07%. Thus, the swap spread 12 months ago was 0.71% – (–0.07%) = 0.78%.

46 A is correct. Country A's yield curve is upward sloping—a condition for the strategy—and more so than Country B's.

47 B is correct. The yield curve for Country B is currently upward sloping, but Tyo expects a reversal in the slope of the current yield curve. This means she expects the resulting yield curve for Country B to slope downward, which implies that the resulting forward curve would lie below the spot yield curve. The forward curve lies below the spot curve in scenarios in which the spot curve is downward sloping; the forward curve lies above the spot curve in scenarios in which the spot curve is upward sloping.

A is incorrect because the yield curve for Country A is currently upward sloping and Tyo expects that the yield curve will maintain its shape and level. That expectation implies that the resulting forward curve would be above the spot yield curve.

C is incorrect because the yield curve for Country C is currently downward sloping and Tyo expects a reversal in the slope of the current yield curve. She thus expects the resulting yield curve for Country C to slope upward, which implies that the resulting forward curve would be above the spot yield curve.

48 A is correct. Tyo's projected spot curve assumes that future spot rates reflect, or will be equal to, the current forward rates for all respective maturities. This assumption implies that the bonds for Country A are fairly valued because the market is effectively discounting the bond's payments at spot rates that match those projected by Tyo.

B and C are incorrect because Tyo's projected spot curves for the two countries do not match the current forward rates for all respective maturities. In the case of Country B, she expects future spot rates to be higher (than the current forward rates that the market is using to discount the bond's payments). For Country C, she expects future spot rates to be lower (than the current forward rates). Hence, she perceives the Country B bond to be currently overvalued and the Country C bond to be undervalued.

49 C is correct. Liquidity preference theory suggests that liquidity premiums exist to compensate investors for the added interest rate risk that they face when lending long term and that these premiums increase with maturity. Tyo and her assistant are assuming that liquidity premiums exist.

50 A is correct. From the forward rate model, $f_{3,2}$, is found as follows:

$$[1 + z_5]^5 = [1 + z_3]^3[1 + f_{3,2}]^2$$

Using the three-year and five-year spot rates, we find

$$(1 + 0.107)^5 = (1 + 0.118)^3[1 + f_{3,2}]^2, \text{ so}$$

$$\sqrt{\frac{(1 + 0.107)^5}{(1 + 0.118)^3}} - 1 = f_{3,2} = 9.07\%$$

51 A is correct. This action is most often consistent with bearish flattening, or short-term bond yields rising more than long-term bond yields resulting in a flatter yield curve.

52 C is correct. When government budget deficits fall, fiscal supply-side effects are most likely to result in lower bond yields.

53 B is correct. A flight to quality is most often associated with bullish flattening, in which the yield curve flattens as long term rates fall by more than short-term rates.

READING

29

The Arbitrage-Free Valuation Framework

by Steven V. Mann, PhD

Steven V. Mann, PhD, is at the University of South Carolina (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** explain what is meant by arbitrage-free valuation of a fixed-income instrument;
☐	**b.** calculate the arbitrage-free value of an option-free, fixed-rate coupon bond;
☐	**c.** describe a binomial interest rate tree framework;
☐	**d.** describe the process of calibrating a binomial interest rate tree to match a specific term structure;
☐	**e.** describe the backward induction valuation methodology and calculate the value of a fixed-income instrument given its cash flow at each node;
☐	**f.** compare pricing using the zero-coupon yield curve with pricing using an arbitrage-free binomial lattice;
☐	**g.** describe pathwise valuation in a binomial interest rate framework and calculate the value of a fixed-income instrument given its cash flows along each path;
☐	**h.** describe a Monte Carlo forward-rate simulation and its application.
☐	**i.** describe term structure models and how they are used;

INTRODUCTION TO ARBITRAGE-FREE VALUATION 1

a explain what is meant by arbitrage-free valuation of a fixed-income instrument

The presentation of the binomial trees in this reading was revised to conform with other readings in 2018 by Donald J. Smith, PhD, Boston University (USA).

© 2021 CFA Institute. All rights reserved.

The idea that market prices adjust until there are no arbitrage opportunities forms the basis for valuing fixed-income securities, derivatives, and other financial assets. If both the net proceeds (e.g., buying and selling the same value of an asset) and the risk of an investment are zero, the return on that investment should also be zero.

This reading is designed to equip candidates with a set of bond valuation tools that are consistent with this idea. The remainder of Section 1 further defines the concept of no arbitrage, and Section 2 provides a framework for an arbitrage-free valuation of fixed-income securities. Section 3 introduces the binomial interest rate tree framework based on a lognormal random walk, which is used to value an option-free bond. The binomial tree model is calibrated to the current yield curve in Section 4. This step ensures that the interest rate tree is consistent with pricing using the zero-coupon (i.e., spot) curve as illustrated in Section 5. The reading next turns to an introduction of pathwise valuation, in Section 6. Section 7 describes a Monte Carlo forward-rate simulation and its application. Section 8 goes beyond the lognormal random walk approach to introduce common term structure models. Building on principles established earlier in the reading, these models incorporate assumptions about changes in interest rates and volatility to capture term structure dynamics and are used by practitioners to price and hedge fixed-income securities and derivatives.

1.1 The Meaning of Arbitrage-Free Valuation

Arbitrage-free valuation refers to an approach to security valuation that determines security values that are consistent with the absence of an **arbitrage opportunity**, which is an opportunity for trades that earn riskless profits without any net investment of money. In well-functioning markets, prices adjust until there are no arbitrage opportunities, which is the **principle of no arbitrage** that underlies the practical validity of arbitrage-free valuation. This principle itself can be thought of as an implication of the idea that identical assets should sell at the same price.

These concepts will be explained in greater detail shortly, but to indicate how they arise in bond valuation, consider first an imaginary world in which financial assets are free of risk and the benchmark yield curve is flat. In this reading, the terms yield, interest rate, and discount rate will be used interchangeably. A flat yield curve implies that the relevant yield is the same for all cash flows regardless of when the cash flows are delivered in time. Accordingly, the value of a bond is the present value of its certain future cash flows. In discounting those cash flows—determining their present value—investors would use the risk-free interest rate because the cash flows are certain; because the yield curve is assumed to be flat, one risk-free rate would exist and apply to all future cash flows. This is the simplest case of bond valuation one can envision. When we exit this imaginary world and enter more realistic environs, bonds' cash flows are risky (i.e., there is some chance the borrower will default) and the benchmark yield curve is not flat. How would our approach change?

A fundamental principle of valuation is that the value of any financial asset is equal to the present value of its expected future cash flows. This principle holds for any financial asset, from zero-coupon bonds to interest rate swaps. Thus, the valuation of a financial asset involves the following three steps:

- Step 1 Estimate the future cash flows.
- Step 2 Determine the appropriate discount rate or discount rates that should be used to discount the cash flows.
- Step 3 Calculate the present value of the expected future cash flows found in Step 1 by applying the appropriate discount rate or rates determined in Step 2.

The traditional approach to valuing bonds is to discount all cash flows with the same discount rate as if the yield curve were flat. However, a bond is properly thought of as a package or portfolio of zero-coupon bonds, also referred to as zeros or discount instruments. Each zero-coupon bond in such a package can be valued separately at a discount rate that depends on the shape of the yield curve and when its single cash flow is delivered in time. The term structure of these discount rates is referred to as the spot curve. Bond values derived by summing the present values of the individual zeros (cash flows) determined by such a procedure can be shown to be arbitrage free. Ignoring transaction costs for the moment, if the bond's value were much less than the sum of the values of its cash flows individually, a trader would perceive an arbitrage opportunity and buy the bond while selling claims to the individual cash flows and pocketing the excess value. Although the details bear further discussion, the valuation of a bond as a portfolio of zeros based on using the spot curve is an example of arbitrage-free valuation. Regardless of the complexity of the bond, each component must have an arbitrage-free value. A bond with embedded options can be valued in parts as the sum of the arbitrage-free bond without options (that is, a bond with no embedded options) and the arbitrage-free value of each of the options.

1.2 The Law of One Price

The central idea of financial economics is that market prices will adjust until there are no opportunities for arbitrage. We will define shortly what is meant by an arbitrage opportunity, but for now think of it as "free money." Prices will adjust until there is no free money to be acquired. Arbitrage opportunities arise from violations of the **law of one price**. The law of one price states that two goods that are perfect substitutes must sell for the same current price in the absence of transaction costs. Two goods that are identical, trading side by side, are priced the same. Otherwise, if it were costless to trade, one would simultaneously buy at the lower price and sell at the higher price. The riskless profit is the difference in the prices. An individual would repeat this transaction without limit until the two prices converge. An implication of these market forces is deceptively straightforward and basic. If you do not put up any of your own money and take no risk, your expected return should be zero.

1.3 Arbitrage Opportunity

With this background, let us define arbitrage opportunity more precisely. An arbitrage opportunity is a transaction that involves no cash outlay that results in a riskless profit. There are two types of arbitrage opportunities. The first type of arbitrage opportunity is often called **value additivity**; put simply, the value of the whole equals the sum of the values of the parts. Consider two risk-free investments with payoffs one year from today and the prices today provided in Exhibit 1. Asset A is a simple risk-free zero-coupon bond that pays off one dollar and is priced today at 0.952381 (= 1/1.05). Asset B is a portfolio of 105 units of Asset A that pays off 105 one year from today and is priced today at 97. The portfolio does not equal the sum of the parts. The portfolio (Asset B) is cheaper than buying 105 units of Asset A at a price of 100 and then combining. An astute investor would sell 105 units of Asset A for 105 × 0.952381 = 100 while simultaneously buying the portfolio, Asset B, for 97. This position generates a certain 3 today (100 – 97) and generates net 0 one year from today because cash inflow for Asset B matches the amount for the 105 units of Asset A sold. An investor would repeat this trade until the prices are equal.

The second type of arbitrage opportunity is often called **dominance**. A financial asset with a risk-free payoff in the future must have a positive price today. Consider two assets, C and D, that are risk-free zero-coupon bonds. Payoffs in one year and prices today are displayed in Exhibit 1. On careful review, it appears that Asset D is cheap relative to Asset C. If both assets are risk-free, they should have the same discount rate. To make money, sell two units of Asset C at a price of 200 and use the proceeds to purchase one unit of Asset D for 200. The construction of the portfolio involves no net cash outlay today. Although it requires zero dollars to construct today, the portfolio generates 10 one year from today. Asset D will generate a 220 cash inflow, whereas the two units of Asset C sold will produce a cash outflow of 210.

Exhibit 1 Price Today and Payoffs in One Year for Sample Assets

Asset	Price Today	Payoff in One Year
A	0.952381	1
B	97	105
C	100	105
D	200	220

This existence of both types of arbitrage opportunity is transitory. Investors aware of this mispricing will demand the securities in question in unlimited quantities. Something must change to restore stability. Prices will adjust until there are no arbitrage opportunities.

EXAMPLE 1

Arbitrage Opportunities

Which of the following investment alternatives includes an arbitrage opportunity?

Bond A: The yield for a 3% annual coupon 10-year bond is 2.5% in New York City. The same bond sells for $104.376 per $100 face value in Chicago.

Bond B: The yield for a 3% annual coupon 10-year bond is 3.2% in Hong Kong SAR. The same bond sells for RMB97.220 per RMB100 face value in Shanghai.

Solution:

Bond B is correct. Bond B's arbitrage-free price may be solved for using a financial calculator or Microsoft Excel as $3/1.032 + 3/1.032^2 + \ldots + 103/1.032^{10} = 98.311$, which is higher than the price in Shanghai. Therefore, an arbitrage opportunity exists. Buy bonds in Shanghai for RMB97.220 and sell them in Hong Kong SAR for RMB98.311. You make RMB1.091 per RMB100 of bonds traded.

Bond A's arbitrage-free price is $3/1.025 + 3/1.025^2 + \ldots + 103/1.025^{10} = 104.376$, which matches the price in Chicago. Therefore, no arbitrage opportunity exists in this market.

1.4 Implications of Arbitrage-Free Valuation for Fixed-Income Securities

Using the arbitrage-free approach, any fixed-income security should be thought of as a package or portfolio of zero-coupon bonds. Thus, a five-year 2% coupon Treasury issue should be viewed as a package of 11 zero-coupon instruments (10 semiannual coupon payments, 1 of which is made at maturity, and 1 principal value payment at maturity). The market mechanism for US Treasuries that enables this approach is the dealer's ability to separate the bond's individual cash flows and trade them as zero-coupon securities. This process is called **stripping**. In addition, dealers can recombine the appropriate individual zero-coupon securities and reproduce the underlying coupon Treasury. This process is called **reconstitution**. Dealers in sovereign debt markets around the globe are free to engage in the same process.

Arbitrage profits are possible when value additivity does not hold. The arbitrage-free valuation approach does not allow a market participant to realize an arbitrage profit through stripping and reconstitution. By viewing any security as a package of zero-coupon securities, a consistent and coherent valuation framework can be developed. Viewing a security as a package of zero-coupon bonds means that two bonds with the same maturity and different coupon rates are viewed as different packages of zero-coupon bonds and valued accordingly. Moreover, two cash flows with identical risks delivered at the same time will be valued using the same discount rate even though they are attached to two different bonds.

ARBITRAGE-FREE VALUATION FOR AN OPTION-FREE BOND 2

b calculate the arbitrage-free value of an option-free, fixed-rate coupon bond

The goal of this section is to develop a method to produce an arbitrage-free value for an option-free bond and to provide a framework—based on interest rate trees—that is rich enough to be applied to the valuation of bonds with embedded options.

For bonds that are option-free, the simplest approach to arbitrage-free valuation involves determining the arbitrage-free value as the sum of the present values of expected future values using the benchmark spot rates. Benchmark securities are liquid, safe securities whose yields serve as building blocks for other interest rates in a country or currency. Sovereign debt is the benchmark in many countries. For example, on-the-run Treasuries serve as benchmark securities in the United States. Par rates derived from the Treasury yield curve can be used to obtain spot rates by means of bootstrapping. Gilts are the benchmark in the United Kingdom, while German bunds serve as the benchmark for euro-denominated bonds. In markets where the sovereign debt market is not sufficiently liquid, the swap curve is a viable alternative.

In this reading, benchmark bonds are assumed to be correctly priced by the market. The valuation model we develop will be constructed to reproduce exactly the prices of the benchmark bonds.

EXAMPLE 2

The Arbitrage-Free Value of an Option-Free Bond

The yield-to-maturity ("par rate") for a benchmark one-year annual coupon bond is 2%, for a benchmark two-year annual coupon bond is 3%, and for a benchmark three-year annual coupon bond is 4%. A three-year, 5% annual coupon bond with the same risk and liquidity as the benchmarks is selling for 102.7751 today (t = 0) to yield 4%. Is this value correct for the bond given the current term structure?

Solution:

The first step in the solution is to find the correct spot rate (zero-coupon rates) for each year's cash flow. The spot rates may be determined using bootstrapping, which is an iterative process. Using the bond valuation equation below, one can solve iteratively for the spot rates, z_t (rate on a zero-coupon bond of maturity t), given the periodic payment, PMT, on the relevant benchmark bond.

$$100 = \frac{PMT}{(1+z_1)^1} + \frac{PMT}{(1+z_2)^2} + \cdots + \frac{PMT + 100}{(1+z_N)^N}.$$

A revised equation, which uses the par rate rather than PMT, may also be used to calculate the spot rates. The revised equation is

$$1 = \frac{\text{Par rate}}{(1+z_1)} + \frac{\text{Par rate}}{(1+z_2)^2} + \ldots + \frac{\text{Par rate} + 1}{(1+z_N)^N}$$

where par rate is PMT divided by 100 and represents the par rate on the benchmark bond and z_t is the t-period zero-coupon rate.

In this example, the one-year spot rate, z_1, is 2%, which is the same as the one-year par rate. To solve for z_2,

$$1 = \frac{0.03}{(1+z_1)} + \frac{0.03+1}{(1+z_2)^2} = \frac{0.03}{(1+0.02)} + \frac{0.03+1}{(1+z_2)^2}.$$

$$z_2 = 3.015\%.$$

To solve for z_3,

$$1 = \frac{0.04}{(1+z_1)} + \frac{0.04}{(1+z_2)^2} + \frac{0.04+1}{(1+z_3)^3} = \frac{0.04}{(1+0.02)} + \frac{0.04}{(1+0.03015)^2} + \frac{0.04+1}{(1+z_3)^3}.$$

$$z_3 = 4.055\%$$

The spot rates are 2%, 3.015%, and 4.055%. The correct arbitrage-free price for the bond, then, is

$$P_0 = 5/1.02 + 5/1.03015^2 + 105/1.04055^3 = 102.8102.$$

To be arbitrage free, each cash flow of a bond must be discounted by the spot rate for zero-coupon bonds maturing on the same date as the cash flow. Discounting early coupons by the bond's yield-to-maturity gives too much discounting with an upward sloping yield curve and too little discounting for a downward sloping yield curve. The bond is mispriced by 0.0351 per 100 of par value.

For option-free bonds, performing valuation discounting with spot rates produces an arbitrage-free valuation. For bonds that have embedded options, we need a different approach. The challenge one faces when developing a framework for valuing bonds with embedded options is that their expected future cash flows are interest rate dependent. If the bonds are option-free, changes in interest rates have no impact on the size and timing of the bond's cash flows. For bonds with options attached, changes in future interest rates impact the likelihood the option will be exercised and in so doing impact the cash flows. Therefore, to develop a framework that values bonds both without and with embedded options, we must allow interest rates to take on different potential values in the future based on some assumed level of volatility. The vehicle to portray this information is an interest rate "tree" representing possible future interest rates consistent with the assumed volatility. Because the interest rate tree resembles a lattice, these models are often called "lattice models." The interest rate tree performs two functions in the valuation process: (1) Generate the cash flows that are interest rate dependent, and (2) supply the interest rates used to determine the present value of the cash flows. This approach will be used in later readings when considering learning outcome statements involving callable bonds.

An interest rate model seeks to identify the elements or *factors* that are believed to explain the dynamics of interest rates. These factors are random or *stochastic* in nature, so we cannot predict the path of any factor. An interest rate model must, therefore, specify a statistical process that describes the stochastic property of these factors to arrive at a reasonably accurate representation of the behavior of interest rates. What is important to understand is that the interest rate models commonly used are based on how short-term interest rates can evolve (i.e., change) over time. Consequently, these interest rate models are referred to as one-factor models because only one interest rate is being modeled over time. More complex models consider how more than one interest rate changes over time (e.g., the short rate and the long rate) and are referred to as two-factor models.

Our task at hand is to describe the binomial interest rate tree framework. The valuation model we are attempting to build is the binomial lattice model. It is so named because the short interest rate can take on one of two possible values consistent with the volatility assumption and an interest rate model. As we will soon discover, the two possible interest rates next period will be consistent with the following three conditions: (1) an interest rate model that governs the random process of interest rates, (2) the assumed level of interest rate volatility, and (3) the current benchmark yield curve. We take the prices of the benchmark bonds as given so that our model recovers the market values for each benchmark bond. In this way, we tie the model to the current yield curve that reflects the underlying economic reality.

2.1 The Binomial Interest Rate Tree

The first step for demonstrating the binomial valuation method is to present the benchmark par curve by using bonds of a country or currency. For simplicity in our illustration, we will use US dollars. The same principles hold with equal force regardless of the country or currency. The benchmark par curve is presented in Exhibit 2. For simplicity, we assume that all bonds have annual coupon payments. Benchmark bonds are conveniently priced at par so the yields-to-maturity and the coupon rates on the bonds are the same. From these par rates, we use the bootstrapping methodology to uncover the underlying spot rates shown in Exhibit 3. Because the par curve is upward sloping, it comes as no surprise that after Year 1 the spot rates are higher than the par rates. In Exhibit 4 we present the one-year implied forward rates derived from the spot curve using no arbitrage. Because the par, spot, and forward curves reflect the same information about interest rates, if one of the three curves is known, it is possible to generate the other two curves. The three curves are identical only if the yield curve is flat.

Exhibit 2 Benchmark Par Curve

Maturity (Years)	Par Rate	Bond Price
1	1.00%	100
2	1.20%	100
3	1.25%	100
4	1.40%	100
5	1.80%	100

Exhibit 3 Underlying One-Year Spot Rates of Par Rates

Maturity (Years)	One-Year Spot Rate
1	1.0000%
2	1.2012%
3	1.2515%
4	1.4045%
5	1.8194%

Exhibit 4 One-Year Implied Forward Rates

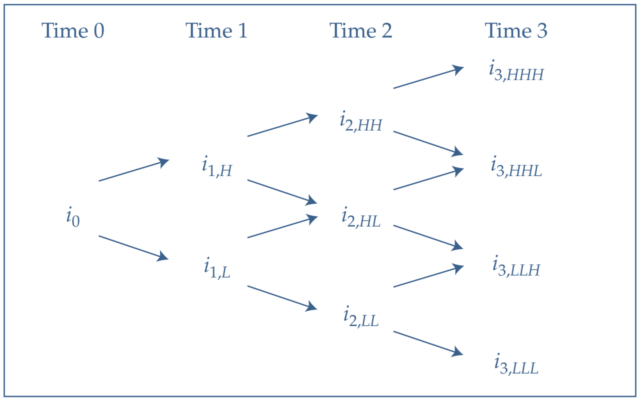
Maturity (Years)	Forward Rate
Current one-year rate	1.0000%
One-year rate, one year forward	1.4028%
One-year rate, two years forward	1.3521%
One-year rate, three years forward	1.8647%
One-year rate, four years forward	3.4965%

Recall from our earlier discussion that if we value the benchmark bonds using rates derived from these curves, we will recover the market price of par for all five bonds in Exhibit 2. Specifically, par rates represent the single interest applied to all the cash flows that will produce the market prices. Discounting each cash flow separately with the set of spot rates will also give the same answer. Finally, forward rates are the discount rates of a single cash flow over a single period. If we discount each cash flow with the appropriate discount rate for each period, the computed values will match the observed prices.

When we approach the valuation of bonds with cash flows that are interest rate dependent, we must explicitly allow interest rates to change. We accomplish this task by introducing interest rate volatility and generating an interest rate tree later in this reading. An interest rate tree is simply a visual representation of the possible values of interest rates based on an interest rate model and an assumption about interest rate volatility.

A binomial interest rate tree is presented in Exhibit 5. Our goal is to learn how to populate this structure with interest rates. Notice the *i*'s, which represent different potential values the one-year interest rates may take over time. As we move from left to right on the tree, the number of possible interest rates increases. The first is the current time (in years), or formally, Time 0. The interest rate displayed at Time 0 is the discount rate that converts Time 1 payments to Time 0 present values. At the bottom of the graph, time is the unit of measurement. Notice that there is one year between possible interest rates. This is called the "time step," and in our illustration, it matches the frequency of the annual cash flows. The *i*'s in Exhibit 5 are called nodes. The first node is called the root of the tree and is simply the current one-year rate at Time 0. Each node thereafter is represented by a both time element and a rate change component.

Exhibit 5 Binomial Interest Rate Tree



We now turn to the question of how to obtain the two possible values for the one-year interest rate one year from today. Two assumptions are required: an interest rate model and a volatility of interest rates. Recall an interest rate model puts structure on the randomness. We are going to use the lognormal random walk, and the resulting tree structure is often referred to as a lognormal tree. A lognormal model of interest rates insures two appealing properties: (1) non-negativity of interest rates and (2) higher volatility at higher interest rates. At each node, there are two possible rates one year forward at Time 1. We will assume for the time being that each has an equal probability of occurring. The two possible rates we will calculate are going to be higher and lower than the one-year forward rate at Time 1 one year from now.

We denote i_L to be the rate lower than the implied forward rate and i_H to be the higher forward rate. The lognormal random walk posits the following relationship between $i_{1,L}$ and $i_{1,H}$:

$$i_{1,H} = i_{1,L}e^{2\sigma},$$

where σ is the standard deviation and *e* is Euler's number, the base of natural logarithms, which is a constant 2.7183. The random possibilities each period are (nearly) centered on the forward rates calculated from the benchmark curve. The intuition of this relationship is deceptively quick and simple. Think of the one-year forward implied interest rate from the yield curve as the average of possible values for the one-year rate at Time 1. The lower of the two rates, i_L, is one standard deviation below the mean (one-year implied forward rate), and i_H is one standard deviation above the mean. Thus, the higher and lower values (i_L and i_H) are multiples of each other, and the multiplier is $e^{2\sigma}$. Note that as the standard deviation (i.e., volatility) increases, the multiplier increases, and the two rates will grow farther apart but will still be (nearly) centered on the implied forward rate derived from the spot curve. We will demonstrate this soon.

We use the following notation to describe the tree at Time 1. Let

σ = assumed volatility of the one-year rate,
$i_{1,L}$ = the lower one-year forward rate one year from now at Time 1, and
$i_{1,H}$ = the higher one-year forward rate one year from now at Time 1.

For example, suppose that $i_{1,L}$ is 1.194% and σ is 15% per year; then $i_{1,H}$ = 1.194%($e^{2\times 0.15}$) = 1.612%.

At Time 2, there are three possible values for the one-year rate, which we will denote as follows:

$i_{2,LL}$ = one-year forward rate at Time 2 assuming the lower rate at Time 1 and the lower rate at Time 2.

$i_{2,HH}$ = one-year forward rate at Time 2 assuming the higher rate at Time 1 and the higher rate at Time 2.

$i_{2,HL}$ = one-year forward rate at Time 2 assuming the higher rate at Time 1 and the lower rate at Time 2, or equivalently, the lower rate at Time 1 and the higher rate at Time 2.

The middle rate will be close to the implied one-year forward rate two years from now derived from the spot curve, whereas the other two rates are two standard deviations above and below this value. (Recall that the multiplier for adjacent rates on the tree differs by a multiple of e raised to the 2σ.) This type of tree is called a recombining tree because there are two paths to get to the middle rate. This feature of the model results in faster computation because the number of possible outcomes each period grows linearly rather than exponentially.

The relationship between $i_{2,LL}$ and the other two one-year rates is as follows:

$i_{2,HH} = i_{2,LL}(e^{4\sigma})$, and $i_{2,HL} = i_{2,LL}(e^{2\sigma})$.

In a given period, adjacent possible outcomes in the tree are two standard deviations apart. So, for example, if $i_{2,LL}$ is 0.980%, and assuming once again that σ is 15%, we calculate

$$i_{2,HH} = 0.980\%(e^{4\times0.15}) = 1.786\%$$

and

$$i_{2,HL} = 0.980\%(e^{2\times0.15}) = 1.323\%.$$

There are four possible values for the one-year forward rate at Time 3. These are represented as follows: $i_{3,HHH}$, $i_{3,HHL}$, $i_{3,LLH}$ and $i_{3,LLL}$. Once again, all the forward rates in the tree are multiples of the lowest possible rates each year. The lowest possible forward rate at Time 3 is $i_{3,LLL}$ and is related to the other three as given below:

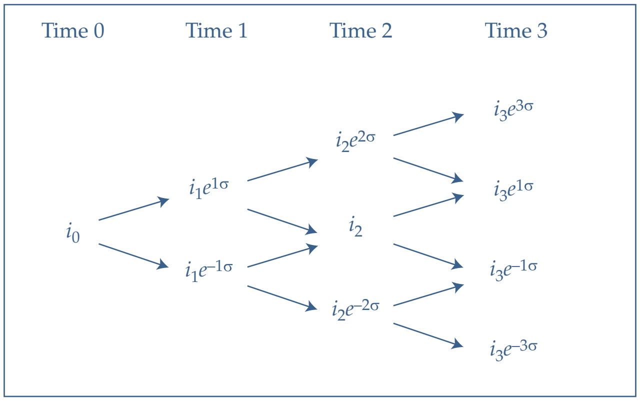
$$i_{3,HHH} = (e^{6\sigma})i_{3,LLL}.$$

$$i_{3,HHL} = (e^{4\sigma})i_{3,LLL}.$$

$$i_{3,LLH} = (e^{2\sigma})i_{3,LLL}.$$

Exhibit 6 shows the notation for a four-year binomial interest rate tree. We can simplify the notation by centering the one-year rates on the tree on implied forward rates on the benchmark yield curve, so i_t is the one-year rate t years from now and the centering rate. The subscripts indicate the rates at the end of the year, so in the second year, it is the rate at the end of Time 2 to the end of Time 3. Exhibit 6 uses this uniform notation. Note that adjacent forward rates in the tree are two standard deviations (σ) apart.

Exhibit 6 Four-Year Binomial Tree



Before we attempt to build an interest rate tree, two additional tools are needed. These tools are introduced in the next two sections.

3 THE BASICS OF CREATING A BINOMIAL INTEREST RATE TREE

c describe a binomial interest rate tree framework

Recall that variance is a measure of dispersion of a probability distribution. The standard deviation is the square root of the variance and is measured in the same units as the mean. With a simple lognormal distribution, the changes in interest rates are proportional to the level of the one-period interest rates each period. Volatility is measured relative to the current level of rates. It can be shown that for a lognormal distribution the standard deviation of the one-year rate is equal to $i_0\sigma$. For example, if σ is 10% and the one-year rate (i_0) is 2%, then the standard deviation of the one-year rate is 2% × 10% = 0.2%, or 20 bps. As a result, interest rate moves are larger when interest rates are high and are smaller when interest rates are low. One of the characteristics of a lognormal distribution is that negative interest rates are not possible, since as rates approach zero, the absolute change in interest rates becomes smaller and smaller.

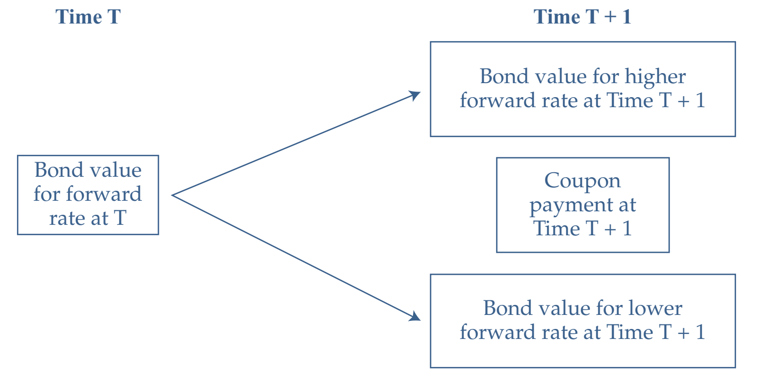
There are two methods commonly used to estimate interest rate volatility. The first method uses historical interest rate volatility based on data from the recent past, which is assumed to be indicative of the future. A second method to estimate interest rate volatility is that derived from observed market prices of interest rate derivatives (e.g., swaptions, caps, floors) known as implied volatility.

3.1 Determining the Value of a Bond at a Node

To find the value of the bond at a node, we use the backward induction valuation methodology. Barring default, we know that at maturity the bonds will be valued at par. So, we start at maturity, fill in those values, and work back from right to left to find the bond's value at the desired node. Suppose we want to determine the bond's value at the lowest node at Time 1. To find this value, we must first calculate the bond's value at the two nodes to the right of the node we selected. The bond's value at the two nodes immediately to the right must be available.

A bond's value at any node will depend on the future coupon payment, *C*, and the expected future value for the bond. This expected value is the average of the value for the forward rate being higher, to be denoted below by *VH*, and the value for the forward rate being lower, *VL*. It is a simple average because in the lognormal model the probabilities for the rate going up or down are equal. This is illustrated in Exhibit 7. Notice that the coupon payment due at the end of the period, at Time T + 1, is placed directly to the right of the node for Time T. The arrows point to the two possible future bond values, one for the forward rate going up at Time T + 1 and the other for the rate going down.

Exhibit 7 Finding a Bond's Value at Any Node



The next step is to determine the present value of the coupon payment and the expected future bond value. The relevant discount rate is the one-year forward rate prevailing at the beginning of the time period, *i*, at Time T. The bond's value at any node is determined by the following expression:

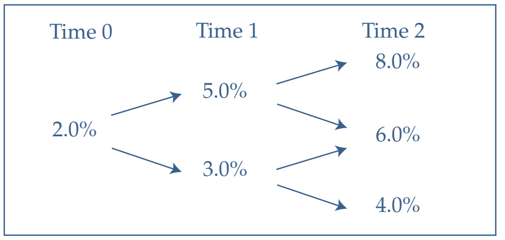
$$\text{Bond value at a node} = \frac{C + \left(0.5 \times VH + 0.5 \times VL\right)}{1 + i}.$$

EXAMPLE 3

Pricing a Bond Using a Binomial Tree

Using the interest rate tree in Exhibit 8, find the correct price for a three-year, annual pay bond with a coupon rate of 5%.

Exhibit 8 Three-Year Binomial Interest Rate Tree



Solution:

Exhibit 9 shows the binomial tree to value the three-year, 5% bond. We start with Time 3. The cash flow is 105, the redemption of par value (100) plus the final coupon payment (5), regardless of the level of the forward rate at Time 2. Using backward induction, we next calculate the present value of the bond as of Time 2 for the three possible forward rates:

105/1.08 = 97.2222.

105/1.06= 99.0566.

105/1.04 = 100.9615.

Working back to Time 1 requires the use of the general expression above for the value at any node. If the forward rate is 5.0% at Time 1, the bond value is 98.2280:

$$\frac{5 + (0.5 \times 97.2222 + 0.5 \times 99.0566)}{1.05} = 98.2280.$$

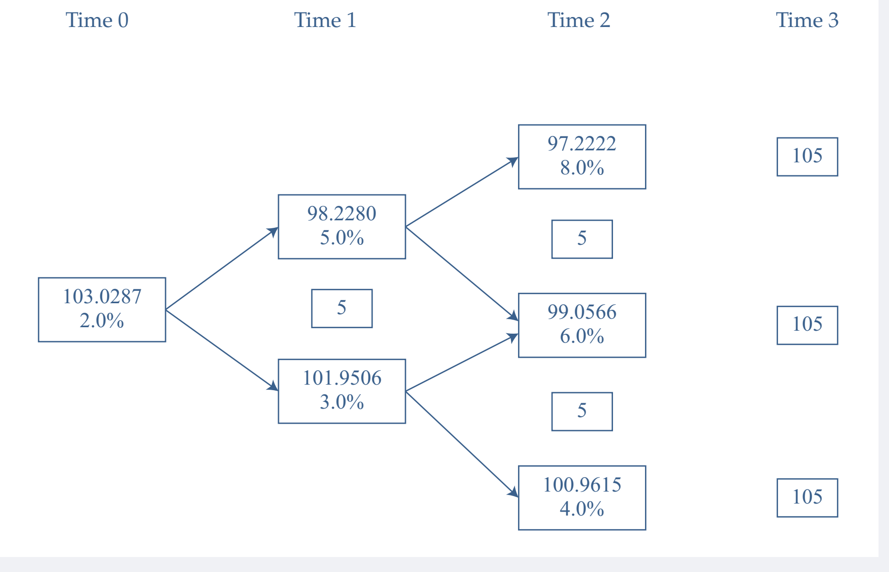
If the forward rate instead is 3.0%, the bond value is 101.9506:

$$\frac{5 + (0.5 \times 99.0566 + 0.5 \times 100.9615)}{1.03} = 101.9506.$$

Finally, the value of the bond at Time 0 is 103.0287:

$$\frac{5 + (0.5 \times 98.2280 + 0.5 \times 101.9506)}{1.02} = 103.0287.$$

Exhibit 9 Three-Year Binomial Tree



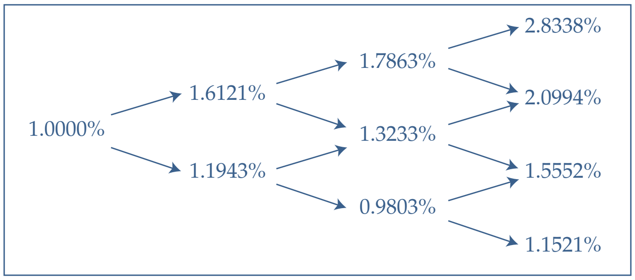
4 CALIBRATING THE BINOMIAL INTEREST RATE TREE TO THE TERM STRUCTURE

d describe the process of calibrating a binomial interest rate tree to match a specific term structure

The construction of a binomial interest rate tree requires multiple steps, but keep in mind what we are trying to accomplish. We assume a process that generates interest rates and volatility. The first step is to describe the calibration of a binomial interest rate tree to match a specific term structure. We do this to ensure that the model is arbitrage free. We fit the interest rate tree to the current yield curve by choosing interest rates such that the model produces the benchmark bond values reported earlier. By doing this, we tie the model to the underlying economic reality.

Recall from Exhibits 2, 3, and 4 the benchmark bond price information and the relevant par, spot, and forward curves. We will assume that volatility, σ, is 15% and construct a four-year tree starting with the two-year bond that carries a coupon rate of 1.20%. A complete four-year binomial interest rate tree is presented in Exhibit 10. We will demonstrate how these rates are determined. The current one-year rate is 1%, i_0.

Exhibit 10 Four-Year Binomial Interest Rate Tree



Finding the rates in the tree is an iterative process, and the interest rates are found numerically. There are two possible rates at Time 1—the higher rate and the lower rate. We observe these rates one year from today. These two rates must be consistent with the volatility assumption, the interest rate model, and the observed market value of the benchmark bond. Assume that the interest rate volatility is 15%. From our discussion earlier, we know that at Time 1 the lower one-year rate is lower than the implied one-year forward rate and the higher rate is a multiple of the lower rate. We iterate to a solution with constraints in mind. Once we select these rates, how will we know the rates are correct? The answer is when we discount the cash flows using the tree and produce a value that matches the price of the two-year benchmark bond. If the model does not produce the correct price with this result, we need to select another forward rate and repeat the process. The process of calibrating a binomial interest rate tree to match a specific term structure is illustrated in the following paragraphs.

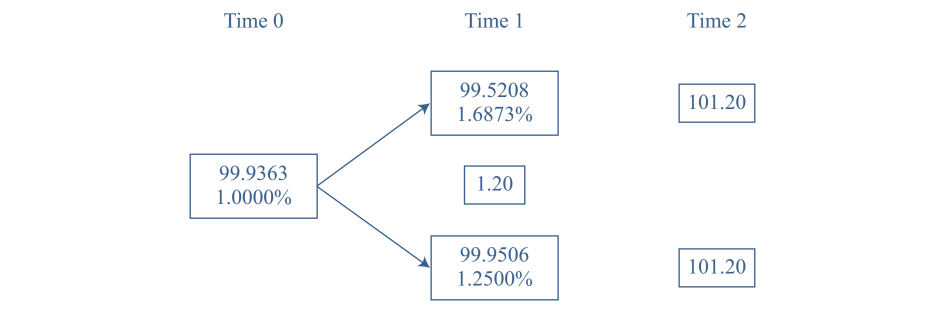
The procedure starts with the selection of a trial rate for one of the Time 1 forward rates—for instance, $i_{1,L}$. This rate should be lower than the implied forward rate from Exhibit 4 of 1.4028%. Suppose that we select 1.2500%. The other forward rate will be 1.6873% [= 1.2500% × ($e^{2\times0.15}$)]. Exhibit 11 shows that the Time 0 value for the 1.20%, two-year bond is 99.9363. The redemption of principal and the final interest payment are placed across from the two nodes for the forward rates. At Time 1, the interest payment due is placed across from the initial rate for Time 0. These are the calculations:

101.20/1.016873 = 99.5208.

$101.20/1.012500 = 99.9506.$

$$\frac{1.20 + \left(0.5 \times 99.5208 + 0.5 \times 99.9506\right)}{1.01} = 99.9363.$$

Exhibit 11 Calibrating the Two-Year Binomial Tree



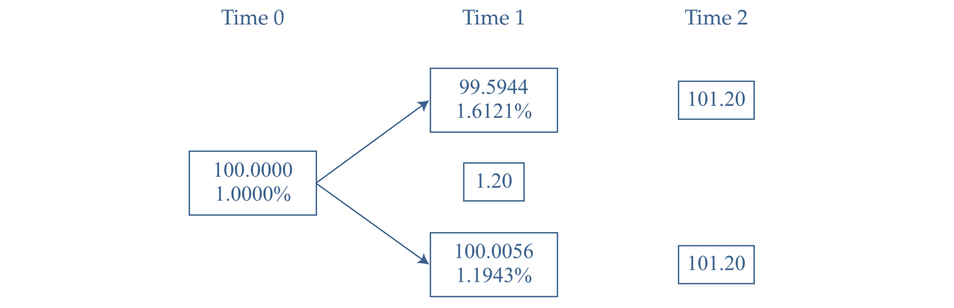
These two trial rates are clearly too high. They need to be lowered somewhat to raise the bond value to attain a Time 0 price for the bond of 100.0000. We could proceed with further trial-and-error search or use an analytic tool, such as Solver in Excel, to carry out this calculation. Essentially, we need to set the cell for the Time 0 bond price to a value of 100.0000 by changing the cell containing the initial lower forward rate for Time 1.

This procedure eventually obtains a value for $i_{1,L}$ of 1.1943%. This is the lower one-year rate. The higher one-year rate is 1.6121% [= 1.1943% × ($e^{2\times0.15}$)]. Notice that the average of these two forward rates is 1.4032% [= (1.6121% + 1.1943%)/2], slightly above the implied forward rate of 1.4028% from Exhibit 4. The binomial tree spreads out around the forward rate curve. The average is slightly higher than the implied forward rate because of the assumption of lognormality.

Recall from the information on the benchmark bonds that the two-year bond will pay its maturity value of 100 at Time 2 and an annual coupon payment of 1.20. The bond's value at Time 2 is 101.20. The present value of the coupon payment plus the bond's maturity value if the higher one-year rate is realized, *VH*, is 99.5944 (= 101.20/1.016121). Alternatively, the present value of the coupon payment plus the bond's maturity value if the lower one-year rate is realized, *VL*, is 100.0056 (= 101.20/1.011943). These two calculations determine the bond's value one year forward. Effectively, the forward rates move the bond's value from Time 2 to Time 1. Exhibit 12 demonstrates that the arbitrage-free forward rates for Time 1 are 1.6121% and 1.1943%. The value for the bond at Time 0 is 100.0000, confirming the calibration:

$$\frac{1.20 + \left(0.5 \times 99.5944 + 0.5 \times 100.0056\right)}{1.010000} = 100.0000.$$

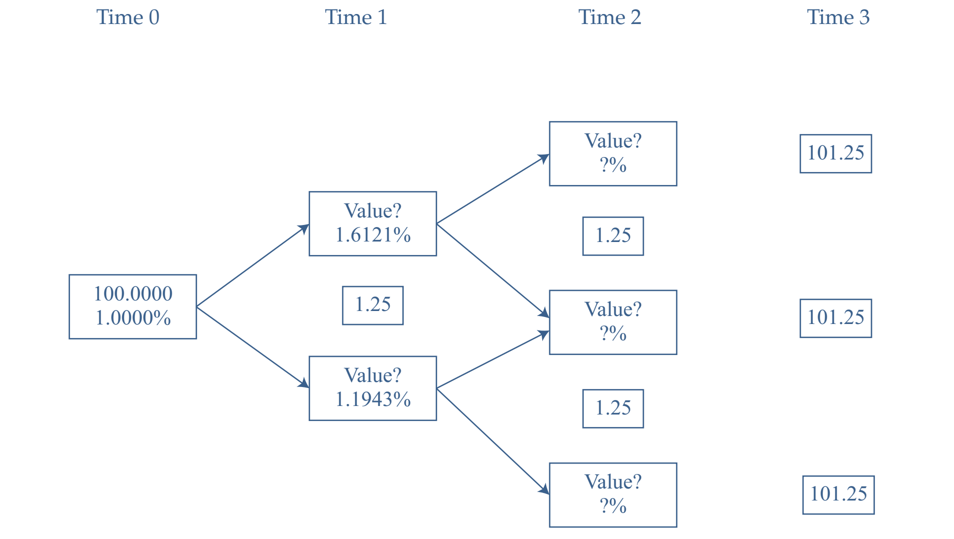
Exhibit 12 Building the Two-Year Binomial Tree



To build out the tree one more year, we repeat the same process, this time using a three-year benchmark bond with a coupon rate of 1.25%. Now, we are looking for three forward rates that are consistent with (1) the interest rate model assumed, (2) the assumed volatility of 15%, (3) a current one-year rate of 1.0%, and (4) the two possible forward rates one year from now (at Time 1) of 1.1943% (the lower rate) and 1.6121% (the higher rate).

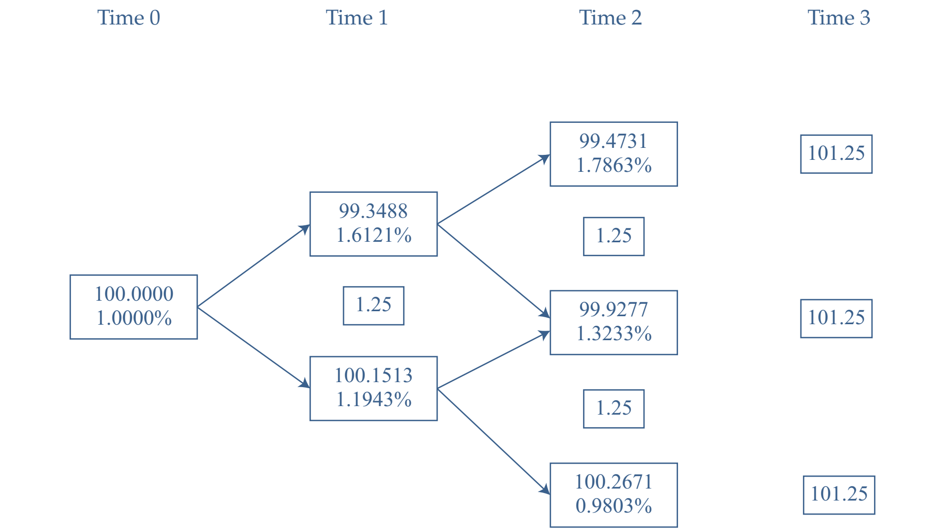
At Time 3, we receive the final coupon payment and maturity value of 101.25. In Exhibit 13, we see the known coupon payments of 1.25 for Times 1 and 2. Also entered are the Time 1 forward rates and the target price of par value for the three-year bond. The unknown items to determine are the Time 1 and Time 2 bond values (Value?) and the Time 2 forward rates (?%).

Exhibit 13 Finding the Time 2 Forward Rates



We need to select a trial value for the middle rate, $i_{2,HL}$. A good choice is the implied forward rate of 1.3521%. The trial value for the upper rate, $i_{2,HH}$, would need to be 1.3521% × ($e^{2\times0.15}$), and the trial value for the lower rate, $i_{2,LL}$, would need to be 1.3521%/($e^{2\times0.15}$). The middle rate is then changed, changing the others as well, until the value for the 1.25% three-year bond is 100.0000. It turns out that the three forward rates are 1.7863%, 1.3233%, and 0.9803%. To demonstrate that these are the correct values, we simply work backward from the cash flows at Time 3 of the tree in Exhibit 13. The same procedure is used to obtain the values at the other nodes. The completed tree is shown in Exhibit 14.

Exhibit 14 Completed Binomial Tree with Calculated Forward Rates



Let us focus on the impact of volatility on the possible forward rates in the tree. If we were to use a higher estimate of volatility—say, 20%—the possible forward rates should spread farther out around the forward curve. If we were to use a lower estimate of volatility—say, 0.01%—the rates should collapse to the implied forward rates from the current yield curve. Exhibits 15 and 16 depict the interest rate trees for the volatilities of 20% and 0.01%, respectively, and confirm the expected outcome. Notice that in Exhibit 16 for 0.01% volatility, the Time 1 forward rates are very close to the implied forward rate of 1.4028% shown in Exhibit 4. Likewise, the Time 2 and Time 3 rates are a small range around the forward rates of 1.3521% and 1.8647%, respectively. In fact, if $\sigma = 0$, the binomial tree is simply the implied forward curve.

Exhibit 15 Completed Tree with σ = 20%

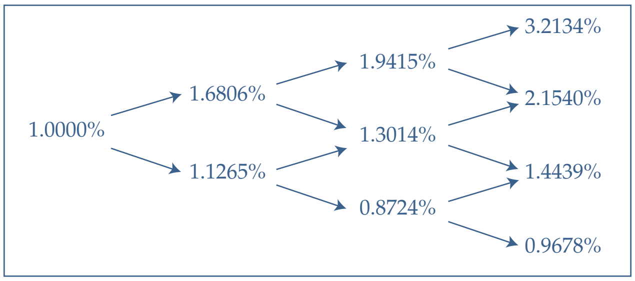
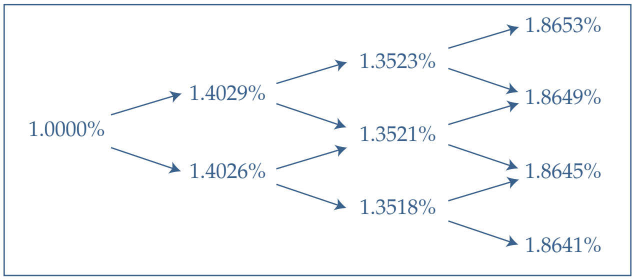


Exhibit 16 Completed Tree with σ = 0.01%



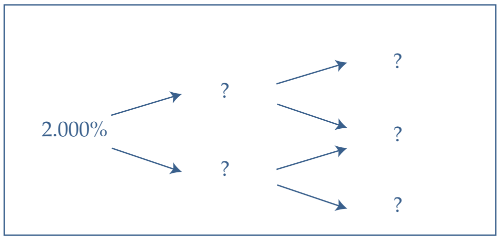
EXAMPLE 4

Calibrating a Binomial Tree to Match a Specific Term Structure

As in Example 2, the one-year par rate is 2.000%, the two-year par rate is 3.000%, and the three-year par rate is 4.000%. Consequently, the spot rates are S_0 = 2.000%, S_1 = 3.015%, and S_2 = 4.055%. The forward rates are F_0 = 2.000%, F_1 = 4.040%, and F_2 = 6.166%. Interest volatility is 15% for all years.

Calibrate the binomial tree in Exhibit 17.

Exhibit 17 Binomial Tree to Calibrate



Solution:

Time 0

The par, spot, and forward rates are all the same for the first period in a binomial tree. Consequently, $Y_0 = S_0 = F_0$ = 2.000%.

Time 1

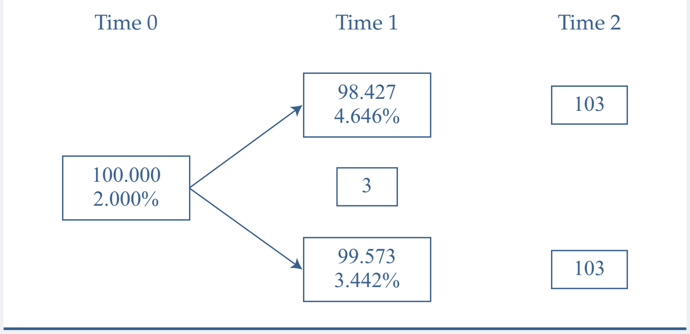
We need to use trial-and-error search (or Solver in Excel) to find the two forward rates that produce a value of 100.000 for the 3%, two-year bond. The lower trial rate needs to be lower than the implied forward rate of 4.040%—for instance, 3.500%. The higher trial rate would be 3.500% × ($e^{2\times0.15}$) = 4.725%. These lead to a Time 0 value for the bond of 99.936. Therefore, the next stage in the procedure lowers the trial rates. Finally, the calibrated forward rates are 4.646% and 3.442%. Exhibit 18 shows that these are the correct rates because the value of the bond at Time 0 is 100.000. These are the calculations:

103/1.04646 = 98.427.

103/1.03442 = 99.573.

$$\frac{3 + \left(0.5 \times 98.427 + 0.5 \times 99.573\right)}{1.02} = 100.0000.$$

Exhibit 18 Calibration of Time 1 Forward Rates



Time 2

The initial trial rate for the middle node for Time 2 is the implied forward rate of 6.166%. The rate for the upper node is 8.323% [= 6.166% × ($e^{2\times0.15}$)], and the rate for the lower node is 4.568% [= 6.166%/($e^{2\times0.15}$)]. Exhibit 19 shows that these rates for Time 2 and the already calibrated rates for Time 1 lead to a value of 99.898 for the 4% three-year bond as of Time 0. These are not the arbitrage-free rates: The Time 2 rates need to be lowered slightly to get the price up to 100.000.

Exhibit 19 Calibration of Time 2 Forward Rates

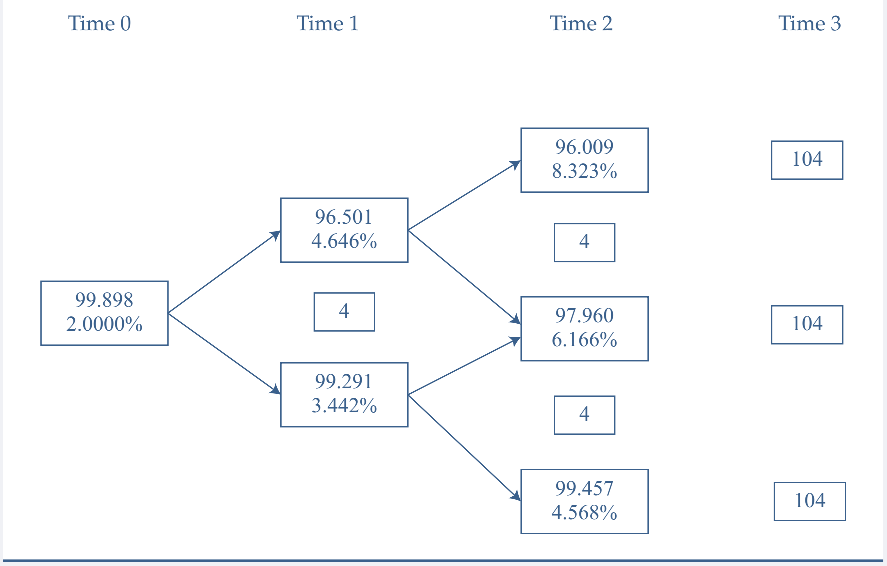


Exhibit 20 displays the completed binomial tree. The calibrated forward rates for Time 2 are 8.167%, 6.050%, and 4.482%. These are the calculations:

104/1.08167 = 96.148.

104/1.06050 = 98.067.

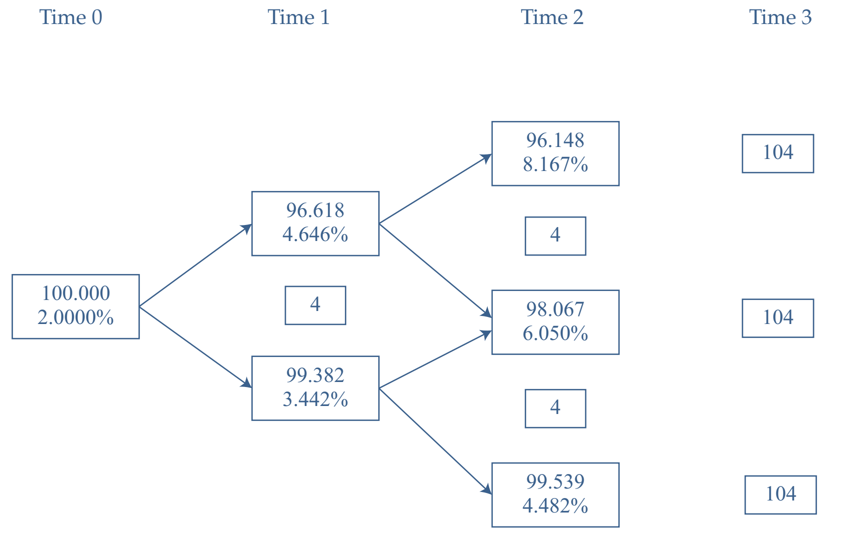
104/1.04482 = 99.538.

$$\frac{4 + (0.5 \times 96.148 + 0.5 \times 98.067)}{1.04646} = 96.618.$$

$$\frac{4 + (0.5 \times 98.067 + 0.5 \times 99.539)}{1.03442} = 99.382.$$

$$\frac{4 + (0.5 \times 96.618 + 0.5 \times 99.382)}{1.02000} = 100.000.$$

Exhibit 20 Completed Binomial Tree



Now that our tree gives the correct prices for the underlying par bonds maturing in one, two, and three years, we say that our tree is calibrated to be arbitrage free. It will price option-free bonds correctly, including prices for the zero-coupon bonds used to find the spot rates, and to the extent that we have chosen an appropriate interest rate process and interest rate volatility, it will provide insights into the value of bonds with embedded options and their risk parameters.

5 VALUING AN OPTION-FREE BOND WITH A BINOMIAL TREE

e describe the backward induction valuation methodology and calculate the value of a fixed-income instrument given its cash flow at each node

f compare pricing using the zero-coupon yield curve with pricing using an arbitrage-free binomial lattice

Our next task is twofold. First, we calculate the arbitrage-free value of an option-free, fixed-rate coupon bond. Second, we compare the pricing using the zero-coupon yield curve with the pricing using an arbitrage-free binomial lattice. Because these two valuation methods are arbitrage free, these two values must be the same.

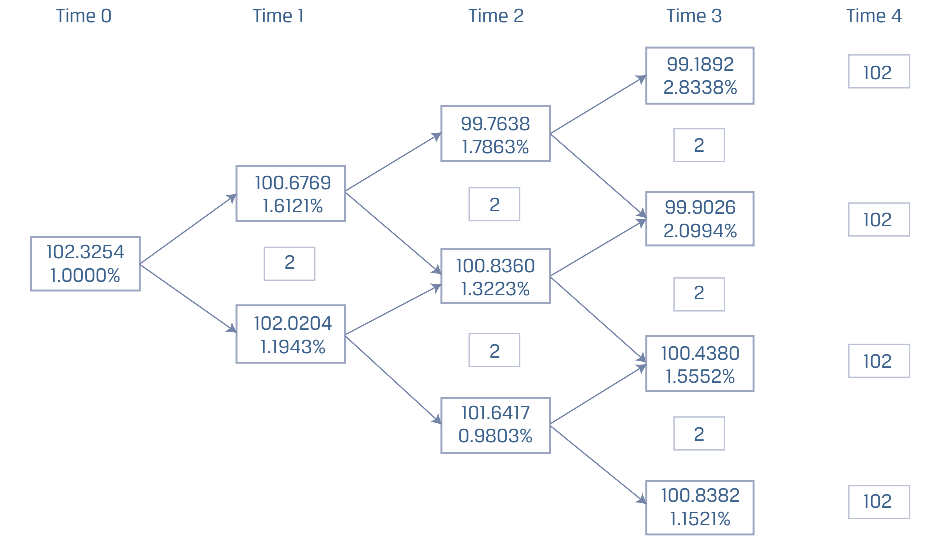
Now, consider an option-free bond with four years remaining to maturity and a coupon rate of 2%. Note that this is not a benchmark bond and it carries a higher coupon and price than the four-year benchmark bond, which is priced at par. The value of this bond can be calculated by discounting the cash flow at the spot rates in Exhibit 3 as shown in the following equation:

$$\frac{2}{(1.01)^1} + \frac{2}{(1.012012)^2} + \frac{2}{(1.012515)^3} + \frac{102}{(1.014044)^4} = 102.3254.$$

The binomial interest rate tree should produce the same value as when discounting the cash flows with the spot rates. An option-free bond that is valued by using the binomial interest rate tree should have the same value as when discounting by the spot rates, which is true because the binomial interest rate tree is arbitrage free.

Let us give the tree a test run and use the 2% option-free bond with four years remaining to maturity. Also assume that the issuer's benchmark yield curve is the one given in Exhibit 2; hence the appropriate binomial interest rate tree is the one in Exhibit 10. Exhibit 21 shows the various values in the discounting process and obtains a bond value of 102.3254. The tree produces the same value for the bond as the spot rates produce and is therefore consistent with our standard valuation model.

Exhibit 21 Sample Valuation for an Option-Free Bond using a Binomial Tree



EXAMPLE 5

Confirming the Arbitrage-Free Value of a Bond

Using the par curve from Example 2 and Example 4, the yield-to -maturity for a one-year annual coupon bond is 2%, for a two-year annual coupon bond is 3%, and for a three-year annual coupon bond is 4%. Because this is the same curve as that used in Example 4, we can use the calibrated tree from that example to price a bond. Let us use a three-year annual coupon bond with a 5% coupon, just as we did in Example 2. We know that if the calibrated tree was built correctly and we perform calculations to value the bond with the tree shown in Exhibit 22, its price should be 102.8105.

Exhibit 22 Binomial Tree from Example 5

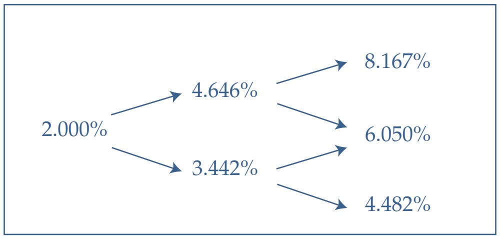
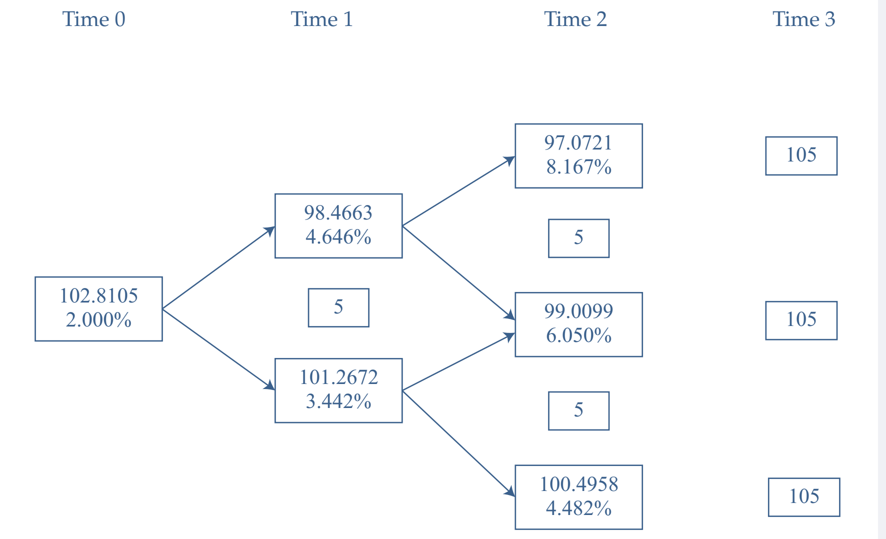


Exhibit 23 Valuing a 5%, Three-Year Bond



Because the tree was calibrated to the same par curve (and spot curve) that was used to price this option-free bond using spot rates only, the tree gives the same price as the spot rate pricing (the small difference is due to rounding).

6 VALUING AN OPTION-FREE BOND WITH PATHWISE VALUATION

g describe pathwise valuation in a binomial interest rate framework and calculate the value of a fixed-income instrument given its cash flows along each path

Pathwise valuation is an alternative approach to backward induction in a binomial tree. The binomial interest rate tree specifies all potential rate paths in the model, whereas an interest rate path is the route an interest rate takes from the current time to the security's maturity. Pathwise valuation calculates the present value of a bond for each possible interest rate path and takes the average of these values across paths. We will use the pathwise valuation approach to produce the same value as the backward induction method for an option-free bond. Pathwise valuation involves the following steps: (1) Specify a list of all potential paths through the tree, (2) determine the present value of a bond along each potential path, and (3) calculate the average across all possible paths.

Determining all potential paths is similar to the following experiment. Suppose you are tossing a fair coin and tracking how many ways heads and tails can be combined. We will use a device called Pascal's Triangle, displayed in Exhibit 24. Pascal's Triangle can be built as follows: Start with the number 1 at the top of the triangle. The numbers in the boxes below are the sum of the two numbers above it except that the edges on each side are all 1. The shaded numbers show that 3 is the sum of 2 and 1. Now toss the coin while keeping track of the possible outcomes. The possible groupings are listed in Exhibit 25, where H stands for heads and T stands for tails.

Exhibit 24 Pascal's Triangle

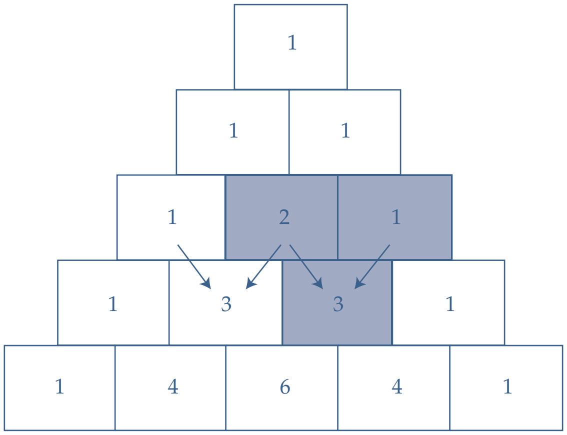


Exhibit 25 Possible Outcomes of Coin Tosses

Number of Tosses	Possible Outcomes	Pascal's Triangle
1	H T	1, 1
2	HH HT TH TT	1,2,1
3	HHH HHT HTH THH HTT THT TTH TTT	1, 3, 3, 1

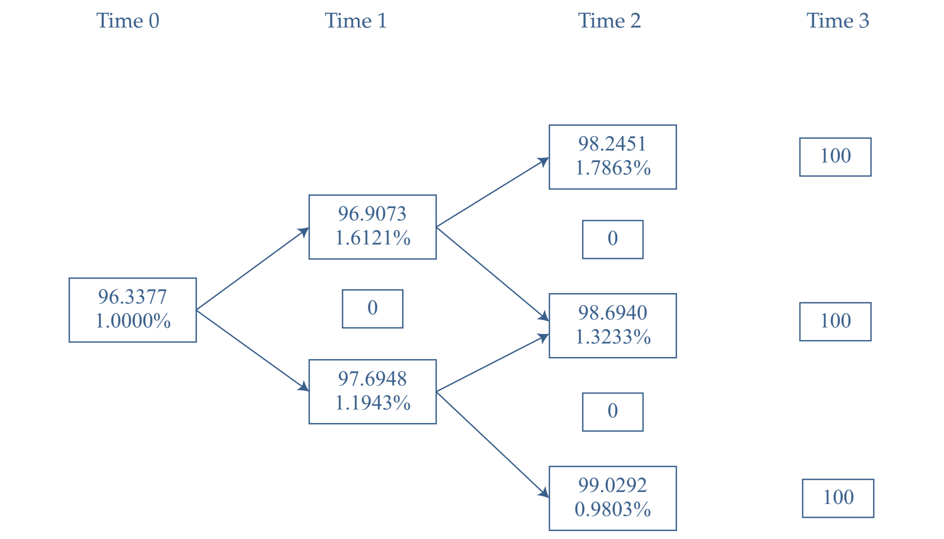
This experiment mirrors exactly the number of interest rate paths in our binomial interest rate tree. The total number of paths for each period/year can be easily determined by using Pascal's Triangle. Let us work through an example for a three-year zero-coupon bond. From Pascal's Triangle, there are four possible paths to arrive at Year 3: HH, HT, TH, TT. Using the same binomial tree from Exhibit 21, we specify the four paths as well as the possible forward rates along those paths. In Exhibit 26, the last column on the right shows the present value for each path. For example, 100/(1.01000 × 1.016121 × 1.017863) = 95.7291. In the bottom right corner is the average present value across all paths.

Exhibit 26 Four Interest Rate Paths for a Three-Year Zero-Coupon Bond

Path	Forward Rate Year 1	Forward Rate Year 2	Forward Rate Year 3	Present Value
1	1.0000%	1.6121%	1.7863%	95.7291
2	1.0000%	1.6121%	1.3233%	96.1665
3	1.0000%	1.1943%	1.3233%	96.5636
4	1.0000%	1.1943%	0.9803%	96.8916
				96.3377

Now, we can use the binomial tree to confirm our calculations for the three-year zero-coupon bond. The analysis is presented in Exhibit 27. The interest rate tree does indeed produce the same value.

Exhibit 27 Binomial Tree to Confirm Bond's Value



EXAMPLE 6

Pathwise Valuation Based on a Binomial Interest Rate Tree

Using the par curve from Example 2, Example 4, and Example 5, the yield-to-maturity for a one-year annual coupon bond is 2%, for a two-year annual coupon bond is 3%, and for a three-year annual coupon bond is 4%. We know that if we generate the paths in the tree correctly and discount the cash flows directly, the three-year, 5% annual coupon bond should still be priced at 102.8105, as calculated in Example 5.

There are four paths through the three-year tree. We discount the cash flows along each of the four paths and take their average, as shown in Exhibits 28, 29, and 30.

Exhibit 28 Cash Flows

Path	Time 0	Time 1	Time 2	Time 3
1	0	5	5	105
2	0	5	5	105
3	0	5	5	105
4	0	5	5	105

Exhibit 29 Discount Rates

Path	Time 0	Time 1	Time 2	Time 3
1	2.000%	4.646%	8.167%	
2	2.000%	4.646%	6.050%	
3	2.000%	3.442%	6.050%	
4	2.000%	3.442%	4.482%	

Exhibit 30 Present Values

Path	Time 0
1	100.5298
2	102.3452
3	103.4794
4	104.8877
Average	**102.8105**

The present values are calculated by discounting the cash flows in Exhibit 28 by the forward rates in Exhibit 29. For example, the present value for the bond along Path 1 is 100.5298:

$$\frac{5}{1.02}+\frac{5}{(1.02)(1.04646)}+\frac{105}{(1.02)(1.04646)(1.08167)}=100.5298.$$

The present value along Path 3 is 103.4794:

$$\frac{5}{1.02}+\frac{5}{(1.02)(1.03442)}+\frac{105}{(1.02)(1.03442)(1.06050)}=103.4794.$$

The average for the bond prices using pathwise valuation is 102.8105, which matches the result obtained using backward induction in Exhibit 23.

7 THE MONTE CARLO METHOD

h describe a Monte Carlo forward-rate simulation and its application

The Monte Carlo method is an alternative method for simulating a sufficiently large number of potential interest rate paths to discover how the value of a security is affected. This method involves randomly selecting paths to approximate the results of a complete pathwise valuation. Monte Carlo methods are often used when a security's cash flows are path dependent. Cash flows are path dependent when the cash flow to be received depends on the path followed to reach its current level as well as the current level itself. For example, the valuation of mortgage-backed securities depends to a great extent on the level of prepayments. As mentioned in an earlier reading, prepayments tend to increase when interest rates fall, because borrowers are more likely to pay off mortgage loans and refinance at lower interest rates. Interest rate paths are generated on the basis of some probability distribution and a volatility assumption, and the model is fit to the current benchmark term structure of interest rates. The benchmark term structure is represented by the current spot rate curve such that the average present value across all scenario interest rate paths for each benchmark bond equals its actual market value. By using this approach, the model is rendered arbitrage free, which is equivalent to calibrating the interest rate tree as discussed in Section 3.

Suppose we intend to value with the Monte Carlo method a 30-year bond that has monthly coupon payments (e.g., mortgage-backed securities). The following steps are taken: (1) Simulate numerous (say, 500) paths of one-month interest rates under a volatility assumption and probability distribution, (2) generate spot rates from the simulated future one-month interest rates, (3) determine the cash flow along each interest rate path, (4) calculate the present value for each path, and (5) calculate the average present value across all interest rate paths.

Using the procedure just described, the model will produce benchmark bond values equal to the market prices only by chance. We want to ensure this is the case; otherwise the model will neither fit the current spot curve nor be arbitrage free. A constant is added to all interest rates on all paths such that the average present value for each benchmark bond equals its market value. The constant added to all short interest rates is called a drift term. When this technique is used, the model is said to be drift adjusted.

How many paths are appropriate for the Monte Carlo method? More paths increase the accuracy of the estimate in a statistical sense, but this does not mean the model is closer to the true fundamental value of the security. The Monte Carlo method is only as good as the valuation model used and the accuracy of the inputs.

Yield curve modelers also often include mean reversion in their Monte Carlo estimation. Mean reversion starts with the common-sense notion that history suggests that interest rates almost never get "too high" or "too low." What is meant by "too high" and "too low" is left to the discretion of the modeler. We implement mean reversion by implementing upper and lower bounds on the random process generating future interest rates. Mean reversion has the effect of moving the interest rate toward the implied forward rates from the yield curve.

EXAMPLE 7

The Application of Monte Carlo Simulation to Bond Pricing

Replace the interest rate paths from Example 6 with randomly generated paths calibrated to the same initial par and spot curves, as shown in Exhibit 31.

Exhibit 31 Discount Rates

Path	Time 0	Time 1	Time 2
1	2.000%	2.500%	4.548%
2	2.000%	3.600%	6.116%
3	2.000%	4.600%	7.766%
4	2.000%	5.500%	3.466%
5	2.000%	3.100%	8.233%
6	2.000%	4.500%	6.116%
7	2.000%	3.800%	5.866%
8	2.000%	4.000%	8.233%

Exhibit 32 Present Values

Path	Time 0
1	105.7459
2	103.2708
3	100.9104
4	103.8543
5	101.9075
6	102.4236
7	103.3020
8	101.0680
Average	**102.8103**

Because we continue to get 102.8103, as shown in Exhibit 32, as the price for our three-year, 5% annual coupon bond, we know that the Monte Carlo simulation has been calibrated correctly. The paths are now different enough such that path-dependent securities, such as mortgage-backed securities, can be analyzed in ways that provide insights not possible in binomial trees, because Monte Carlo techniques provide greater flexibility to change parameters over time.

TERM STRUCTURE MODELS 8

i. describe term structure models and how they are used

Term structure models provide quantitatively precise descriptions of how interest rates evolve. A model provides a simplified description of a real-world phenomenon on the basis of a set of assumptions. These assumptions cannot be completely accurate in depicting the real world but are necessary for analytical tractability. Despite simplifying assumptions, models explain real-world phenomena sufficiently well to be useful for pricing and hedging.

The binomial tree and Monte Carlo simulation valuation approaches for complex fixed-income instruments described earlier rely on specific assumptions about the underlying asset properties. For example, how do we establish the node values in the binomial trees, and what determines the dispersion in rates from the top to the bottom nodes? This answer comes from term structure models, which make assumptions about the properties of rates over time and then use those properties to "fit," or determine the values of the rates at each node, binomial lattices used for pricing and risk management applications. The following section introduces common term structure models, with an emphasis on the underlying assumptions about the statistical properties of interest rates. Each of the models can be "fit" to lattice models for valuation and risk management applications.

Modeling the future path of interest rates is not only critical for scenario analysis and stress testing individual bonds and bond portfolio values but also important in the valuation of complex fixed-income instruments. A detailed description of these models depends on mathematical and statistical knowledge beyond the scope of this reading, but fixed-income practitioners will often find that these or other term structure models are embedded in many of the desktop tools and data analytics software they may use during their investment industry career. Thus, we provide a broad overview of these models in this reading.

8.1 Model Choice

Term structure models go beyond the lognormal random walk approach used earlier to describe the dynamics of the term structure for the purpose of pricing and hedging fixed-income securities and derivatives. All term structure models make simplifying assumptions about the evolution of rates over time. Many different interest rate models that differ in their assumptions exist. Arguably, there are many models, since no one model perfectly captures interest rate dynamics. Modelers face a trade-off between simplicity and accuracy when selecting a term structure model. Practitioners should be aware of the categories of models and their important features (which stem from their assumptions) as well as how those features affect pricing and hedging.

8.1.1 *Interest rate factors*

The valuation and hedging of fixed-income securities and their derivatives require information across the entire term structure. To develop a term structure model useful for pricing and hedging applications, we focus on modeling the factors that determine the term structure. The simplest class of models use one factor—the short rate, or the one-period rate—as the factor that drives the term structure. Although the use of one factor may seem limiting, because it implies all rates move in the same direction during any short time interval, it does not mean they have to move by the same amounts. Multi-factor models incorporate additional factors, such as the slope of the term structure, with the complexity of the models increasing in the number of factors.

8.1.2 *Interest rate process*

Term structure models use stochastic processes to describe interest rate dynamics. These stochastic processes have two components: a drift term and an uncertain, or stochastic, term. Although the stochastic processes are continuous time, the models can be "fit" to binomial or trinomial interest rate lattices using a discrete version of the models (integrating over time to obtain rates that span time intervals).

For a one-factor model, the general form of the process describing the short rate's (r) dynamics is

$$dr = \theta_t dt + \sigma_t dZ.$$

The drift term, $\theta_t dt$, describes the expected (zero-volatility) rate path. For example, in a one-factor model of the short rate, the drift describes the expected evolution of the short rate over time. The drift term may be constant or mean reverting.

The second term, $\sigma_t dZ$, adds randomness, or volatility, to the process. This dispersion term allows for the pricing of bonds with option features as well as interest rate derivatives and may take a variety of forms. The term Z is a Weiner process that is distributed normally. Given the symmetry of the normal distribution, it is possible and quite common for these models to produce interest rate paths with negative rates.

Within classes of models, such as one-factor no-arbitrage models, the key differences between the various models involve the stochastic difference equation.

8.1.3 *Class of model*

One class of models uses the arbitrage-free approach combined with assumptions about the statistical properties of interest rates. This class of models is referred to as no-arbitrage term structure models, where no-arbitrage is synonymous with arbitrage free. No-arbitrage term structure models begin with a set of assumptions about the term structure—a factor (or factors) and the stochastic process describing the factor evolution(s)—and take the term structure as given, assuming that both bond prices and the term structure bootstrapped from those prices are correct. The no-arbitrage models are "parameterized," which is the process of determining the values of the variables in the model such that those parameters produce bond prices that match current market prices. These models are used widely in practice and are often favored by practitioners since their pricing results are consistent with market prices.

Equilibrium term structure models seek to describe term structure dynamics using fundamental economic variables that are assumed to affect interest rates. The modeling process imposes restrictions that allow for the derivation of equilibrium prices for bonds and interest rate options.

Although equilibrium models use similar continuous stochastic difference equations to describe interest rate changes, equilibrium model parameters are not forced to values that produce bond prices consistent with current market prices. This property is seen by some market participants as a significant drawback in a static setting, such as pricing and hedging for the current time. However, other practitioners prefer equilibrium models since they capture not just the current market environment as reflected in the term structure but also the possibility of many different future paths. For more dynamic applications, equilibrium models may be preferred.

The best-known equilibrium models are the **Cox-Ingersoll-Ross model** (Cox, Ingersoll, and Ross 1985) and the **Vasicek model** (Vasicek 1977), discussed in the next two sections. Both the Vasicek and Cox–Ingersoll–Ross (CIR) models assume a single factor, the short-term interest rate, r_t. This approach is plausible because empirically, parallel shifts are often found to explain more than 90% of yield changes. In contrast, multifactor models may be able to model the curvature of a yield curve more accurately, but at the cost of greater complexity.

The reason that no-arbitrage models fit the current term structure is their greater number of parameters. These added parameters increase the computational requirements for estimation, which some practitioners find to be undesirable.

Other contrasts are more technical. They include that equilibrium models use real probabilities, whereas arbitrage-free models use so-called risk-neutral probabilities. An excellent example of an equilibrium term structure model is the Cox–Ingersoll–Ross model, discussed next.

8.2 Equilibrium Models

This section introduces the Cox–Ingersoll–Ross and Vasicek interest rate models.

8.2.1 *The Cox–Ingersoll–Ross model*

The Cox–Ingersoll–Ross (CIR) model assumes interest rates follow a mean-reverting process. However, the variance of rate changes differs depending on the level of rates. The CIR model uses the following formula to describe the interest rate process:

$$dr_t = k(\theta - r_t)dt + \sigma\sqrt{r_t}\,dZ.$$

Note that the drift term has three components. The level of rates at time t is r_t, and θ is the long-run mean rate, so their difference is the distance of the rate from its mean. The drift term equals zero if the rate is at the long-run mean, or $r_t = \theta$. The remaining drift term parameter, k, modulates the speed at which the rate reverts to its mean.

Another important feature of the CIR model is that the random component varies as rates change. In other words, the short-rate volatility is a function of the short rate. Importantly, at low rates, r_t, the term becomes small, which prevents rates from turning negative.

8.2.2 *The Vasicek model*

Although not developed in the context of a general equilibrium of individuals seeking to optimize consumption and investment decisions, as was the case for the CIR model, the Vasicek model is viewed as an equilibrium term structure model. Similar to the CIR model, the Vasicek model includes mean reversion. The Vasicek model uses the following equation to describe the interest rate process:

$$dr_t = k\left(\theta - r_t\right)dt + \sigma dZ.$$

The Vasicek model has the same drift term as the CIR model and thus tends toward mean reversion in the short rate. The stochastic or volatility term follows a random normal distribution for which the mean is zero and the standard deviation is 1. Unlike the CIR model, interest rates are calculated assuming constant volatility over the period of analysis. As with the CIR model, there is only one stochastic driver of the interest rate process. A key characteristic of the Vasicek model worth noting is that it is theoretically possible for the interest rate to become negative.

8.3 Arbitrage-Free Models

We will next illustrate two foundational no-arbitrage term structure models. There are many additional no-arbitrage models, but the basic features are similar, with differences stemming from different assumed interest rate processes.

8.3.1 *The Ho–Lee model*

In **arbitrage-free models**, the analysis begins with the current term structure, extrapolated from the market prices of a reference set of financial instruments. A maintained assumption is that the reference bonds are priced correctly. Unlike general equilibrium models, which have only a few parameters and can thus match only a few term structure points, arbitrage-free models allow the parameters to vary deterministically with time, creating a greater number of parameters and thus more points of match. As a result, the market yield curve can be modeled with the accuracy needed for such applications as valuing derivatives and bonds with embedded options.

The first arbitrage-free model was introduced by Ho and Lee (1986). The model is calibrated to market data and uses a binomial lattice approach to generate a distribution of possible future interest rates. In the **Ho–Lee model**, the short rate follows a normal process, as follows:

$$dr_t = \theta_t dt + \sigma dZ.$$

We see that the drift term, θ_t, is time dependent. This time dependency means there is a value for θ_t at each time step, which is critical for the model to produce prices that match market prices.

The Ho–Lee model, similar to the Vasicek model, has constant volatility, and interest rates may become negative because of the symmetry of the normal distribution and the model's use of constant volatility.

8.3.2 *The Kalotay–Williams–Fabozzi model*

The **Kalotay–Williams–Fabozzi (KWF) model** is analogous to the Ho–Lee model in that it assumes constant drift, no mean reversion, and constant volatility. However, the stochastic differential equation describes the dynamics of the log of the short rate, and as a result, the log of the short rate is distributed normally, meaning the short rate itself is distributed lognormally.

The differential process for the KWF model is

$$d\ln(r_t) = \theta_t dt + \sigma dZ.$$

At first glance, the main implication of modeling the log of the short rate is that it will prevent negative rates. After further analysis, it becomes evident that there are pricing implications where interest rate option values are influenced by the tails of the rate distributions. Exhibit 33 summarizes the key differences between these term structure models.

Exhibit 33 Term Structure Model Summary

Model	Type	Short Rate	Drift Term	Volatility
CIR	Equilibrium	dr_t	Mean reversion at speed k	Varies with $\sqrt{r_t}$
Vasicek	Equilibrium	dr_t	Mean reversion at speed k	Constant
Ho–Lee	Arbitrage free	dr_t	Time dependent	Constant
KWF	Arbitrage free	$d\ln(r_t)$	Time dependent	Constant

8.4 Modern Models

The one-factor models presented thus far are the building blocks on which modern interest rate models rely. Some current models extend those models to include multiple factors, while others use sophisticated approaches that combine observed forward curves with volatilities extracted from interest rate option prices.

The Gauss+ model is a multi-factor interest rate model used extensively in valuation and hedging. The Gauss+ model incorporates short-, medium- and long-term rates. The long-term factor is mean reverting and reflects trends in macroeconomic variables. The medium-term rate also reverts to the long-run rate. The short-term rate does not exhibit a random component, which is consistent with the central bank controlling the short end of the rate curve. This results in a hump-shaped volatility curve across tenors, with medium-term rates being the most volatile.

Although there are many different term structure models, knowledge of the basic assumptions and design of the classic models helps professionals understand and adapt more sophisticated modern models.

Example 8 addresses several basic points about modern term structure models.

EXAMPLE 8

Term Structure Models

1 Which of the following would be expected to provide the *most* accurate modeling with respect to the observed term structure?

   A CIR model

   B Ho–Lee model

   C Vasicek model

2 Which of the following statements about the Vasicek model is *most* accurate? It has:

   A a single factor, the long rate.

   B a single factor, the short rate.

   C two factors, the short rate and the long rate.

3 The CIR model:

   A assumes interest rates are not mean reverting.

   B has a drift term that differs from that of the Vasicek model.

   C assumes interest rate volatility increases with increases in the level of interest rates.

Solution to 1:

B is correct. The CIR model and the Vasicek model are examples of equilibrium term structure models, whereas the Ho–Lee model is an example of an arbitrage-free term structure model. A benefit of arbitrage-free term structure models is that they are calibrated to the current term structure. In other words, the starting prices ascribed to securities are those currently found in the market. In contrast, equilibrium term structure models frequently generate term structures that are inconsistent with current market data.

Solution to 2:

B is correct. Use of the Vasicek model requires assumptions for the short-term interest rate, which are usually derived from more general assumptions about the state variables that describe the overall economy. Using the assumed process for the short-term rate, one can determine the yield on longer-term bonds by looking at the expected path of interest rates over time.

Solution to 3:

C is correct. The drift term of the CIR model is identical to that of the Vasicek model, and both models assume that interest rates are mean reverting. The major difference between the two models is that the CIR model assumes a rise in interest rate volatility as rates increase, while the Vasicek model assumes interest rate volatility is constant.

SUMMARY

This reading presents the principles and tools for arbitrage valuation of fixed-income securities. Much of the discussion centers on the binomial interest rate tree, which can be used extensively to value both option-free bonds and bonds with embedded options. The following are the main points made in the reading:

- A fundamental principle of valuation is that the value of any financial asset is equal to the present value of its expected future cash flows.
- A fixed-income security is a portfolio of zero-coupon bonds, each with its own discount rate that depends on the shape of the yield curve and when the cash flow is delivered in time.
- In well-functioning markets, prices adjust until there are no opportunities for arbitrage, or a transaction that involves no cash outlay yet results in a riskless profit.
- Using the arbitrage-free approach, viewing a security as a package of zero-coupon bonds means that two bonds with the same maturity and different coupon rates are viewed as different packages of zero-coupon bonds and valued accordingly.
- For bonds that are option-free, an arbitrage-free value is simply the present value of expected future values using the benchmark spot rates.
- A binomial interest rate tree permits the short interest rate to take on one of two possible values consistent with the volatility assumption and an interest rate model based on a lognormal random walk.
- An interest rate tree is a visual representation of the possible values of interest rates (forward rates) based on an interest rate model and an assumption about interest rate volatility.
- The possible interest rates for any following period are consistent with the following three assumptions: (1) an interest rate model that governs the random process of interest rates, (2) the assumed level of interest rate volatility, and (3) the current benchmark yield curve.
- From the lognormal distribution, adjacent interest rates on the tree are multiples of e raised to the 2σ power, with the absolute change in interest rates becoming smaller and smaller as rates approach zero.
- We use the backward induction valuation methodology that involves starting at maturity, filling in those values, and working back from right to left to find the bond's value at the desired node.
- The interest rate tree is fit to the current yield curve by choosing interest rates that result in the benchmark bond value. By doing this, the bond value is arbitrage free.

- An option-free bond that is valued by using the binomial interest rate tree should have the same value as when discounting by the spot rates.
- Pathwise valuation calculates the present value of a bond for each possible interest rate path and takes the average of these values across paths.
- The Monte Carlo method is an alternative method for simulating a sufficiently large number of potential interest rate paths in an effort to discover how the value of a security is affected, and it involves randomly selecting paths in an effort to approximate the results of a complete pathwise valuation.
- Term structure models seek to explain the yield curve shape and are used to value bonds (including those with embedded options) and bond-related derivatives. General equilibrium and arbitrage-free models are the two major types of such models.
- Arbitrage-free models are frequently used to value bonds with embedded options. Unlike equilibrium models, arbitrage-free models begin with the observed market prices of a reference set of financial instruments, and the underlying assumption is that the reference set is correctly priced.

REFERENCES

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross. 1985. “An Intertemporal General Equilibrium Model of Asset Prices.” Econometrica 53 (March): 363–84. doi:10.2307/1911241

Ho, Thomas S. Y., and Bin Lee Sang. 1986. “Term Structure Movements and Pricing Interest Rate Contingent Claims.” Journal of Finance 41 (December): 1011–29. doi:10.1111/j.1540-6261.1986.tb02528.x

Vasicek, Oldrich. 1977. “An Equilibrium Characterization of the Term Structure.” Journal of Financial Economics 5 (November): 177–88. doi:10.1016/0304-405X(77)90016-2

PRACTICE PROBLEMS

The following information relates to Questions 1–6

Katrina Black, a portfolio manager at Coral Bond Management, Ltd., is conducting a training session with Alex Sun, a junior analyst in the fixed-income department. Black wants to explain to Sun the arbitrage-free valuation framework used by the firm. Black presents Sun with Exhibit 1, showing a fictitious bond being traded on three exchanges, and asks Sun to identify the arbitrage opportunity of the bond. Sun agrees to ignore transaction costs in his analysis.

Exhibit 1 Three-Year, €100 par, 3.00% Coupon, Annual Pay Option-Free Bond

	Eurex	NYSE Euronext	Frankfurt
Price	€103.7956	€103.7815	€103.7565

Black shows Sun some exhibits that were part of a recent presentation. Exhibit 3 presents most of the data of a binomial lognormal interest rate tree fit to the yield curve shown in Exhibit 2. Exhibit 4 presents most of the data of the implied values for a four-year, option-free, annual pay bond with a 2.5% coupon based on the information in Exhibit 3.

Exhibit 2 Yield-to-Maturity Par Rates for One-, Two-, and Three-Year Annual Pay Option-Free Bonds

One-year	Two-year	Three-year
1.25%	1.50%	1.70%

© 2021 CFA Institute. All rights reserved.

Exhibit 3 Binomial Interest Rate Tree Fit to the Yield Curve (Volatility = 10%)

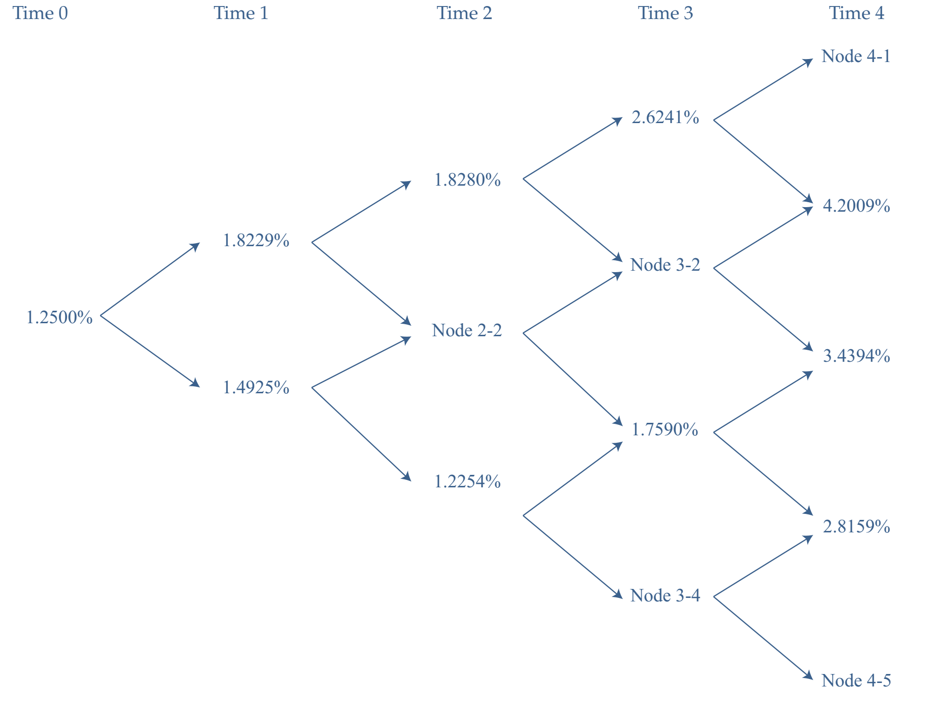
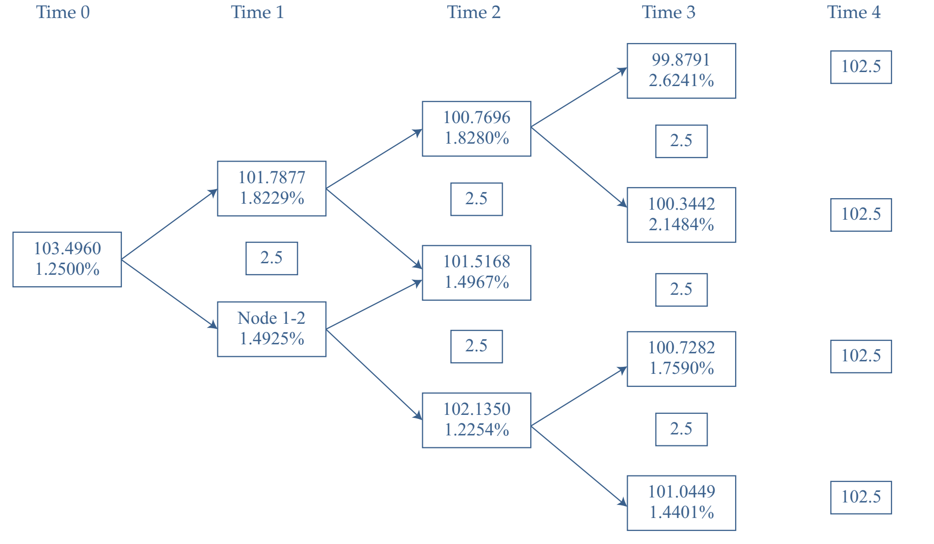


Exhibit 4 Implied Values (in Euros) for a 2.5%, Four-Year, Option-Free, Annual Pay Bond Based on Exhibit 3



Black asks about the missing data in Exhibits 3 and 4 and directs Sun to complete the following tasks related to those exhibits:

Task 1	Test that the binomial interest tree has been properly calibrated to be arbitrage free.
Task 2	Develop a spreadsheet model to calculate pathwise valuations. To test the accuracy of the spreadsheet, use the data in Exhibit 3 and calculate the value of the bond if it takes a path of lowest rates in Year 1 and Year 2 and the second lowest rate in Year 3.
Task 3	Identify a type of bond where the Monte Carlo calibration method should be used in place of the binomial interest rate method.
Task 4	Update Exhibit 3 to reflect the current volatility, which is now 15%.

1 Based on Exhibit 1, the *best* action that an investor should take to profit from the arbitrage opportunity is to:

A buy on Frankfurt, sell on Eurex.

B buy on NYSE Euronext, sell on Eurex.

C buy on Frankfurt, sell on NYSE Euronext.

2 Based on Exhibits 1 and 2, the exchange that reflects the arbitrage-free price of the bond is:

A Eurex.

B Frankfurt.

C NYSE Euronext.

3 Recall from the reading that each node is represented by both a time element and a rate change component. Which of the following statements about the missing data in Exhibit 3 is correct?

A Node 3–2 can be derived from Node 2–2.

B Node 4–1 should be equal to Node 4–5 multiplied by $e^{0.4}$.

C Node 2–2 approximates the implied one-year forward rate two years from now.

4 Based on the information in Exhibits 3 and 4, the bond price in euros at Node 1–2 in Exhibit 4 is *closest* to:

A 102.7917.

B 104.8640.

C 105.2917.

5 A benefit of performing Task 1 is that it:

A enables the model to price bonds with embedded options.

B identifies benchmark bonds that have been mispriced by the market.

C allows investors to realize arbitrage profits through stripping and reconstitution.

6 If the assumed volatility is changed as Black requested in Task 4, the forward rates shown in Exhibit 3 will *most likely*:

A spread out.

B remain unchanged.

C converge to the spot rates.

The following information relates to Questions 7–10

Betty Tatton is a fixed-income analyst with the hedge fund Sailboat Asset Management (SAM). SAM invests in a variety of global fixed-income strategies, including fixed-income arbitrage. Tatton is responsible for pricing individual investments and analyzing market data to assess the opportunity for arbitrage. She uses two methods to value bonds:

Method 1 Discount each year's cash flow separately using the appropriate interest rate curve.

Method 2 Build and use a binomial interest rate tree.

Tatton compiles pricing data for a list of annual pay bonds (Exhibit 1). Each of the bonds will mature in two years, and Tatton considers the bonds risk-free; both the one-year and two-year benchmark spot rates are 2%. Tatton calculates the arbitrage-free prices and identifies an arbitrage opportunity to recommend to her team.

Exhibit 1 Market Data for Selected Bonds

Asset	Coupon	Market Price
Bond A	1%	98.0584
Bond B	3%	100.9641
Bond C	5%	105.8247

Next, Tatton uses the benchmark yield curve provided in Exhibit 2 to consider arbitrage opportunities of both option-free corporate bonds and corporate bonds with embedded options. The benchmark bonds in Exhibit 2 pay coupons annually, and the bonds are priced at par.

Exhibit 2 Benchmark Par Curve

Maturity (years)	Yield-to-Maturity (YTM)
1	3.0%
2	4.0%
3	5.0%

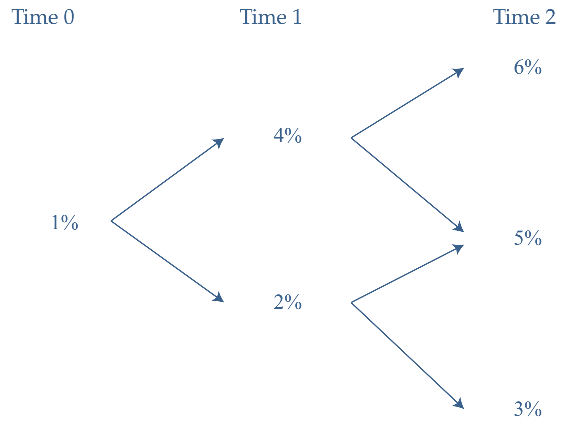
Tatton then identifies three mispriced three-year annual coupon bonds and compiles data on the bonds (see Exhibit 3).

Exhibit 3 Market Data of Annual Pay Corporate Bonds

Company	Coupon	Market Price	Yield	Embedded Option?
Hutto-Barkley Inc.	3%	94.9984	5.6%	No
Luna y Estrellas Intl.	0%	88.8996	4.0%	Yes
Peaton Scorpio Motors	0%	83.9619	6.0%	No

Lastly, Tatton identifies two mispriced Swiss bonds, Bond X, a three-year bond, and Bond Y, a five-year bond. Both are 6% annual coupon bonds. To calculate the bonds' values, Tatton devises the first three years of the interest rate lognormal tree presented in Exhibit 4 using historical interest rate volatility data. Tatton considers how these data would change if implied volatility, which is higher than historical volatility, were used instead.

Exhibit 4 Interest Rate Tree—Forward Rates Based on Swiss Market



7 Based on Exhibit 1, which of the following bonds *most likely* includes an arbitrage opportunity?

A Bond A

B Bond B

C Bond C

8 Based on Exhibits 2 and 3 and using Method 1, the amount (in absolute terms) by which the Hutto-Barkley Inc. corporate bond is mispriced is *closest* to:

A 0.3368 per 100 of par value.

B 0.4682 per 100 of par value.

C 0.5156 per 100 of par value.

9 Method 1 would *most likely* not be an appropriate valuation technique for the bond issued by:

A Hutto-Barkley Inc.

B Luna y Estrellas Intl.

C Peaton Scorpio Motors.

10 Based on Exhibit 4 and using Method 2, the correct price for Bond X is *closest* to:

A 97.2998.

B 109.0085.

C 115.0085.

The following information relates to Questions 11–18

Meredith Alvarez is a junior fixed-income analyst with Canzim Asset Management. Her supervisor, Stephanie Hartson, asks Alvarez to review the asset price and payoff data shown in Exhibit 1 to determine whether an arbitrage opportunity exists.

Exhibit 1 Price and Payoffs for Two Risk-Free Assets

Asset	Price Today	Payoff in One Year
Asset A	$500	$525
Asset B	$1,000	$1,100

Hartson also shows Alvarez data for a bond that trades in three different markets in the same currency. These data appear in Exhibit 2.

Exhibit 2 2% Coupon, Five-Year Maturity, Annual Pay Bond

	New York	Hong Kong	Mumbai
Yield-to-Maturity	1.9%	2.3%	2.0%

Hartson asks Alvarez to value two bonds (Bond C and Bond D) using the binomial tree in Exhibit 3. Exhibit 4 presents selected data for both bonds.

Exhibit 3 Binomial Interest Rate Tree with Volatility = 25%

Time 0	Time 1	Time 2
		2.7183%
	2.8853%	
1.500%		1.6487%
	1.7500%	
		1.0000%

Exhibit 4 Selected Data on Annual Pay Bonds

Bond	Maturity	Coupon Rate
Bond C	2 years	2.5%
Bond D	3 years	3.0%

Hartson tells Alvarez that she and her peers have been debating various viewpoints regarding the conditions underlying binomial interest rate trees. The following statements were made in the course of the debate.

Statement 1 The only requirements needed to create a binomial interest rate tree are current benchmark interest rates and an assumption about interest rate volatility.

Statement 2 Potential interest rate volatility in a binomial interest rate tree can be estimated using historical interest rate volatility or observed market prices from interest rate derivatives.

Statement 3 A bond value derived from a binomial interest rate tree with a relatively high volatility assumption will be different from the value calculated by discounting the bond's cash flows using current spot rates.

Based on data in Exhibit 5, Hartson asks Alvarez to calibrate a binomial interest rate tree starting with the calculation of implied forward rates shown in Exhibit 6.

Exhibit 5 Selected Data for a Binomial Interest Rate Tree

Maturity	Par Rate	Spot Rate
1	2.5000%	2.5000%
2	3.5000%	3.5177%

Exhibit 6 Calibration of Binomial Interest Rate Tree with Volatility = 25%

Time 0	Time 1
	5.8365%
2.500%	
	Lower one-period forward rate

Hartson mentions pathwise valuations as another method to value bonds using a binomial interest rate tree. Using the binomial interest rate tree in Exhibit 3, Alvarez calculates the possible interest rate paths for Bond D shown in Exhibit 7.

Exhibit 7 Interest Rate Paths for Bond D

Path	Time 0	Time 1	Time 2
1	1.500%	2.8853%	2.7183%
2	1.500	2.8853	1.6487
3	1.500	1.7500	1.6487
4	1.500	1.7500	1.0000

Before leaving for the day, Hartson asks Alvarez about the value of using the Monte Carlo method to simulate a large number of potential interest rate paths to value a bond. Alvarez makes the following statements.

Statement 4	Increasing the number of paths increases the estimate's statistical accuracy.
Statement 5	The bond value derived from a Monte Carlo simulation will be closer to the bond's true fundamental value.

11 Based on Exhibit 1, Alvarez finds that an arbitrage opportunity is:

A not available.

B available based on the dominance principle.

C available based on the value additivity principle.

12 Based on the data in Exhibit 2, the *most* profitable arbitrage opportunity would be to buy the bond in:

A Mumbai and sell it in Hong Kong.

B Hong Kong and sell it in New York.

C New York and sell it in Hong Kong.

13 Based on Exhibits 3 and 4, the value of Bond C at the upper node at Time 1 is *closest* to:

A 97.1957.

B 99.6255.

C 102.1255.

14 Based on Exhibits 3 and 4, the price for Bond D is *closest* to:

A 97.4785.

B 103.3230.

C 106.3230.

15 Which of the various statements regarding binomial interest rate trees is correct?

A Statement 1

B Statement 2

C Statement 3

16 Based on Exhibits 5 and 6, the value of the lower one-period forward rate is *closest to*:

A 3.5122%.

B 3.5400%.

C 4.8037%.

17 Based on Exhibits 4 and 7, the present value of Bond D's cash flows following Path 2 is *closest* to:

A 97.0322.

B 102.8607.

C 105.8607.

18 Which of the statements regarding Monte Carlo simulation is correct?

A Only Statement 4 is correct.

B Only Statement 5 is correct.

C Both Statement 4 and Statement 5 are correct.

19 Which term structure model can be calibrated to closely fit an observed yield curve?

A The Ho–Lee model

B The Vasicek model

C The Cox–Ingersoll–Ross model

The following information relates to Questions 20–21

Keisha Jones is a junior analyst at Sparling Capital. Julie Anderson, a senior partner and Jones's manager, meets with Jones to discuss interest rate models used for the firm's fixed-income portfolio.

Anderson begins the meeting by asking Jones to describe features of equilibrium and arbitrage-free term structure models. Jones responds by making the following statements:

Statement 1 Equilibrium term structure models are factor models that use the observed market prices of a reference set of financial instruments, assumed to be correctly priced, to model the market yield curve.

Statement 2 In contrast, arbitrage-free term structure models seek to describe the dynamics of the term structure by using fundamental economic variables that are assumed to affect interest rates.

Anderson then asks Jones about her preferences concerning term structure models. Jones states:

> I prefer arbitrage-free models. Even though equilibrium models require fewer parameters to be estimated relative to arbitrage-free models, arbitrage-free models allow for time-varying parameters. In general, this allowance leads to arbitrage-free models being able to model the market yield curve more precisely than equilibrium models.

20 Which of Jones's statements regarding equilibrium and arbitrage-free term structure models is *incorrect*?

A Statement 1 only

B Statement 2 only

C Both Statement 1 and Statement 2

21 Is Jones correct in describing key differences in equilibrium and arbitrage-free models as they relate to the number of parameters and model accuracy?

A Yes

B No, she is incorrect about which type of model requires fewer parameter estimates.

C No, she is incorrect about which type of model is more precise at modeling market yield curves.

22 Which of the following statements comparing the Ho–Lee and Kalotay–Williams–Fabozzi (KWF) equilibrium term structure models is *correct*?

A The Ho–Lee model assumes constant volatility, while the KWF model does not.

B The KWF model incorporates the possibility of negative rates, while the Ho–Lee model does not.

C The KWF model describes the log of the dynamics of the short rate, while the Ho–Lee model does not.

SOLUTIONS

1 A is correct. This is the same bond being sold at three different prices, so an arbitrage opportunity exists by buying the bond from the exchange where it is priced lowest and immediately selling it on the exchange that has the highest price. Accordingly, an investor would maximize profit from the arbitrage opportunity by buying the bond on the Frankfurt exchange (which has the lowest price, €103.7565) and selling it on the Eurex exchange (which has the highest price, €103.7956) to generate a risk-free profit of €0.0391 (as mentioned, ignoring transaction costs) per €100 par.

B is incorrect because buying on NYSE Euronext and selling on Eurex would result in a €0.0141 profit per €100 par (€103.7956 – €103.7815 = €0.0141), which is not the maximum arbitrage profit available. A greater profit would be realized if the bond were purchased in Frankfurt and sold on Eurex.

C is incorrect because buying on Frankfurt and selling on NYSE Euronext would result in an €0.0250 profit per €100 par (€103.7815 – €103.7565 = €0.0250). A greater profit would be realized if the bond were purchased in Frankfurt and sold on Eurex.

2 C is correct. The bond from Exhibit 1 is selling for its calculated value on the NYSE Euronext exchange. The arbitrage-free value of a bond is the present value of its cash flows discounted by the spot rate for zero-coupon bonds maturing on the same date as each cash flow. The value of this bond, 103.7815, is calculated as follows:

	Year 1	Year 2	Year 3	Total PV
Yield-to-maturity	1.2500%	1.500%	1.700%	
Spot rate[1]	1.2500%	1.5019%	1.7049%	
Cash flow	3.00	3.00	103.00	
Present value of payment[2]	2.9630	2.9119	97.9066	103.7815

	Eurex	NYSE Euronext	Frankfurt
Price	€103.7956	€103.7815	€103.7565
Mispricing (per 100 par value)	0.141	0	–0.025

Notes:
(1) Spot rates are calculated using bootstrapping. For example, Year 2 spot rate (z_2): $100 = 1.5/1.0125 + 101.5/(1 + z_2)^2$; $z_2 = 0.015019$.
(2) Present value calculated using the formula $PV = FV/(1 + r)^n$, where n = number of years until cash flow, FV = cash flow amount, and r = spot rate.

A is incorrect because the price on the Eurex exchange, €103.7956, was calculated using the yield-to-maturity rate to discount the cash flows when the spot rates should have been used. C is incorrect because the price on the Frankfurt exchange, €103.7565, uses the Year 3 spot rate to discount all the cash flows.

3 C is correct. Because Node 2–2 is the middle node rate in Year 2, it will be close to the implied one-year forward rate two years from now (as derived from the spot curve). Node 4–1 should be equal to the product of Node 4–5 and $e^{0.8}$. Lastly, Node 3–2 cannot be derived from Node 2–2; it can be derived from any other Year 3 node; for example, Node 3–2 can be derived from Node 3–4 (equal to the product of Node 3–4 and $^{e4\sigma}$).

4 A is correct. The value of a bond at a particular node, in this case Node 1–2, can be derived by determining the present value of the coupon payment and expected future bond values to the right of that node on the tree. In this case, those two nodes are the middle node in Year 2, equal to 101.5168, and the lower node in Year 2, equal to 102.1350. The coupon payment is 2.5. The bond value at Node 1–2 is calculated as follows:

$$\text{Value} = \frac{2.5 + (0.5 \times 101.5168 + 0.5 \times 102.1350)}{1.014925}$$

$$= 102.7917.$$

5 A is correct. Calibrating a binomial interest rate tree to match a specific term structure is important because we can use the known valuation of a benchmark bond from the spot rate pricing to verify the accuracy of the rates shown in the binomial interest rate tree. Once its accuracy is confirmed, the interest rate tree can then be used to value bonds with embedded options. While discounting with spot rates will produce arbitrage-free valuations for option-free bonds, this spot rate method will not work for bonds with embedded options where expected future cash flows are interest-rate dependent (because rate changes impact the likelihood of options being exercised). The interest rate tree allows for the alternative paths that a bond with embedded options might take.

B is incorrect because calibration does not identify mispriced benchmark bonds. In fact, benchmark bonds are employed to prove the accuracy of the binomial interest rate tree, because they are assumed to be correctly priced by the market.

C is incorrect because the calibration of the binomial interest rate tree is designed to produce an arbitrage-free valuation approach and such an approach does not allow a market participant to realize arbitrage profits through stripping and reconstitution.

6 A is correct. Volatility is one of the two key assumptions required to estimate rates for the binomial interest rate tree. Increasing the volatility from 10% to 15% would cause the possible forward rates to spread out on the tree because it increases the exponent in the relationship multiple between nodes ($e^{x\sigma}$, where x = 2 times the number of nodes above the lowest node in a given year in the interest rate tree). Conversely, using a lower estimate of volatility would cause the forward rates to narrow or converge to the implied forward rates from the prevailing yield curve.

B is incorrect because volatility is a key assumption in the binomial interest rate tree model. Any change in volatility will cause a change in the implied forward rates.

C is incorrect because increasing the volatility from 10% to 15% causes the possible forward rates to spread out on the tree, not converge to the implied forward rates from the current yield curve. Rates will converge to the implied forward rates when lower estimates of volatility are assumed.

7 B is correct. Bond B's arbitrage-free price is calculated as follows:

$$\frac{3}{1.02} + \frac{103}{1.02^2} = 101.9416,$$

which is higher than the bond's market price of 100.9641. Therefore, an arbitrage opportunity exists. Since the bond's value (100.9641) is less than the sum of the values of its discounted cash flows individually (101.9416), a trader would perceive an arbitrage opportunity and could buy the bond while selling claims to the individual cash flows (zeros), capturing the excess value. The arbitrage-free prices of Bond A and Bond C are equal to the market prices of the respective bonds, so there is no arbitrage opportunity for these two bonds:

$$\text{Bond A: } \frac{1}{1.02} + \frac{101}{1.02^2} = 98.0584.$$

$$\text{Bond C: } \frac{5}{1.02} + \frac{105}{1.02^2} = 105.8247.$$

8 C is correct. The first step in the solution is to find the correct spot rate (zero-coupon rates) for each year's cash flow. The benchmark bonds in Exhibit 2 are conveniently priced at par so the yields-to-maturity and the coupon rates on the bonds are the same. Because the one-year issue has only one cash flow remaining, the YTM equals the spot rate of 3% (or z_1 = 3%). The spot rates for Year 2 (z_2) and Year 3 (z_3) are calculated as follows:

$$100 = \frac{4}{1.0300} + \frac{104}{(1 + z_2)^2}; z_2 = 4.02\%.$$

$$100 = \frac{5}{1.0300} + \frac{5}{(1.0402)^2} + \frac{105}{(1 + z_3)^3}; z_3 = 5.07\%.$$

The correct arbitrage-free price for the Hutto-Barkley Inc. bond is

$$P_0 = \frac{3}{(1.0300)} + \frac{3}{(1.0402)^2} + \frac{103}{(1.0507)^3} = 94.4828.$$

Therefore, the bond is mispriced by 94.9984 – 94.4828 = 0.5156 per 100 of par value.

A is incorrect because the correct spot rates are not calculated and instead the Hutto-Barkley Inc. bond is discounted using the respective YTM for each maturity. Therefore, this leads to an incorrect mispricing of 94.6616 – 94.9984 = –0.3368 per 100 of par value.

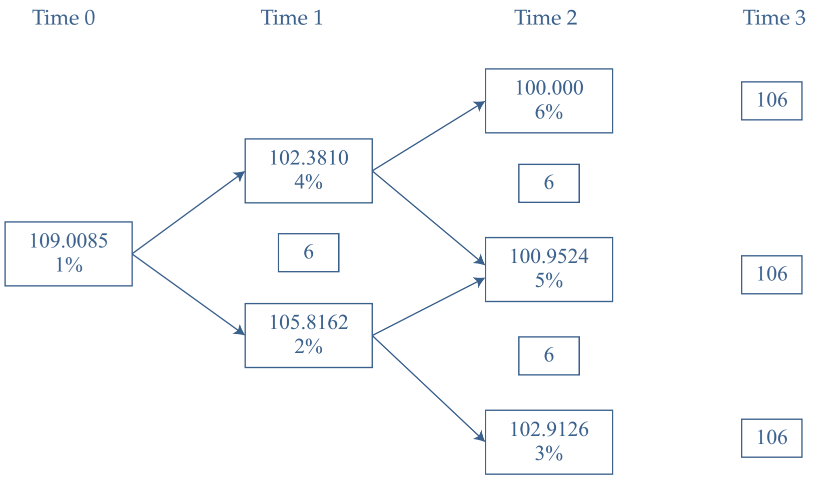
B is incorrect because the spot rates are derived using the coupon rate for Year 3 (maturity) instead of using each year's respective coupon rate to employ the bootstrap methodology. This leads to an incorrect mispricing of 94.5302 – 94.9984 = –0.4682 per 100 of par value.

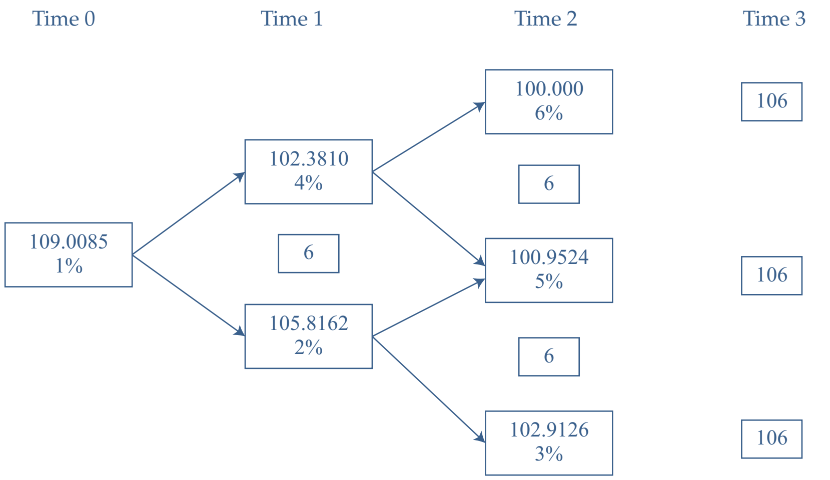
9 B is correct. The Luna y Estrellas Intl. bond contains an embedded option. Method 1 will produce an arbitrage-free valuation for option-free bonds; however, for bonds with embedded options, changes in future interest rates impact the likelihood the option will be exercised and so impact future cash flows. Therefore, to develop a framework that values bonds with embedded options, interest rates must be allowed to take on different potential values in the future based on some assumed level of volatility (Method 2).

A and C are incorrect because the Hutto-Barkley Inc. bond and the Peaton Scorpio Motors bond are both option-free bonds and can be valued using either Method 1 or Method 2 to produce an arbitrage-free valuation.

10 B is correct. This is the binomial tree that obtains a bond value of 109.0085.

Valuing a 6%, Three-Year Bond





These are the calculations:

$106/1.06 = 100.0000.$

$106/1.05 = 100.9524.$

$106/1.03 = 102.9126.$

$$\frac{6 + \left(0.5 \times 100.0000 + 0.5 \times 100.9524\right)}{1.04} = 102.3810.$$

$$\frac{6 + \left(0.5 \times 100.9524 + 0.5 \times 102.9126\right)}{1.02} = 105.8162.$$

$$\frac{6 + \left(0.5 \times 102.3810 + 0.5 \times 105.8162\right)}{1.01} = 109.0085.$$

A is incorrect because the Time T coupon payment is subtracted from the value in each node calculation for Time T. C is incorrect because it assumes that a coupon is paid at Time 0.

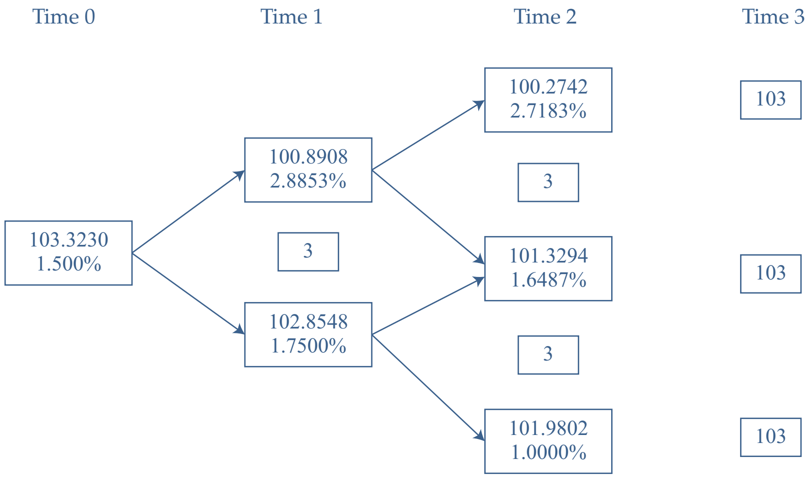
11 B is correct. Based on the dominance principle, an arbitrage opportunity exists. The dominance principle asserts that a financial asset with a risk-free payoff in the future must have a positive price today. Because Asset A and Asset B are both risk-free assets, they should have the same discount rate. Relative to its payoff, Asset A is priced at \$500/525, or 0.95238, and Asset B is priced at \$1,000/1,100, or 0.90909. Given its higher implied discount rate (10%) and lower corresponding price, Asset B is cheap relative to Asset A, which has a lower implied discount rate (5%) and a higher corresponding price.

The arbitrage opportunity based on dominance is to sell two units of Asset A for \$1,000 and buy one unit of Asset B. There is no cash outlay today, and in one year, the portfolio delivers a net cash inflow of \$50 [= \$1,100 – (2 × \$525)].

12 B is correct. Of the three markets, the New York bond has the lowest yield-to-maturity and, correspondingly, the highest bond price. Similarly, the Hong Kong bond has the highest yield-to-maturity and the lowest bond price of the three markets. Therefore, the most profitable arbitrage trade would be to buy the bond in Hong Kong and sell it in New York.

13 B is correct. The bond value at the upper node at Time 1 is closest to 99.6255. The cash flow at Time 2 is 102.5, the redemption of par value (100) plus the final coupon payment (2.5). Using backward induction, we calculate the present value of the bond at the upper node of Time 1 as 102.5/1.028853 = 99.6255.

14 B is correct. The price of Bond D is closest to 103.3230 and can be calculated using backward induction.



$$\text{Bond value at a node} = \frac{C + \left(0.5 \times VH + 0.5 \times VL\right)}{1 + i}.$$

Calculations:

The cash flow at Time 3 is 103, the redemption of par value (100) plus the final coupon payment (3).

Time 2 node values:

Upper node: 103/1.027183 = 100.2742.

Middle node: 103/1.016487 = 101.3294.

Lower node: 103/1.010000 = 101.9802.

Working back to Time 1 requires the use of the general expression above.

Time 1 node values:

$$\text{Upper node: } \frac{3 + \left(0.5 \times 100.2742 + 0.5 \times 101.3294\right)}{1.028853} = 100.8908.$$

$$\text{Lower node: } \frac{3 + (0.5 \times 101.3294 + 0.5 \times 101.9802)}{1.0175} = 102.8548.$$

Time 0 node value:

$$\frac{3 + (0.5 \times 100.8908 + 0.5 \times 102.8548)}{1.015} = 103.3230.$$

Therefore, the price of the bond is 103.3230.

15 B is correct. Two methods are commonly used to estimate potential interest rate volatility in a binomial interest rate tree. The first method bases estimates on historical interest rate volatility. The second method uses observed market prices of interest rate derivatives.

Statement 1 is incorrect because there are three requirements to create a binomial interest rate tree, not two. The third requirement is an assumption regarding the interest rate model. Statement 3 is incorrect because the valuation of a bond using spot rates and the valuation of a bond from an interest rate tree will be the same regardless of the volatility assumption used in the model.

16 B is correct. The value of the lower one-period forward rate is closest to 3.5400%. Since the higher one-period forward rate is 5.8365% and interest rate volatility is 25%, the lower rate equals the higher rate multiplied by $e^{-2\sigma}$. This is calculated as $0.058365 \times e^{-0.50} = 0.035400$.

17 B is correct. The present value of Bond D's cash flows following Path 2 is 102.8607 and can be calculated as follows:

$$\frac{3}{1.015} + \frac{3}{(1.015)(1.028853)} + \frac{103}{(1.015)(1.028853)(1.016487)} = 102.8607.$$

18 A is correct. Increasing the number of paths using the Monte Carlo method does increase the estimate's statistical accuracy. It does not, however, provide a value that is closer to the bond's true fundamental value.

19 A is correct. The Ho–Lee model is arbitrage free and can be calibrated to closely match the observed term structure.

20 C is correct. Both statements are incorrect because Jones incorrectly describes both types of model. Equilibrium term structure models are factor models that seek to describe the dynamics of the term structure by using fundamental economic variables that are assumed to affect interest rates. Arbitrage-free term structure models use observed market prices of a reference set of financial instruments, assumed to be correctly priced, to model the market yield curve.

21 A is correct. Consistent with Jones's statement, equilibrium term structure models require fewer parameters to be estimated relative to arbitrage-free models, and arbitrage-free models allow for time-varying parameters. Consequently, arbitrage-free models can model the market yield curve more precisely than equilibrium models.

22 C is correct. The Kalotay–Williams–Fabozzi equilibrium term structure model is similar to the Ho–Lee model in that it assumes constant drift, no mean reversion, and constant volatility, but the KWF model describes the log of the dynamics of the short rate, while the Ho–Lee model does not.

FIXED INCOME

STUDY SESSION 12

Fixed Income (2)

This study session continues use of the binomial valuation method to value bonds with embedded options. Sensitivity to interest rates and interest rate volatility are key considerations. Option-adjusted spreads are introduced for the evaluation of risky bonds. Credit analysis concepts, tools, and applications are then discussed along with the term structure of credit spreads. The study session concludes with credit default swaps and their use in managing credit exposure.

READING ASSIGNMENTS

© 2021 CFA Institute. All rights reserved.

READING

30

Valuation and Analysis of Bonds with Embedded Options

by Leslie Abreo, Ioannis Georgiou, CFA, and Andrew Kalotay, PhD

Leslie Abreo, MFE, is at Andrew Kalotay Associates, Inc. (USA). Ioannis Georgiou, CFA, is at Finovex.com (Cyprus). Andrew Kalotay, PhD, is at Andrew Kalotay Associates, Inc. (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe fixed-income securities with embedded options;
☐	**b.** explain the relationships between the values of a callable or putable bond, the underlying option-free (straight) bond, and the embedded option;
☐	**c.** describe how the arbitrage-free framework can be used to value a bond with embedded options;
☐	**d.** explain how interest rate volatility affects the value of a callable or putable bond;
☐	**e.** explain how changes in the level and shape of the yield curve affect the value of a callable or putable bond;
☐	**f.** calculate the value of a callable or putable bond from an interest rate tree;
☐	**g.** explain the calculation and use of option-adjusted spreads;
☐	**h.** explain how interest rate volatility affects option-adjusted spreads;
☐	**i.** calculate and interpret effective duration of a callable or putable bond;
☐	**j.** compare effective durations of callable, putable, and straight bonds;
☐	**k.** describe the use of one-sided durations and key rate durations to evaluate the interest rate sensitivity of bonds with embedded options;
☐	**l.** compare effective convexities of callable, putable, and straight bonds;

(continued)

The presentation of the binomial trees in this reading was revised to conform with other readings in 2018 & 2019 by Donald J. Smith, PhD, Boston University (USA).

© 2021 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**m.** calculate the value of a capped or floored floating-rate bond;
☐	**n.** describe defining features of a convertible bond;
☐	**o.** calculate and interpret the components of a convertible bond's value;
☐	**p.** describe how a convertible bond is valued in an arbitrage-free framework;
☐	**q.** compare the risk–return characteristics of a convertible bond with the risk–return characteristics of a straight bond and of the underlying common stock.

1 INTRODUCTION AND OVERVIEW OF EMBEDDED OPTIONS

a describe fixed-income securities with embedded options

The valuation of a fixed-rate, option-free bond generally requires determining its future cash flows and discounting them at the appropriate rates. Valuation becomes more complicated when a bond has one or more embedded options because the values of embedded options are typically contingent on interest rates.

Understanding how to value and analyze bonds with embedded options is important for practitioners. Issuers of bonds often manage interest rate exposure with embedded options, such as call provisions. Investors in callable bonds must appreciate the risk of being called. The perception of this risk is collectively represented by the premium, in terms of increased coupon or yield, that the market demands for callable bonds relative to otherwise identical option-free bonds. Issuers and investors must also understand how other types of embedded options—such as put provisions, conversion options, caps, and floors—affect bond values and the sensitivity of these bonds to interest rate movements.

We first provide a brief overview of various types of embedded options. We then discuss bonds that include a call or put provision. Taking a building-block approach, we show how the arbitrage-free valuation framework discussed earlier can be applied to the valuation of callable and putable bonds—first in the absence of interest rate volatility, and then when interest rates fluctuate. We also discuss how option-adjusted spreads are used to value risky callable and putable bonds. We then turn to interest rate sensitivity. It highlights the need to use effective duration, including one-sided durations and key rate durations, as well as effective convexity to assess the effect of interest rate movements on the value of callable and putable bonds. We also explain the valuation of capped and floored floating-rate bonds (floaters) and convertible bonds.

1.1 Overview of Embedded Options

The term "embedded bond options" or **embedded options** refers to contingency provisions found in the bond's indenture or offering circular. These options represent rights that enable their holders to take advantage of interest rate movements. They can be exercised by the issuer or the bondholder, or they may be exercised automatically depending on the course of interest rates. For example, a call option allows the issuer to benefit from lower interest rates by retiring the bond issue early and refinancing at a lower cost. In contrast, a put option allows the bondholder to benefit from higher interest rates by putting back the bonds to the issuer and reinvesting the proceeds of the retired bond at a higher yield. These options are not independent of the bond and thus cannot be traded separately—hence the adjective "embedded." In this section, we provide a review of familiar embedded options.

Corresponding to every embedded option, or combination of embedded options, is an underlying bond with a specified issuer, issue date, maturity date, principal amount and repayment structure, coupon rate and payment structure, and currency denomination. We also refer to this underlying option-free bond as the **straight bond**. The coupon of an underlying bond can be fixed or floating. Fixed-coupon bonds may have a single rate for the life of the bond, or the rate may step up or step down according to a coupon schedule. The coupons of floaters are reset periodically according to a formula based on a reference rate plus a credit spread—for example, Market reference rate + 100 basis points (bps). Except when we discuss capped and floored floaters, our focus is on fixed-coupon, single-rate bonds, also referred to as fixed-rate bonds.

1.1.1 *Simple Embedded Options*

Call and put options are standard examples of embedded options. In fact, the vast majority of bonds with embedded options are callable, putable, or both. The call provision is by far the most prevalent type of embedded option.

1.1.1.1 Call Options A **callable bond** is a bond that includes an embedded call option. The call option is an issuer option; that is, the right to exercise the option is at the discretion of the bond's issuer. The call provision allows the issuer to redeem the bond issue prior to maturity. Early redemption usually happens when the issuer has the opportunity to replace a high-coupon bond with another bond that has more favorable terms, typically when interest rates have fallen or when the issuer's credit quality has improved.

Until the 1990s, most long-term corporate bonds in the United States were callable after either 5 or 10 years. The initial call price (exercise price) was typically at a premium above par, the premium depended on the coupon, and the call price gradually declined to par a few years prior to maturity. Today, most investment-grade corporate bonds are essentially non-refundable. They may have a "make-whole call," so named because the call price is such that the bondholders are more than "made whole" (compensated) in exchange for surrendering their bonds. The call price is calculated at a narrow spread to a benchmark security—usually an on-the-run sovereign bond, such as Treasuries in the United States or gilts in the United Kingdom. Thus, economical refunding is virtually out of the question. Investors need have no fear of receiving less than their bonds are worth.

Most callable bonds include a call **protection period** during which the issuer cannot call the bond. For example, a 10-year callable bond may have a call protection period of three years, meaning that the first potential call date is three years after the bond's issue date. Call protection periods may be as short as one month or extend to several years. For example, high-yield corporate bonds are often callable a few years after issuance. Holders of such bonds are usually less concerned about early redemption than about possible default. Of course, this perspective can change over the life of the bond—for example, if the issuer's credit quality improves.

Callable bonds include different types of call features. The issuer of a European-style callable bond can exercise the call option only once on the call date. An American-style callable bond is continuously callable at any time starting on the first call date. A Bermudan-style call option can be exercised only on a predetermined schedule on specified dates following the call protection period. These dates are specified in the bond's indenture or offering circular.

With a few exceptions, bonds issued by government-sponsored enterprises in the United States (e.g., Fannie Mae, Freddie Mac, Federal Home Loan Banks, and Federal Farm Credit Banks) are callable. These bonds tend to have relatively short maturities (5–10 years) and very short call protection periods (three months to one year). The call price is almost always at 100% of par, and the call option is often Bermudan style.

Tax-exempt municipal bonds (often called "munis"), a type of non-sovereign (local) government bond issued in the United States, are almost always callable at 100% of par any time after the end of the 10th year. They may also be eligible for advance refunding—a highly specialized topic that is not discussed here.

Although the bonds of US government-sponsored enterprises and municipal issuers account for most of the callable bonds issued and traded globally, bonds that include call provisions are also found in other countries in Asia Pacific, Europe, Canada, and Central and South America. The vast majority of callable bonds are denominated in US dollars or euros because of investors' demand for securities issued in these currencies. Australia, the United Kingdom, Japan, and Norway are examples of countries that have a market for callable bonds denominated in local currency.

1.1.1.2 Put Options and Extension Options A **putable bond** is a bond that includes an embedded put option. The put option is an investor option; that is, the right to exercise the option is at the discretion of the bondholder. The put provision allows the bondholders to put back the bonds to the issuer prior to maturity, usually at par. This usually happens when interest rates have risen and higher-yielding bonds are available.

Similar to callable bonds, most putable bonds include protection periods. They can be European or, rarely, Bermudan style, but there are no American-style putable bonds.

Another type of embedded option that resembles a put option is an extension option. At maturity, the holder of an **extendible bond** (sometimes spelled "extendable") has the right to keep the bond for a number of years after maturity, possibly with a different coupon. In this case, the terms of the bond's indenture or offering circular are modified, but the bond remains outstanding. An example of a corporate extendible is an offering from Heathrow Funding Ltd. It pays a 0.50% coupon and matures on 17 May 2024. However, it is extendible to 7 May 2026 as a floating-rate note paying 12-month MRR plus 4.00%. We will discuss the resemblance between a putable and an extendible bond later.

1.1.2 *Complex Embedded Options*

Although callable and putable bonds are the most common types of bonds with embedded options, there are bonds with other types of options or combinations of options. For instance, some bonds can be both callable and putable. These bonds can be either called by the issuer or put by the bondholders.

Convertible bonds are another type of bond with an embedded option. The conversion option allows bondholders to convert their bonds into the issuer's common stock. Convertible bonds are usually also callable by the issuer; the call provision enables the issuer to take advantage of lower interest rates or to force conversion.

Another layer of complexity is added when the option is contingent on some particular event. An example is the estate put or survivor's option that may be available to retail investors. In the event of the holder's death, this bond can be put at par by the heir(s). Because the estate put comes into play only in the event of the bondholder's death, the value of a bond with an estate put is contingent on the life expectancy of its holder, which is uncertain.

Bonds may contain several interrelated issuer options without any investor option. A prime example is a **sinking fund bond** (sinker). A sinker requires the issuer to set aside funds over time to retire the bond issue, thus reducing credit risk. Such a bond may be callable and may also include options unique to sinking fund bonds, such as an acceleration provision and a delivery option.

SINKING FUND BONDS

The underlying bond has an amortizing structure—for example, a 30-year maturity with level annual principal repayments beginning at the end of the 11th year. In this case, each payment is 5% of the original principal amount. A typical sinking fund bond may include the following options:

- A standard *call option* above par, with declining premiums, starting at the end of Year 10. Thus, the entire bond issue could be called from Year 10 onward.
- An *acceleration provision*, such as a "triple up." Such a provision allows the issuer to repurchase at par three times the mandatory amount, or in this case 15% of the original principal amount, on any scheduled sinking fund date. Assume that the issuer wants to retire the bonds at the end of Year 11. Instead of calling the entire outstanding amount at a premium, it would be more cost effective to "sink" 15% at par and call the rest at a premium. Thus, the acceleration provision provides an additional benefit to the issuer if interest rates decline.
- A *delivery option*, which allows the issuer to satisfy a sinking fund payment by delivering bonds to the bond's trustee in lieu of cash. The bond's trustee is appointed by the issuer but acts in a fiduciary capacity with the bondholders. If the bonds are currently trading below par, say at 90% of par, it is more cost effective for the issuer to buy back bonds from investors to meet the sinking fund requirements than to pay par. The delivery option benefits the issuer if interest rates rise. Of course, the benefit can be materialized only if there is a liquid market for the bonds. Investors can take defensive action by accumulating the bonds and refusing to sell them at a discount.

From the issuer's perspective, the combination of the call option and the delivery option is effectively a "long straddle"—an option strategy involving the purchase of a put option and a call option on the same underlying with the same exercise price and expiration date. At expiration, if the underlying price is above the exercise price, the put option is worthless but the call option is in the money. In contrast, if the underlying price is below the exercise price, the call option is worthless but the put option is in the money. Thus, a long straddle benefits the investor when the underlying price moves up or down. The greater the move up or down (i.e., the greater the volatility), the greater the benefit for the investor. As a consequence, a sinking fund bond benefits the issuer not only if interest rates decline but also if they rise. Determining the combined value of the underlying bond and the three options is quite challenging.

EXAMPLE 1

Types of Embedded Options

1 Investors in putable bonds *most likely* seek to take advantage of:

A higher interest rates.

B improvements in the issuer's credit rating.

C movements in the price of the issuer's common stock.

2 The conversion option in a convertible bond is a right held by:

A the issuer.

B the bondholders.

C the issuer and the bondholders jointly.

Solution to 1:

A is correct. A putable bond offers the bondholder the ability to take advantage of a rise in interest rates by putting back the bond to the issuer and reinvesting the proceeds of the retired bond in a higher-yielding bond.

Solution to 2:

B is correct. A conversion option is a call option that gives the bondholders the right to convert their bonds into the issuer's common stock.

The presence of embedded options affects a bond's value. To quantify this effect, financial theory and financial technology come into play. The following section presents basic valuation and analysis concepts for bonds with embedded options.

2 VALUATION AND ANALYSIS OF CALLABLE AND PUTABLE BONDS

b explain the relationships between the values of a callable or putable bond, the underlying option-free (straight) bond, and the embedded option

c describe how the arbitrage-free framework can be used to value a bond with embedded options

Under the arbitrage-free framework, the value of a bond with embedded options is equal to the sum of the arbitrage-free values of its parts. We first identify the relationships between the values of a callable or putable bond, the underlying option-free (straight) bond, and the call or put option. We then discuss how to value callable and putable bonds under different risk and interest rate volatility scenarios.

2.1 Relationships between the Values of a Callable or Putable Bond, Straight Bond, and Embedded Option

The value of a bond with embedded options is equal to the sum of the arbitrage-free value of the straight bond and the arbitrage-free values of the embedded options.

For a callable bond, the decision to exercise the call option is made by the issuer. Thus, the investor is long the bond but short the call option. From the investor's perspective, therefore, the value of the call option *decreases* the value of the callable bond relative to the value of the straight bond:

Value of callable bond = Value of straight bond – Value of issuer call option.

The value of the straight bond can be obtained by discounting the bond's future cash flows at the appropriate rates. The hard part is valuing the call option because its value is contingent on future interest rates. Specifically, the issuer's decision to call the bond depends on its ability to refinance at a lower cost. In practice, the value of the call option is often calculated as the difference between the value of the straight bond and the value of the callable bond:

Value of issuer call option
= Value of straight bond – Value of callable bond. **(1)**

For a putable bond, the decision to exercise the put option is made by the investor. Thus, the investor has a long position in both the bond and the put option. As a consequence, the value of the put option *increases* the value of the putable bond relative to the value of the straight bond.

Value of putable bond = Value of straight bond + Value of investor put option.

It follows that

Value of investor put option
= Value of putable bond − Value of straight bond. (2)

Although most investment professionals do not need to be experts in bond valuation, they should have a solid understanding of the basic analytical approach, which is presented in the following sections.

2.2 Valuation of Default-Free and Option-Free Bonds: A Refresher

An asset's value is the present value of the cash flows the asset is expected to generate in the future. In the case of a default-free and option-free bond, the future cash flows are, by definition, certain. Thus, the question is, at which rates should these cash flows be discounted? The answer is that each cash flow should be discounted at the spot rate corresponding to the cash flow's payment date. Although spot rates might not be directly observable, they can be inferred from readily available information, usually from the market prices of actively traded on-the-run sovereign bonds of various maturities. These prices can be transformed into spot rates, par rates (i.e., coupon rates of hypothetical bonds of various maturities selling at par), or forward rates. Recall from Level I that spot rates, par rates, and forward rates are equivalent ways of conveying the same information; knowing any one of them is sufficient to determine the others.

Suppose we want to value a three-year 4.25% annual coupon bond. Exhibit 1 provides the equivalent forms of a yield curve with maturities of one, two, and three years.

Exhibit 1 Equivalent Forms of a Yield Curve

Maturity (year)	Par Rate (%)	Spot Rate (%)	One-Year Forward Rate (%)	
1	2.500	2.500	0 years from now	2.500
2	3.000	3.008	1 year from now	3.518
3	3.500	3.524	2 years from now	4.564

We start with the par rates provided in the second column of Exhibit 1. Because we are assuming annual coupons and annual compounding, the one-year spot rate is simply the one-year par rate. The hypothetical one-year par bond implied by the given par rate has a single cash flow of 102.500 (principal plus coupon) in Year 1. In order to have a present value of par, this future cash flow must be divided by 1.025. Thus, the one-year spot rate or discount rate is 2.500% (*Note:* All cash flows and values are expressed as a percentage of par).

A two-year 3.000% par bond has two cash flows: 3 in Year 1 and 103 in Year 2. By definition, the sum of the two discounted cash flows must equal 100. We know that the discount rate appropriate for the first cash flow is the one-year spot rate (2.500%). We now solve the following equation to determine the two-year spot rate (z_2):

$$\frac{3}{(1.025)} + \frac{103}{(1 + z_2)^2} = 100.$$

We can follow a similar approach to determine the three-year spot rate (z_3):

$$\frac{3.500}{(1.02500)} + \frac{3.500}{(1.03008)^2} + \frac{103.500}{(1 + z_3)^3} = 100.$$

The one-year forward rates are determined by using indifference equations. Assume an investor has a two-year horizon. She could invest for two years either at the two-year spot rate or at the one-year spot rate for one year and then reinvest the proceeds at the one-year forward rate one year from now ($F_{1,1}$). The result of investing using either of the two approaches should be the same. Otherwise, there would be an arbitrage opportunity. Thus,

$$(1 + 0.03008)^2 = (1 + 0.02500) \times (1 + F_{1,1}).$$

Similarly, the one-year forward rate two years from now ($F_{2,1}$) can be calculated using the following equation:

$$(1 + 0.03524)^3 = (1 + 0.03008)^2 \times (1 + F_{2,1}).$$

The three-year 4.25% annual coupon bond can now be valued using the spot rates:

$$\frac{4.25}{(1.02500)} + \frac{4.25}{(1.03008)^2} + \frac{104.25}{(1.03524)^3} = 102.114.$$

An equivalent way to value this bond is to discount its cash flows one year at a time using the one-year forward rates:

$$\frac{4.25}{(1.02500)} + \frac{4.25}{(1.02500)(1.03518)} + \frac{104.25}{(1.02500)(1.03518)(1.04564)} = 102.114.$$

2.3 Valuation of Default-Free Callable and Putable Bonds in the Absence of Interest Rate Volatility

When valuing bonds with embedded options, the approach relying on one-period forward rates provides a better framework than that relying on the spot rates because we need to know the value of the bond at different points in time in the future to determine whether the embedded option will be exercised at those points in time.

2.3.1 *Valuation of a Callable Bond at Zero Volatility*

Let us apply this framework to the valuation of a Bermudan-style three-year 4.25% annual coupon bond that is callable at par one year and two years from now. The decision to exercise the call option is made by the issuer. Because the issuer borrowed money, it will exercise the call option when the value of the bond's future cash flows is higher than the call price (exercise price). Exhibit 2 shows how to calculate the value of this callable bond using the one-year forward rates calculated in Exhibit 1.

Exhibit 2 Valuation of a Default-Free Three-Year 4.25% Annual Coupon Bond Callable at Par One Year and Two Years from Now at Zero Volatility

	Today	Year 1	Year 2	Year 3
Cash flow		4.250	4.250	104.250
Discount rate		2.500%	3.518%	4.564%
Value of the callable bond	$\frac{100 + 4.250}{1.02500} = 101.707$	$\frac{99.700 + 4.250}{1.03518} = \cancel{100.417}$ Called at 100	$\frac{104.250}{1.04564} = 99.700$ Not called	

We start by discounting the bond's cash flow at maturity (104.250) to Year 2 using the one-year forward rate two years from now (4.564%). The present value at Year 2 of the bond's future cash flows is 99.700. This value is lower than the call price of 100, so a rational borrower will not call the bond at that point in time. Next, we add the cash flow in Year 2 (4.250) to the present value of the bond's future cash flows at Year 2 (99.700) and discount the sum to Year 1 using the one-year forward rate one year from now (3.518%). The present value at Year 1 of the bond's future cash flows is 100.417. Here, a rational borrower will call the bond at 100 because leaving it outstanding would be more expensive than redeeming it. Last, we add the cash flow in Year 1 (4.250) to the present value of the bond's future cash flows at Year 1 (100.000) then discount the sum to today at 2.500%. The result (101.707) is the value of the callable bond (*Note:* For the purpose of coverage of this topic, all cash flows and values are expressed as a percentage of par).

We can apply Equation 1 to calculate the value of the call option embedded in this callable bond. The value of the straight bond is the value of the default-free and option-free three-year 4.25% annual coupon bond calculated earlier (102.114). Thus,

Value of issuer call option = 102.114 – 101.707 = 0.407.

Recall from the earlier discussion about the relationships between the value of a callable bond, straight bond, and call option that the investor is long the bond and short the call option. Thus, the value of the call option decreases the value of the callable bond relative to that of an otherwise identical option-free bond.

2.3.2 *Valuation of a Putable Bond at Zero Volatility*

We now apply this framework to the valuation of a Bermudan-style three-year 4.25% annual coupon bond that is putable at par one year and two years from now. The decision to exercise the put option is made by the investor. Because the investor lent money, he will exercise the put option when the value of the bond's future cash flows is lower than the put price (exercise price). Exhibit 3 shows how to calculate the value of the three-year 4.25% annual coupon bond putable at par one year and two years from today.

Exhibit 3 Valuation of a Default-Free Three-Year 4.25% Annual Coupon Bond Putable at Par One Year and Two Years from Now at Zero Volatility

	Today	Year 1	Year 2	Year 3
Cash flow		4.250	4.250	104.250
Discount rate		2.500%	3.518%	4.564%
Value of the putable bond	$\frac{100.707 + 4.250}{1.02500} = 102.397$	$\frac{100 + 4.250}{1.03518} = 100.707$ Not put	$\frac{104.250}{1.04564} = \cancel{99.700}$ Put at 100	

We can apply Equation 2 to calculate the value of the put option:

Value of investor put option = 102.397 – 102.114 = 0.283.

Because the investor is long the bond and the put option, the value of the put option increases the value of the putable bond relative to that of an otherwise identical option-free bond.

OPTIMAL EXERCISE OF OPTIONS

The holder of an embedded bond option can extinguish (or possibly modify the terms of) the bond. Assuming that the option is currently exercisable, the obvious question is, does it pay to exercise? Assuming that the answer is affirmative, the follow-up question is whether it is better to exercise the option at present or to wait.

Let us consider the first question: Would it be profitable to exercise the option? The answer is usually straightforward: Compare the value of exercising with the value of not exercising. For example, suppose that a bond is currently putable at 100. If the bond's market price is above 100, putting the bond makes no sense because the cash value from selling the bond would exceed 100. In contrast, if the bond's market price is 100, putting the bond should definitely be considered. Note that the market price of the bond cannot be less than 100 because such a situation creates an arbitrage opportunity: Buy the bond below 100 and immediately put it at 100.

The logic of a call decision by the issuer is similar. If a bond's market price is significantly less than the call price, calling is foolish because the bond could be simply repurchased in the market at a lower price. Alternatively, if the price is very close to the call price, calling may make sense.

Assume that we have determined that exercising the option would be profitable. If the option under consideration is European style, it is obvious that it should in fact be exercised: There is no justification for not doing so. But if it is an American-style or Bermudan-style option, the challenge is to determine whether it is better to act now or to wait for a better opportunity. The problem is that although circumstances may become more favorable, they may also get worse. So, option holders must consider the odds and decide to act or wait, depending on their risk preference.

The approach presented here for valuing bonds with embedded options assumes that the option holders, be they issuers or investors, are risk neutral. They exercise if, and only if, the benefit from exercise exceeds the expected benefit from waiting. In reality, option holders may be risk averse and may exercise early even if the option is worth more alive than dead.

EXAMPLE 2

Valuation of Default-Free Callable and Putable Bonds

George Cahill, a portfolio manager, has identified three five-year annual coupon bonds issued by a sovereign government. The three bonds have identical characteristics. The exceptions are that Bond A is an option-free bond; Bond B is callable at par two years and three years from today; and Bond C is also callable at par two years and three years from today as well as putable at par one year from today.

1 Relative to the value of Bond A, the value of Bond B is:

  A lower.

  B the same.

  C higher.

2 Relative to the value of Bond B, the value of Bond C is:

  A lower.

  B the same.

  C higher.

3 Given an anticipation of rising interest rates, Bond C will be expected to:

  A be called by the issuer.

  B be put by the bondholders.

  C mature without exercise of any of the embedded options.

Solution to 1:

A is correct. Bond B is a callable bond, and Bond A is the underlying option-free (straight) bond. The call option embedded in Bond B is an issuer option that decreases the bond's value for the investor. If interest rates decline, bond prices usually increase; however, the price appreciation of Bond B will be capped relative to the price appreciation of Bond A because the issuer will call the bond to refinance at a lower cost.

Solution to 2:

C is correct. Relative to Bond B, Bond C includes a put option. A put option is an investor option that increases the bond's value for the investor. Thus, the value of Bond C is higher than that of Bond B.

Solution to 3:

B is correct. As interest rates rise, bond prices decrease. Thus, the bondholders will have an incentive to exercise the put option so that they can reinvest the proceeds of the retired bond at a higher yield.

Exhibits 2 and 3 show how callable and putable bonds are valued in the absence of interest rate volatility. In real life, however, interest rates do fluctuate. Thus, the option holder must consider possible evolutions of the yield curve over time.

3 EFFECT OF INTEREST RATE VOLATILITY ON THE VALUE OF CALLABLE AND PUTABLE BONDS

d explain how interest rate volatility affects the value of a callable or putable bond

e explain how changes in the level and shape of the yield curve affect the value of a callable or putable bond

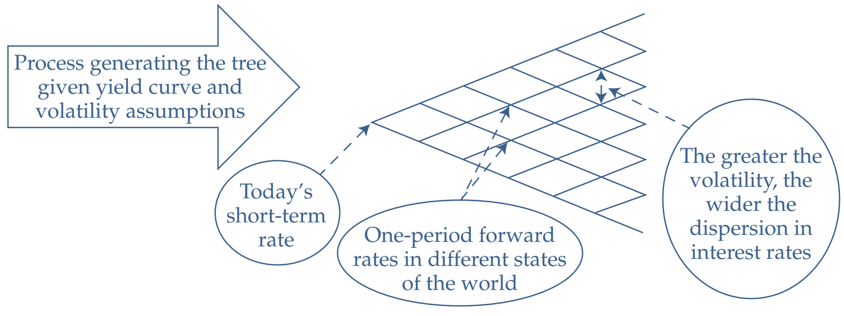
In this section, we discuss the effects of interest rate volatility as well as the level and shape of the yield curve on the value of embedded options.

3.1 Interest Rate Volatility

The value of any embedded option, regardless of the type of option, increases with interest rate volatility. The greater the volatility, the more opportunities for the embedded option to be exercised. Thus, it is critical for issuers and investors to understand the effect of interest rate volatility on the value of bonds with embedded options.

The effect of interest rate volatility is represented in an interest rate tree or lattice, as illustrated in Exhibit 4. From each node on the tree starting from today, interest rates could go up or down. From these two states, interest rates could again go up or down. The dispersion between these up and down states anywhere on the tree is determined by the process generating interest rates based on a given yield curve and interest rate volatility assumptions.

Exhibit 4 Building an Interest Rate Tree



Exhibits 5 and 6 show the effect of interest rate volatility on the value of a callable bond and putable bond, respectively.

Exhibit 5 Value of a 30-Year 4.50% Bond Callable at Par in 10 Years under Different Volatility Scenarios Assuming a 4% Flat Yield Curve

The stacked bars in Exhibit 5 represent the value of the straight bond, which is unaffected by interest rate volatility. The white component is the value of the call option; taking it away from the value of the straight bond gives the value of the callable bond—the shaded component. All else being equal, the call option increases in value with interest rate volatility. At zero volatility, the value of the call option is 4.60% of par; at 30% volatility, it is 14.78% of par. Thus, as interest rate volatility increases, the value of the callable bond decreases.

Exhibit 6 Value of a 30-Year 3.75% Bond Putable at Par in 10 Years under Different Volatility Scenarios Assuming a 4% Flat Yield Curve

In Exhibit 6, the shaded component is the value of the straight bond, and the white component is the value of the put option; thus, the stacked bars represent the value of the putable bond. All else being equal, the put option increases in value with interest rate volatility. At zero volatility, the value of the put option is 2.30% of par; at 30% volatility, it is 10.54% of par. Thus, as interest rate volatility increases, the value of the putable bond increases.

3.2 Level and Shape of the Yield Curve

The value of a callable or putable bond is also affected by changes in the level and shape of the yield curve.

3.2.1 *Effect on the Value of a Callable Bond*

Exhibit 7 shows the value of the same callable bond as in Exhibit 5 under different flat yield curve levels assuming an interest rate volatility of 15%.

Exhibit 7 Value of a 30-Year 4.50% Bond Callable at Par in 10 Years under Different Flat Yield Curve Levels at 15% Interest Rate Volatility

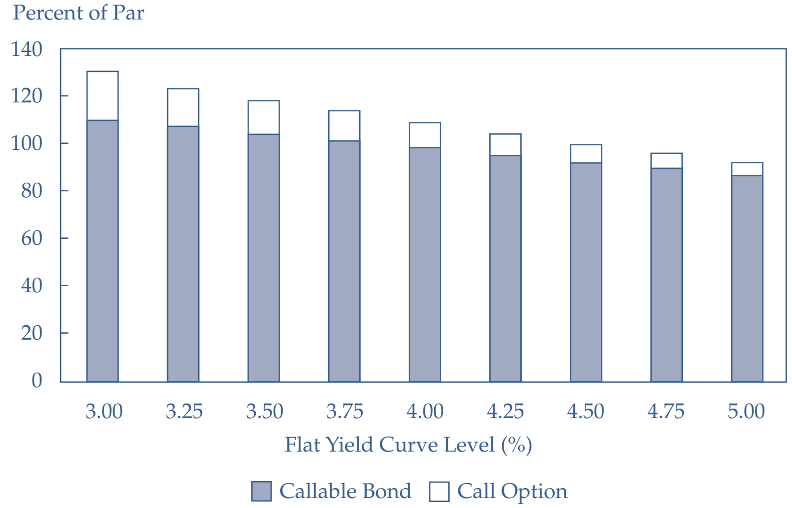
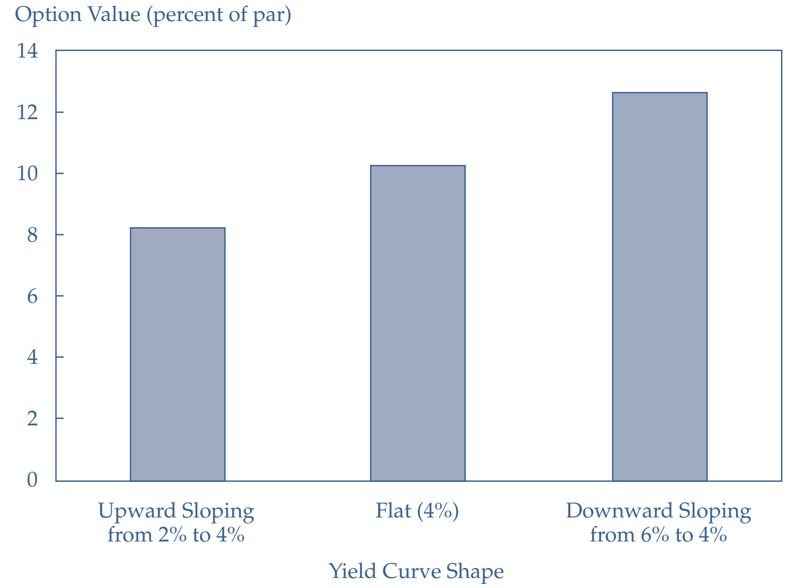


Exhibit 7 shows that as interest rates decline, the value of the straight bond rises; however, the rise is partially offset by the increase in the value of the call option. For example, if the yield curve is 5% flat, the value of the straight bond is 92.27% of par and the value of the call option is 5.37% of par; thus, the value of the callable bond is 86.90% of par. If the yield curve declines to 3% flat, the value of the straight bond rises by 40% to 129.54% of par, but the value of the callable bond increases by only 27% to 110.43% of par. Thus, the value of the callable bond rises less rapidly than the value of the straight bond, limiting the upside potential for the investor.

The value of a call option, and thus the value of a callable bond, is also affected by changes in the shape of the yield curve, as illustrated in Exhibit 8.

Exhibit 8 Value of a Call Option Embedded in a 30-Year 4.50% Bond Callable at Par in 10 Years under Different Yield Curve Shapes at 15% Interest Rate Volatility



All else being equal, the value of the call option increases as the yield curve flattens. If the yield curve is upward sloping with short-term rates at 2% and long-term rates at 4% (the first bar), the value of the call option represents approximately 8% of par. It rises to approximately 10% of par if the yield curve flattens to 4% (the second bar). The value of the call option increases further if the yield curve actually inverts. Exhibit 8 shows that it exceeds 12% of par if the yield curve is downward sloping with short-term rates at 6% and long-term rates at 4% (the third bar). An inverted yield curve is rare but does happen from time to time.

The intuition to explain the effect of the shape of the yield curve on the value of the call option is as follows. When the yield curve is upward sloping, the one-period forward rates on the interest rate tree are high and opportunities for the issuer to call the bond are fewer. When the yield curve flattens or inverts, many nodes on the tree have lower forward rates that increase the opportunities to call.

Assuming a normal, upward-sloping yield curve at the time of issue, the call option embedded in a callable bond issued at par is out of the money. It would not be called if the arbitrage-free forward rates at zero volatility prevailed. Callable bonds issued at a large premium, as happens frequently in the municipal sector in the United States, are in the money. They will be called if the arbitrage-free forward rates prevail.

3.2.2 *Effect on the Value of a Putable Bond*

Exhibits 9 and 10 show how changes in the level and shape of the yield curve affect the value of the putable bond used in Exhibit 6.

Exhibit 9 Value of a 30-Year 3.75% Bond Putable at Par in 10 Years under Different Flat Yield Curve Levels at 15% Interest Rate Volatility

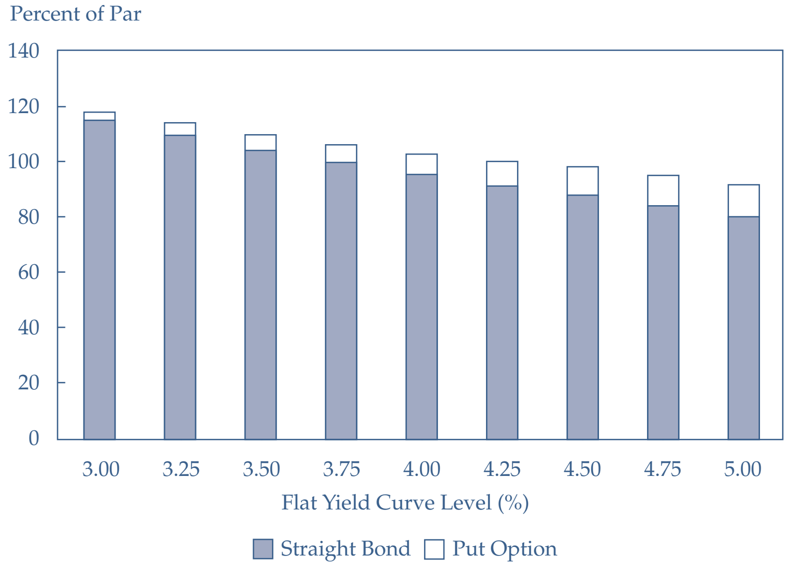
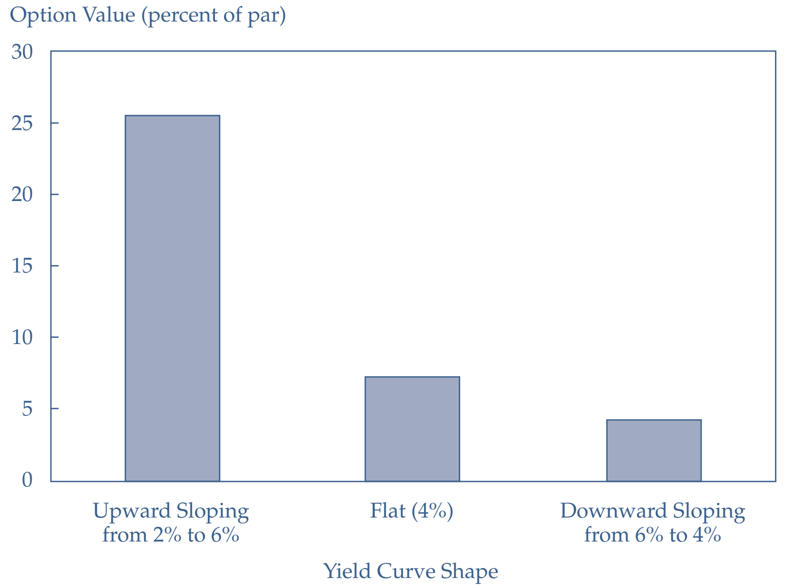


Exhibit 9 illustrates why the put option is considered a hedge against rising interest rates for investors. As interest rates rise, the value of the straight bond declines; however, the decline is partially offset by the increase in the value of the put option. For example, if the yield curve moves from 3% flat to 5% flat, the value of the straight bond falls by 30% while the fall in the value of the putable bond is limited to 22%.

Exhibit 10 Value of the Put Option Embedded in a 30-Year 3.75% Bond Putable at Par in 10 Years under Different Yield Curve Shapes at 15% Interest Rate Volatility



All else being equal, the value of the put option decreases as the yield curve moves from being upward sloping, to flat, to downward sloping. When the yield curve is upward sloping, the one-period forward rates in the interest rate tree are high, which creates more opportunities for the investor to put the bond. As the yield curve flattens or inverts, the number of opportunities declines.

VALUATION OF DEFAULT-FREE CALLABLE AND PUTABLE BONDS IN THE PRESENCE OF INTEREST RATE VOLATILITY

4

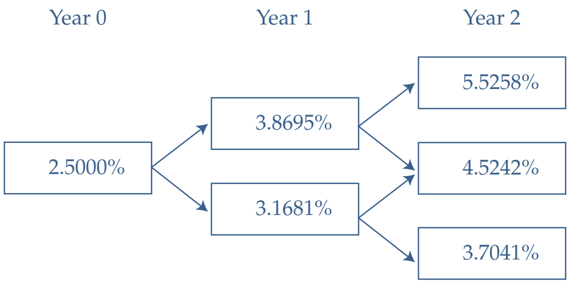
f calculate the value of a callable or putable bond from an interest rate tree

The procedure to value a bond with an embedded option in the presence of interest rate volatility is as follows:

- Generate a tree of interest rates based on the given yield curve and interest rate volatility assumptions.
- At each node of the tree, determine whether the embedded option will be exercised.
- Apply the backward induction valuation methodology to calculate the bond's present value. This methodology involves starting at maturity and working back from right to left to find the bond's present value.

Let us return to the default-free three-year 4.25% annual coupon bonds discussed earlier to illustrate how to apply this valuation procedure. The bonds' characteristics are identical. The yield curve given in Exhibit 1 remains the same—with one-year, two-year, and three-year par yields of 2.500%, 3.000%, and 3.500%, respectively. But we now assume an interest rate volatility of 10% instead of 0%. The resulting binomial interest rate tree showing the one-year forward rates zero, one, and two years from now is shown in Exhibit 11. The branching from each node to an up state and a down state is assumed to occur with equal probability.

Exhibit 11 Binomial Interest Rate Tree at 10% Interest Rate Volatility



The calibration of a binomial interest rate tree was discussed in earlier coverage of fixed-income concepts. As mentioned before, the one-year par rate, the one-year spot rate, and the one-year forward rate zero years from now are identical (2.500%). Because there is no closed-form solution, the one-year forward rates one year from now in the two states are determined iteratively by meeting the following two constraints:

1 The rate in the up state (R_u) is given by

$$R_u = R_d \times e^{2\sigma\sqrt{t}}$$

where R_d is the rate in the down state, σ is the interest rate volatility (10% here), and t is the time in years between "time slices" (a year, so here $t = 1$).

2 The discounted value of a two-year par bond (bearing a 3.000% coupon rate in this example) equals 100.

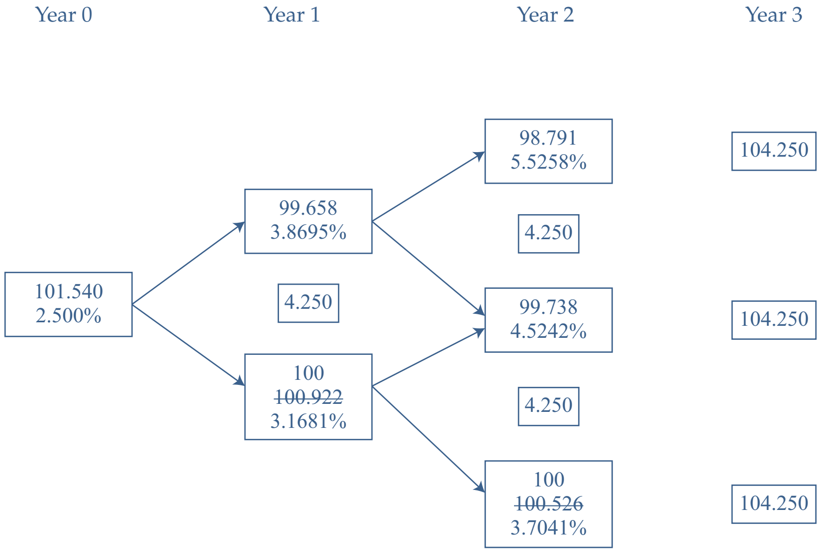
In Exhibit 11 at the one-year time slice, R_d is 3.1681% and R_u is 3.8695%. Having established the rates that correctly value the one-year and two-year par bonds implied by the given par yield curve, we freeze these rates and proceed to iterate the rates in the next time slice to determine the one-year forward rates in the three states two years from now. The same constraints as before apply: (1) Each rate must be related to its neighbor by the factor $e^{2\sigma\sqrt{t}}$, and (2) the rates must discount a three-year par bond (bearing a 3.500% coupon rate in this example) to a value of 100.

Now that we have determined all the one-year forward rates, we can value the three-year 4.25% annual coupon bonds that are either callable or putable at par one year and two years from now.

4.1 Valuation of a Callable Bond with Interest Rate Volatility

Exhibit 12 depicts the valuation of a callable bond at 10% volatility.

Exhibit 12 Valuation of a Default-Free Three-Year 4.25% Annual Coupon Bond Callable at Par One Year and Two Years from Now at 10% Interest Rate Volatility



The coupon and principal cash flows are placed directly to the right of the interest rate nodes. The calculated bond values at each node are placed above the interest rate. We start by calculating the bond values at Year 2 by discounting the cash flow for Year 3 with the three possible rates.

$$98.791 = \frac{104.250}{1.055258}$$

$$99.738 = \frac{104.250}{1.045242}$$

$$100.526 = \frac{104.250}{1.037041}$$

Because the bond is callable at par in Year 2, we check each scenario to determine whether the present value of the future cash flows is higher than the call price, in which case the issuer calls the bond. Exercise happens only at the bottom of the tree, where the rate is 3.7041%, and so we reset the value from 100.526 to 100 in that state.

The value in each state of Year 1 is calculated by discounting the values in the two future states emanating from the present state plus the coupon at the appropriate rate in the present state:

$$99.658 = \frac{4.250 + (0.5 \times 98.791 + 0.5 \times 99.738)}{1.038695}.$$

The first term in the numerator is the coupon payment, and the second term is the expected bond value at Year 2. In this model, the probabilities for moving to the higher and lower node are the same (0.5):

$$100.922 = \frac{4.250 + (0.5 \times 99.738 + 0.5 \times 100)}{1.031681}.$$

Notice that the reset value of 100 is used to get the expected bond value. Once again the bond will be callable at the lower node where the interest rate is 3.1681%.

At Year 0, the value of the callable bond is 101.540:

$$101.540 = \frac{4.250 + \left(0.5 \times 99.658 + 0.5 \times 100\right)}{1.025000}.$$

The value of the call option, obtained by taking the difference between the value of the straight bond and the value of the callable bond, is now 0.574 (102.114 – 101.540). The fact that the value of the call option is larger at 10% volatility than at 0% volatility (0.407) is consistent with our earlier discussion that option value increases with interest rate volatility.

EXAMPLE 3

Valuation of a Callable Bond Assuming Interest Rate Volatility

Return to the valuation of the Bermudan-style three-year 4.25% annual coupon bond callable at par one year and two years from now as depicted in Exhibit 12. The one-year, two-year, and three-year par yields are 2.500%, 3.000%, and 3.500%, respectively, and the interest rate volatility is 10%.

1 Assume that nothing changes relative to the initial setting except that the interest rate volatility is now 15% instead of 10%. The new value of the callable bond is:

A less than 101.540.

B equal to 101.540.

C more than 101.540.

2 Assume that nothing changes relative to the initial setting except that the bond is now callable at 102 instead of 100. The new value of the callable bond is *closest to*:

A 100.000.

B 102.000.

C 102.114.

Solution to 1:

A is correct. A higher interest rate volatility increases the value of the call option. Because the value of the call option is subtracted from the value of the straight bond to obtain the value of the callable bond, a higher value for the call option leads to a lower value for the callable bond. Thus, the value of the callable bond at 15% volatility is less than that at 10% volatility—that is, less than 101.540.

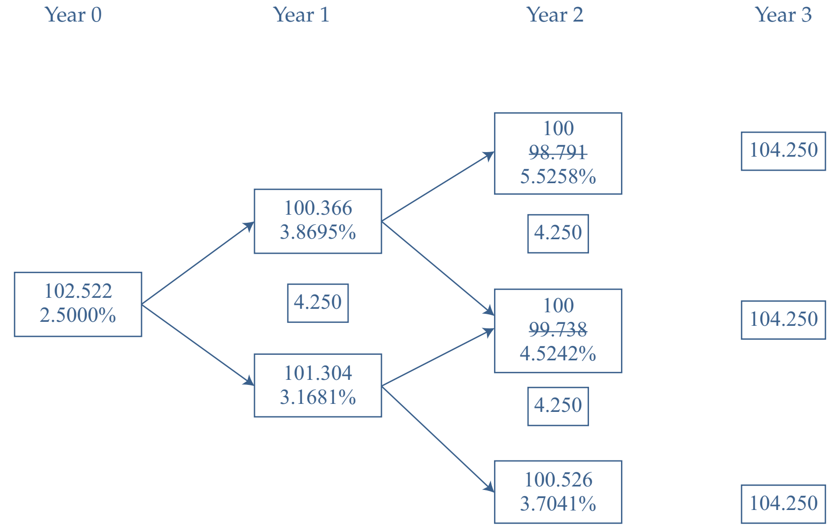
Solution to 2:

C is correct. Looking at Exhibit 12, the call price is too high for the call option to be exercised in any scenario. Thus, the value of the call option is zero, and the value of the callable bond is equal to the value of the straight bond—that is, 102.114.

4.2 Valuation of a Putable Bond with Interest Rate Volatility

The valuation of the three-year 4.25% annual coupon bond putable at par one year and two years from now at 10% volatility is depicted in Exhibit 13. The procedure for valuing a putable bond is very similar to that described earlier for valuing a callable bond, except that in each state, the bond's value is compared with the put price. The investor puts the bond only when the present value of the bond's future cash flows is lower than the put price. In this case, the value is reset to the put price (100). It happens twice in Year 2, in the states where the interest rates are 5.5258% and 4.5242%. The investor would not exercise the put option in Year 1 because the values for the bond exceed the put price.

Exhibit 13 Valuation of a Default-Free Three-Year 4.25% Annual Coupon Bond Putable at Par One Year and Two Years from Now at 10% Interest Rate Volatility



The value of the putable bond is 102.522. The value of the put option, obtained by taking the difference between the value of the putable bond and the value of the straight bond, is now 0.408 (102.522 − 102.114). As expected, the value of the put option is larger at 10% volatility than at 0% volatility (0.283).

EXAMPLE 4

Valuation of a Putable Bond Assuming Interest Rate Volatility

Return to the valuation of the Bermudan-style three-year 4.25% annual coupon bond putable at par one year and two years from now, as depicted in Exhibit 13. The one-year, two-year, and three-year par yields are 2.500%, 3.000%, and 3.500%, respectively, and the interest rate volatility is 10%.

1 Assume that nothing changes relative to the initial setting except that the interest rate volatility is now 20% instead of 10%. The new value of the putable bond is:

A less than 102.522.

B equal to 102.522.

C more than 102.522.

2 Assume that nothing changes relative to the initial setting except that the bond is now putable at 95 instead of 100. The new value of the putable bond is *closest to*:

A 97.522.

B 102.114.

C 107.522.

Solution to 1:

C is correct. A higher interest rate volatility increases the value of the put option. Because the value of the put option is added to the value of the straight bond to obtain the value of the putable bond, a higher value for the put option leads to a higher value for the putable bond. Thus, the value of the putable bond at 20% volatility is more than that at 10% volatility—that is, more than 102.522.

Solution to 2:

B is correct. Looking at Exhibit 13, the put price is too low for the put option to be exercised in any scenario. Thus, the value of the put option is zero, and the value of the putable bond is equal to the value of the straight bond—that is, 102.114.

PUTABLE VS. EXTENDIBLE BONDS

Putable and extendible bonds are equivalent, except that their underlying option-free bonds are different. Consider a three-year 3.30% bond putable in Year 2. Its value should be exactly the same as that of a two-year 3.30% bond extendible by one year. Otherwise, there would be an arbitrage opportunity. Clearly, the cash flows of the two bonds are identical up to Year 2. The cash flows in Year 3 are dependent on the one-year forward rate two years from now. These cash flows will also be the same for both bonds regardless of the level of interest rates at the end of Year 2.

If the one-year forward rate at the end of Year 2 is higher than 3.30%, the putable bond will be put because the bondholder can reinvest the proceeds of the retired bond at a higher yield and the extendible bond will not be extended for the same reason. So, both bonds pay 3.30% for two years and are then redeemed. Alternatively, if the one-year forward rate at the end of Year 2 is lower than 3.30%, the putable bond will not be put because the bondholder would not want to reinvest at a lower yield and the extendible bond will be extended to hold onto the higher interest rate. Thus, both bonds pay 3.30% for three years and are then redeemed.

EXAMPLE 5

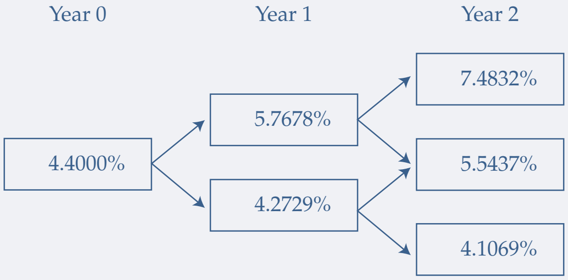
Valuation of Bonds with Embedded Options Assuming Interest Rate Volatility

Sidley Brown, a fixed-income associate at KMR Capital, is analyzing the effect of interest rate volatility on the values of callable and putable bonds issued by Weather Analytics (WA). WA is owned by the sovereign government, so its bonds are considered default free. Brown is currently looking at three of WA's bonds and has gathered the following information about them:

Characteristic	Bond X	Bond Y	Bond Z
Time to maturity	Three years from today	Three years from today	Three years from today
Coupon	5.2% annual	Not available	4.8% annual

Characteristic	Bond X	Bond Y	Bond Z
Type of bond	Callable at par one year and two years from today	Callable at par one year and two years from today	Putable at par two years from today
Price (as a % of par)	Not available	101.325	Not available

The one-year, two-year, and three-year par rates are 4.400%, 4.700%, and 5.000%, respectively. Based on an estimated interest rate volatility of 15%, Brown has constructed the following binomial interest rate tree:



1 The price of Bond X is *closest to*:

A 96.057% of par.

B 99.954% of par.

C 100.547% of par.

2 The coupon rate of Bond Y is *closest to*:

A 4.200%.

B 5.000%.

C 6.000%.

3 The price of Bond Z is *closest to*:

A 99.638% of par.

B 100.340% of par.

C 100.778% of par.

Brown is now analyzing the effect of interest rate volatility on the price of WA's bonds.

4 Relative to its price at 15% interest rate volatility, the price of Bond X at a lower interest rate volatility will be:

A lower.

B the same.

C higher.

5 Relative to its price at 15% interest rate volatility, the price of Bond Z at a higher interest rate volatility will be:

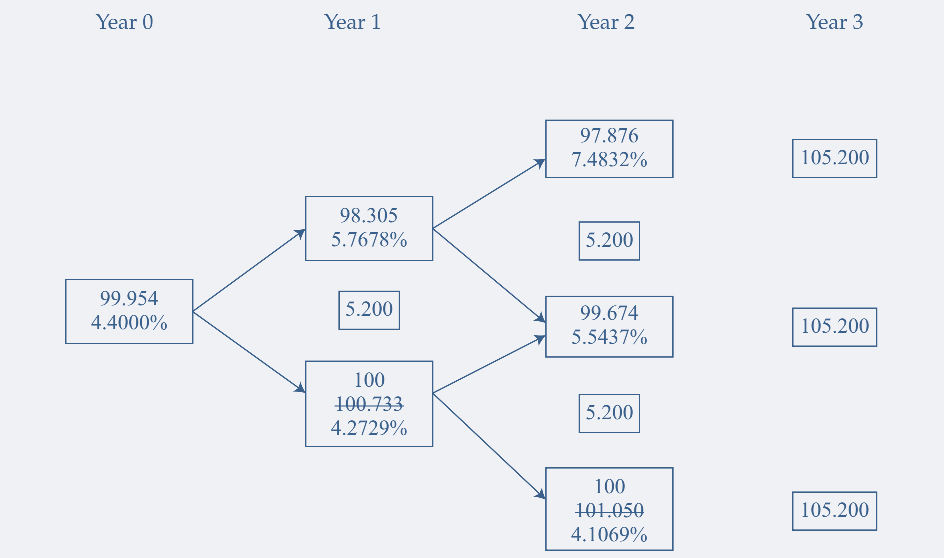
A lower.

B the same.

C higher.

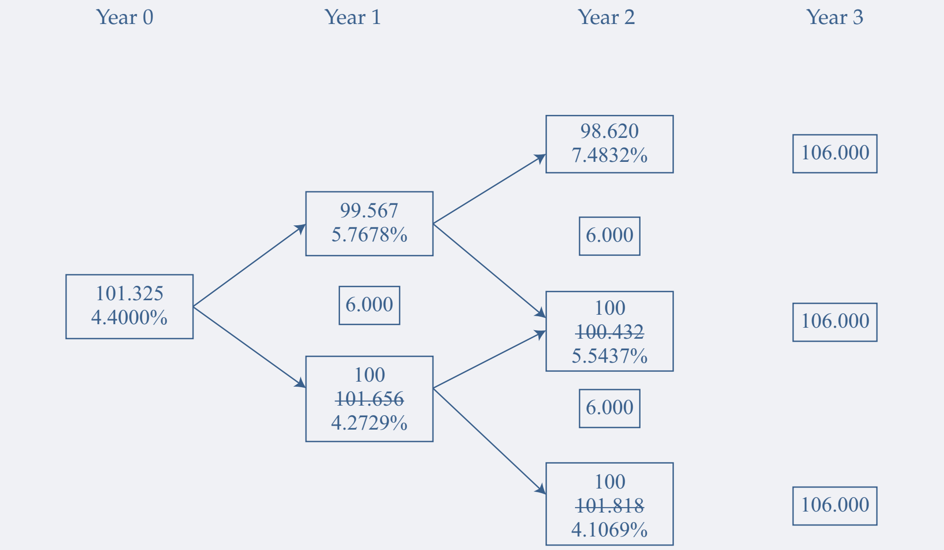
Solution to 1:

B is correct.



Solution to 2:

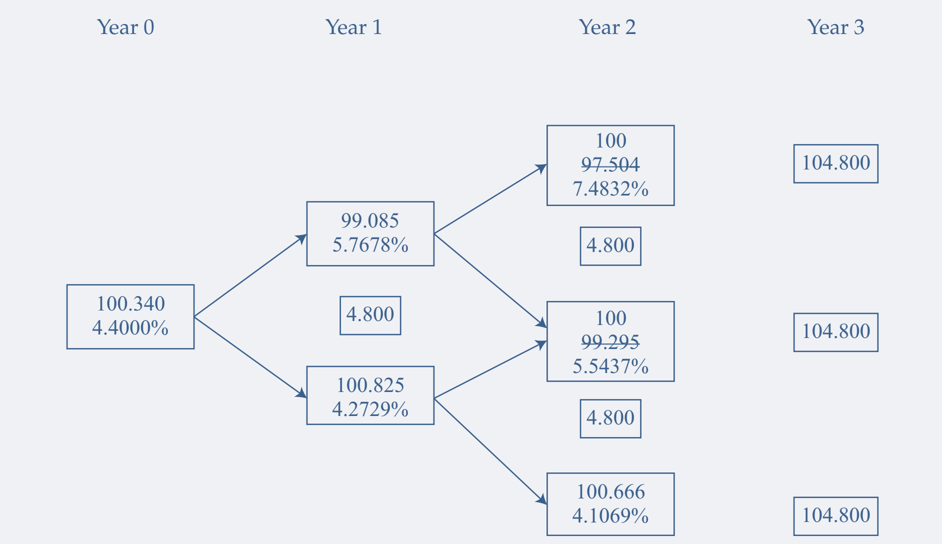
C is correct.



Although the correct answer can be found by using the interest rate tree depicted, it is possible to identify it by realizing that the other two answers are clearly incorrect. The three-year 5% straight bond is worth par given that the three-year par rate is 5%. Because the presence of a call option reduces the price of a callable bond, a three-year 5% bond callable at par can only be worth less than par—and certainly less than 101.325 given the yield curve and interest rate volatility assumptions—so B is incorrect. The value of a bond with a coupon rate of 4% is even less, so A is incorrect. Thus, C must be the correct answer.

Solution to 3:

B is correct.



Solution to 4:

C is correct. Bond X is a callable bond. As shown in Equation 1, the value of the call option decreases the value of Bond X relative to the value of the underlying option-free bond. As interest rate volatility decreases, the value of the call option decreases; thus, the value of Bond X increases.

Solution to 5:

C is correct. Bond Z is a putable bond. As shown in Equation 2, the value of the put option increases the value of Bond Z relative to the value of the underlying option-free bond. As interest rate volatility increases, the value of the put option increases; thus, the value of Bond Z increases.

5 VALUATION OF RISKY CALLABLE AND PUTABLE BONDS

g explain the calculation and use of option-adjusted spreads

h explain how interest rate volatility affects option-adjusted spreads

Although the approach described earlier for default-free bonds may apply to securities issued by sovereign governments in their local currency, the fact is that most bonds are subject to default. Accordingly, we have to extend the framework to the valuation of risky bonds.

Two distinct approaches to valuing bonds are subject to default risk. The industry-standard approach is to increase the discount rates above the default-free rates to reflect default risk. Higher discount rates imply lower present values, and thus the value of a risky bond will be lower than that of an otherwise identical default-free bond.

The second approach to valuing risky bonds is to make the default probabilities explicit—that is, assigning a probability to each time period going forward. For example, the probability of default in Year 1 may be 1%; the probability of default in Year 2, conditional on surviving Year 1, may be 1.25%; and so on. This approach requires specifying the recovery value given default (e.g., 40% of par). Information about default probabilities and recovery values may be accessible from credit default swaps. This important topic is covered elsewhere.

5.1 Option-Adjusted Spread

Depending on available information, two standard approaches are used to construct a suitable yield curve for a risky bond. The more satisfactory but less convenient one is to use an issuer-specific curve, which represents the issuer's borrowing rates over the relevant range of maturities. Unfortunately, most bond professionals do not have access to such a level of detail. A more convenient and relatively satisfactory alternative is to uniformly raise the one-year forward rates derived from the default-free benchmark yield curve by a fixed spread, which is estimated from the market prices of suitable bonds of similar credit quality. This fixed spread is known as the zero-volatility spread, or Z-spread.

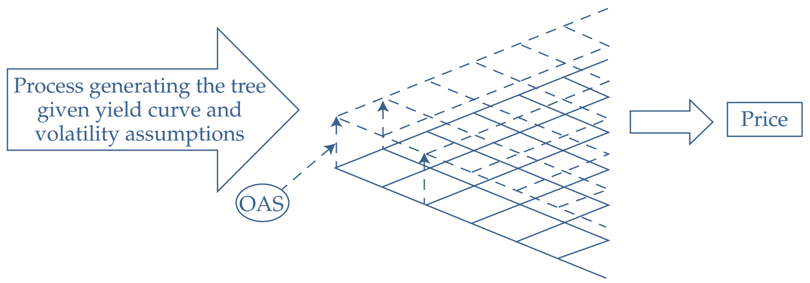
To illustrate, we return to the three-year 4.25% option-free bond introduced earlier, but now we assume that it is a risky bond and that the appropriate Z-spread is 100 bps. To calculate the arbitrage-free value of this bond, we have to increase each of the one-year forward rates given in Exhibit 1 by the Z-spread of 100 bps:

$$\frac{4.25}{(1.03500)} + \frac{4.25}{(1.03500)(1.04518)} + \frac{104.25}{(1.03500)(1.04518)(1.05564)} = 99.326.$$

As expected, the value of this risky bond (99.326) is considerably lower than the value of an otherwise identical but default-free bond (102.114).

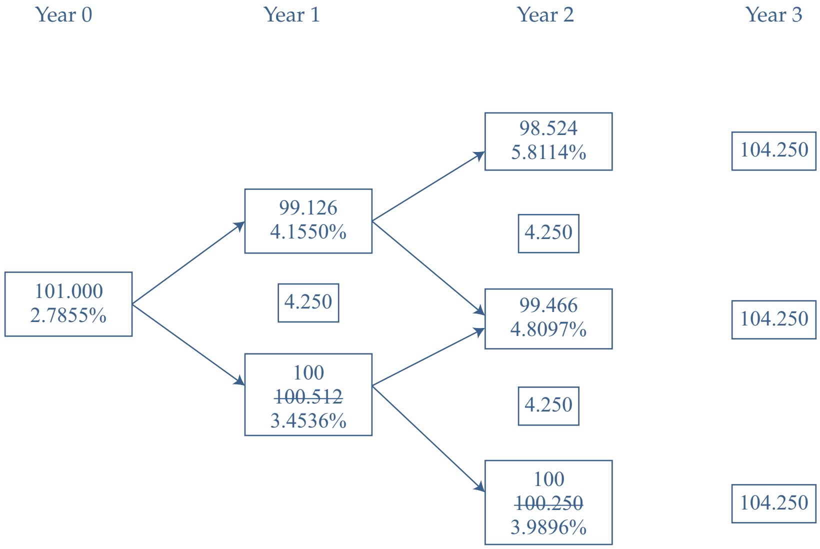
The same approach can be applied to the interest rate tree when valuing risky bonds with embedded options. In this case, an **option-adjusted spread** (OAS) is used. As depicted in Exhibit 14, the OAS is the constant spread that when added to all the one-period forward rates on the interest rate tree, makes the arbitrage-free value of the bond equal to its market price. Note that the Z-spread for an option-free bond is simply its OAS at zero volatility.

Exhibit 14 Interest Rate Tree and OAS



If the bond's price is given, the OAS is determined by trial and error. For example, suppose that the market price of a three-year 4.25% annual coupon bond callable in one year and two years from now (identical to the one valued in Exhibit 12 except that it is risky instead of default-free) is 101.000. To determine the OAS, we try shifting all the one-year forward rates in each state by adding a constant spread. For example, when we add 30 bps to all the one-year forward rates, we obtain a value for the callable bond of 100.973, which is lower than the bond's price. Because of the inverse relationship between a bond's price and its yield, this result means that the discount rates are too high, so we try a slightly lower spread. Adding 28 bps results in a value for the callable bond of 101.010, which is slightly too high. As illustrated in Exhibit 15, the constant spread added uniformly to all the one-period forward rates that justifies the given market price of 101.000 is 28.55 bps; this number is the OAS.

Exhibit 15 OAS of a Risky Three-Year 4.25% Annual Coupon Bond Callable at Par One Year and Two Years from Now at 10% Interest Rate Volatility



As illustrated in Exhibit 15, the value at each node is adjusted based on whether the call option is exercised. Thus, the OAS removes the amount that results from the option risk, which is why this spread is called "option adjusted."

OAS is often used as a measure of value relative to the benchmark. An OAS lower than that for a bond with similar characteristics and credit quality indicates that the bond is likely overpriced (rich) and should be avoided. A larger OAS than that of a bond with similar characteristics and credit quality means that the bond is likely underpriced (cheap). If the OAS is close to that of a bond with similar characteristics and credit quality, the bond looks fairly priced. In our example, the OAS at 10% volatility is 28.55 bps. This number should be compared with the OAS of bonds with similar characteristics and credit quality to make a judgment about the bond's attractiveness.

5.2 Effect of Interest Rate Volatility on Option-Adjusted Spread

The dispersion of interest rates on the tree is volatility dependent, and so is the OAS. Exhibit 16 shows the effect of volatility on the OAS for a callable bond. The bond is a 5% annual coupon bond with 23 years left to maturity, callable in three years, priced at 95% of par, and valued assuming a flat yield curve of 4%.

Exhibit 16 Effect of Interest Rate Volatility on the OAS for a Callable Bond

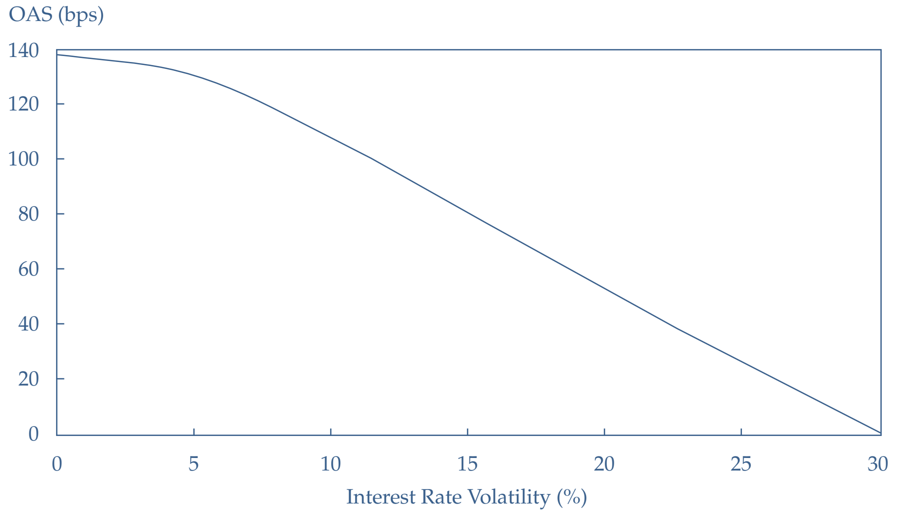
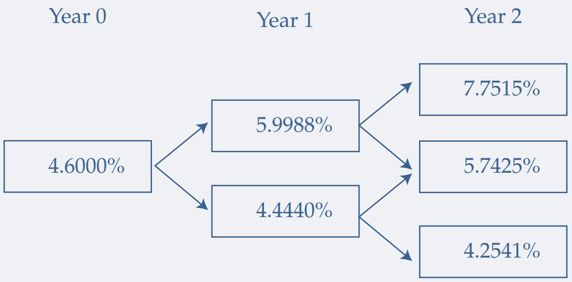


Exhibit 16 shows that as interest rate volatility increases, the OAS for the callable bond decreases. The OAS drops from 138.2 bps at 0% volatility to 1.2 bps at 30% volatility. This exhibit clearly demonstrates the importance of the interest rate volatility assumption. Returning to the example in Exhibit 15, the callable bond may look underpriced at 10% volatility. If an investor assumes a higher volatility, however, the OAS and thus relative cheapness will decrease.

EXAMPLE 6

Option-Adjusted Spread

Robert Jourdan, a portfolio manager, has just valued a 7% annual coupon bond that was issued by a French company and has three years remaining until maturity. The bond is callable at par one year and two years from now. In his valuation, Jourdan used the yield curve based on the on-the-run French government bonds. The one-year, two-year, and three-year par rates are 4.600%, 4.900%, and 5.200%, respectively. Based on an estimated interest rate volatility of 15%, Jourdan constructed the following binomial interest rate tree:



Jourdan valued the callable bond at 102.294% of par. However, Jourdan's colleague points out that because the corporate bond is riskier than French government bonds, the valuation should be performed using an OAS of 200 bps.

1 To update his valuation of the French corporate bond, Jourdan should:

A subtract 200 bps from the bond's annual coupon rate.

B add 200 bps to the rates in the binomial interest rate tree.

C subtract 200 bps from the rates in the binomial interest rate tree.

2 All else being equal, the value of the callable bond at 15% volatility is *closest to*:

A 99.198% of par.

B 99.247% of par.

C 104.288% of par.

3 Holding the price calculated in the previous question, the OAS for the callable bond at 20% volatility will be:

A lower.

B the same.

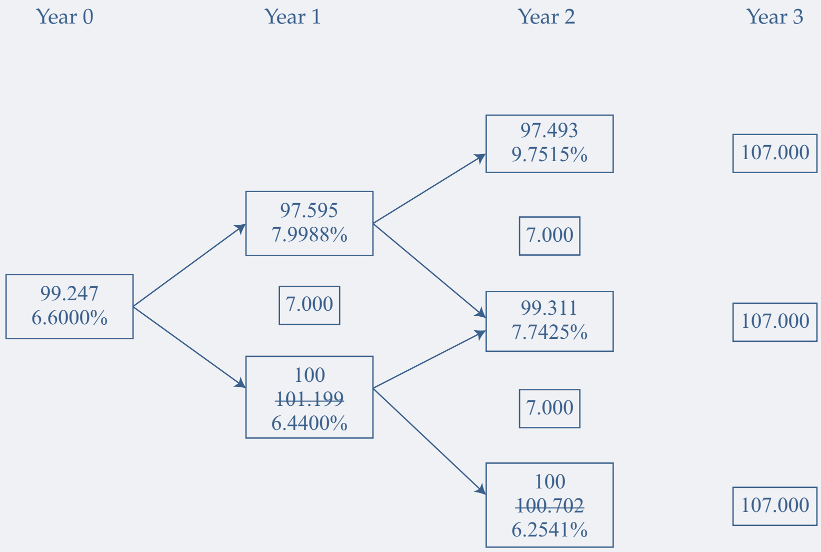
C higher.

Solution to 1:

B is correct. The OAS is the constant spread that must be *added* to all the one-period forward rates given in the binomial interest rate tree to justify a bond's given market price.

Solution to 2:

B is correct.



Solution to 3:

A is correct. If interest rate volatility increases from 15% to 20%, the OAS for the callable bond will decrease.

SCENARIO ANALYSIS OF BONDS WITH OPTIONS

Another application of valuing bonds with embedded options is scenario analysis over a specified investment horizon. In addition to reinvestment of interest and principal, option valuation comes into play in that callable and putable bonds can be redeemed and their proceeds reinvested during the holding period. Making scenario-dependent, optimal option-exercise decisions involves computationally intensive use of OAS technology because the call or put decision must be evaluated considering the evolution of interest rate scenarios during the holding period.

Performance over a specified investment horizon entails a trade-off between reinvestment of cash flows and change in the bond's value. Let us take the example of a 4.5% bond with five years left to maturity and assume that the investment horizon is one year. If the bond is option free, higher interest rates increase the reinvestment income but result in lower principal value at the end of the investment horizon. Because the investment horizon is short, reinvestment income is relatively insignificant and performance will be dominated by the change in the value of the principal. Accordingly, lower interest rates will result in superior performance.

If the bond under consideration is callable, however, it is not at all obvious how the interest rate scenario affects performance. Suppose, for example, that the bond is first callable six months from now and that its current market price is 99.74. Steeply rising interest rates would depress the bond's price, and performance would definitely suffer. But steeply declining interest rates would also be detrimental because the bond would be called and *both interest and principal* would have to be reinvested at lower interest rates. Exhibit 17 shows the return over the one-year investment horizon for the 4.5% bond first callable in six months with five years left to maturity and valued on a 4% flat yield curve.

Exhibit 17 Effect of Interest Rate Changes on a Callable Bond's Total Return

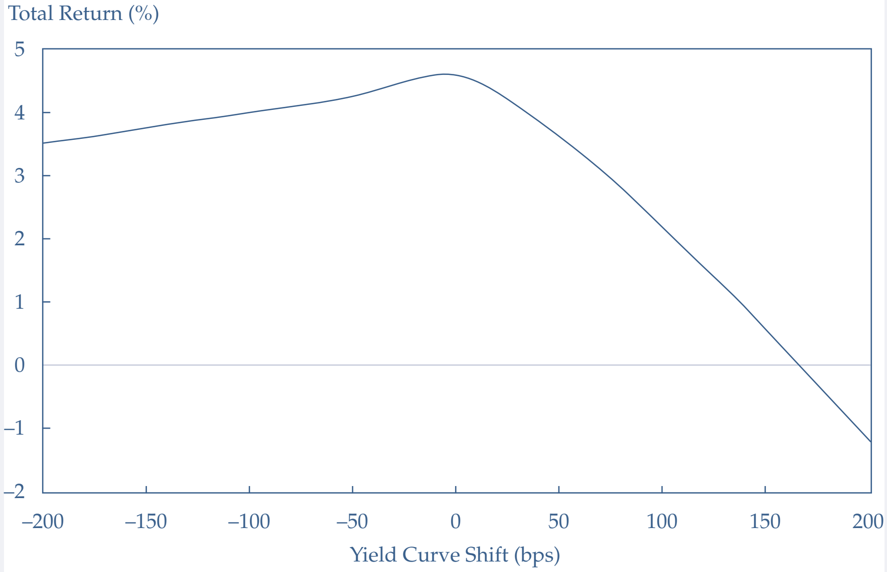


Exhibit 17 clearly shows that lower interest rates do not guarantee higher returns for callable bonds. The point to keep in mind is that the bond may be called long before the end of the investment horizon. Assuming that it is called on the horizon date would overestimate performance. Thus, a realistic prediction of option exercise is essential when performing scenario analysis of bonds with embedded options.

6 BONDS WITH EMBEDDED OPTIONS: EFFECTIVE DURATION

i. calculate and interpret effective duration of a callable or putable bond

j. compare effective durations of callable, putable, and straight bonds

Measuring and managing exposure to interest rate risk are two essential tasks of fixed-income portfolio management. Applications range from hedging a portfolio to asset–liability management of financial institutions. Portfolio managers, whose performance is often measured against a benchmark, also need to monitor the interest rate risk of both their portfolio and the benchmark. In this section, we cover two key measures of interest rate risk: duration and convexity.

6.1 Duration

The duration of a bond measures the sensitivity of the bond's full price (including accrued interest) to changes in the bond's yield to maturity (in the case of *yield* duration measures) or to changes in benchmark interest rates (in the case of yield-curve or *curve* duration measures). Yield duration measures, such as modified duration, can be used only for option-free bonds because these measures assume that a bond's expected cash flows do not change when the yield changes. This assumption is in general false for bonds with embedded options because the values of embedded options are typically contingent on interest rates. Thus, for bonds with embedded options, the only appropriate duration measure is the curve duration measure known as effective (or option-adjusted) duration. Because effective duration works for straight bonds as well as for bonds with embedded options, practitioners tend to use it regardless of the type of bond being analyzed.

6.1.1 *Effective Duration*

Effective duration indicates the sensitivity of the bond's price to a 100 bps parallel shift of the benchmark yield curve—in particular, the government par curve—assuming no change in the bond's credit spread (*Note:* Although it is possible to explore how arbitrary changes in interest rates affect the bond's price, in practice the change is usually specified as a parallel shift of the benchmark yield curve). The formula for calculating a bond's effective duration is

$$\text{EffDur} = \frac{(PV_{-}) - (PV_{+})}{2 \times (\Delta\text{Curve}) \times (PV_0)}, \tag{3}$$

where

ΔCurve = the magnitude of the parallel shift in the benchmark yield curve (in decimal)

PV_{-} = the full price of the bond when the benchmark yield curve is shifted down by ΔCurve

PV_{+} = the full price of the bond when the benchmark yield curve is shifted up by ΔCurve

PV_0 = the current full price of the bond (i.e., with no shift)

How is this formula applied in practice? Without a market price, we would need an issuer-specific yield curve to compute PV_0, PV_{-}, and PV_{+}. But practitioners usually have access to the bond's current price and thus use the following procedure:

1 Given a price (PV_0), calculate the implied OAS to the benchmark yield curve at an appropriate interest rate volatility.

2 Shift the benchmark yield curve down, generate a new interest rate tree, and then revalue the bond using the OAS calculated in Step 1. This value is PV_{-}.

3 Shift the benchmark yield curve up by the same magnitude as in Step 2, generate a new interest rate tree, and then revalue the bond using the OAS calculated in Step 1. This value is PV_{+}.

4 Calculate the bond's effective duration using Equation 3.

Let us illustrate using the same three-year 4.25% bond callable at par one year and two years from now, the same par yield curve (i.e., one-year, two-year, and three-year par yields of 2.500%, 3.000%, and 3.500%, respectively), and the same interest rate volatility (10%) as before. Also as before, we assume that the bond's current full price is 101.000. We apply the procedure just described:

1 As shown in Exhibit 15, given a price (PV_0) of 101.000, the OAS at 10% volatility is 28.55 bps.
2 We shift the par yield curve down by, say, 30 bps, generate a new interest rate tree, and then revalue the bond at an OAS of 28.55 bps. As shown in Exhibit 18, PV_- is 101.599.
3 We shift the par yield curve up by the same 30 bps, generate a new interest rate tree, and then revalue the bond at an OAS of 28.55 bps. As shown in Exhibit 19, PV_+ is 100.407.
4 Thus,

$$\text{EffDur} = \frac{101.599 - 100.407}{2 \times 0.0030 \times 101.000} = 1.97.$$

An effective duration of 1.97 indicates that a 100 bps increase in interest rate would reduce the value of the three-year 4.25% callable bond by 1.97%.

Exhibit 18 Valuation of a Three-Year 4.25% Annual Coupon Bond Callable at Par One Year and Two Years from Now at 10% Interest Rate Volatility with an OAS of 28.55 bps When Interest Rates Are Shifted Down by 30 bps

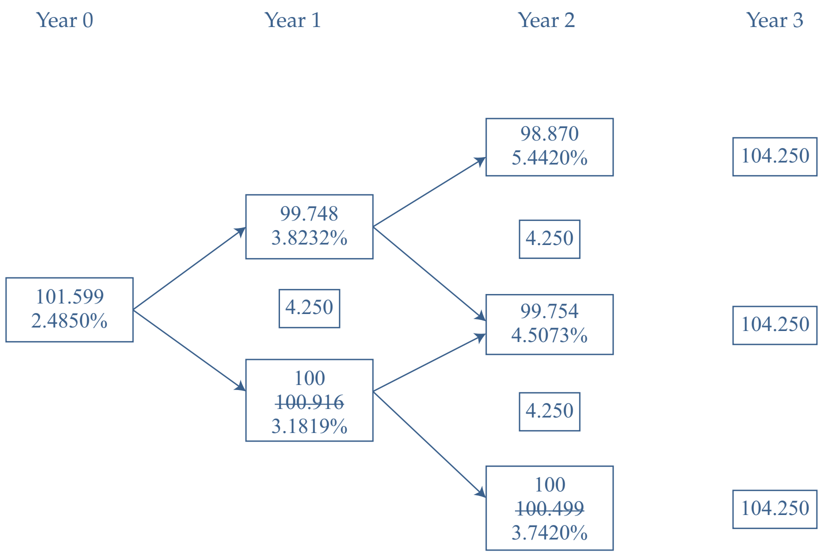
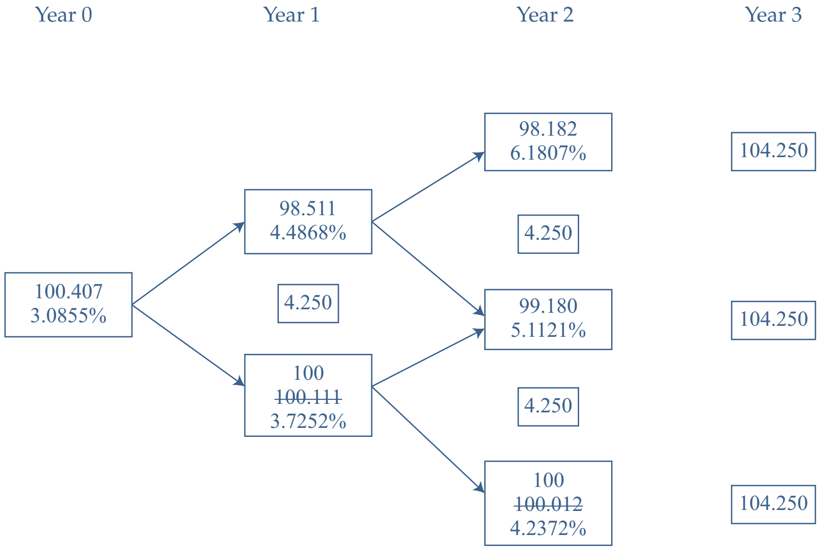


Exhibit 19 Valuation of a Three-Year 4.25% Annual Coupon Bond Callable at Par One Year and Two Years from Now at 10% Interest Rate Volatility with an OAS of 28.55 bps When Interest Rates Are Shifted Up by 30 bps



The effective duration of a callable bond cannot exceed that of the straight bond. When interest rates are high relative to the bond's coupon, the call option is out of the money so the bond is unlikely to be called. Thus, the effect of an interest rate change on the price of a callable bond is very similar to that on the price of an otherwise identical option-free bond; the callable and straight bonds have very similar effective durations. In contrast, when interest rates fall, the call option moves into the money. Remember that the call option gives the issuer the right to retire the bond at the call price and thus limits the price appreciation when interest rates decline. As a consequence, the call option reduces the effective duration of the callable bond relative to that of the straight bond.

The effective duration of a putable bond also cannot exceed that of the straight bond. When interest rates are low relative to the bond's coupon, the put option is out of the money so the bond is unlikely to be put. Thus, the effective duration of the putable bond is in this case very similar to that of an otherwise identical option-free bond. In contrast, when interest rates rise, the put option moves into the money and limits the price depreciation because the investor can put the bond and reinvest the proceeds of the retired bond at a higher yield. Thus, the put option reduces the effective duration of the putable bond relative to that of the straight bond.

When the embedded option (call or put) is deep in the money, the effective duration of the bond with an embedded option resembles that of the straight bond maturing on the first exercise date, reflecting the fact that the bond is highly likely to be called or put on that date.

Exhibit 20 compares the effective durations of option-free, callable, and putable bonds. All bonds are 4% annual coupon bonds with a maturity of 10 years. Both the call option and the put option are European-like and exercisable two months from now. The bonds are valued assuming a 4% flat yield curve and an interest rate volatility of 10%.

Exhibit 20 Comparison of the Effective Durations of Option-Free, Callable, and Putable Bonds

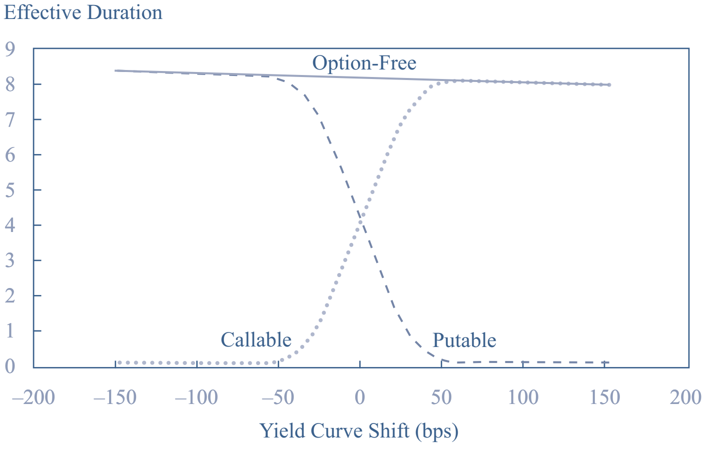


Exhibit 20 shows that the effective duration of an option-free bond changes very little in response to interest rate movements. As expected, when interest rates rise the put option moves into the money, which limits the price depreciation of the putable bond and shortens its effective duration. In contrast, the effective duration of the callable bond shortens when interest rates fall, which is when the call option moves into the money and thus limits the price appreciation of the callable bond.

EFFECTIVE DURATION IN PRACTICE

Effective duration is a concept most practically used in the context of a portfolio. Thus, an understanding of the effective durations of various types of instruments helps manage portfolio duration. In the following table, we show some properties of the effective duration of cash and the common types of bonds:

Type of Bond	Effective Duration
Cash	0
Zero-coupon bond	≈ Maturity
Fixed-rate bond	< Maturity
Callable bond	≤ Duration of straight bond
Putable bond	≤ Duration of straight bond
Floater (MRR flat)	≈ Time (in years) to next reset

In general, a bond's effective duration does not exceed its maturity. There are a few exceptions, however, such as tax-exempt bonds when analyzed on an after-tax basis.

Knowing the effective duration of each type of bond is useful when one needs to change portfolio duration. For example, a portfolio manager who wants to shorten the effective duration of a portfolio of fixed-rate bonds can add floaters. For the debt manager of a company or other issuing entity, another way of shortening effective duration is to issue callable bonds. The topic of changing portfolio duration is covered thoroughly in Level III.

ONE-SIDED AND KEY RATE DURATION

7

k describe the use of one-sided durations and key rate durations to evaluate the interest rate sensitivity of bonds with embedded options

Effective durations are normally calculated by averaging the changes resulting from shifting the benchmark yield curve up and down by the same amount. This calculation works well for option-free bonds, but the results can be misleading in the presence of embedded options. The problem is that when the embedded option is in the money, the price of the bond has limited upside potential if the bond is callable or limited downside potential if the bond is putable. Thus, the price sensitivity of bonds with embedded options is not symmetrical to positive and negative changes in interest rates of the same magnitude.

Consider, for example, a 4.5% bond maturing in five years, which is currently callable at 100. On a 4% flat yield curve at 15% volatility, the value of this callable bond is 99.75. If interest rates declined by 30 bps, the price would rise to 100. In fact, no matter how far interest rates decline, the price of the callable bond cannot exceed 100 because no investor will pay more than the price at which the bond can be immediately called. In contrast, the price decline has no limit if interest rates rise. Thus, the average price response to up- and down-shifts of interest rates (effective duration) is not as informative as the price responses to the up-shift (one-sided up-duration) and the down-shift (one-sided down-duration) of interest rates.

Exhibits 21 and 22 illustrate why **one-sided durations**—that is, the effective durations when interest rates go up or down—are better at capturing the interest rate sensitivity of a callable or putable bond than the (two-sided) effective durations, particularly when the embedded option is near the money.

Exhibit 21 Durations for a 4.5% Annual Coupon Bond Maturing in Five Years and Immediately Callable at Par on a 4% Flat Yield Curve at 15% Interest Rate Volatility

	At a 4% Flat Yield Curve	Interest Rate up by 30 bps	Interest Rate down by 30 bps
Value of the bond	99.75	99.17	100.00
Duration measure	Effective duration 1.39	One-sided up-duration 1.94	One-sided down-duration 0.84

Exhibit 21 shows that a 30 bps increase in the interest rate has a greater effect on the value of the callable bond than a 30 bps decrease in the interest rate. The fact that the one-sided up-duration is higher than the one-sided down-duration confirms that the callable bond is more sensitive to interest rate rises than to interest rate declines.

Exhibit 22 Durations for a 4.1% Annual Coupon Bond Maturing in Five Years and Immediately Putable at Par on a 4% Flat Yield Curve at 15% Interest Rate Volatility

	At a 4% Flat Yield Curve	Interest Rate up by 30 bps	Interest Rate down by 30 bps
Value of the bond	100.45	100.00	101.81
Duration measure	Effective duration 3.00	One-sided up-duration 1.49	One-sided down-duration 4.51

The one-sided durations in Exhibit 22 indicate that the putable bond is more sensitive to interest rate declines than to interest rate rises.

7.1 Key Rate Durations

Effective duration is calculated by assuming parallel shifts in the benchmark yield curve. In reality, however, interest rate movements are not as neat. Many portfolio managers and risk managers like to isolate the price responses to changes in the rates of key maturities on the benchmark yield curve. For example, how would the price of a bond be expected to change if only the two-year benchmark rate moved up by 5 bps? The answer is found by using **key rate durations** (also known as partial durations), which reflect the sensitivity of the bond's price to changes in specific maturities on the benchmark yield curve. Thus, key rate durations help portfolio managers and risk managers identify the "shaping risk" for bonds—that is, the bond's sensitivity to changes in the shape of the yield curve (e.g., steepening and flattening).

The valuation procedure and formula applied in the calculation of key rate durations are identical to those used in the calculation of effective duration, but instead of shifting the entire benchmark yield curve, only key points are shifted one at a time. Thus, the effective duration for each maturity point shift is calculated in isolation.

Exhibits 23, 24, and 25 show the key rate durations for bonds valued at a 4% flat yield curve. Exhibit 23 examines option-free bonds (assuming semi-annual coupons), and Exhibits 24 and 25 extend the analysis to callable and putable bonds, respectively.

Exhibit 23 Key Rate Durations of 10-Year Option-Free Bonds Valued at a 4% Flat Yield Curve

Coupon (%)	Price (% of par)	Key Rate Durations				
		Total	2-Year	3-Year	5-Year	10-Year
0	67.30	9.81	–0.07	–0.34	–0.93	11.15
2	83.65	8.83	–0.03	–0.13	–0.37	9.37
4	100.00	8.18	0.00	0.00	0.00	8.18
6	116.35	7.71	0.02	0.10	0.27	7.32
8	132.70	7.35	0.04	0.17	0.47	6.68
10	149.05	7.07	0.05	0.22	0.62	6.18

As shown in Exhibit 23, for option-free bonds not trading at par (the white rows), shifting any par rate has an effect on the value of the bond, but shifting the maturity-matched (10-year in this example) par rate has the greatest effect. This is simply because the largest cash flow of a fixed-rate bond occurs at maturity with the payment of both the final coupon and the principal.

For an option-free bond trading at par (the shaded row), the maturity-matched par rate is the only rate that affects the bond's value. It is a definitional consequence of "par" rates. If the 10-year par rate on a curve is 4%, then a 10-year 4% bond valued on that curve at zero OAS will be worth par regardless of the par rates of the other maturity points on the curve. In other words, shifting any rate other than the 10-year rate on the par yield curve will not change the value of a 10-year bond trading at par. Shifting a par rate up or down at a particular maturity point, however, respectively increases or decreases the *discount rate* at that maturity point. These facts will be useful to remember in the following paragraph.

As illustrated in Exhibit 23, key rate durations can sometimes be negative for maturity points that are shorter than the maturity of the bond being analyzed if the bond is a zero-coupon bond or has a very low coupon. We can explain why this is the case by using the zero-coupon bond (the first row of Exhibit 23). As discussed in the previous paragraph, if we increase the five-year par rate, the value of a 10-year bond trading at par must remain unchanged because the 10-year par rate has not changed. But the five-year zero-coupon rate has increased because of the increase in the five-year par rate. Thus, the value of the five-year coupon of the 10-year bond trading at par will be lower than before the increase. But because the value of the 10-year bond trading at par must remain par, the remaining cash flows, including the cash flow occurring in Year 10, must be discounted at slightly *lower* rates to compensate. This results in a lower 10-year zero-coupon rate, which makes the value of a 10-year zero-coupon bond (whose only cash flow is in Year 10) *rise* in response to an *upward* change in the five-year par rate. Consequently, the five-year key rate duration for a 10-year zero-coupon bond is negative (–0.93).

Unlike for option-free bonds, the key rate durations of bonds with embedded options depend not only on the *time to maturity* but also on the *time to exercise*. Exhibits 24 and 25 illustrate this phenomenon for 30-year callable and putable bonds. Both the call option and the put option are European-like exercisable 10 years from now, and the bonds are valued assuming a 4% flat yield curve and a volatility of 15%.

Exhibit 24 Key Rate Durations of 30-Year Bonds Callable in 10 Years Valued at a 4% Flat Yield Curve with 15% Interest Rate Volatility

Coupon (%)	Price (% of par)	Key Rate Durations					
		Total	2-Year	3-Year	5-Year	10-Year	30-Year
2	64.99	19.73	–0.02	–0.08	–0.21	–1.97	22.01
4	94.03	13.18	0.00	0.02	0.05	3.57	9.54
6	114.67	9.11	0.02	0.10	0.29	6.00	2.70
8	132.27	7.74	0.04	0.17	0.48	6.40	0.66
10	148.95	7.14	0.05	0.22	0.62	6.06	0.19

The bond with a coupon of 2% (the first row of Exhibit 24) is unlikely to be called, and thus it behaves more like a 30-year option-free bond, whose effective duration depends primarily on movements in the 30-year par rate. Therefore, the rate that has the highest effect on the value of the callable bond is the maturity-matched (30-year) rate. As the bond's coupon increases, however, so does the likelihood of the bond being called. Thus, the bond's total effective duration shortens, and the rate that has the highest effect on the callable bond's value gradually shifts from the 30-year rate to the 10-year rate. At the very high coupon of 10%, because of the virtual certainty of being called, the callable bond behaves like a 10-year option-free bond; the 30-year key rate duration is negligible (0.19) relative to the 10-year key rate duration (6.06).

Exhibit 25 Key Rate Durations of 30-Year Bonds Putable in 10 Years Valued at a 4% Flat Yield Curve with 15% Interest Rate Volatility

Coupon (%)	Price (% of par)	Key Rate Durations					
		Total	2-Year	3-Year	5-Year	10-Year	30-Year
2	83.89	9.24	–0.03	–0.14	–0.38	8.98	0.81
4	105.97	12.44	0.00	–0.01	–0.05	4.53	7.97
6	136.44	14.75	0.01	0.03	0.08	2.27	12.37
8	169.96	14.90	0.01	0.06	0.16	2.12	12.56
10	204.38	14.65	0.02	0.07	0.21	2.39	11.96

If the 30-year bond putable in 10 years has a high coupon, its price is more sensitive to the 30-year rate because it is unlikely to be put and thus behaves like an otherwise identical option-free bond. The 10% putable bond (the last row of Exhibit 25), for example, is most sensitive to changes in the 30-year rate, as illustrated by a 30-year key rate duration of 11.96. At the other extreme, a low-coupon bond is most sensitive to movements in the 10-year rate. It is almost certain to be put and so behaves like an option-free bond maturing on the put date.

8 EFFECTIVE CONVEXITY

l. compare effective convexities of callable, putable, and straight bonds

Duration is an approximation of the expected bond price responses to changes in interest rates because actual changes in bond prices are not linear, particularly for bonds with embedded options. Thus, it is useful to measure **effective convexity**—that is, the sensitivity of duration to changes in interest rates—as well. The formula to calculate a bond's effective convexity is

$$\text{EffCon} = \frac{(PV_{-}) + (PV_{+}) - [2 \times (PV_0)]}{(\Delta\text{Curve})^2 \times (PV_0)}, \quad (4)$$

where

ΔCurve = the magnitude of the parallel shift in the benchmark yield curve (in decimal)

PV_{-} = the full price of the bond when the benchmark yield curve is shifted down by ΔCurve

PV_{+} = the full price of the bond when the benchmark yield curve is shifted up by ΔCurve

PV_0 = the current full price of the bond (i.e., with no shift)

Let us return to the three-year 4.25% bond callable at par one year and two years from now. We still use the same par yield curve (i.e., one-year, two-year, and three-year par yields of 2.500%, 3.000%, and 3.500%, respectively) and the same interest rate volatility (10%) as before, but we now assume that the bond's current full price is 100.785 instead of 101.000. Thus, the implied OAS is 40 bps. Given 30 bps shifts in the benchmark yield curve, the resulting PV_- and PV_+ are 101.381 and 100.146, respectively. Using Equation 4, the effective convexity is:

$$\text{EffCon} = \frac{101.381 + 100.146 - 2 \times 100.785}{(0.003)^2 \times 100.785} = -47.41.$$

[Note that there are two different conventions for reporting convexity in practice; "raw" convexity figures, such as in this example, are sometimes scaled (divided) by 100.]

Exhibit 20, shown earlier, displays effective durations but also illustrates the effective convexities of callable and putable bonds. When interest rates are high and the value of the call option is low, the callable and straight bond experience very similar effects from changes in interest rates. They both have positive convexity. However, the effective convexity of the callable bond turns negative when the call option is near the money, as in the example just presented, which indicates that the upside for a callable bond is much smaller than the downside. The reason is because when interest rates decline, the price of the callable bond is capped by the price of the call option if it is near the exercise date.

Conversely, putable bonds always have positive convexity. When the option is near the money, the upside for a putable bond is much larger than the downside because the price of a putable bond is floored by the price of the put option if it is near the exercise date.

Compared side by side, putable bonds have more upside potential than otherwise identical callable bonds when interest rates decline. Putable bonds also have less downside risk than otherwise identical callable bonds when interest rates rise.

EXAMPLE 7

Interest Rate Sensitivity

Erna Smith, a portfolio manager, has two fixed-rate bonds in her portfolio: a callable bond (Bond X) and a putable bond (Bond Y). She wants to examine the interest rate sensitivity of these two bonds to a parallel shift in the benchmark yield curve. Assuming an interest rate volatility of 10%, her valuation software shows how the prices of these bonds change for 30 bps shifts up or down:

	Bond X	**Bond Y**
Time to maturity	Three years from today	Three years from today
Coupon	3.75% annual	3.75% annual
Type of bond	Callable at par one year from today	Putable at par one year from today
Current price (% of par)	100.594	101.330
Price (% of par) when shifting the benchmark yield curve down by 30 bps	101.194	101.882
Price (% of par) when shifting the benchmark yield curve up by 30 bps	99.860	100.924

1 The effective duration for Bond X is *closest* to:

A 0.67.

B 2.21.

C 4.42.

2 The effective duration for Bond Y is *closest* to:

A 0.48.

B 0.96.

C 1.58.

3 When interest rates rise, the effective duration of:

A Bond X shortens.

B Bond Y shortens.

C the underlying option-free (straight) bond corresponding to Bond X lengthens.

4 When the option embedded in Bond Y is in the money, the one-sided durations *most likely* show that the bond is:

A more sensitive to a decrease in interest rates.

B more sensitive to an increase in interest rates.

C equally sensitive to a decrease or to an increase in interest rates.

5 The price of Bond X is affected:

A only by a shift in the one-year par rate.

B only by a shift in the three-year par rate.

C by all par rate shifts but is most sensitive to shifts in the one-year and three-year par rates.

6 The effective convexity of Bond X:

A cannot be negative.

B turns negative when the embedded option is near the money.

C turns negative when the embedded option moves out of the money.

7 Which of the following statements is *most* accurate?

A Bond Y exhibits negative convexity.

B For a given decline in interest rate, Bond X has less upside potential than Bond Y.

C The underlying option-free (straight) bond corresponding to Bond Y exhibits negative convexity.

Solution to 1:

B is correct. The effective duration for Bond X is

$$\text{EffDur} = \frac{101.194 - 99.860}{2 \times 0.003 \times 100.594} = 2.21.$$

A is incorrect because the duration of a bond with a single cash flow one year from now is approximately one year, so 0.67 is too low—even assuming that the bond will be called in one year with certainty. C is incorrect because 4.42 exceeds the maturity of Bond X (three years).

Solution to 2:

C is correct. The effective duration for Bond Y is

$$\text{EffDur} = \frac{101.882 - 100.924}{2 \times 0.003 \times 101.330} = 1.58.$$

Solution to 3:

B is correct. When interest rates rise, a put option moves into the money and the putable bond is more likely to be put. Thus, it behaves like a shorter-maturity bond, and its effective duration shortens. A is incorrect because when interest rates rise, a call option moves out of the money; so, the callable bond is less likely to be called. C is incorrect because the effective duration of an option-free bond goes down as interest rates rise.

Solution to 4:

A is correct. If interest rates rise, the investor's ability to put the bond at par limits the price depreciation. In contrast, the increase in the bond's price has no limit when interest rates decline. Thus, the price of a putable bond whose embedded option is in the money is more sensitive to a decrease in interest rates.

Solution to 5:

C is correct. The main driver of the call decision is the two-year forward rate one year from now. This rate is most significantly affected by changes in the one-year and three-year par rates.

Solution to 6:

B is correct. The effective convexity of a callable bond turns negative when the call option is near the money because the price response of a callable bond to lower interest rates is capped by the call option. That is, in case of a decline in interest rates, the issuer will call the bonds and refund at lower rates, thus limiting the upside potential for the investor.

Solution to 7:

B is correct. As interest rates decline, the value of a call option increases whereas the value of a put option decreases. The call option embedded in Bond X limits its price appreciation, but Bond Y has no such cap. Thus, Bond X has less upside potential than Bond Y. A is incorrect because a putable bond always has positive convexity; that is, Bond Y has more upside than downside potential. C is incorrect because an option-free bond exhibits low positive convexity.

9 VALUATION AND ANALYSIS OF CAPPED AND FLOORED FLOATING-RATE BONDS

m calculate the value of a capped or floored floating-rate bond

Options in floating-rate bonds (floaters) are exercised automatically depending on the course of interest rates; if the coupon rate rises or falls below the threshold, the cap or floor automatically applies. Similar to callable and putable bonds, capped and floored floaters can be valued by using the arbitrage-free framework.

9.1 Valuation of a Capped Floater

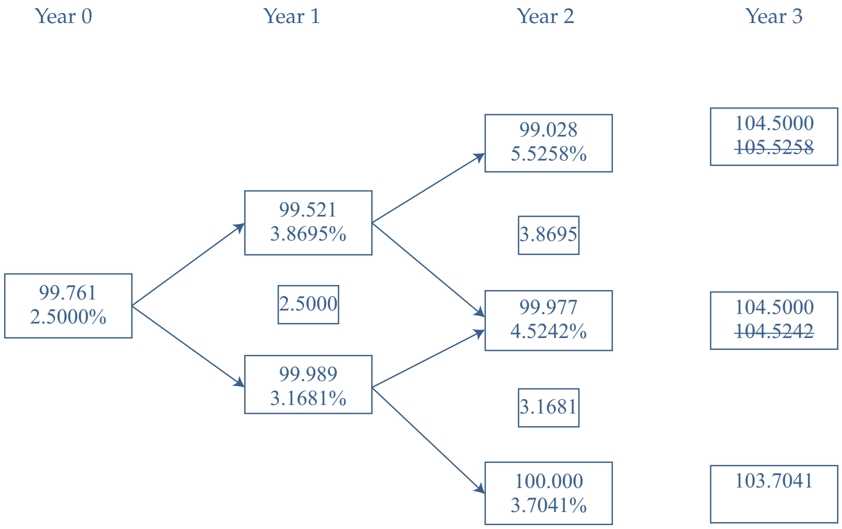
The cap provision in a floater prevents the coupon rate from increasing above a specified maximum rate. As a consequence, a **capped floater** protects the issuer against rising interest rates and is thus an issuer option. Because the investor is long the bond but short the embedded option, the value of the cap decreases the value of the capped floater relative to the value of the straight bond:

$$\text{Value of capped floater} = \text{Value of straight bond} - \text{Value of embedded cap}. \quad (5)$$

To illustrate how to value a capped floater, consider a floating-rate bond that has a three-year maturity. The floater's coupon pays the one-year reference rate annually, set in arrears, and is capped at 4.500%. The term "set in arrears" means that the coupon rate is set at the *end* of the coupon period; the payment date and the setting date are one and the same. For simplicity, we assume that the issuer's credit quality closely matches the reference rate swap curve (i.e., there is no credit spread) and that the reference rate swap curve is the same as the par yield curve given in Exhibit 1 (i.e., one-year, two-year, and three-year par yields of 2.500%, 3.000%, and 3.500%, respectively). We also assume that the interest rate volatility is 10%.

The valuation of the capped floater is depicted in Exhibit 26.

Exhibit 26 Valuation of a Three-Year Reference Rate Floater Capped at 4.500% at 10% Interest Rate Volatility



Without a cap, the value of this floater would be 100 because in every scenario, the coupon paid would be equal to the discount rate. But because the coupon rate is capped at 4.500%, which is lower than the highest interest rates in the tree, the value of the capped floater will be lower than the value of the straight bond.

For each scenario, we check whether the cap applies; if it does, the cash flow is adjusted accordingly. For example, at the top of the tree at Year 2, the reference rate (5.5258%) is higher than the 4.500% cap. Thus, the coupon payment at Year 3 is capped at the 4.500 maximum amount, and the cash flow is adjusted downward from the uncapped amount (105.5258) to the capped amount (104.5000). The coupon is also capped when the reference rate is 4.5242% at Year 2.

As expected, the value of the capped floater is lower than 100 (99.761). The value of the cap can be calculated by using Equation 5:

Value of embedded cap = 100 – 99.761 = 0.239.

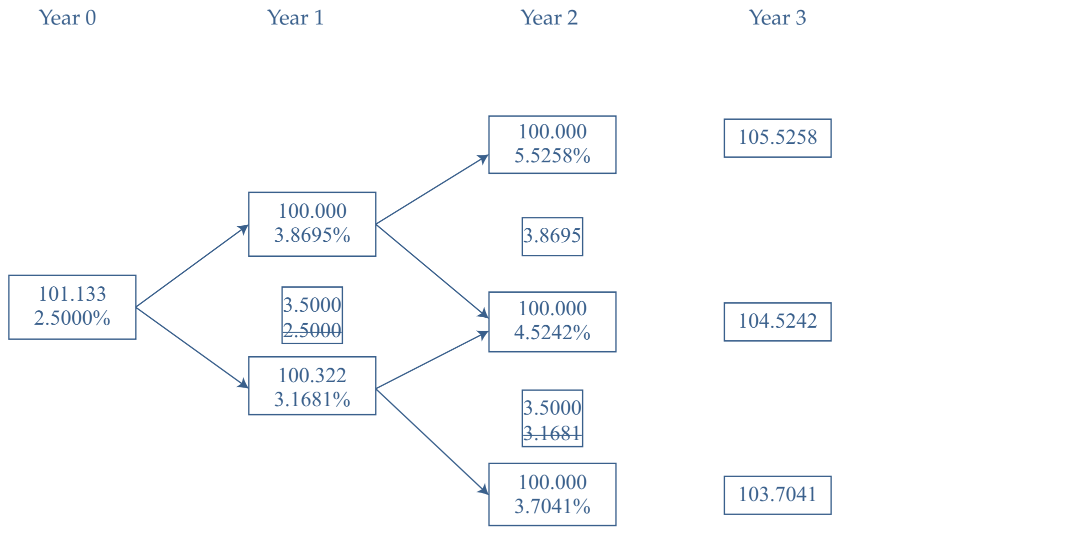
9.2 Valuation of a Floored Floater

The floor provision in a floater prevents the coupon rate from decreasing below a specified minimum rate. As a consequence, a **floored floater** protects the investor against declining interest rates and is thus an investor option. Because the investor is long both the bond and the embedded option, the value of the floor increases the value of the floored floater relative to the value of the straight bond:

Value of floored floater
= Value of straight bond + Value of embedded floor. **(6)**

To illustrate how to value a floored floater, we return to the example we used for the capped floater but assume that the embedded option is now a 3.500% floor instead of a 4.500% cap. The other assumptions remain the same. The valuation of the floored floater is depicted in Exhibit 27.

Exhibit 27 Valuation of a Three-Year Reference Rate Floater Floored at 3.500% at 10% Interest Rate Volatility



Recall from the discussion about the capped floater that if there were no cap, the value of the floater would be 100 because the coupon paid would equal the discount rate. The same principle applies here: If there were no floor, the value of this floater would be 100. Because the presence of the floor potentially increases the cash flows, however, the value of the floored floater must be equal to or higher than the value of the straight bond.

Exhibit 27 shows that the floor is binding at Year 0 because the reference rate (2.5000%) is less than the cap rate (3.5000%) and at Year 1 at the lower node where the reference rate is 3.1681%. Thus, the corresponding interest payments at Year 1 and 2 are increased to the minimum amount of 3.5000. As a consequence, the value of the floored floater exceeds 100 (101.133). The value of the floor can be calculated by using Equation 6:

Value of embedded floor = 101.133 – 100 = 1.133.

EXAMPLE 8

Valuation of Capped and Floored Floaters

1 A three-year floating rate bond pays annual coupons of one-year reference rate (set in arrears) and is capped at 5.600%. The reference rate swap curve is as given in Exhibit 1 (i.e., the one-year, two-year, and three-year par yields are 2.500%, 3.000%, and 3.500%, respectively), and interest rate volatility is 10%. The value of the capped floater is *closest to*:

A 100.000.

B 105.600.

C 105.921.

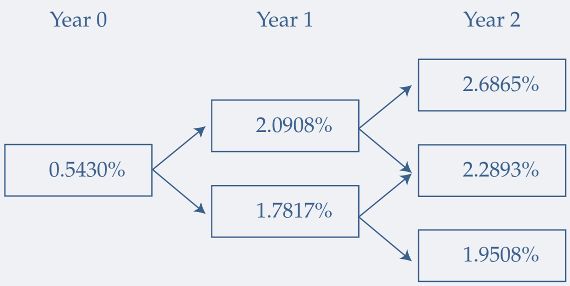
2 A three-year floating-rate bond pays annual coupons of one-year reference rate (set in arrears) and is floored at 3.000%. The reference swap curve is as given in Exhibit 1 (i.e., the one-year, two-year, and three-year par yields are 2.500%, 3.000%, and 3.500%, respectively), and interest rate volatility is 10%. The value of the floored floater is *closest to*:

A 100.000.

B 100.488.

C 103.000.

3 An issuer in the eurozone wants to sell a three-year floating-rate note at par with an annual coupon based on the 12-month Euribor + 300 bps. Because the 12-month Euribor is currently at a historic low and the issuer wants to protect itself against a sudden increase in interest cost, the issuer's advisers recommend increasing the credit spread to 320 bps and capping the coupon at 5.50%. Assuming an interest rate volatility of 8%, the advisers have constructed the following binomial interest rate tree:



The value of the capped floater is *closest to*:

A 92.929.

B 99.916.

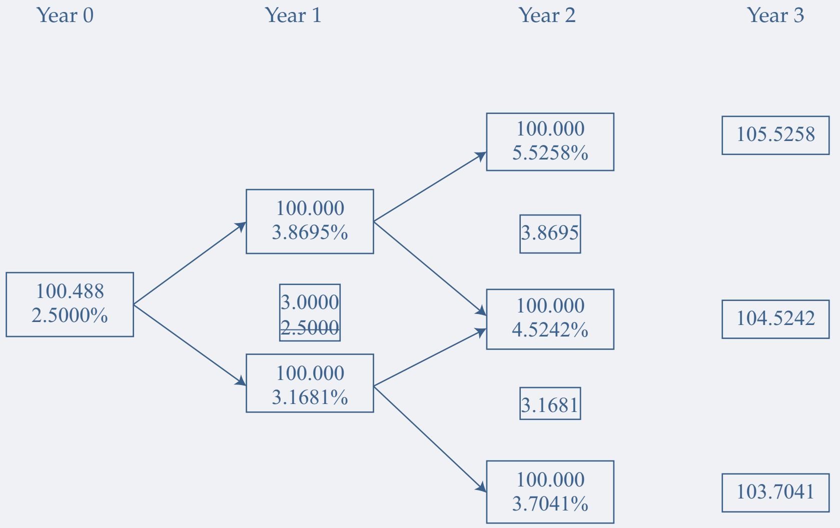
C 109.265.

Solution to 1:

A is correct. As illustrated in Exhibit 26, the cap is higher than any of the rates at which the floater is reset on the interest rate tree. Thus, the value of the bond is the same as if it had no cap—that is, 100.

Solution to 2:

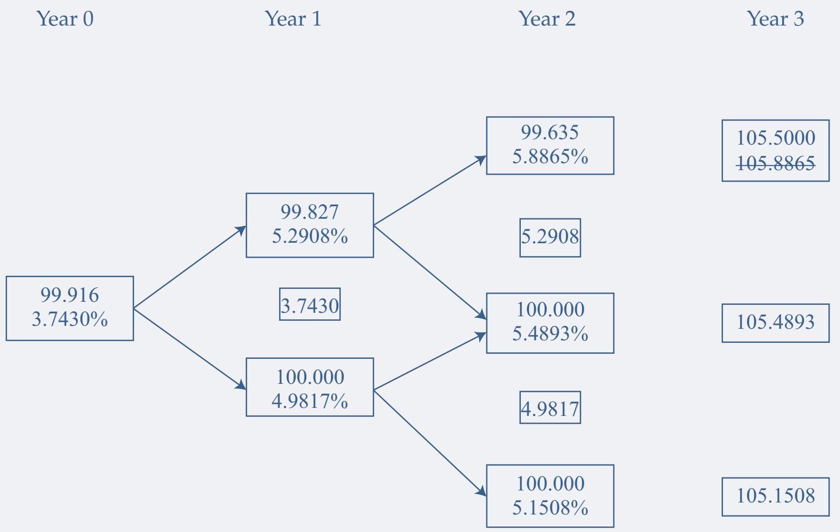
B is correct. One can eliminate C because as illustrated in Exhibit 27, all else being equal, the bond with a higher floor (3.500%) has a value of 101.133. The value of a bond with a floor of 3.000% cannot be higher. Intuitively, B is the likely correct answer because the straight bond is worth 100. However, it is still necessary to calculate the value of the floored floater because if the floor is low enough, it could be worthless.



Here, it turns out that the floor adds 0.488 in value to the straight bond. Had the floor been 2.500%, the floored floater and the straight bond would both be worth par.

Solution to 3:

B is correct.



VALUATION AND ANALYSIS OF CONVERTIBLE BONDS: DEFINING FEATURES AND ANALYSIS OF A CONVERTIBLE BOND

10

- **n** describe defining features of a convertible bond
- **o** calculate and interpret the components of a convertible bond's value

So far, we have discussed bonds for which the exercise of the option is at the discretion of the issuer (callable bond), at the discretion of the bondholder (putable bond), or set through a pre-defined contractual arrangement (capped and floored floaters). What distinguishes a convertible bond from the bonds discussed earlier is that exercising the option results in the change of the security from a bond to a common stock. This section describes defining features of convertible bonds and discusses how to analyze and value these bonds.

10.1 Defining Features of a Convertible Bond

A **convertible bond** presents the characteristics of an option-free bond and an embedded conversion option, which gives bondholders the right to convert their debt into equity during the **conversion period** at a pre-determined **conversion price**.

Investors usually accept a lower coupon for convertible bonds than for otherwise identical non-convertible bonds because they can participate in the potential upside through the conversion mechanism that allows the bondholders to convert their bonds into shares at a cost lower than market value. The issuer benefits from paying a lower coupon. In case of conversion, an added benefit for the issuer is that it no longer has to repay the debt that was converted into equity.

However, what might appear as a win–win situation for both the issuer and the investors is not a "free lunch" because the issuer's existing shareholders face dilution in case of conversion. In addition, if the underlying share price remains below the conversion price and the bond is not converted, the issuer must repay the debt or refinance it, potentially at a higher cost. If conversion is not achieved, the bondholders will have lost interest income relative to an otherwise identical non-convertible bond that would have been issued with a higher coupon and would have thus offered investors an additional spread.

We will use the information provided in Exhibit 28 to describe the features of a convertible bond and then illustrate how to analyze it. Exhibit 28 is based on a \$1 billion convertible bond issued in June 2018 by Twitter, Inc. (TWTR), a company listed on the New York Stock Exchange. Some features of the actual convertible bond, such as the presence of a make-whole call option, have been dropped for simplicity.

Exhibit 28 Twitter, Inc., \$1 billion, 0.25% Convertible Bonds Due 15 June 2024

- **Issue Date:** 11 June 2018
- **Ranking:** Senior unsecured
- **Interest:** 0.25% per year. Interest will accrue from 11 June 2018 and will be payable semiannually in arrears on 15 June and 15 December of each year, beginning on 15 December 2018.
- **Issue Price:** 100% of par value
- **Maturity:** 15 June 2024
- **Conversion Rate:** Each bond of par value of \$1,000 is convertible to 17.5 shares of common stock.
- **Conversion Price:** \$57.14 per share
- **Share Price at Issuance:** \$40.10
- (Assumed) **Share Price on 15 June 2019:** \$35.14

Exhibit 28 (Continued)

- (Assumed) **Convertible Bond Price on 15 June 2019:** 95.225% of par value
- **Conversion Premium:** 42.5%

The applicable share price at which the investor can convert the bonds into ordinary (common) shares is called the conversion price. In the Twitter example provided in Exhibit 28, the conversion price is $57.14 per share.

The number of shares of common stock that the bondholder receives from converting the bonds into shares is called the **conversion rate (or ratio)**. In the Twitter example, bondholders who hold $10,000 in par value can convert their bonds into shares and receive 175 shares ($10,000/$57.14). The conversion rate is 17.5 per $1,000 in par value. The conversion may be exercised during a particular period or at set intervals during the life of the bond.

The conversion price in Exhibit 28 is referred to as the *initial* conversion price because it reflects the conversion price *at issuance*. Corporate actions—such as stock splits, bonus share issuances, and rights or warrants issuances—affect a company's share price and may reduce the benefit of conversion for the convertible bondholders. Thus, the terms of issuance of the convertible bond contain detailed information defining how the conversion price and conversion ratio are adjusted should such a corporate action occur during the life of the bond. For example, suppose that Twitter performs a 2:1 stock split to its common shareholders. In this case, the conversion price would be adjusted to $28.57 (i.e., $57.14/2) per share and the conversion rate adjusted to 35 (i.e., 17.5 × 2) shares per $1,000 of nominal value.

As long as the convertible bond is still outstanding and has not been converted, the bondholders receive interest payments (semiannually in the Twitter example). Meanwhile, if the issuer declares and pays dividends, common shareholders receive dividend payments. The terms of issuance may offer no compensation to convertible bondholders for dividends paid out during the life of the bond at one extreme, or they may offer full protection by adjusting the conversion price downward for any dividend payments at the other extreme. Typically, a threshold dividend is defined in the terms of issuance. Annual dividend payments below the threshold dividend have no effect on the conversion price. In contrast, the conversion price is adjusted downward for annual dividend payments above the threshold dividend to offer compensation to convertible bondholders.

Should the issuer be acquired by or merged with another company during the life of the bond, bondholders might no longer be willing to continue lending to the new entity. Change-of-control events are defined in the prospectus or offering circular, and if such an event occurs, convertible bondholders usually have the choice between

- a put option that can be exercised during a specified period following the change-of-control event and that provides full redemption of the nominal value of the bond; or
- an adjusted conversion price that is lower than the initial conversion price. This downward adjustment gives the convertible bondholders the opportunity to convert their bonds into shares earlier and at more advantageous terms—thus allowing them to participate in the announced merger or acquisition as common shareholders.

In addition to a put option in case of a change-of-control event, it is not unusual for a convertible bond to include a put option that convertible bondholders can exercise during specified periods. Put options can be classified as "hard" puts or "soft" puts. In the case of a hard put, the issuer must redeem the convertible bond for cash. In the case of a soft put, the investor has the right to exercise the put but the issuer chooses how the payment will be made. The issuer may redeem the convertible bond for cash, common stock, subordinated notes, or a combination of the three.

It is more frequent for convertible bonds to include a call option that gives the issuer the right to call the bond during a specified period and at specified times. As discussed earlier, the issuer may exercise the call option and redeem the bond early if interest rates are falling or if its credit rating is revised upward—thus enabling the issuance of debt at a lower cost. The issuer may also believe that its share price will increase significantly in the future because of its performance or because of events that will take place in the economy or in its sector. In this case, the issuer may try to maximize the benefit to its existing shareholders relative to convertible bondholders and call the bond. To offer convertible bondholders protection against early repayment, convertible bonds usually have a protection period. Subsequently, they can be called but at a premium, which decreases as the maturity of the bond approaches.

If a convertible bond is callable, the issuer has an incentive to call the bond when the underlying share price increases above the conversion price in order to avoid paying further coupons. Such an event is called **forced conversion** because it forces bondholders to convert their bonds into shares. Otherwise, the redemption value that bondholders would receive from the issuer calling the bond would result in a disadvantageous position and a loss compared with conversion. Even if interest rates have not fallen or the issuer's credit rating has not improved, thus not allowing refinancing at a lower cost, the issuer might still proceed with calling the bond when the underlying share price exceeds the conversion price. Doing so allows the issuer to take advantage of the favorable equity market conditions and force the bondholders to convert their bonds into shares. The forced conversion strengthens the issuer's capital structure and eliminates the risk that a subsequent correction in equity prices prevents conversion and requires redeeming the convertible bonds at maturity.

10.2 Analysis of a Convertible Bond

A number of investment metrics and ratios help analyze and value a convertible bond.

10.2.1 *Conversion Value*

The **conversion value**, or parity value, of a convertible bond indicates the value of the bond if it is converted at the market price of the shares.

Conversion value = Underlying share price × Conversion ratio.

Based on the information provided in Exhibit 28, we can calculate the conversion value for Twitter's convertible bonds at the issuance date and on 15 June 2019 (*Note:* The assumed prices actually pertain to 11 April 2019 to simplify the calculation of the straight bond values as there are then five full years to maturity):

Conversion value at the issuance date = \$40.10 × 17.5 = \$701.75.

Conversion value on 15 June 2019 = \$35.14 × 17.5 = \$614.95.

10.2.2 *Minimum Value of a Convertible Bond*

The minimum value of a convertible bond is equal to the greater of

- the conversion value and
- the value of the underlying option-free bond. Theoretically, the value of the straight bond (straight value) can be estimated by using the market value of a non-convertible bond of the issuer with the same characteristics as the

convertible bond but without the conversion option. In practice, such a bond rarely exists. Thus, the straight value is found by using the arbitrage-free framework and by discounting the bond's future cash flows at the appropriate rates.

The minimum value of a convertible bond can also be described as a floor value. It is a *moving* floor, however, because the straight value is not fixed; it changes with fluctuations in interest rates and credit spreads. If interest rates rise, the value of the straight bond falls, making the floor fall. Similarly, if the issuer's credit spread increases—as a result, for example, of a downgrade of its credit rating from investment grade to non-investment grade—the floor value will fall too.

Using the conversion values calculated earlier, the minimum value of Twitter's convertible bonds at the issuance date is

$$\begin{aligned}\text{Minimum value at the issuance date} &= \text{Maximum } (\$701.75; \$1{,}000)\\ &= \$1{,}000.\end{aligned}$$

The straight value at the issuance date is $1,000 because the issue price is set at 100% of par. But after this date, this value will fluctuate. Thus, to calculate the minimum value of Twitter's convertible bond on 15 June 2019, it is first necessary to calculate the value of the straight bond that day using the arbitrage-free framework. From Exhibit 28, the coupon is 0.25%, paid semiannually. Assuming a 2.5% flat yield curve, the straight value on 15 June 2019 when five years remain until maturity is $894.86 per $1,000 in par value:

$$\frac{\$1.25}{\left(1+\dfrac{0.025}{2}\right)^1}+\frac{\$1.25}{\left(1+\dfrac{0.025}{2}\right)^2}+\ldots+\frac{\$1{,}001.25}{\left(1+\dfrac{0.025}{2}\right)^{10}}=\$894.86.$$

It follows that the minimum value of Twitter's convertible bonds on 15 June 2019 is:

$$\text{Minimum value} = \text{Maximum } (\$614.95; \$894.86) = \$894.86.$$

If the value of the convertible bond were lower than the greater of the conversion value and the straight value, an arbitrage opportunity would ensue. Two scenarios help illustrate this concept. Returning to the Twitter example, suppose that the convertible bond is selling for $850.00 on 15 June 2019—that is, at a price that is lower than the straight value of $894.86. In this scenario, the convertible bond is cheap relative to the straight bond; put another way, the convertible bond offers a higher yield than an otherwise identical non-convertible bond. Thus, investors will find the convertible bond attractive, buy it, and push its price up until the convertible bond price returns to the straight value and the arbitrage opportunity disappears.

Alternatively, assume that on 15 June 2019 the yield on otherwise identical non-convertible bonds is 12.00% instead of 2.50%. Using the arbitrage-free framework, the straight value is $567.59 per $1,000 in par value. Suppose that the convertible bond is selling at this straight value—that is, at a price that is lower than its conversion value of $614.95. In this case, an arbitrageur can buy the convertible bond for $567.59, convert it into 17.5 shares, and sell the shares at $35.14 each or $614.95 in total. The arbitrageur makes a profit equal to the difference between the conversion value and the straight value—that is, $47.36 ($614.95 – $567.59). As more arbitrageurs follow the same strategy, the convertible bond price will increase until it reaches the conversion value and the arbitrage opportunity disappears.

10.2.3 *Market Conversion Price, Market Conversion Premium per Share, and Market Conversion Premium Ratio*

Many investors do not buy a convertible bond at issuance on the primary market but instead buy such a bond later in its life on the secondary market. The **market conversion premium per share** allows investors to identify the premium or discount payable when buying the convertible bond rather than the underlying common stock:

Market conversion premium per share
= Market conversion price – Underlying share price,

where

$$\text{Market conversion price} = \frac{\text{Convertible bond price}}{\text{Conversion ratio}}.$$

The market conversion price represents the price that investors effectively pay for the underlying common stock if they buy the convertible bond and then convert it into shares. It can be viewed as a break-even price. Once the underlying share price exceeds the market conversion price, any further rise in the underlying share price is certain to increase the value of the convertible bond by at least the same percentage (we will discuss why at a later stage).

Based on the information provided in Exhibit 28,

$$\text{Market conversion price on 15 June 2019} = \frac{\$952.25}{17.5} = \$54.40$$

and

Market conversion premium per share on 15 June 2019

= \$57.14 – \$54.40

= \$2.74.

The **market conversion premium ratio** expresses the premium, or discount, investors have to pay as a percentage of the current market price of the shares:

$$\text{Market conversion premium ratio} = \frac{\text{Market conversion premium per share}}{\text{Underlying share price}}.$$

In the Twitter example,

$$\text{Market conversion premium ratio on 15 June 2019} = \frac{\$2.74}{\$35.14}$$

= 7.80%.

Why would investors be willing to pay a premium to buy the convertible bond? Recall that the straight value acts as a floor for the convertible bond price. Thus, as the underlying share price falls, the convertible bond price will not fall below the straight value. Viewed in this context, the market conversion premium per share resembles the price of a call option. Investors who buy a call option limit their downside risk to the price of the call option (premium). Similarly, the premium paid when buying a convertible bond allows investors to limit their downside risk to the straight value. There is a fundamental difference, however, between the buyers of a call option and the buyers of a convertible bond. The former know exactly the amount of the downside risk, whereas the latter know only that the most they can lose is the difference between the convertible bond price and the straight value because the straight value is not fixed.

Market conversion discounts per share are rare, but they can theoretically happen given that the convertible bond and the underlying common stock trade in different markets with different types of market participants. For example, highly volatile share prices may result in the market conversion price being lower than the underlying share price.

10.2.4 *Downside Risk with a Convertible Bond*

Many investors use the straight value as a measure of the downside risk of a convertible bond and calculate the following metric:

$$\text{Premium over straight value} = \frac{\text{Convertible bond price}}{\text{Straight value}} - 1.$$

All else being equal, the higher the premium over straight value, the less attractive the convertible bond. In the Twitter example,

$$\text{Premium over straight value} = \frac{\$952.25}{\$894.86}$$

=6.41%

Despite its use in practice, the premium over straight value is a flawed measure of downside risk because, as mentioned earlier, the straight value is not fixed but rather fluctuates with changes in interest rates and credit spreads.

10.2.5 *Upside Potential of a Convertible Bond*

The upside potential of a convertible bond depends primarily on the prospects of the underlying common stock. Thus, convertible bond investors should be familiar with the techniques used to value and analyze common stocks. These techniques are covered elsewhere.

11 VALUATION OF A CONVERTIBLE BOND AND COMPARISON OF RISK–RETURN CHARACTERISTICS

p describe how a convertible bond is valued in an arbitrage-free framework

q compare the risk–return characteristics of a convertible bond with the risk–return characteristics of a straight bond and of the underlying common stock

Historically, the valuation of convertible bonds has been challenging because these securities combine characteristics of bonds, stocks, and options—thus requiring an understanding of what affects the value of fixed income, equity, and derivatives. The complexity of convertible bonds has also increased over time as a result of market innovations and additions to the terms and conditions of these securities. For example, there are now contingent convertible bonds and convertible contingent convertible bonds, which are even more complex to value and analyze.

CONTINGENT CONVERTIBLES

Contingent convertible bonds, or "CoCos," pay a higher coupon than otherwise identical non-convertible bonds; however, they usually are deeply subordinated and may be converted into equity or face principal write-downs if regulatory capital ratios are breached. Convertible contingent convertible bonds, or "CoCoCos," combine a traditional convertible bond and a CoCo. They are convertible at the discretion of the investor, thus offering upside potential if the share price increases. They are also converted into equity or face principal write-downs in the event of a regulatory capital breach. CoCos and CoCoCos are usually issued by financial institutions, particularly in Europe.

The fact that many bond's prospectuses or offering circulars frequently provide for an independent financial valuer to determine the conversion price (and, in essence, the value of the convertible bond) under different scenarios is evidence of the complexity associated with valuing convertible bonds. Because of this complexity, convertible bonds in many markets come with selling restrictions. They are typically offered in very high denominations and only to professional or institutional investors. Regulators perceive them as securities that are too risky for retail investors to invest in directly.

As with any fixed-income instrument, convertible bond investors should perform a diligent risk–reward analysis of the issuer, including its ability to service the debt and repay the principal, as well as a review of the bond's terms of issuance (e.g., collateral, credit enhancements, covenants, and contingent provisions). In addition, convertible bond investors must analyze the factors that typically affect bond prices, such as interest rate movements. Because most convertible bonds have lighter covenants than otherwise similar non-convertible bonds and are frequently issued as subordinated securities, the valuation and analysis of some convertible bonds can be complex.

The investment characteristics of a convertible bond depend on the underlying share price, so convertible bond investors must also analyze factors that may affect the issuer's common stock, including dividend payments and the issuer's actions (e.g., acquisitions or disposals, rights issues). Even if the issuer is performing well, adverse market conditions might depress share prices and prevent conversion. Thus, convertible bond investors must also identify and analyze the exogenous reasons that might ultimately have a negative effect on convertible bonds.

Academics and practitioners have developed advanced models to value convertible bonds, but the most commonly used model remains the arbitrage-free framework. A traditional convertible bond can be viewed as a straight bond and a call option on the issuer's common stock, so

Value of convertible bond = Value of straight bond
+ Value of call option on the issuer's stock.

Many convertible bonds include a call option that gives the issuer the right to call the bond during a specified period and at specified times. The value of such bonds is

Value of callable convertible bond = Value of straight bond + Value of call option on the issuer's stock – Value of issuer call option.

Suppose that the callable convertible bond also includes a put option that gives the bondholder the right to require that the issuer repurchase the bond. The value of such a bond is

Value of callable putable convertible bond = Value of straight bond + Value of call option on the issuer's stock – Value of issuer call option + Value of investor put option.

No matter how many options are embedded into a bond, the valuation procedure remains the same. It relies on generating a tree of interest rates based on the given yield curve and interest rate volatility assumptions, determining at each node of the tree whether the embedded options will be exercised, and then applying the backward induction valuation methodology to calculate the present value of the bond.

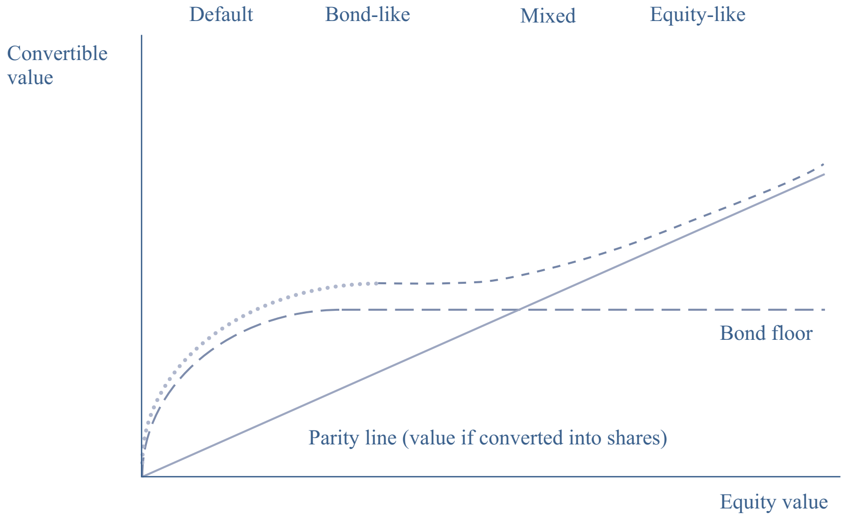
11.1 Comparison of the Risk–Return Characteristics of a Convertible Bond, the Straight Bond, and the Underlying Common Stock

In its simplest form, a convertible bond can be viewed as a straight bond and a call option on the issuer's common stock. When the underlying share price is well below the conversion price, the convertible bond is described as "busted convertible" and exhibits mostly bond risk–return characteristics. That is, the risk–return characteristics of the convertible bond resemble those of the underlying option-free (straight) bond. In this case, the call option is out of the money, so share price movements do not significantly affect the price of the call option and, thus, the price of the convertible bond. Consequently, the price movement of the convertible bond closely follows that of the straight bond, and such factors as interest rate movements and credit spreads significantly affect the convertible bond price. As the share price approaches zero, the value of the bond will fall to approach the present value of the recovery rate in bankruptcy. The convertible bond exhibits even stronger bond risk–return characteristics when the call option is out of the money and the conversion period is approaching its end because the time value component of the option decreases toward zero, making it highly likely that the conversion option will expire worthless. This scenario is shown in Exhibit 29 on the left.

In contrast, when the underlying share price is above the conversion price, a convertible bond exhibits mostly stock risk–return characteristics (see the right-hand side of Exhibit 29). That is, the risk–return characteristics of the convertible bond resemble those of the underlying common stock. In this case, the call option is in the money, so the price of the call option—and thus the price of the convertible bond—is significantly affected by share price movements but mostly unaffected by factors driving the value of an otherwise identical option-free bond, such as interest rate movements. When the call option is in the money, it is more likely to be exercised by the bondholder and the value of the shares resulting from the conversion is higher than the redemption value of the bond. Such convertible bonds trade at prices that closely follow the conversion value of the convertible bond, and their price exhibits similar movements to that of the underlying stock.

In between the bond and the stock extremes, the call option component increases in value as the underlying share price approaches the conversion price. The return on the convertible bond during such periods increases significantly but at a lower rate than the increase in the underlying share price because the conversion price has not yet been reached. When the share price exceeds the conversion price and goes higher, the change in the convertible bond price converges toward the change in the underlying share price. This is why we noted earlier that when the underlying share price exceeds the market conversion price, any further rise in the underlying share price is certain to increase the value of the convertible bond by at least the same percentage.

Exhibit 29 Price Behavior of a Convertible Bond and the Underlying Common Stock



Why would an investor not exercise the conversion option when the underlying share price is above the conversion price? The call option on the issuer's common stock may be a European-style option that cannot be exercised now but only at the end of a pre-determined period. Even if the call option is an American-style option, making it possible to convert the bond into equity, it may not be optimal for the convertible bondholder to exercise prior to the expiry of the conversion period. As discussed earlier, it is sometimes better to wait than to exercise an option that is in the money. The investor may also prefer to sell the convertible bond instead of exercising the conversion option.

Except for busted convertibles, the most important factor in the valuation of convertible bonds is the underlying share price. However, it is worth mentioning that large movements in interest rates or in credit spreads may significantly affect the value of convertible bonds. For a convertible bond with a fixed coupon, all else being equal, a significant fall in interest rates would result in an increase in its value and price, whereas a significant rise in interest rates would lead in a decrease in its value and price. Similarly, all else being equal, a significant improvement in the issuer's credit quality would result in an increase in the value and price of its convertible bonds, whereas a deterioration of the issuer's credit quality would lead to a decrease in the value and price of its convertible bonds.

EXAMPLE 9

Valuation of Convertible Bonds

Nick Andrews, a fixed-income investment analyst, has been asked by his supervisor to prepare an analysis of the convertible bond issued by Heavy Element Inc., a chemical industry company, for presentation to the investment committee. Andrews has gathered the following data from the convertible bond's prospectus and market information:

Issuer: Heavy Element Inc.

Issue Date: 15 September 2020

Maturity Date: 15 September 2025

Interest: 3.75% payable annually

Issue Size: $100,000,000

Issue Price: $1,000 at par

Conversion Ratio: 23.26

Convertible Bond Price on 16 September 2022: $1,230

Share Price on 16 September 2022: $52

The conversion price is *closest to*:

A $19.

B $43.

C $53.

The conversion value on 16 September 2022 is *closest to*:

A $24.

B $230.

C $1,209.

The market conversion premium per share on 16 September 2022 is *closest to*:

A $0.88.

B $2.24.

C $9.00.

The risk–return characteristics of the convertible bond on 16 September 2022 *most likely* resemble that of:

A a busted convertible.

B Heavy Element's common stock.

C a bond of Heavy Element that is identical to the convertible bond but without the conversion option.

As a result of favorable economic conditions, credit spreads for the chemical industry narrow, resulting in lower interest rates for the debt of such companies as Heavy Element. All else being equal, the price of Heavy Element's convertible bond will *most likely*:

A decrease significantly.

B not change significantly.

C increase significantly.

Suppose that on 16 September 2022 the convertible bond is available in the secondary market at a price of $1,050. An arbitrageur can make a risk-free profit by:

A buying the underlying common stock and shorting the convertible bond.

B buying the convertible bond, exercising the conversion option, and selling the shares resulting from the conversion.

C shorting the convertible bond and buying a call option on the underlying common stock exercisable at the conversion price on the conversion date.

A few months have passed. Because of chemical spills in lake water at the site of a competing facility, the government has introduced very costly environmental legislation. As a result, share prices of almost all publicly traded chemical companies, including Heavy Element, have decreased sharply. Heavy Element's share price is now $28. Now, the risk–return characteristics of the convertible bond *most likely* resemble that of:

A a bond.

B a hybrid instrument.

C Heavy Element's common stock.

Solution to 1:

B is correct. The conversion price is equal to the par value of the convertible bond divided by the conversion ratio—that is, $1,000/23.26 = $43 per share.

Solution to 2:

C is correct. The conversion value is equal to the underlying share price multiplied by the conversion ratio—that is, $52 × 23.26 = $1,209.

Solution to 3:

A is correct. The market conversion premium per share is equal to the convertible bond price divided by the conversion ratio, minus the underlying share price—that is, ($1,230/23.26) – $52 = $52.88 – $52 = $0.88.

Solution to 4:

B is correct. The underlying share price ($52) is well above the conversion price ($43). Thus, the convertible bond exhibits risk–return characteristics that are similar to those of the underlying common stock. A is incorrect because a busted convertible is a convertible bond for which the underlying common stock trades at a significant discount relative to the conversion price. C is incorrect because it describes a busted convertible.

Solution to 5:

B is correct. The underlying share price ($52) is well above the conversion price ($43). Thus, the convertible bond exhibits mostly stock risk–return characteristics, and its price is mainly driven by the underlying share price. Consequently, the decrease in credit spreads will have little effect on the convertible bond price.

Solution to 6:

B is correct. The convertible bond price ($1,050) is lower than its minimum value ($1,209). Thus, the arbitrageur can buy the convertible bond for $1,050; convert it into 23.26 shares; and sell the shares at $52 each, or $1,209 in total, making a profit of $159. A and C are incorrect because in both scenarios, the arbitrageur is short the underpriced asset (convertible bond) and long an overpriced asset, resulting in a loss.

Solution to 7:

A is correct. The underlying share price ($28) is now well below the conversion price ($43), so the convertible bond is a busted convertible and exhibits mostly bond risk–return characteristics. B is incorrect because the underlying share price would have to be close to the conversion price for the risk–return characteristics of the convertible bond to resemble that of a hybrid instrument. C is incorrect because the underlying share price would have to be in excess of the conversion price for the risk–return characteristics of the convertible bond to resemble that of the company's common stock.

SUMMARY

- An embedded option represents a right that can be exercised by the issuer, by the bondholder, or automatically depending on the course of interest rates. It is attached to, or embedded in, an underlying option-free bond called a straight bond.
- Simple embedded option structures include call options, put options, and extension options. Callable and putable bonds can be redeemed prior to maturity, at the discretion of the issuer in the former case and of the bondholder in the latter case. An extendible bond gives the bondholder the right to keep the bond for a number of years after maturity. Putable and extendible bonds are equivalent, except that their underlying option-free bonds are different.
- Complex embedded option structures include bonds with other types of options or combinations of options. For example, a convertible bond includes a conversion option that allows the bondholders to convert their bonds into the issuer's common stock. A bond with an estate put can be put by the heirs of a deceased bondholder. Sinking fund bonds make the issuer set aside funds over time to retire the bond issue and are often callable, may have an acceleration provision, and may also contain a delivery option. Valuing and analyzing bonds with complex embedded option structures is challenging.
- According to the arbitrage-free framework, the value of a bond with an embedded option is equal to the arbitrage-free values of its parts—that is, the arbitrage-free value of the straight bond and the arbitrage-free values of each of the embedded options.
- Because the call option is an issuer option, the value of the call option decreases the value of the callable bond relative to an otherwise identical but non-callable bond. In contrast, because the put option is an investor option, the value of the put option increases the value of the putable bond relative to an otherwise identical but non-putable bond.
- In the absence of default and interest rate volatility, the bond's future cash flows are certain. Thus, the value of a callable or putable bond can be calculated by discounting the bond's future cash flows at the appropriate one-period forward rates, taking into consideration the decision to exercise the option. If a bond is callable, the decision to exercise the option is made by the issuer, which will exercise the call option when the value of the bond's future cash flows is higher than the call price. In contrast, if the bond is putable, the decision to exercise the option is made by the bondholder, who will exercise the put option when the value of the bond's future cash flows is lower than the put price.
- In practice, interest rates fluctuate and interest rate volatility affects the value of embedded options. Thus, when valuing bonds with embedded options, it is important to consider the possible evolution of the yield curve over time.
- Interest rate volatility is modeled using a binomial interest rate tree. The higher the volatility, the lower the value of the callable bond and the higher the value of the putable bond.
- Valuing a bond with embedded options assuming an interest rate volatility requires three steps: (1) Generate a tree of interest rates based on the given yield curve and volatility assumptions; (2) at each node of the tree, determine whether the embedded options will be exercised; and (3) apply the backward induction valuation methodology to calculate the present value of the bond.

- The option-adjusted spread is the single spread added uniformly to the one-period forward rates on the tree to produce a value or price for a bond. OAS is sensitive to interest rate volatility: The higher the volatility, the lower the OAS for a callable bond.
- For bonds with embedded options, the best measure to assess the sensitivity of the bond's price to a parallel shift of the benchmark yield curve is effective duration. The effective duration of a callable or putable bond cannot exceed that of the straight bond.
- When the option is near the money, the convexity of a callable bond is negative, indicating that the upside for a callable bond is much smaller than the downside, whereas the convexity of a putable bond is positive, indicating that the upside for a putable bond is much larger than the downside.
- Because the prices of callable and putable bonds respond asymmetrically to upward and downward interest rate changes of the same magnitude, one-sided durations provide a better indication regarding the interest rate sensitivity of bonds with embedded options than (two-sided) effective duration.
- Key rate durations show the effect of shifting only key points, one at a time, rather than the entire yield curve.
- The arbitrage-free framework can be used to value capped and floored floaters. The cap provision in a floater is an issuer option that prevents the coupon rate from increasing above a specified maximum rate. Thus, the value of a capped floater is equal to or less than the value of the straight bond. In contrast, the floor provision in a floater is an investor option that prevents the coupon from decreasing below a specified minimum rate. Thus, the value of a floored floater is equal to or higher than the value of the straight bond.
- The characteristics of a convertible bond include the conversion price, which is the applicable share price at which the bondholders can convert their bonds into common shares, and the conversion ratio, which reflects the number of shares of common stock that the bondholders receive from converting their bonds into shares. The conversion price is adjusted in case of corporate actions, such as stock splits, bonus share issuances, and rights and warrants issuances. Convertible bondholders may receive compensation when the issuer pays dividends to its common shareholders, and they may be given the opportunity to either put their bonds or convert their bonds into shares earlier and at more advantageous terms in the case of a change of control.
- A number of investment metrics and ratios help analyze and value convertible bonds. The conversion value indicates the value of the bond if it is converted at the market price of the shares. The minimum value of a convertible bond sets a floor value for the convertible bond at the greater of the conversion value or the straight value. This floor is moving, however, because the straight value is not fixed. The market conversion premium represents the price investors effectively pay for the underlying shares if they buy the convertible bond and then convert it into shares. Scaled by the market price of the shares, it represents the premium payable when buying the convertible bond rather than the underlying common stock.
- Because convertible bonds combine characteristics of bonds, stocks, and options, as well as potentially other features, their valuation and analysis are challenging. Convertible bond investors should consider the factors that affect not only bond prices but also the underlying share price.

- The arbitrage-free framework can be used to value convertible bonds, including callable and putable ones. Each component (straight bond, call option of the stock, and call and/or put option on the bond) can be valued separately.
- The risk–return characteristics of a convertible bond depend on the underlying share price relative to the conversion price. When the underlying share price is well below the conversion price, the convertible bond is "busted" and exhibits mostly bond risk–return characteristics. Thus, it is mainly sensitive to interest rate movements. In contrast, when the underlying share price is well above the conversion price, the convertible bond exhibits mostly stock risk–return characteristics. Thus, its price follows similar movements to the price of the underlying stock. In between these two extremes, the convertible bond trades like a hybrid instrument.

PRACTICE PROBLEMS

The following information relates to Questions 1–10

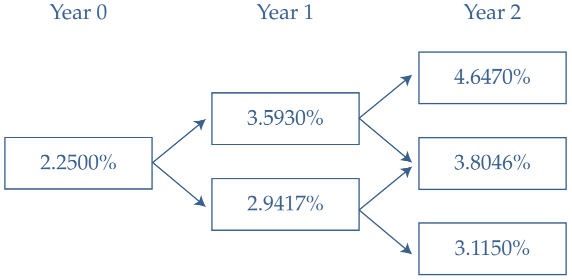
Samuel & Sons is a fixed-income specialty firm that offers advisory services to investment management companies. On 1 October 20X0, Steele Ferguson, a senior analyst at Samuel, is reviewing three fixed-rate bonds issued by a local firm, Pro Star, Inc. The three bonds, whose characteristics are given in Exhibit 1, carry the highest credit rating.

Exhibit 1 Fixed-Rate Bonds Issued by Pro Star, Inc.

Bond	Maturity	Coupon	Type of Bond
Bond #1	1 October 20X3	4.40% annual	Option-free
Bond #2	1 October 20X3	4.40% annual	Callable at par on 1 October 20X1 and on 1 October 20X2
Bond #3	1 October 20X3	4.40% annual	Putable at par on 1 October 20X1 and on 1 October 20X2

The one-year, two-year, and three-year par rates are 2.250%, 2.750%, and 3.100%, respectively. Based on an estimated interest rate volatility of 10%, Ferguson constructs the binomial interest rate tree shown in Exhibit 2.

Exhibit 2 Binomial Interest Rate Tree



On 19 October 20X0, Ferguson analyzes the convertible bond issued by Pro Star given in Exhibit 3. That day, the option-free value of Pro Star's convertible bond is \$1,060 and its stock price \$37.50.

© 2020 CFA Institute. All rights reserved.

Exhibit 3 Convertible Bond Issued by Pro Star, Inc.	
Issue Date:	**6 December 20X0**
Maturity Date:	6 December 20X4
Coupon Rate:	2%
Issue Price:	$1,000
Conversion Ratio:	31

1 The call feature of Bond #2 is *best* described as:
- A European style.
- B American style.
- C Bermudan style.

2 The bond that would *most likely* protect investors against a significant increase in interest rates is:
- A Bond #1.
- B Bond #2.
- C Bond #3.

3 A fall in interest rates would *most likely* result in:
- A a decrease in the effective duration of Bond #3.
- B Bond #3 having more upside potential than Bond #2.
- C a change in the effective convexity of Bond #3 from positive to negative.

4 The value of Bond #2 is *closest* to:
- A 102.103% of par.
- B 103.121% of par.
- C 103.744% of par.

5 The value of Bond #3 is *closest* to:
- A 102.103% of par.
- B 103.688% of par.
- C 103.744% of par.

6 All else being equal, a rise in interest rates will *most likely* result in the value of the option embedded in Bond #3:
- A decreasing.
- B remaining unchanged.
- C increasing.

7 All else being equal, if Ferguson assumes an interest rate volatility of 15% instead of 10%, the bond that would *most likely* increase in value is:
- A Bond #1.
- B Bond #2.
- C Bond #3.

8 All else being equal, if the shape of the yield curve changes from upward sloping to flattening, the value of the option embedded in Bond #2 will *most likely*:
- A decrease.
- B remain unchanged.
- C increase.

9 The conversion price of the bond in Exhibit 3 is closest to:
- A $26.67.
- B $32.26.
- C $34.19.

10 If the market price of Pro Star's common stock falls from its level on 19 October 20X0, the price of the convertible bond will *most likely*:
- A fall at the same rate as Pro Star's stock price.
- B fall but at a slightly lower rate than Pro Star's stock price.
- C be unaffected until Pro Star's stock price reaches the conversion price.

The following information relates to Questions 11–19

Rayes Investment Advisers specializes in fixed-income portfolio management. Meg Rayes, the owner of the firm, would like to add bonds with embedded options to the firm's bond portfolio. Rayes has asked Mingfang Hsu, one of the firm's analysts, to assist her in selecting and analyzing bonds for possible inclusion in the firm's bond portfolio.

Hsu first selects two corporate bonds that are callable at par and have the same characteristics in terms of maturity, credit quality, and call dates. Hsu uses the option adjusted spread (OAS) approach to analyze the bonds, assuming an interest rate volatility of 10%. The results of his analysis are presented in Exhibit 1.

Exhibit 1 Summary Results of Hsu's Analysis Using the OAS Approach

Bond	OAS (in bps)
Bond #1	25.5
Bond #2	30.3

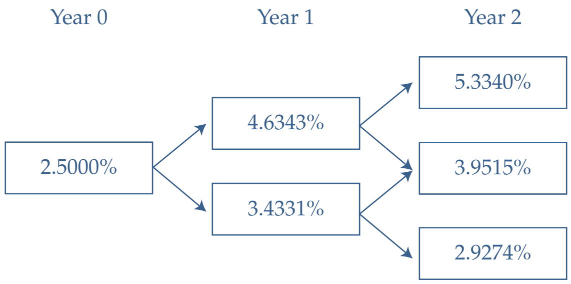
Hsu then selects the four bonds issued by RW, Inc., given in Exhibit 2. These bonds all have a maturity of three years and the same credit rating. Bonds #4 and #5 are identical to Bond #3, an option-free bond, except that they each include an embedded option.

Exhibit 2 Bonds Issued by RW, Inc.

Bond	Coupon	Special Provision
Bond #3	4.00% annual	
Bond #4	4.00% annual	Callable at par at the end of years 1 and 2
Bond #5	4.00% annual	Putable at par at the end of years 1 and 2
Bond #6	One-year reference rate annually, set in arrears	

To value and analyze RW's bonds, Hsu uses an estimated interest rate volatility of 15% and constructs the binomial interest rate tree provided in Exhibit 3.

Exhibit 3 Binomial Interest Rate Tree Used to Value RW's Bonds



Rayes asks Hsu to determine the sensitivity of Bond #4's price to a 20 bps parallel shift of the benchmark yield curve. The results of Hsu's calculations are shown in Exhibit 4.

Exhibit 4 Summary Results of Hsu's Analysis about the Sensitivity of Bond #4's Price to a Parallel Shift of the Benchmark Yield Curve

Magnitude of the Parallel Shift in the Benchmark Yield Curve	+20 bps	–20 bps
Full Price of Bond #4 (% of par)	100.478	101.238

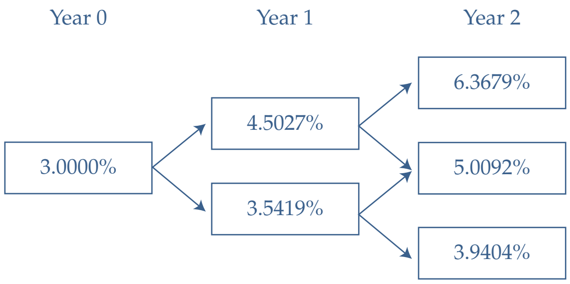
Hsu also selects the two floating-rate bonds issued by Varlep, plc, given in Exhibit 5. These bonds have a maturity of three years and the same credit rating.

Exhibit 5 Floating-Rate Bonds Issued by Varlep, plc

Bond	Coupon
Bond #7	One-year reference rate annually, set in arrears, capped at 5.00%
Bond #8	One-year reference rate annually, set in arrears, floored at 3.50%

To value Varlep's bonds, Hsu constructs the binomial interest rate tree provided in Exhibit 6.

Exhibit 6 Binomial Interest Rate Tree Used to Value Varlep's Bonds



Last, Hsu selects the two bonds issued by Whorton, Inc., given in Exhibit 7. These bonds are close to their maturity date and are identical, except that Bond #9 includes a conversion option. Whorton's common stock is currently trading at $30 per share.

Exhibit 7 Bonds Issued by Whorton, Inc.

Bond	Type of Bond
Bond #9	Convertible bond with a conversion price of $50
Bond #10	Identical to Bond #9 except that it does not include a conversion option

11 Based on Exhibit 1, Rayes would *most likely* conclude that relative to Bond #1, Bond #2 is:

- **A** overpriced.
- **B** fairly priced.
- **C** underpriced.

12 The effective duration of Bond #6 is:

- **A** close to 1.
- **B** higher than 1 but lower than 3.
- **C** higher than 3.

13 In Exhibit 2, the bond whose effective duration might lengthen if interest rates rise is:

- **A** Bond #3.
- **B** Bond #4.
- **C** Bond #5.

14 The effective duration of Bond #4 is *closest* to:

- **A** 0.76.
- **B** 1.88.
- **C** 3.77.

15 The value of Bond #7 is *closest* to:

- **A** 99.697% of par.
- **B** 99.936% of par.
- **C** 101.153% of par.

16 The value of Bond #8 is *closest* to:

- **A** 98.116% of par.

B 100.000% of par.

C 100.485% of par.

17 The value of Bond #9 is equal to the value of Bond #10:

A plus the value of a put option on Whorton's common stock.

B plus the value of a call option on Whorton's common stock.

C minus the value of a call option on Whorton's common stock.

18 The minimum value of Bond #9 is equal to the *greater* of:

A the conversion value of Bond #9 and the current value of Bond #10.

B the current value of Bond #10 and a call option on Whorton's common stock.

C the conversion value of Bond #9 and a call option on Whorton's common stock.

19 The factor that is currently *least likely* to affect the risk–return characteristics of Bond #9 is:

A interest rate movements.

B Whorton's credit spreads.

C Whorton's common stock price movements.

The following information relates to Questions 20–27

John Smith, an investment adviser, meets with Lydia Carter to discuss her pending retirement and potential changes to her investment portfolio. Domestic economic activity has been weakening recently, and Smith's outlook is that equity market values will be lower during the next year. He would like Carter to consider reducing her equity exposure in favor of adding more fixed-income securities to the portfolio.

Government yields have remained low for an extended period, and Smith suggests considering investment-grade corporate bonds to provide additional yield above government debt issues. In light of recent poor employment figures and two consecutive quarters of negative GDP growth, the consensus forecast among economists is that the central bank, at its next meeting this month, will take actions that will lead to lower interest rates.

Smith and Carter review par, spot, and one-year forward rates (Exhibit 1) and four fixed-rate investment-grade bonds issued by Alpha Corporation that are being considered for investment (Exhibit 2).

Exhibit 1 Par, Spot, and One-Year Forward Rates (annual coupon payments)

Maturity (Years)	Par Rate (%)	Spot Rate (%)	One-Year Forward (%)
1	1.0000	1.0000	1.0000
2	1.2000	1.2012	1.4028
3	1.2500	1.2515	1.3522

Exhibit 2 Selected Fixed-Rate Bonds of Alpha Corporation

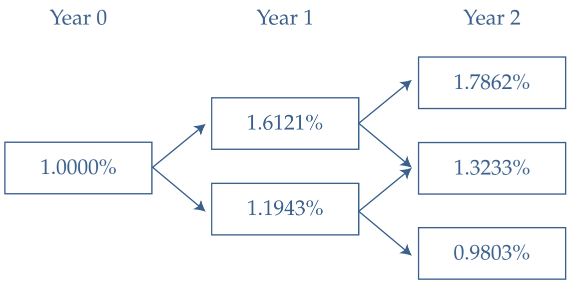
Bond	Annual Coupon	Type of Bond
Bond 1	1.5500%	Straight bond
Bond 2	1.5500%	Convertible bond: currently trading out of the money
Bond 3	1.5500%	Putable bond: putable at par one year and two years from now
Bond 4	1.5500%	Callable bond: callable at par without any protection periods

Note: All bonds in Exhibit 2 have remaining maturities of exactly three years.

Carter tells Smith that the local news media have been reporting that housing starts, exports, and demand for consumer credit are all relatively strong, even in light of other poor macroeconomic indicators. Smith explains that the divergence in economic data leads him to believe that volatility in interest rates will increase. Smith also states that he recently read a report issued by Brown and Company forecasting that the yield curve could invert within the next six months.

Smith develops a binomial interest rate tree with a 15% interest rate volatility assumption to assess the value of Alpha Corporation's bonds. Exhibit 3 presents the interest rate tree.

Exhibit 3 Binomial Interest Rate Tree for Alpha Corporation with 15% Interest Rate Volatility



Carter asks Smith about the possibility of analyzing bonds that have lower credit ratings than the investment-grade Alpha bonds. Smith discusses four other corporate bonds with Carter. Exhibit 4 presents selected data on the four bonds.

Exhibit 4 Selected Information on Fixed-Rate Bonds for Beta, Gamma, Delta, and Rho Corporations

Bond	Issuer	Bond Features	Credit Rating
Bond 5	Beta Corporation	Coupon 1.70% Callable in Year 2 OAS of 45 bps	B
Bond 6	Gamma Corporation	Coupon 1.70% Callable in Year 2 OAS of 65 bps	B

Exhibit 4 (Continued)

Bond	Issuer	Bond Features	Credit Rating
Bond 7	Delta Corporation	Coupon 1.70% Callable in Year 2 OAS of 85 bps	B
Bond 8	Rho Corporation	Coupon 1.70% Callable in Year 2 OAS of 105 bps	CCC

Notes: All bonds have remaining maturities of three years. OAS stands for option-adjusted spread.

20 Based on Exhibit 2, and assuming that the forecast for interest rates and Smith's outlook for equity returns are validated, which bond's option is *most likely* to be exercised?

A Bond 2

B Bond 3

C Bond 4

21 Based on Exhibit 2, the current price of Bond 1 is *most likely* greater than the current price of:

A Bond 2.

B Bond 3.

C Bond 4.

22 Assuming the forecast for interest rates is proven accurate, which bond in Exhibit 2 will likely experience the smallest price increase?

A Bond 1

B Bond 3

C Bond 4

23 Based on the information in Exhibit 1 and Exhibit 2, the value of the embedded option in Bond 4 is *closest* to:

A nil.

B 0.1906.

C 0.8789.

24 If Smith's interest rate volatility forecast turns out to be true, which bond in Exhibit 2 is likely to experience the greatest price increase?

A Bond 2

B Bond 3

C Bond 4

25 If the Brown and Company forecast comes true, which of the following is *most* likely to occur? The value of the embedded option in:

A Bond 3 decreases.

B Bond 4 decreases.

C both Bond 3 and Bond 4 increases.

26 Based on Exhibit 2 and Exhibit 3, the market price of Bond 4 is *closest* to:

A 100.0000.

B 100.5123.

C 100.8790.

27 Which of the following conclusions regarding the bonds in Exhibit 4 is correct?

A Bond 5 is relatively cheaper than Bond 6.

B Bond 7 is relatively cheaper than Bond 6.

C Bond 8 is relatively cheaper than Bond 7.

The following information relates to Questions 28–36

Jules Bianchi is a bond analyst for Maneval Investments, Inc. Bianchi gathers data on three corporate bonds, as shown in Exhibit 1.

Exhibit 1 Selected Bond Data

Issuer	Coupon Rate	Price	Bond Description
Ayrault, Inc. (AI)	5.25%	100.200	Callable at par in one year and two years from today
Blum, Inc. (BI)	5.25%	101.300	Option-free
Cresson Enterprises (CE)	5.25%	102.100	Putable at par in one year from today

Note: Each bond has a remaining maturity of three years, annual coupon payments, and a credit rating of BBB.

To assess the interest rate risk of the three bonds, Bianchi constructs two binomial interest rate trees based on a 10% interest rate volatility assumption and a current one-year rate of 4%. Panel A of Exhibit 2 provides an interest rate tree assuming the benchmark yield curve shifts down by 30 bps, and Panel B provides an interest rate tree assuming the benchmark yield curve shifts up by 30 bps. Bianchi determines that the AI bond is currently trading at an option-adjusted spread (OAS) of 13.95 bps relative to the benchmark yield curve.

Exhibit 2 Binomial Interest Rate Trees

A Interest Rates Shift Down by 30 bps

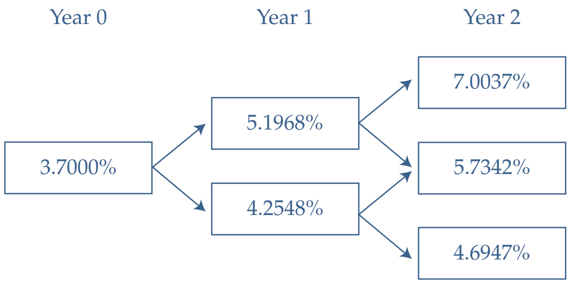
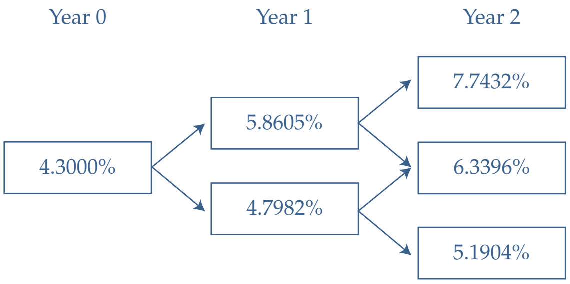


Exhibit 2 (Continued)

B Interest Rates Shift Up by 30 bps



Armand Gillette, a convertible bond analyst, stops by Bianchi's office to discuss two convertible bonds. One is issued by DeLille Enterprises (DE), and the other is issued by Raffarin Incorporated (RI). Selected data for the two bonds are presented in Exhibits 3 and 4.

Exhibit 3 Selected Data for DE Convertible Bond

Issue price	€1,000 at par
Conversion period	13 September 20X5 to 12 September 20X8
Initial conversion price	€10.00 per share
Threshold dividend	€0.50 per share
Change of control conversion price	€8.00 per share
Common stock share price on issue date	€8.70
Share price on 17 September 20X5	€9.10
Convertible bond price on 17 September 20X5	€1,123

Exhibit 4 Selected Data for RI Convertible Bond

Straight bond value	€978
Value of embedded issuer call option	€43
Value of embedded investor put option	€26
Value of embedded call option on issuer's stock	€147
Conversion price	€12.50
Current common stock share price	€11.75

Gillette makes the following comments to Bianchi:

- "The DE bond does not contain any call or put options, but the RI bond contains both an embedded call option and put option. I expect that DeLille Enterprises will soon announce a common stock dividend of €0.70 per share."
- "My belief is that, over the next year, Raffarin's share price will appreciate toward the conversion price but not exceed it."

28 Based on Exhibits 1 and 2, the effective duration for the AI bond is *closest to*:

A 1.98.

B 2.15.

C 2.73.

29 If benchmark yields were to fall, which bond in Exhibit 1 would *most likely* experience a decline in effective duration?

A AI bond

B BI bond

C CE bond

30 Based on Exhibit 1, for the BI bond, one-sided:

A up-duration will be greater than one-sided down-duration.

B down-duration will be greater than one-sided up-duration.

C up-duration and one-sided down-duration will be about equal.

31 Based on Exhibit 1, which key rate duration is the largest for the BI bond?

A One-year key rate duration

B Two-year key rate duration

C Three-year key rate duration

32 Which bond in Exhibit 1 *most likely* has the lowest effective convexity?

A AI bond

B BI bond

C CE bond

33 Based on Exhibit 3, if DeLille Enterprises pays the dividend expected by Gillette, the conversion price of the DE bond will:

A be adjusted downward.

B not be adjusted.

C be adjusted upward.

34 Based on Exhibit 3, the market conversion premium per share for the DE bond on 17 September 20X5 is *closest* to:

A €0.90.

B €2.13.

C €2.53.

35 Based on Exhibit 4, the arbitrage-free value of the RI bond is *closest* to:

A €814.

B €1,056.

C €1,108.

36 Based on Exhibit 4 and Gillette's forecast regarding Raffarin's share price, the return on the RI bond over the next year is *most likely* to be:

A lower than the return on Raffarin's common shares.

B the same as the return on Raffarin's common shares.

C higher than the return on Raffarin's common shares.

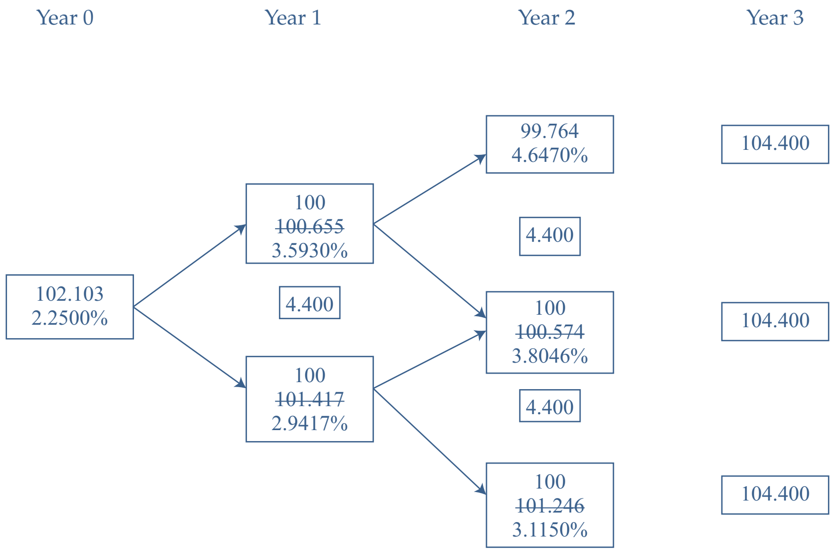
SOLUTIONS

1 C is correct. The call option embedded in Bond #2 can be exercised only at two predetermined dates: 1 October 20X1 and 1 October 20X2. Thus, the call feature is Bermudan style.

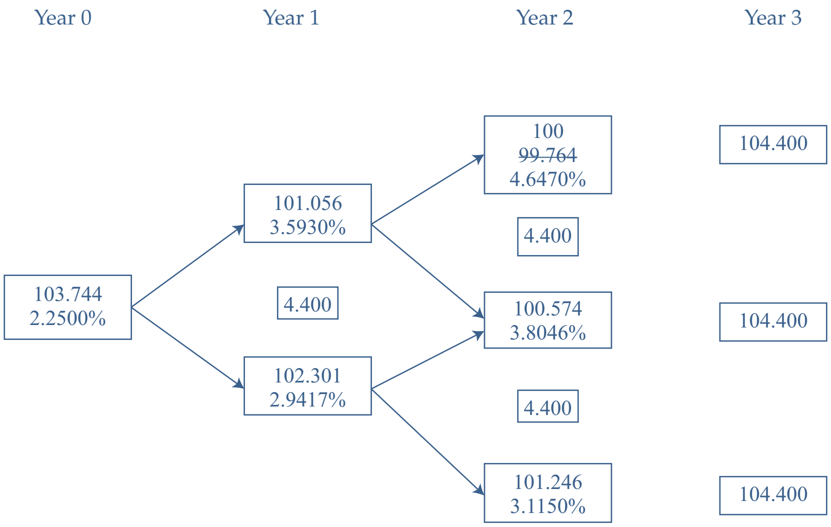
2 C is correct. The bond that would most likely protect investors against a significant increase in interest rates is the putable bond (i.e., Bond #3). When interest rates have risen and higher-yield bonds are available, a put option allows the bondholders to put back the bonds to the issuer prior to maturity and to reinvest the proceeds of the retired bonds in higher-yielding bonds.

3 B is correct. A fall in interest rates results in a rise in bond values. For a callable bond, such as Bond #2, the upside potential is capped because the issuer is more likely to call the bond. In contrast, the upside potential for a putable bond, such as Bond #3, is uncapped. Thus, a fall in interest rates would result in a putable bond having more upside potential than an otherwise identical callable bond. Note that A is incorrect because the effective duration of a putable bond increases, not decreases, with a fall in interest rates; the bond is less likely to be put and thus behaves more like an option-free bond. C is also incorrect because the effective convexity of a putable bond is always positive. It is the effective convexity of a callable bond that will change from positive to negative if interest rates fall and the call option is near the money.

4 A is correct:



5 C is correct:



6 C is correct. Bond #3 is a putable bond, and the value of a put option increases as interest rates rise. At higher interest rates, the value of the underlying option-free bond (straight bond) declines, but the decline is offset partially by the increase in the value of the embedded put option, which is more likely to be exercised.

7 C is correct. Regardless of the type of option, an increase in interest rate volatility results in an increase in option value. Because the value of a putable bond is equal to the value of the straight bond *plus* the value of the embedded put option, Bond #3 will increase in value if interest rate volatility increases. Put another way, an increase in interest rate volatility will most likely result in more scenarios where the put option is exercised, which increases the values calculated in the interest rate tree and, thus, the value of the putable bond.

8 C is correct. Bond #2 is a callable bond, and the value of the embedded call option increases as the yield curve flattens. When the yield curve is upward sloping, the one-period forward rates on the interest rate tree are high and opportunities for the issuer to call the bond are fewer. When the yield curve flattens or inverts, many nodes on the tree have lower forward rates, which increase the opportunities to call and, thus, the value of the embedded call option.

9 B is correct. The conversion price of a convertible bond is equal to the par value divided by the conversion ratio—that is, $1,000/31 = $32.26 per share.

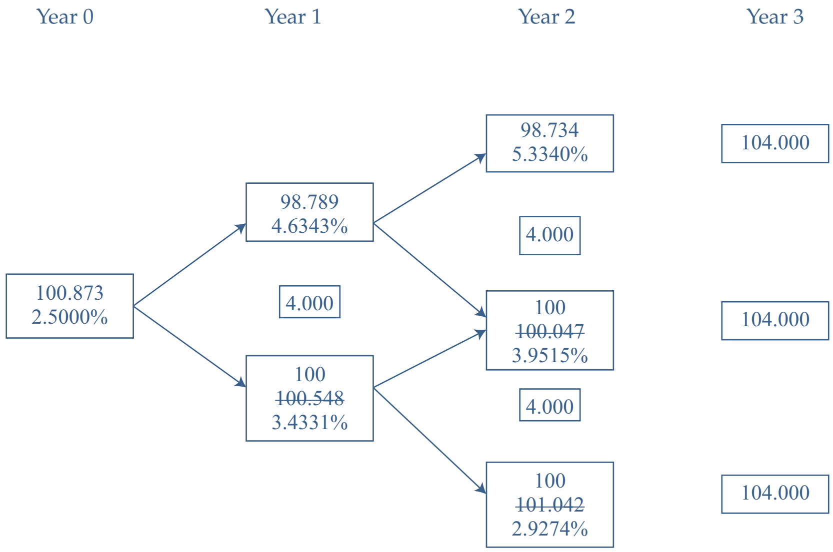
10 B is correct. The conversion value of the bond is 31 × $37.50 or $1,162.50, which represents its minimum value. Thus, the convertible bond exhibits mostly stock risk–return characteristics; a fall in the stock price will result in a fall in the convertible bond price. However, the change in the convertible bond price is less than the change in the stock price because the convertible bond has a floor. That floor is the value of the straight (option-free) bond.

11 C is correct. The option-adjusted spread (OAS) is the constant spread added to all the one-period forward rates that makes the arbitrage-free value of a risky bond equal to its market price. The OAS approach is often used to assess bond relative values. If two bonds have the same characteristics and credit quality, they should have the same OAS. If this is not the case, the bond with the largest OAS (i.e., Bond #2) is likely to be underpriced (cheap) relative to the bond with the smallest OAS (i.e., Bond #1).

12 A is correct. The effective duration of a floating-rate bond is close to the time to next reset. As the reset for Bond #6 is annual, the effective duration of this bond is close to 1.

13 B is correct. Effective duration indicates the sensitivity of a bond's price to a 100 bps parallel shift of the benchmark yield curve assuming no change in the bond's credit spread. The effective duration of an option-free bond, such as Bond #3, goes down as interest rates rise. As interest rates rise, a call option moves out of the money, which increases the value of the callable bond and lengthens its effective duration. In contrast, as interest rates rise, a put option moves into the money, which limits the price depreciation of the putable bond and shortens its effective duration. Thus, the bond whose effective duration might lengthen if interest rates rise is the callable bond (i.e., Bond #4).

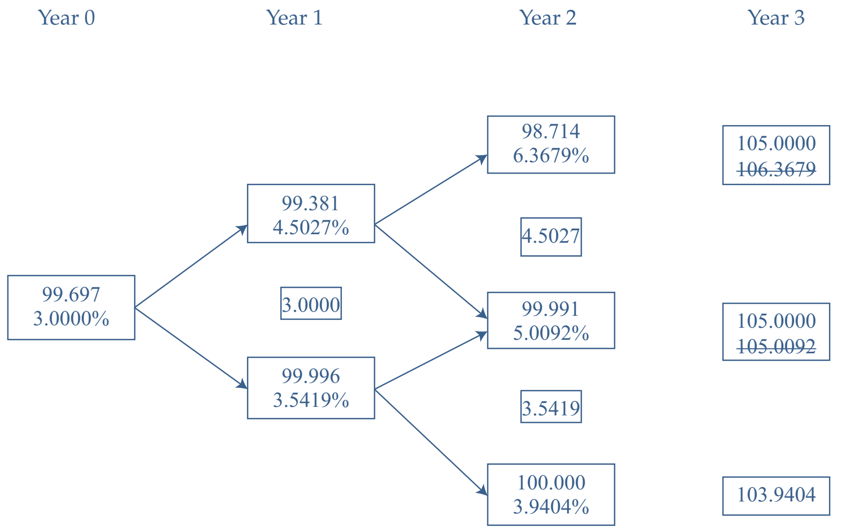
14 B is correct. The effective duration of Bond #4 can be calculated using Equation 3, where ΔCurve is 20 bps, PV_- is 101.238, and PV_+ is 100.478. PV_0, the current full price of the bond (i.e., with no shift), is not given but can be calculated using Exhibit 3 as follows:



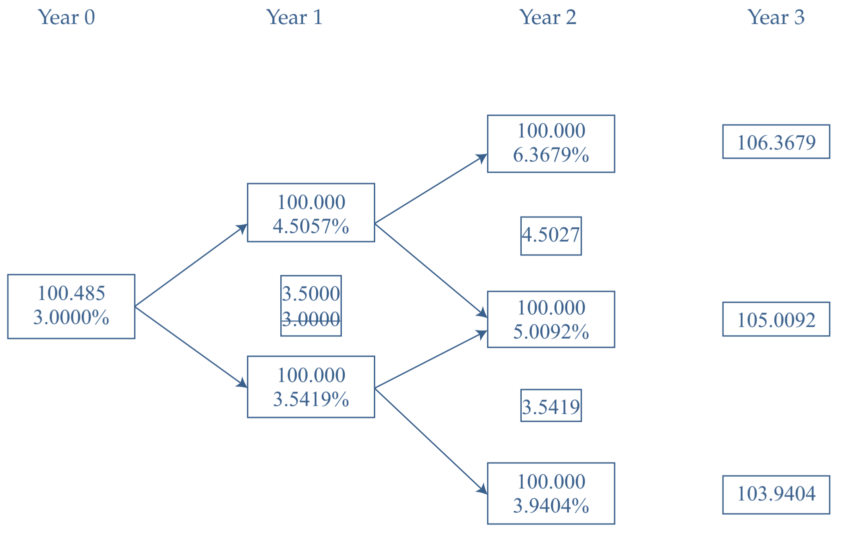
Thus, the effective duration of Bond #4 is:

$$\text{EffDur} = \frac{101.238 - 100.478}{2 \times (0.0020) \times (100.873)} = 1.88.$$

15 A is correct:



16 C is correct:



17 B is correct. A convertible bond includes a conversion option, which is a call option on the issuer's common stock. This conversion option gives the bondholders the right to convert their debt into equity. Thus, the value of Bond #9, the convertible bond, is equal to the value of Bond #10, the underlying option-free bond (straight bond), plus the value of a call option on Whorton's common stock.

18 A is correct. The minimum value of a convertible bond is equal to the greater of the conversion value of the convertible bond (i.e., Bond #9) and the current value of the straight bond (i.e., Bond #10).

19 C is correct. The risk–return characteristics of a convertible bond depend on the market price of the issuer's common stock (underlying share price) relative to the bond's conversion price. When the underlying share price is well below the conversion price, the convertible bond exhibits mostly bond risk–return characteristics. In this case, the price of the convertible bond is mainly affected by interest rate movements and the issuer's credit spreads. In contrast, when the underlying share price is above the conversion price, the convertible bond exhibits mostly stock risk–return characteristics. In this case, the price of the convertible bond is mainly affected by the issuer's common stock price movements. The underlying share price ($30) is lower than the conversion price of Bond #9 ($50). Thus, Bond #9 exhibits mostly bond risk–return characteristics and is least affected by Whorton's common stock price movements.

20 C is correct. If the central bank takes actions that lead to lower interest rates, the yields on Alpha's bonds are likely to decrease. If the yield to maturity on Bond 4 (callable) falls below the 1.55% coupon rate, the call option will become valuable and Alpha may call the bond because it is in the money.

A is incorrect because if the equity market declines, the market value of Alpha stock will also likely decrease. Therefore, Bond 2 (convertible) would have a lower conversion value; hence, the conversion option likely would not be exercised. Because Bond 2 is currently trading out of the money, it will likely trade further out of the money once the price of Alpha stock decreases.

B is incorrect because Bond 3 (putable) is more likely to be exercised in an increasing rather than a decreasing interest rate environment.

21 C is correct. All four bonds in Exhibit 2 issued by Alpha Corporation offer the same coupon rate and have the same remaining term to maturity. Bond 4 (callable) most likely has a current price that is less than Bond 1 (straight or option free) because investors are short the call option and must be compensated for bearing call risk. Bond 2 (convertible) most likely has a current price that is greater than Bond 1 because investors are paying for the conversion option embedded in Bond 2 and the option has time value associated with it, even though the option is trading out of the money. Similarly, Bond 3 (putable) most likely has a current price that is greater than Bond 1 because investors are paying for the put option.

22 C is correct. The consensus economic forecast is for interest rates to decrease. In an environment of decreasing interest rates, all bond prices should rise, ignoring any price impact resulting from any embedded options. When interest rates fall, the value of the embedded call option in Bond 4 (callable) increases, causing an opposing effect on price. The put option of putable bonds, by contrast, increases in value when interest rates rise rather than decline.

23 C is correct. Bond 4 is a callable bond. Value of an issuer call option = Value of straight bond – Value of callable bond. The value of the straight bond may be calculated using the spot rates or the one-year forward rates.

Value of an option-free (straight) bond with a 1.55% coupon using spot rates:

$$1.55/(1.0100)^1 + 1.55/(1.012012)^2 + 101.55/(1.012515)^3 = 100.8789.$$

The value of a callable bond (at par) with no call protection period cannot exceed 100, as at that price or higher the bond would be called. The value of the call option = 100.8789 – 100 = 0.8789.

24 B is correct. An increase in interest rate volatility will cause the value of the put and call options embedded in Bond 3 and Bond 4 to increase. Bond 3 (putable) would experience an increase in price because the increased value of the put option increases the bond's value. In contrast, Bond 4 (callable) will experience a price decrease because the increased value of the call option reduces the callable bond's value. Bond 2, an out-of-the-money convertible, will resemble the risk–return characteristics of a straight bond and will thus be unaffected by interest rate volatility.

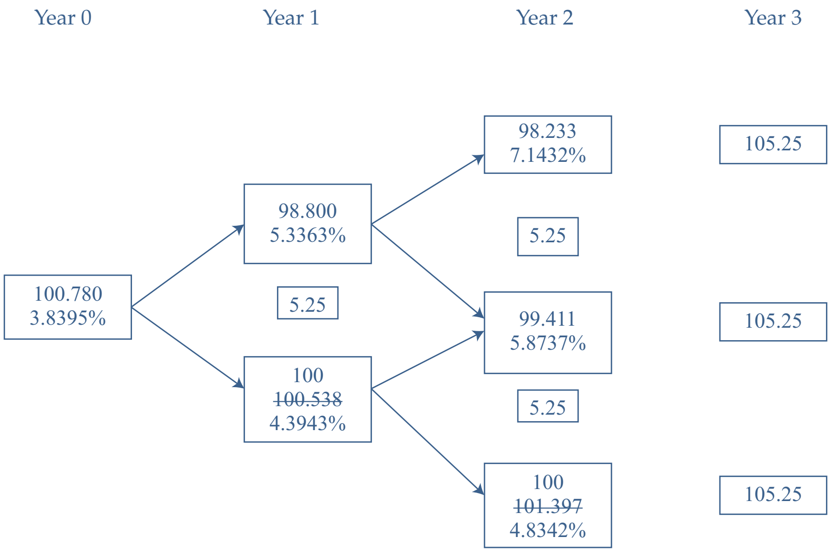
25 A is correct. All else being equal, the value of a put option decreases as the yield curve moves from being upward sloping to flat to downward sloping (inverted). Alternatively, a call option's value increases as the yield curve flattens and increases further if the yield curve inverts. Therefore, if the yield curve became inverted, the value of the embedded option in Bond 3 (putable) would decrease and the value of the embedded option in Bond 4 (callable) would increase.

26 A is correct. The market price of callable Bond 4 with no protection period cannot exceed 100.

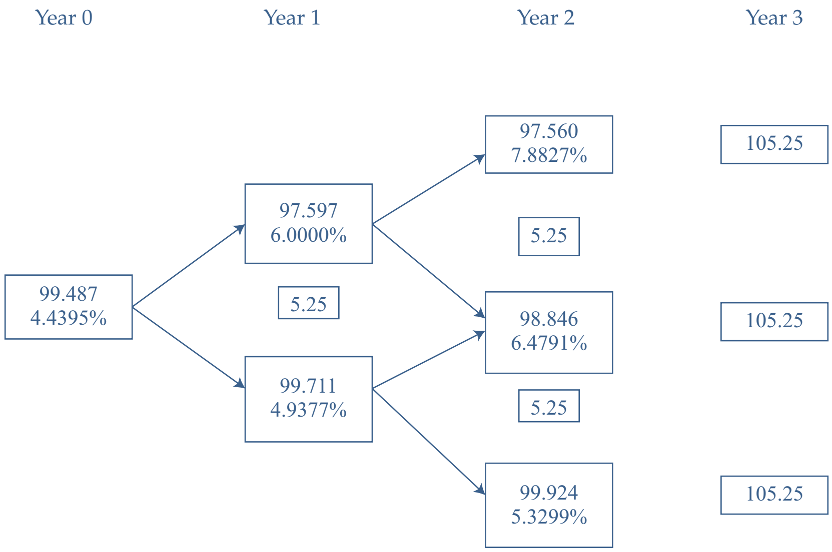
27 B is correct. A bond with a larger option-adjusted spread (OAS) than that of a bond with similar characteristics and credit quality means that the bond is likely underpriced (cheap). Bond 7 (OAS 85 bps) is relatively cheaper than Bond 6 (OAS 65 bps).

C is incorrect because Bond 8 (CCC) has a lower credit rating than Bond 7 (B) and the OAS alone cannot be used for the relative value comparison. The larger OAS (105 bps) incorporates compensation for the difference between the B and CCC bond credit ratings. Therefore, there is not enough information to draw a conclusion about relative value.

28 B is correct. The AI bond's value if interest rates shift down by 30 bps (PV_-) is 100.78:



The AI bond's value if interest rates shift up by 30 bps (PV_+) is 99.487:



$$\text{EffDur} = \frac{(\text{PV}_-) - (\text{PV}_+)}{2 \times (\Delta\text{Curve}) \times (\text{PV}_0)} = \frac{100.780 - 99.487}{2 \times 0.003 \times 100.200} = 2.15.$$

29 A is correct. The AI bond is a callable bond, and the effective duration of a callable bond decreases when interest rates fall. The reason is because a decline in interest rates may result in the call option moving into the money, which limits the price appreciation of the callable bond. Exhibit 1 also shows that the price of the AI bond is 100.200 and that it is callable at par in one year and two years. Thus, the call option is already in the money and would likely be exercised in response to increases in the AI bond's price.

30 C is correct. The BI bond is an option-free bond, and one-sided up-duration and one-sided down-duration will be about equal for option-free bonds.

31 C is correct. The BI bond is an option-free bond. Its longest key rate duration will be in the year of its maturity because the largest cash flow (payment of both coupon and principal) occurs in that year.

32 A is correct. All else being equal, a callable bond will have lower effective convexity than an option-free bond when the call option is in the money. Similarly, when the call option is in the money, a callable bond will also have lower effective convexity than a putable bond if the put option is out of the money. Exhibit 1 shows that the callable AI bond is currently priced slightly higher than its call price of par value, which means the embedded call option is in the money. The put option embedded in the CE bond is not in the money; the bond is currently priced 2.1% above par value. Thus, at the current price, the putable CE bond is more likely to behave like the option-free BI bond. Consequently, the effective convexity of the AI bond will likely be lower than the option-free BI bond and the putable CE bond.

33 A is correct. The conversion price would be adjusted downward because Gillette's expected dividend payment of €0.70 is greater than the threshold dividend of €0.50.

34 B is correct. The market conversion premium per share is equal to the market conversion price minus the underlying share price. The market conversion price is calculated as follows:

$$\text{Market conversion price} = \frac{\text{Convertible bond price}}{\text{Conversion ratio}}$$

$$= \frac{€1{,}123}{€1{,}000/€10 \text{ per share}} = €11.23 \text{ per share.}$$

The market conversion premium per share is then calculated as follows:

$$\begin{aligned}\text{Market conversion premium per share} &= \text{Market conversion price} - \\ &\quad \text{Underlying share price.} \\ &= €11.23 - €9.10 = €2.13.\end{aligned}$$

35 C is correct. The value of a convertible bond with both an embedded call option and a put option can be determined using the following formula:

$$\begin{aligned}\text{Value of callable putable convertible bond} &= \text{Value of straight bond} \\ &\quad + \text{Value of call option on the} \\ &\quad \text{issuer's stock} - \text{Value of issuer} \\ &\quad \text{call option} + \text{Value of investor} \\ &\quad \text{put option.} \\ \text{Value of callable putable bond} &= €978 + €147 - €43 + \\ &\quad €26 = €1{,}108.\end{aligned}$$

36 A is correct. Over the next year, Gillette believes that Raffarin's share price will continue to increase toward the conversion price but not exceed it. If Gillette's forecast becomes true, the return on the RI bond will increase but at a lower rate than the increase in Raffarin's share price because the conversion price is not expected to be reached.

READING

31

Credit Analysis Models

by James F. Adams, PhD, CFA, and Donald J. Smith, PhD

James Adams, PhD, CFA, is at New York University (USA). Donald J. Smith, PhD, is at Boston University Questrom School of Business (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** explain expected exposure, the loss given default, the probability of default, and the credit valuation adjustment;
☐	**b.** explain credit scores and credit ratings;
☐	**c.** calculate the expected return on a bond given transition in its credit rating;
☐	**d.** explain structural and reduced-form models of corporate credit risk, including assumptions, strengths, and weaknesses;
☐	**e.** calculate the value of a bond and its credit spread, given assumptions about the credit risk parameters;
☐	**f.** interpret changes in a credit spread;
☐	**g.** explain the determinants of the term structure of credit spreads and interpret a term structure of credit spreads;
☐	**h.** compare the credit analysis required for securitized debt to the credit analysis of corporate debt.

INTRODUCTION

1

Credit analysis plays an important role in the broader fixed-income space. Our coverage will go over important concepts, tools, and applications of credit analysis. We first look at modeling credit risk. The inputs to credit risk modeling are the expected exposure to default loss, the loss given default, and the probability of default. We explain these terms and use a numerical example to illustrate the calculation of the credit valuation adjustment for a corporate bond and its credit spread over a government bond yield taken as a proxy for a default-risk-free rate (or default-free rate).

We then discuss credit scoring and credit ratings. Credit scoring is a measure of credit risk used in retail loan markets, and ratings are used in the wholesale bond market. We explain two types of credit analysis models used in practice—structural

© 2021 CFA Institute. All rights reserved.

models and reduced-form models. Both models are highly mathematical and beyond the scope of our coverage. Therefore, we provide only an overview to highlight the key ideas and the similarities and differences between them. We then use the arbitrage-free framework and a binomial interest rate tree to value risky fixed-rate and floating-rate bonds for different assumptions about interest rate volatility. We also build on the credit risk model to interpret changes in credit spreads that arise from changes in the assumed probability of default, the recovery rate, or the exposure to default loss. We also explain the term structure of credit spreads and finally compare the credit analysis required for securitized debt with the credit analysis of corporate bonds.

2 MODELING CREDIT RISK AND THE CREDIT VALUATION ADJUSTMENT

a explain expected exposure, the loss given default, the probability of default, and the credit valuation adjustment

The difference between the yields to maturity on a corporate bond and a government bond with the same maturity is the most commonly used measure of credit risk. It is called the *credit spread* and is also known in practice as the G-spread. It reveals the compensation to the investor for bearing the default risk of the issuer—the possibility that the issuer fails to make a scheduled payment in full on the due date—and for losses incurred in the event of default.

The terms “default risk” and “credit risk” are sometimes used interchangeably in practice, but we will distinguish between the two in our coverage. Default risk is the narrower term because it addresses the likelihood of an event of default. Credit risk is the broader term because it considers both the default probability and how much is expected to be lost if default occurs. For example, it is possible that the default risk on a collateralized loan is high while the credit risk is low, especially if the value of the collateral is high relative to the amount that is owed.

We assume that the corporate bond and the default-risk-free government bond have the same taxation and liquidity. This is a simplifying assumption, of course. In reality, government bonds typically are more liquid than corporate bonds. Also, differences in liquidity within the universe of corporate bonds are great. Government bonds are available in greater supply than even the most liquid corporates and have demand from a wider set of institutional investors. In addition, government bonds can be used more readily as collateral in repo transactions and for centrally cleared derivatives. Also, there are differences in taxation in some markets. For example, interest income on US corporate bonds is taxable by both the federal and state governments. Government debt, however, is exempt from taxes at the state level. Disregarding tax and liquidity differences allows us to focus on default risk and expected loss as the determining factors for the credit spread.

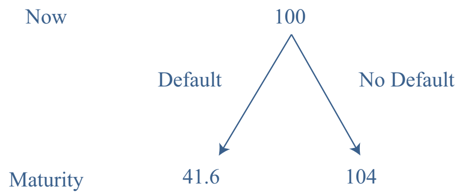
The first factor to consider in modeling credit risk is the **expected exposure** to default loss. This quantity is the projected amount of money the investor could lose if an event of default occurs, before factoring in possible recovery. Although the most common event of default is nonpayment leading to bankruptcy proceedings, the bond prospectus might identify other events of default, such as the failure to meet a different obligation or the violation of a financial covenant.

Consider a one-year, 4% annual payment corporate bond priced at par value. The expected exposure to default loss at the end of the year is simply 104 (per 100 of par value). Later, we will include multiple time periods and volatility in interest rates. That complicates the calculation of expected exposure because we will need to consider the likelihood that the bond price varies as interest rates vary. In this initial example, the exposure is simply the final coupon payment plus the redemption of principal.

The second factor is the assumed **recovery rate**, which is the percentage of the loss recovered from a bond in default. The recovery rate varies by industry, the degree of seniority in the capital structure, the amount of leverage in the capital structure in total, and whether a particular security is secured or otherwise collateralized. We assume a 40% recovery rate for this corporate bond, which is a common baseline assumption in practice. Given the recovery rate assumption, we can determine the assumed **loss given default** (the amount of loss if a default occurs). This is 62.4 per 100 of par value: 104 × (1 – 0.40) = 62.4. A related term is *loss severity*; if the recovery rate is 40%, the assumed loss severity is 60%.

Exhibit 1 illustrates the projected cash flows on the corporate bond. If there is no default, the investor receives 104. If default occurs, the investor receives 41.6: 104 – 62.4 = 41.6. We assume instantaneous recovery, which surely is another simplifying assumption. In practice, lengthy time delays can occur between the event of default and eventual recovery of cash. Notice that we assume that the recovery rate applies to interest as well as principal. One last note is that in the exhibits that we use, calculations may slightly differ on occasion due to rounding at intermediate steps.

Exhibit 1 A Simple Credit Risk Example



The third factor is the assumed **probability of default**, which is the probability that a bond issuer will not meet its contractual obligations on schedule. It is important in credit risk modeling to distinguish *risk-neutral* probabilities of default and *actual* (or historical) default probabilities. "Risk-neutral" follows the usage of the term in option pricing. In the risk-neutral option pricing methodology, the expected value for the payoffs is discounted using the risk-free interest rate. The key point is that in getting the expected value of the option, the risk-neutral probabilities associated with the payoffs need to be used. The same idea applies to valuing corporate bonds.

Suppose that a credit rating agency has collected an extensive dataset on the historical default experience for one-year corporate bonds issued by companies having the same business profile as the issuer in this example. It is observed that 99% of the bonds survive and make the full coupon and principal payment at maturity. Just 1% of the bonds default, resulting in an average recovery rate of 40%. Based on these data, the actual default probability for the corporate bond can reasonably be assumed to be 1%.

If the actual probability of default is used to get the expected future value for the corporate bond, the result is 103.376: (104 × 0.99) + (41.6 × 0.01) = 103.376. Discounting that amount at an assumed risk-free rate of 3% gives a present value of 100.365: 103.376/1.03 = 100.365. Note that in risk-neutral valuation, the expected value is discounted using the risk-free rate and not the bond's yield to maturity. The key point is that 100.365 overstates the observed value of the bond, which is 100. The issue is to determine the default probability that does produce a value of 100.

Denote the risk-neutral default probability to be P^*. The probability of survival is $1 - P^*$. Given that the corporate bond is priced at 100, $P^* = 1.60\%$. This is found as the solution to P^* in

$$100 = \frac{\left[104 \times \left(1 - P^*\right)\right] + \left(41.6 \times P^*\right)}{1.03}.$$

One reason for the difference between actual (or historical) and risk-neutral default probabilities is that actual default probabilities do not include the default risk premium associated with uncertainty over the timing of possible default loss. Another reason is that the observed spread over the yield on a risk-free bond in practice also includes liquidity and tax considerations in addition to credit risk.

To further see the interaction between the credit risk parameters—the expected exposure, the loss given default, and the probability of default—we consider a five-year, zero-coupon corporate bond. Our goal is to determine the fair value for the bond given its credit risk, its yield to maturity, and its spread over a maturity-matching government bond.

Exhibit 2 displays the calculation of the **credit valuation adjustment** (CVA). The CVA is the value of the credit risk in present value terms. In Exhibit 2, LGD stands for the loss given default, POD stands for the probability of default on the given date, POS stands for the probability of survival as of the given date, DF stands for the discount factor, and PV stands for the present value.

Exhibit 2 A Five-Year, Zero-Coupon Corporate Bond

Date (1)	Exposure (2)	Recovery (3)	LGD (4)	POD (5)	POS (6)	Expected Loss (7)	DF (8)	PV of Expected Loss (9)
0								
1	88.8487	35.5395	53.3092	1.2500%	98.7500%	0.6664	0.970874	0.6470
2	91.5142	36.6057	54.9085	1.2344%	97.5156%	0.6778	0.942596	0.6389
3	94.2596	37.7038	56.5558	1.2189%	96.2967%	0.6894	0.915142	0.6309
4	97.0874	38.8350	58.2524	1.2037%	95.0930%	0.7012	0.888487	0.6230
5	100.0000	40.0000	60.0000	1.1887%	93.9043%	0.7132	0.862609	0.6152
				6.0957%			CVA =	3.1549

The first step is to get the exposures to default loss. These are shown in Column 2 of Exhibit 2. We assume a flat government bond yield curve at 3.00%. Also, we assume that default occurs only at year-end—on Dates 1, 2, 3, 4, and 5—and that default will not occur on Date 0, the current date. The exposure on Date 5 is 100. For the other dates, we discount using the risk-free rate and the remaining number of years until maturity. For example, exposure at Date 1 is $100/(1.0300)^4 = 88.8487$.

Note that there is no interest rate volatility in this example. In a later section, we will use the arbitrage-free framework to build a binomial interest rate tree for a specified level of volatility. Then, knowing the probability of attaining each node in the tree, we will calculate the *expected exposure* for each date.

Column 3 of Exhibit 2 projects the assumed recovery if default occurs. Here, the recovery rate is a percentage of the exposure. In general, it will be a percentage of the expected exposure, including coupon interest payments, when the model allows for interest rate volatility. We assume for this example that the recovery rate is 40%. The amounts shown in Column 3 are the exposures in Column 2 times 0.40.

Column 4 shows the loss given default. It is the exposure for each date minus the assumed recovery. If the issuer defaults on Date 4, the investor's loss is projected to be 58.2524 (= 97.0874 – 38.8350) per 100 of par value.

The next parameter is the risk-neutral probability of default for each date. In Column 5 of Exhibit 2, we assume that the POD on Date 1 is 1.25%. We use *conditional probabilities of default*, meaning that each year-by-year POD assumes no prior default. These are called hazard rates in statistics. Column 6 reports the probability of survival for each year. The probability of surviving past Date 1 and arriving at Date 2 is 98.75% (= 100% – 1.25%). Therefore, the POD for Date 2 is 1.2344% (= 1.25% × 98.75%), and the POS is 97.5156% (= 98.75% – 1.2344%). The POD for Date 3 is 1.2189% (= 1.25% × 97.5156%), and the POS is 96.2967% (= 97.5156% – 1.2189%). The cumulative probability of default over the five-year lifetime of the corporate bond is 6.0957%, the sum of the PODs in Column 5.

Another method to calculate the POS for each year—a method that is used later in our discussion—is 100% minus the annual default probability raised to the power of the number of years. For example, the probability of the bond surviving until maturity is $(100\% - 1.25\%)^5 = 93.9043\%$. Note that 6.0957% plus 93.9043% equals 100%.

The assumed annual default probability does not need to be the same each year. Later we will show some examples of it changing over the lifetime of the bond.

Column 7 gives the *expected loss* for each date. This is the LGD times the POD. For example, if default occurs on Date 3, the expected loss is 0.6894 per 100 of par value. The exposure is 94.2596. At 40% recovery, the LGD is 56.5558. Assuming no prior default, the POD for that date is 1.2189%. The expected loss of 0.6894 is calculated as 56.5558 times 1.2189%.

Column 8 presents the default-risk-free *discount factors* based on the flat government bond yield curve at 3.00%. The Date 5 discount factor is 0.862609 $[= 1/(1.0300)^5]$. Finally, Column 9 shows the present value of the expected loss for each year. This is the expected loss times the discount factor. The present value of the expected Date 5 loss is 0.6152 per 100 of par value, the expected loss of 0.7132 times 0.862609.

The sum of Column 9 is 3.1549. This amount is known as the credit valuation adjustment. It allows us to calculate the *fair value* of the five-year, zero-coupon corporate bond. If the bond were default free, its price would be 86.2609—that is, the par value of 100 times the Date 5 discount factor. Subtracting the CVA from this amount gives a fair value of 83.1060 (= 86.2609 – 3.1549).

We can now calculate the credit spread on the corporate bond. Given a price of 83.1060, its yield to maturity is 3.77%. The solution for *yield* in this expression is

$$\frac{100}{(1 + \text{Yield})^5} = 83.1060.$$

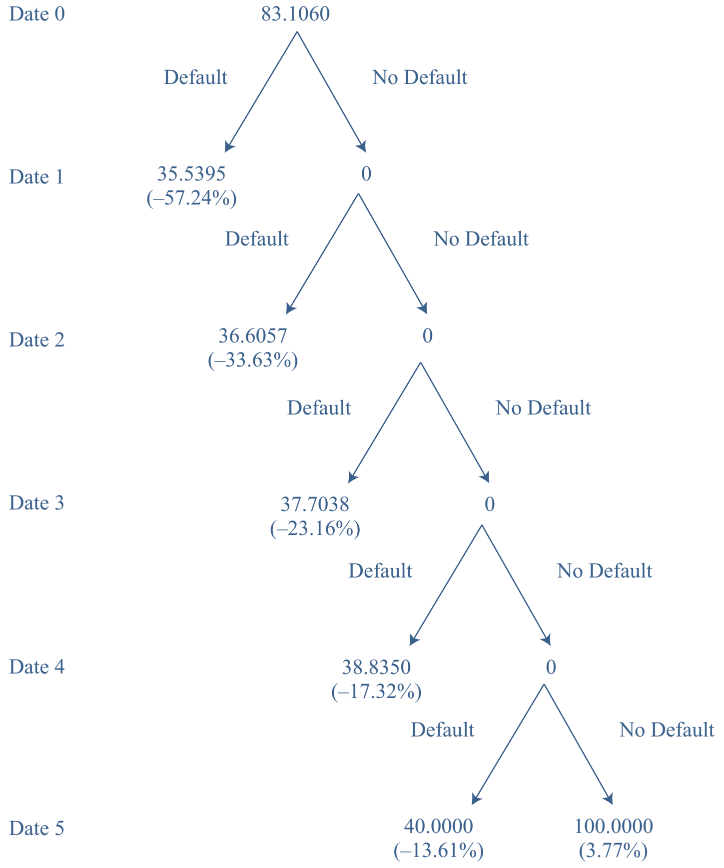
The yield on the five-year, zero-coupon government bond is 3.00%. Therefore, the credit spread is 77 bps: 3.77% – 3.00% = 0.77%. (Note that an approximation for the credit spread commonly used in practice is the annual default probability times 1 minus the recovery rate. In this case, the approximate credit spread is 0.75% [= 1.25% × (1 – 0.40)].) A key point is that the compensation for credit risk received by the investor can be expressed in two ways: (1) as the CVA of 3.1549 in terms of a present value per 100 of par value on Date 0 and (2) as a credit spread of 77 bps in terms of an annual percentage rate for five years.

Exhibit 3 provides a display of the projected cash flows and annual rates of return depending on when and if default occurs. On Date 0, the five-year, zero-coupon corporate bond is worth its fair value, 83.1060 per 100 of par value. If on Date 1 the issuer defaults, the investor gets the recoverable amount of 35.5395. The annual rate of return is −57.24%, the solution for the internal rate of return (IRR):

$$83.1060 = \frac{35.5395}{1 + \text{IRR}}.$$
$$\text{IRR} = -0.5724.$$

If there is no default, the investor receives the coupon payment on that date, which in this case is zero.

Exhibit 3 Projected Annual Rates of Return



If the issuer defaults on Date 2, the annual rate of return is −33.63%.

$$83.1060 = \frac{0}{(1 + \text{IRR})^1} + \frac{36.6057}{(1 + \text{IRR})^2}.$$
$$\text{IRR} = -0.3363.$$

If the default occurs on the maturity date, the annual rate of return "improves" to −13.61%:

$$83.1060 = \frac{0}{(1 + \text{IRR})^1} + \frac{0}{(1 + \text{IRR})^2} + \frac{0}{(1 + \text{IRR})^3} + \frac{0}{(1 + \text{IRR})^4} + \frac{40.0000}{(1 + \text{IRR})^5}.$$
$$\text{IRR} = -0.1361.$$

If there is no default, which is most likely because the probability of survival to Date 5 is 93.9043%, the realized rate of return is 3.77%. This reminds us that a yield to maturity on a risky bond is a measure of return to the investor, assuming no default.

The key observation from this example is that the investor faces a wide range of outcomes on the bond depending critically on the *timing* of default. This is a source of the default risk premium that typically is built into the pricing of the bond. Stated differently, the probability of default in credit risk models incorporates the likely time of incidence of default events as well as uncertainty over the timing of the events.

Although this is clearly a simple example of a credit risk model, it does serve to illustrate the interaction between the exposure to default loss for each date, the recovery rate, the loss given default, the probability of default, the expected loss, and the present value of expected loss. It can be made more complex and realistic. Here, the initial probability of default (the hazard rate) used to calculate the conditional PODs and the recovery rate is the same for each year, but these parameters could vary year by year. The government bond yield curve is flat, but it could be upward or downward sloping. Then, the discount factors would need to be calculated sequentially by a process known as "bootstrapping." An example of this process is included later.

In this example, we assume an annual default probability and a recovery rate to get the fair value for the risky corporate bond. This could be reversed. Suppose that we observe that the market price for the five-year, zero-coupon bond is 83.1060 and its credit spread is 77 bps. Then, the same table could be used to get—by trial-and-error search—the annual probability of default that is consistent with the bond price and a recovery rate of 40%. That default probability, which is used to calculate the year-by-year PODs, would be 1.25%. Another possibility is to change the assumed recovery rate. Suppose it is 30% of the exposure. Given the observed bond price and credit spread, the default probability would turn out to be 1.0675%. In that case, the lower recovery rate is offset by the lower probability of default. A higher recovery rate would need to be offset by a higher default probability. In general, for a given price and credit spread, the assumed probability of default and the recovery rate are positively correlated.

EXAMPLE 1

Analysis of Credit Risk (1)

A fixed-income analyst is considering the credit risk over the next year for three corporate bonds currently held in her bond portfolio. Her assessment for the exposure, probability of default, and recovery is summarized in this table:

Corporate Bond	Exposure (per 100 of par value)	Probability of Default	Recovery (per 100 of par value)
A	104	0.75%	40
B	98	0.90%	35
C	92	0.80%	30

Although all three bonds have very similar yields to maturity, the differences in the exposures arise because of differences in their coupon rates.

Based on these assumptions, how would she rank the three bonds, from highest to lowest, in terms of credit risk over the next year?

Solution:

She needs to get the loss given default for each bond and multiply that by the probability of default to get the expected loss. The LGD is the exposure minus the assumed recovery.

Corporate Bond	LGD (per 100 of par value)	POD	Expected Loss
A	64	0.75%	0.480
B	63	0.90%	0.567
C	62	0.80%	0.496

Based on the expected losses, Bond B has the highest credit risk and Bond A, the lowest. The ranking is B, C, and A. Note that there is not enough information to recommend a trading strategy because the current prices of the bonds are not given.

EXAMPLE 2

Analysis of Credit Risk (2)

A fixed-income trader at a hedge fund observes a three-year, 5% annual payment corporate bond trading at 104 per 100 of par value. The research team at the hedge fund determines that the risk-neutral annual probability of default used to calculate the conditional POD for each date for the bond, given a recovery rate of 40%, is 1.50%. The government bond yield curve is flat at 2.50%.

Based on these assumptions, does the trader deem the corporate bond to be overvalued or undervalued? By how much? If the trader buys the bond at 104, what are the projected annual rates of return?

Solution:

The trader needs to build a table similar to that shown in Exhibit 2; this table is presented in Exhibit 4.

Exhibit 4 CVA Calculation for Example 2

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0								
1	109.8186	43.9274	65.8911	1.5000%	98.5000%	0.9884	0.975610	0.9643
2	107.4390	42.9756	64.4634	1.4775%	97.0225%	0.9524	0.951814	0.9066
3	105.0000	42.0000	63.0000	1.4553%	95.5672%	0.9169	0.928599	0.8514
				4.4328%			CVA =	2.7222

The exposures are the values for the bond plus the coupon payment for each date assuming a yield to maturity of 2.50%. The exposure is 109.8186 for Date 1 when two years to maturity remain:

$$5 + \frac{5}{(1.0250)^1} + \frac{105}{(1.0250)^2} = 109.8186.$$

The assumed recovery for Date 1 is 43.9274 (= 109.8186 × 0.40) for a loss given default of 65.8911 (= 109.8186 – 43.9274). (Note that all calculations are carried out on spreadsheets to preserve precision. The rounded results are reported in the text.) The expected loss is 0.9884 (= 65.8911 × 0.0150). The discount factor for Date 1 is 0.975610 = $1/(1.0250)^1$. The present value of the expected loss is 0.9643 (= 0.9884 × 0.975610).

The credit valuation adjustment for the bond is 2.7222, the sum of the present values of expected loss. If this five-year, 5% bond were default free, its price would be 107.1401.

$$\frac{5}{(1.0250)^1} + \frac{5}{(1.0250)^2} + \frac{105}{(1.0250)^3} = 107.1401.$$

Therefore, the fair value of the bond given the assumed credit risk parameters is 104.4178 (= 107.1401 – 2.7222). If this three-year, 5% bond were default free, its price would be 107.1401.

The projected annual rates of return for default on Dates 1, 2, and 3 are –57.76%, –33.27%, and –22.23%, respectively. If there is no default, the rate of return is 3.57%, which is the yield to maturity. Note that these rates of return neglect coupon reinvestment risk because internal rate of return calculations implicitly assume reinvestment at the same rate. The calculations are as follows:

$$104 = \frac{43.9274}{(1 + \text{IRR})^1}.$$

$$\text{IRR} = -0.5776.$$

$$104 = \frac{5}{(1 + \text{IRR})^1} + \frac{42.9756}{(1 + \text{IRR})^2}.$$

$$\text{IRR} = -0.3327.$$

$$104 = \frac{5}{(1 + \text{IRR})^1} + \frac{5}{(1 + \text{IRR})^2} + \frac{42.0000}{(1 + \text{IRR})^3}.$$

$$\text{IRR} = -0.2223.$$

$$104 = \frac{5}{(1 + \text{IRR})^1} + \frac{5}{(1 + \text{IRR})^2} + \frac{105}{(1 + \text{IRR})^3}.$$

$$\text{IRR} = 0.0357.$$

Environmental, social, and governance (ESG) considerations may also play a role in credit risk assessment. For example, companies responsible for pollution run the risk of fines or other business sanctions, those with poor labor practices risk their reputation and may face customer boycotts or lawsuits, and firms with weak governance are more likely to engage in aggressive or even fraudulent accounting. Estimated probabilities of default and loss given default should incorporate these potential impacts.

Recent years have also seen several types of bond with explicit links to ESG matters. Climate, or green, bonds are typically issued with proceeds earmarked for environmentally beneficial purposes and may come with tax incentives to enhance their attractiveness to investors.

Another category of fixed-income instruments whose special features affect credit risk assessment are catastrophe and pandemic bonds. They resemble an insurance product, rather than a traditional debt instrument. For example, the World Bank issued pandemic bonds in 2017, offering investors high interest payments in return for taking on the risk of losing capital should a pandemic occur, in which case they would pay out aid to poor nations suffering from a serious outbreak of infectious disease. At the time of this writing (July 2020), nearly all the principal from those bonds has been wiped out because caseloads and deaths from COVID-19 have exceeded the bonds' thresholds.

3 CREDIT SCORES AND CREDIT RATINGS

b explain credit scores and credit ratings

c calculate the expected return on a bond given transition in its credit rating

Credit scores and ratings are used by lenders in deciding to extend credit to a borrower and in determining the terms of the contract. Credit scores are used primarily in the retail lending market for small businesses and individuals. Credit ratings are used in the wholesale market for bonds issued by corporations and government entities, as well as for asset-backed securities (ABS).

Credit scoring methodologies can vary. In some countries, only negative information, such as delinquent payments or outright default, is included. Essentially, everyone has a good credit score until proven otherwise. In other countries, a broader set of information is used to determine the score. A score reflects actual observed factors. In general, credit reporting agencies are national in scope because of differences in legal systems and privacy concerns across countries.

The FICO score, which is the federally registered trademark of the Fair Isaac Corporation, is used in the United States by about 90% of lenders to retail customers. FICO scores are computed using data from consumer credit files collected by three national credit bureaus: Experian, Equifax, and TransUnion. Five primary factors are included in the proprietary algorithm used to get the score:

- 35% for the payment history: This includes the presence or lack of such information as delinquency, bankruptcy, court judgments, repossessions, and foreclosures.
- 30% for the debt burden: This includes credit card debt-to-limit ratios, the number of accounts with positive balances, and the total amount owed.
- 15% for the length of credit history: This includes the average age of accounts on the credit file and the age of the oldest account.
- 10% for the types of credit used: This includes the use of installment payments, consumer finance, and mortgages.
- 10% for recent searches for credit: This includes "hard" credit inquiries when consumers apply for new loans but not "soft" inquiries, such as for employee verification or self-checking one's score.

Fair Isaac Corporation, on its website, notes items that are not included in the FICO credit score: race, color, national origin, sex, marital status, age, salary, occupation, employment history, home address, and child/family support obligations. The company also reports from time to time the distribution across scores, which range from a low of 300 to a perfect score of 850. Exhibit 5 shows the distribution for three particular months: October 2005, before the global financial crisis; April 2009, in the depths of the crisis; and April 2017, well after the crisis. It is evident that the percentage of weak scores increased as economic conditions worsened but has gone down since then. The average FICO score varied from 688 to 687 to 700 during these months.

Exhibit 5 Distribution of FICO Scores

FICO Score	October 2005	April 2009	April 2017
300–499	6.6%	7.3%	4.7%
500–549	8.0%	8.7%	6.8%
550–599	9.0%	9.1%	8.5%
600–649	10.2%	9.5%	10.0%
650–699	12.8%	12.0%	13.2%
700–749	16.4%	15.9%	17.1%
750–799	20.1%	19.3%	19.0%
800–850	16.9%	18.2%	20.7%

Source: Fair Isaac Corporation.

EXAMPLE 3

Credit Scoring

Tess Waresmith is a young finance professional who plans to eventually buy a two-family house, live in one unit, and rent the other to help cover the mortgage payments. She is a careful money manager and every year checks her FICO credit score. She is pleased to see that it has improved from 760 last year to 775 this year. Which of these factors can explain the improvement?

A She is now one year older and has not had any late payments on credit cards during the year.

B Her bank on its own raised her limit on a credit card from $1,000 to $2,500, but she has maintained the same average monthly balance.

C She applied for and received a new car loan from her credit union.

D She refrained from checking her FICO score monthly, which some of her friends do.

Solution:

Factors A, B, and C help explain the improvement. Going down the list:

A Age itself is not a factor used by Fair Isaac to determine the credit score. However, the average age of the accounts is a factor, as is the age of the oldest account. Therefore, other things being equal, the passage of time tends to improve the score. In general, age and credit score are highly correlated.

B The credit card debt-to-limit ratio is a component of the debt burden. Having a higher limit for the same average balance reduces the ratio and improves the credit score.

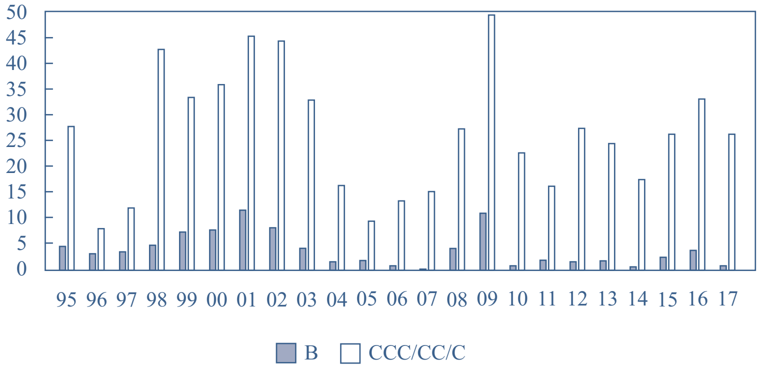
C Because the car loan is a new type of credit usage and thus does not have any late payments, it has a positive impact on the score.

D Refraining from self-checking one's credit score has no impact. Self-checking is deemed to be a "soft inquiry" and does not factor into the calibration of the FICO score.

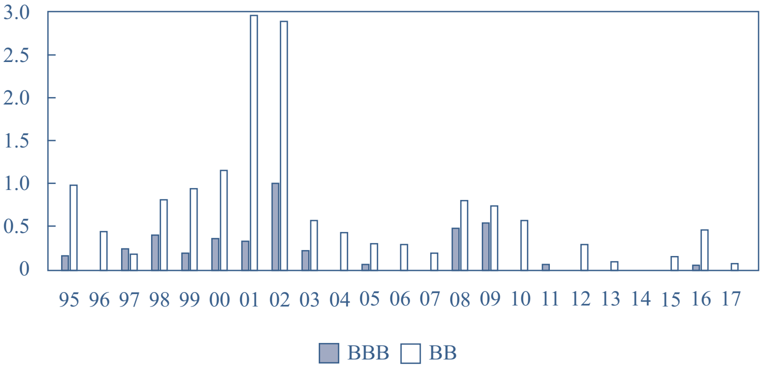
Whereas credit scores are the primary measure of credit risk in retail lending, credit ratings are widely used in corporate and sovereign bond markets. The three major global credit rating agencies are Moody's Investors Service, Standard & Poor's, and Fitch Ratings. Each provides quality ratings for issuers as well as specific issues. Similar to credit scores, these are ordinal ratings focusing on the probability of default. The historical corporate default experience by various ratings for 1995 to 2017 is shown in Exhibit 6.

Exhibit 6 Historical Corporate Default Experience by Rating (entries are in %)

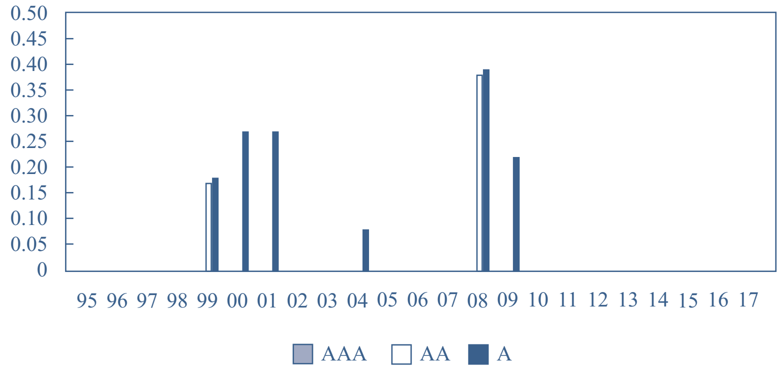
A.



B.



C.



The credit rating agencies consider the expected loss given default by means of *notching*, which is a rating adjustment methodology (covered earlier in the CFA Program curriculum) to reflect the priority of claim for specific debt issues of that issuer and to reflect any subordination. The issuer rating is typically for senior unsecured debt. The rating on subordinated debt is then adjusted, or "notched," by lowering it one or two levels. This inclusion of loss given default in addition to the probability of default explains why they are called "credit ratings" and not just "default ratings."

In addition to the "letter grade," the rating agencies provide an outlook (positive, stable, or negative) for the issuer as well as when the issuer is under "watch." For example, what follows is the history of Standard & Poor's issuer rating for RadioShack Corporation as it moved from BBB– in 1969 to BB+ in 1978, to AAA in 1983, to BB in 2006, and finally to default in 2015:

• 2 May 1969	BBB–
• 13 October 1978	BB+
• 12 December 1980	BB
• 1 April 1981	BBB+
• 7 January 1982	A
• 10 January 1983	AAA
• 28 November 1984	A+/Watch Negative
• 8 August 1991	A/Stable
• 4 January 1993	A/Watch Negative
• 25 February 1993	A–/Stable
• 27 May 1993	A–/Watch Positive
• 17 January 1994	A–/Stable
• 17 October 1996	A–/Negative
• 24 February 1999	A–/Stable
• 13 May 2005	A–/Watch Negative
• 8 August 2005	BBB+/Stable
• 21 April 2006	BBB–/Stable
• 24 July 2006	BBB–/Negative
• 25 October 2006	BB/Negative
• 12 August 2008	BB/Stable
• 21 November 2011	BB–/Stable
• 2 March 2012	B+/Negative
• 30 July 2012	B–/Negative
• 21 November 2012	CCC+/Negative

(continued)

- 1 August 2013 CCC/Negative
- 20 December 2013 CCC+/Negative
- 16 June 2014 CCC/Negative
- 11 September 2014 CCC–/Negative
- 6 February 2015 D

Source: Standard & Poor's, "2014 Annual Global Corporate Default Study and Rating Transitions," Table 54 (30 April 2015).

The history of RadioShack illustrates that the rating can remain the same for prolonged periods of time. The company was A+ from 1984 to 1991 and A– from 1993 to 2005. The rating agencies report *transition matrixes* based on their historical experience. Exhibit 7 is a representative example. It shows the probabilities of a particular rating transitioning to a different rating over the course of the following year. An A rated issuer has an 87.50% probability of remaining at that level; a 0.05% probability of moving up to AAA (such as RadioShack did in 1983); a 2.50% probability of moving up to AA; an 8.40% probability of moving down to BBB; 0.75% down to BB; 0.60% to B; 0.12% to CCC, CC, or C; and 0.08% to D, where it is in default.

Exhibit 7 Representative One-Year Corporate Transition Matrix (entries are in %)

From/To	AAA	AA	A	BBB	BB	B	CCC, CC, C	D
AAA	**90.00**	9.00	0.60	0.15	0.10	0.10	0.05	0.00
AA	1.50	**88.00**	9.50	0.75	0.15	0.05	0.03	0.02
A	0.05	2.50	**87.50**	8.40	0.75	0.60	0.12	0.08
BBB	0.02	0.30	4.80	**85.50**	6.95	1.75	0.45	0.23
BB	0.01	0.06	0.30	7.75	**79.50**	8.75	2.38	1.25
B	0.00	0.05	0.15	1.40	9.15	**76.60**	8.45	4.20
CCC, CC, C	0.00	0.01	0.12	0.87	1.65	18.50	**49.25**	29.60
Credit Spread	0.60%	0.90%	1.10%	1.50%	3.40%	6.50%	9.50%	

Exhibit 7 also shows representative credit spreads for a 10-year corporate bond. The credit transition matrix and the credit spreads allow a fixed-income analyst to estimate a one-year rate of return given the possibility of credit rating migration but still no default. Assume that an A rated 10-year corporate bond will have a modified duration of 7.2 at the end of the year given stable yields and spreads. For each possible transition, the analyst can calculate the expected percentage price change as the product of the modified duration and the change in the spread:

From A to AAA:	$-7.2 \times (0.60\% - 1.10\%) = +3.60\%$.
From A to AA:	$-7.2 \times (0.90\% - 1.10\%) = +1.44\%$.
From A to BBB:	$-7.2 \times (1.50\% - 1.10\%) = -2.88\%$.
From A to BB:	$-7.2 \times (3.40\% - 1.10\%) = -16.56\%$.
From A to B:	$-7.2 \times (6.50\% - 1.10\%) = -38.88\%$.
From A to CCC, CC, or C:	$-7.2 \times (9.50\% - 1.10\%) = -60.48\%$.

The probabilities of migration now can be used to calculate the expected percentage change in the bond value over the year. The expected percentage change in bond value for an A rated corporate bond is found by multiplying each expected percentage price change for a possible credit transition by its respective transition probability found in the row associated with the A rating and summing the products:

(0.0005 × 3.60%) + (0.0250 × 1.44%) + (0.8750 × 0%) + (0.0840 × –2.88%) + (0.0075 × –16.56%) + (0.0060 × –38.88%) + (0.0012 × –60.48%) = –0.6342%.

Therefore, the expected return on the bond over the next year is its yield to maturity minus 0.6342%, assuming no default. If the bond was not investment grade, the small probability of a transition to default would need to be taken into consideration.

Credit spread migration typically reduces the expected return for two reasons. First, the probabilities for change are not symmetrically distributed around the current rating. They are skewed toward a downgrade rather than an upgrade. Second, the increase in the credit spread is much larger for downgrades than the decrease in the spread for upgrades.

EXAMPLE 4

The Impact of Credit Migration on Expected Return

Manuel Perello is a wealth manager for several Latin American families who seek to keep a portion of their assets in very high-quality corporate bonds. Mr. Perello explains that the yields to maturity on the bonds should be adjusted for possible *credit spread widening* to measure the expected rate of return over a given time horizon. In his presentation to one of the families, he uses a 10-year, AAA rated corporate bond that would have a modified duration of 7.3 at the end of the year. Using the corporate transition matrix in Exhibit 7, Mr. Perello concludes that the expected return on the bond over the next year can be approximated by the yield to maturity less 32.5 bps to account for a possible credit downgrade even if there is no default. Demonstrate how he arrives at that conclusion.

Solution:

First, calculate the expected percentage price change using the modified duration for the bond and the change in the credit spread:

From AAA to AA:	–7.3 × (0.90% – 0.60%) = –2.19%.
From AAA to A:	–7.3 × (1.10% – 0.60%) = –3.65%.
From AAA to BBB:	–7.3 × (1.50% – 0.60%) = –6.57%.
From AAA to BB:	–7.3 × (3.40% – 0.60%) = –20.44%.
From AAA to B:	–7.3 × (6.50% – 0.60%) = –43.07%.
From AAA to CCC, CC, or C:	–7.3 × (9.50% – 0.60%) = –64.97%.

Second, calculate the expected percentage change in bond value over the year using the probabilities associated with the AAA rating row in the corporate transition matrix:

(0.9000 × 0%) + (0.0900 × –2.19%) + (0.0060 × –3.65%) + (0.0015 × –6.57%) + (0.0010 × –20.44%) + (0.0010 × –43.07%) + (0.0005 × –64.97%) = –0.3249%.

4 STRUCTURAL AND REDUCED-FORM CREDIT MODELS

d explain structural and reduced-form models of corporate credit risk, including assumptions, strengths, and weaknesses

Credit analysis models fall into two broad categories—structural models and reduced-form models (Fabozzi 2013). Structural models of credit risk date back to the 1970s and the seminal contributions to finance theory by Fischer Black, Myron Scholes, and Robert Merton (Black and Scholes 1973; Merton 1974). Their key insights were that a company defaults on its debt if the value of its assets falls below the amount of its liabilities and that the probability of that event has the features of an option.

Reduced-form varieties emerged in the 1990s (Jarrow and Turnbull 1995; Duffie and Singleton 1999) and avoid a fundamental problem with the structural models. The Black–Scholes–Merton option pricing model explicitly assumes that the assets on which the options are written (i.e., the shares of a company) are actively traded. That assumption is fine for stock options; however, the assets of the company typically do not trade. Reduced-form models get around this problem by not treating default as an endogenous (internal) variable. Instead, the default is an exogenous (external) variable that occurs randomly. Unlike structural models that aim to explain *why* default occurs (i.e., when the asset value falls below the amount of liabilities), reduced-form models aim to explain statistically *when*. This is known as the *default time* and can be modeled using a Poisson stochastic process. The key parameter in this process is the *default intensity*, which is the probability of default over the next time increment. Reduced-form credit risk models are thus also called *intensity-based* and *stochastic default rate* models.

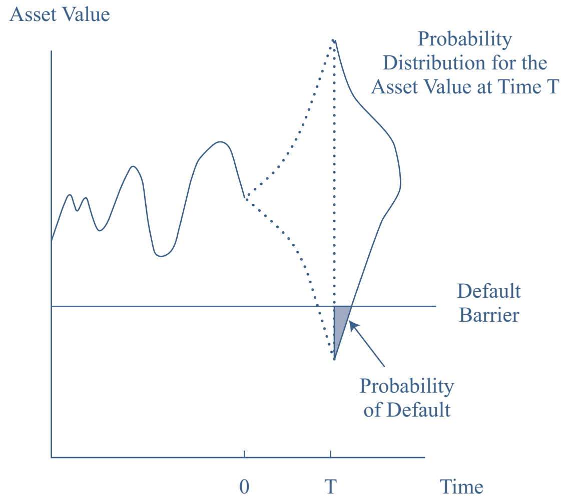
Both types of credit risk model have advantages and disadvantages. Structural models provide insight into the nature of credit risk but can be burdensome to implement. The modeler needs to determine the value of the company, its volatility, and the default barrier that is based on the liabilities of the company. In the model, the company defaults when the value of its assets dips below this default barrier. Although straightforward in theory, it can be difficult in practice because of limitations in available data. Examples of companies hiding debt (Enron Corporation, Tyco International, WorldCom, Parmalat, and Lehman Brothers, to name a few) highlight the challenge to measure the default barrier, especially in times when knowing changes in default probabilities would be most beneficial to investors (Smith 2011).

Reduced-form models have the advantage that the inputs are observable variables, including historical data. The default intensity is estimated using regression analysis on *company-specific* variables (e.g., leverage ratio, net-income-to-assets ratio, and cash-to-assets ratio), and *macroeconomic* variables (e.g., unemployment rate, GDP growth rate, measures of stock market volatility). This flexibility allows the model to directly reflect the business cycle in the credit risk measure.

A disadvantage of reduced-form models is that, unlike structural models, they do not explain the economic reasons for default. Also, reduced-form models assume that default comes as a "surprise" and can occur at any time. In reality, default is rarely a surprise because the issuer usually has been downgraded several times before the final event, as we saw with the RadioShack experience in the previous section.

Exhibit 8 depicts a structural model of default. The vertical axis measures the asset value of the company. It is called a structural model because it depends on the structure of the company's balance sheet—its assets, liabilities, and equity. It also can be called a *company-value* model because the key variable is the asset value of the company. In Exhibit 8, the asset value has been volatile prior to now, time 0, but has remained above the horizontal line that represents the default barrier. If the asset value falls below the barrier, the company defaults on the debt.

Exhibit 8 A Structural Model of Default



Source: This exhibit is adapted from Duffie and Singleton (2003, p. 54).

There is a probability distribution for the asset value as of some future date, time T. The probability of default is endogenous to this structural model. It is the portion of the probability distribution that lies below the default barrier. This default probability increases with the variance of the future asset value, with greater time to T, and with greater financial leverage. Less debt in the capital structure lowers the horizontal line and reduces the probability of default. These factors indicate that credit risk is linked to option pricing theory.

An important feature of the structural credit models is that they allow interpretation of debt and equity values in terms of options. Let $A(T)$ be the random asset value as of time T. To simplify, we can assume that the debt liabilities are zero-coupon bonds that mature at time T. These bonds have a face value of K, which represents the default barrier in Exhibit 8. The values for debt and equity at time T are denoted $D(T)$ and $E(T)$ and depend on the relationship between $A(T)$ and K:

$$D(T) + E(T) = A(T). \tag{1}$$

$$E(T) = \max[A(T) - K, 0]. \tag{2}$$

$$D(T) = A(T) - \max[A(T) - K, 0]. \tag{3}$$

Equation 1 is the balance sheet identity: The market values of debt and equity at time T equal the asset value. Equation 2 indicates that equity is essentially a purchased call option on the assets of the company whereby the strike price is the face value of the debt. It is a long position in a call option because the value of equity goes up when the asset value goes up. Moreover, like options, equity does not take on negative values. Equation 3 shows that in this formulation, the debtholders own the assets of the company and have written the call option held by the shareholders. We can interpret the premium that the debtholders receive for writing the option as the value of having priority of claim in the event that the asset value falls below K. In that case, the value of equity falls to zero and the debtholders own the remaining assets.

Suppose that at time T, $A(T) > K$ so that the call option is in the money to the shareholders. Then, $E(T) = A(T) - K$ and $D(T) = A(T) - [A(T) - K] = K$. Instead, suppose that $A(T) < K$ so that the call option is out of the money and the debt is in default. In this case, $E(T) = 0$ and $D(T) = A(T) - 0 = A(T)$. In both situations, as well as when $A(T) = K$, the balance sheet identity holds. Notice that *limited liability* is an inherent assumption in this model. Equity, like options, does not take on negative values.

EXAMPLE 5

An Equivalent Option Interpretation of Debt and Equity

Carol Feely is a junior credit analyst at one of the major international credit rating agencies. She understands that in the standard structural models, equity is interpreted as a call option on the asset value of the company. However, she is not comfortable with the assumption that it is the debtholders who implicitly own the assets and write a call option on them. She claims that the model should start with the understanding that the shareholders own the net value of the company, which is $A(T) - K$, and that their limited liability is essentially the value of a long position in a put option at a strike price of K. Furthermore, the debtholders own a "risk-free" bond having a value of K at time T and a short position in the put that is held by the shareholders.

Demonstrate that Ms. Feely's "embedded put option" interpretation provides the same values for debt and equity at time T as does the more customary call option structural model.

Solution:

A long position in a put option on the asset value at a strike price of K takes the form max$[K - A(T), 0]$. This put option has intrinsic value to its holder when $K > A(T)$ and is worthless when $K \leq A(T)$. The values for $E(T)$ and $D(T)$ according to Ms. Feely at time T are as follows:

$$E(T) = A(T) - K + \max[K - A(T), 0].$$

$$D(T) = K - \max[K - A(T), 0].$$

If $A(T) > K$ at time T, the put option is out of the money, $E(T) = A(T) - K + 0 = A(T) - K$, and $D(T) = K - 0 = K$. If $A(T) < K$, the put is in the money, $E(T) = A(T) - K + [K - A(T)] = 0$, and $D(T) = K - [K - A(T)] = A(T)$. This interpretation indicates that the value of limited liability to shareholders is the value of the put option that they purchase from the debtholders. Ms. Feely is correct in that the same payoffs as the embedded call option interpretation are obtained.

Although credit risk is inherently linked to option pricing, it is the implementation of structural models that has provided practical value to fixed-income analysis. Many credit rating agencies and consultancies, most notably Moody's KMV Corporation, use option pricing methodologies to estimate such credit risk parameters as the probability of default and the loss given default. Building on the classic Black–Scholes–Merton model and later variants, the model builders use historical data on the company's equity price to estimate volatility, which is a key element in option pricing models.

These advantages and disadvantages indicate that the choice of credit risk model depends on how it is to be used and by whom. Structural models require information best known to the managers of the company (and perhaps their commercial bankers and the credit rating agencies). Therefore, they can be used for internal risk management, for banks' internal credit risk measures, and for publicly available credit ratings. Reduced-form models require only information generally available in financial markets, which suggests that they should be used to value risky debt securities and credit derivatives.

VALUING RISKY BONDS IN AN ARBITRAGE-FREE FRAMEWORK

5

e calculate the value of a bond and its credit spread, given assumptions about the credit risk parameters

In this section, we use the arbitrage-free framework to analyze the credit risk of a corporate bond in the context of volatile interest rates (based on Smith 2017). Earlier, we solved for the credit valuation adjustment and the credit spread under the assumptions of no interest rate volatility and a flat government bond yield curve. A binomial interest rate tree for benchmark bond yields allows us to calculate the *expected exposure* to default loss. In addition, we have an upward-sloping yield curve for benchmark bonds. We take the risk-neutral probability of default as given, as if it has been determined using a structural or reduced-form credit model. We also assume a recovery rate if default were to occur that conforms to the seniority of the debt issue and the nature of the issuer's assets.

The first step is to build the binomial interest rate tree under the assumption of no arbitrage. Exhibit 9 displays the data on annual payment benchmark government bonds that are used to build the binomial interest rate tree. This is the *par curve* because each bond is priced at par value. The coupon rates are equal to the yields to maturity because the years to maturity are whole numbers (integers) so that there is no accrued interest. The one-year government bond has a negative yield to reflect the conditions seen in some financial markets. Note that the actual one-year security is likely to be a zero-coupon bond priced at a premium, at 100.2506 per 100 of par value: (100/100.2506) – 1 = –0.0025. However, on a par curve for which all the bonds are priced at 100, it is shown as having a negative coupon rate.

Exhibit 9 Par Curve for Annual Payment Benchmark Government Bonds, Spot Rates, Discount Factors, and Forward Rates

Maturity	Coupon Rate	Price	Discount Factor	Spot Rate	Forward Rate
1	–0.25%	100	1.002506	–0.2500%	
2	0.75%	100	0.985093	0.7538%	1.7677%
3	1.50%	100	0.955848	1.5166%	3.0596%
4	2.25%	100	0.913225	2.2953%	4.6674%
5	2.75%	100	0.870016	2.8240%	4.9664%

Note: All calculations in this and subsequent exhibits were completed on a spreadsheet; rounded results are reported in the text.

The discount factors and spot rates are bootstrapped using the cash flows on the underlying benchmark bonds in this sequence of equations:

$100 = (100 - 0.25) \times DF_1.$

$DF_1 = 1.002506.$

$100 = (0.75 \times 1.002506) + (100.75 \times DF_2).$

$DF_2 = 0.985093.$

$100 = (1.50 \times 1.002506) + (1.50 \times 0.985093) + (101.50 \times DF_3).$

$DF_3 = 0.955848.$

$100 = (2.25 \times 1.002506) + (2.25 \times 0.985093) + (2.25 \times 0.955848) + (102.25 \times DF_4)$.

$DF_4 = 0.913225$.

$100 = (2.75 \times 1.002506) + (2.75 \times 0.985093) + (2.75 \times 0.955848) + (2.75 \times 0.913225) + (102.75 \times DF_5)$.

$DF_5 = 0.870016$.

The spot (i.e., implied zero-coupon) rates can be calculated from the discount factors; for instance, the two-year spot rate is 0.7538% and the four-year spot rate is 2.2953%:

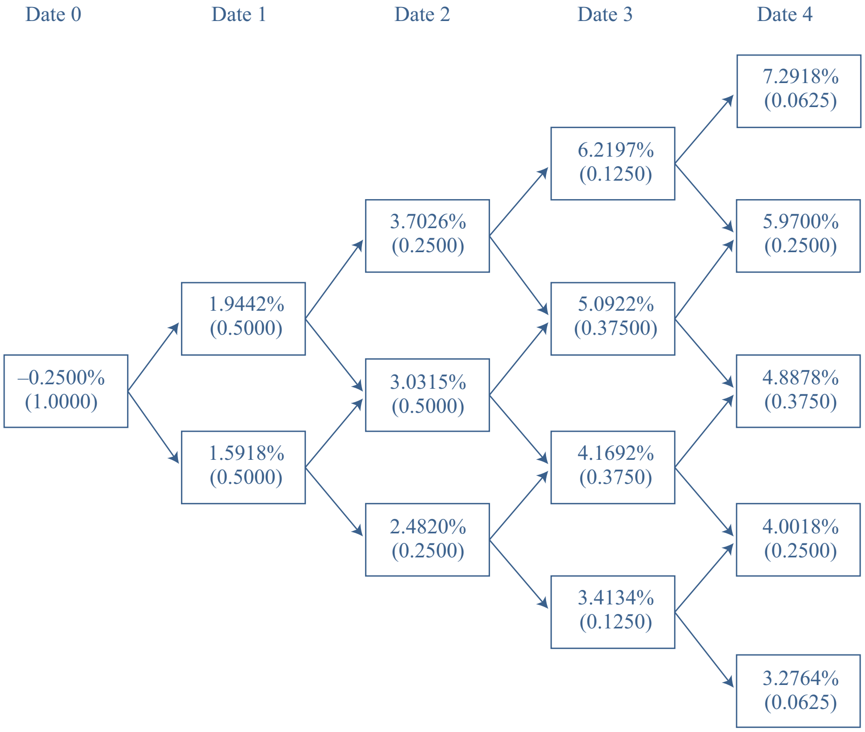
$$\left(\frac{1}{0.985093}\right)^{1/2} - 1 = 0.007538.$$

$$\left(\frac{1}{0.913225}\right)^{1/4} - 1 = 0.022953.$$

The forward rates are calculated as the ratios of the discount factors. The one-year forward rate two years into the future is 3.0596%: 0.985093/0.955848 – 1 = 0.030596. The one-year forward rate four years into the future is 4.9665%: 0.913225/0.870016 – 1 = 0.049665.

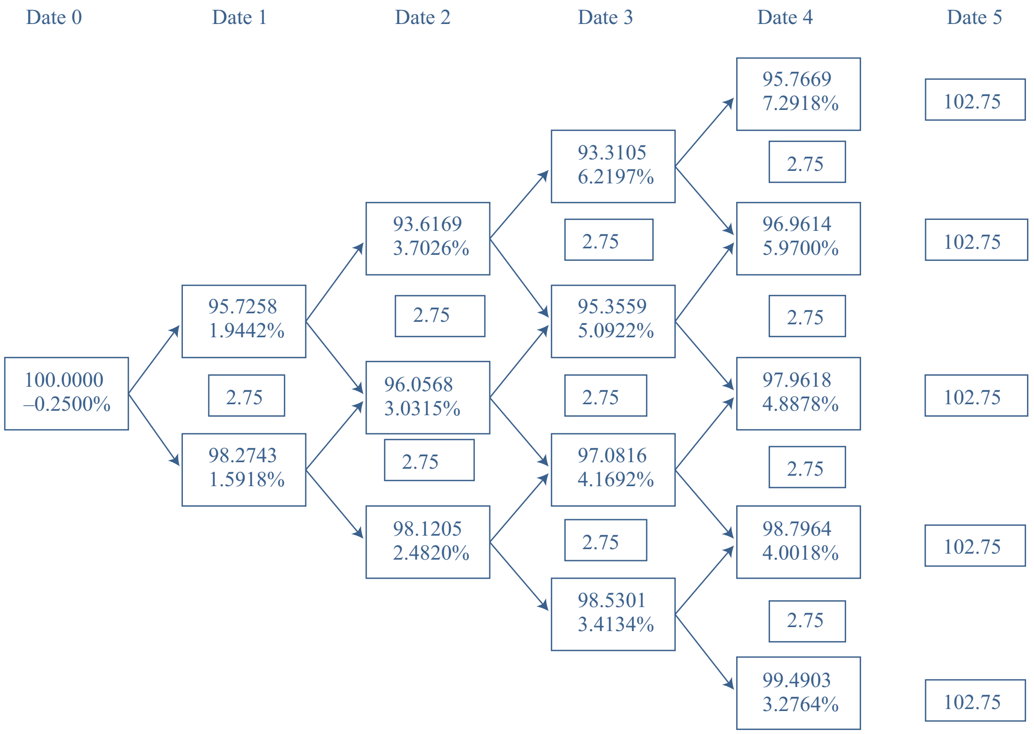
Following the methodology detailed in the "Arbitrage-Free Valuation Framework" topic, we build a binomial interest rate tree for one-year forward rates consistent with the pricing of the benchmark government bonds and an assumption of future interest rate volatility. Here we assume 10% volatility. The resulting binomial interest rate tree is presented in Exhibit 10. Below each rate is the probability of attaining that node in the tree. The current (Date 0) one-year rate of –0.25% will rise to 1.9442% or "fall" to 1.5918% by the end of the year (Date 1) with equal probability. On Date 2, at the end of the second year, the one-year rate will be 3.7026%, 3.0315%, or 2.4820% with probabilities of 0.25, 0.50, and 0.25, respectively. On Date 4, the forward rate will fall within the range of a high of 7.2918% to a low of 3.2764%. For each date, the possible rates are spread out around the forward rates shown in Exhibit 9.

Exhibit 10 One-Year Binomial Interest Rate Tree for 10% Volatility



To demonstrate that this is an arbitrage-free binomial interest rate tree, we calculate the Date 0 value of a 2.75% annual payment government bond. We know from Exhibit 9 that this bond is priced at par value. Exhibit 11 shows that the Date 0 value is indeed 100.0000. Notice that the scheduled year-end coupon and principal payments are placed to the right of each forward rate in the tree.

Exhibit 11 Valuation of a 2.75% Annual Payment Government Bond



These are the five Date 4 values for the government bond, shown above the interest rate at each node:

$$102.75/1.072918 = 95.7669.$$
$$102.75/1.059700 = 96.9614.$$
$$102.75/1.048878 = 97.9618.$$
$$102.75/1.040018 = 98.7964.$$
$$102.75/1.032764 = 99.4903.$$

These are the four Date 3 values:

$$\frac{\left[(0.5 \times 95.7669) + (0.5 \times 96.9614)\right] + 2.75}{1.062197} = 93.3105.$$

$$\frac{\left[(0.5 \times 96.9614) + (0.5 \times 97.9618)\right] + 2.75}{1.050922} = 95.3559.$$

$$\frac{\left[(0.5 \times 97.9618) + (0.5 \times 98.7964)\right] + 2.75}{1.041692} = 97.0816.$$

$$\frac{\left[(0.5 \times 98.7964) + (0.5 \times 99.4903)\right] + 2.75}{1.034134} = 98.5301.$$

Continuing with backward induction, the Date 0 value turns out to be 100.0000, confirming that the binomial interest rate tree has been correctly calibrated.

Now consider a five-year, 3.50% annual payment corporate bond. A fixed-income analyst assigns an annual default probability of 1.25% and a recovery rate of 40% to this bond and assumes 10% volatility in benchmark interest rates. The problem at hand for the analyst is to assess the fair value for the bond under these assumptions. This is done in two steps:

- First, determine the value for the corporate bond assuming no default (VND).
- Second, calculate the credit valuation adjustment.

The fair value of the bond is the VND minus the CVA.

The binomial interest rate tree for the benchmark rates in Exhibit 10 can be used to calculate the VND for the bond. Exhibit 12 shows that the VND is 103.5450 per 100 of par value. This number could also have been obtained more directly by using the benchmark discount factors:

$$(3.50 \times 1.002506) + (3.50 \times 0.985093) + (3.50 \times 0.955848) + (3.50 \times 0.913225) + (103.50 \times 0.870016) = 103.5450.$$

The advantage of using the binomial interest rate tree to get the VND is that the same tree is used to calculate the expected exposure to default loss, which is a key element in the credit risk model.

Exhibit 12 Value of a 3.50% Annual Payment Corporate Bond Assuming No Default

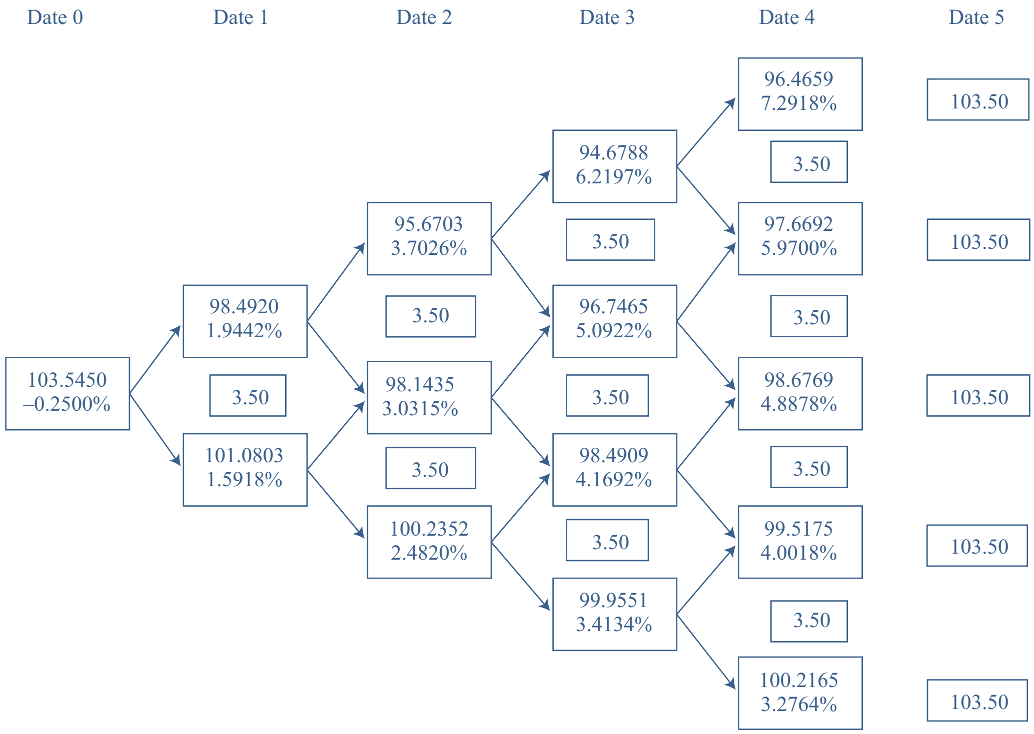


Exhibit 13 shows that the credit valuation adjustment to the value assuming no default is 3.5394 per 100 of par value. The expected exposure for Date 4 is 102.0931, calculated using the bond values at each node, the probability of attaining the node, and the coupon payment:

$$(0.0625 \times 96.4659) + (0.25 \times 97.6692) + (0.375 \times 98.6769) + (0.25 \times 99.5175) + (0.0625 \times 100.2165) + 3.50 = 102.0931.$$

(Note again that all calculations are done on a spreadsheet to maintain precision; only the rounded results are reported in the text.) The loss given default for Date 4 is 61.2559 [= 102.0931 × (1 – 0.40)] because the assumed recovery rate is 40% of the exposure. The probability of default at Date 4 is 1.2037%, assuming no prior default. This is based on the probability of survival into the fourth year. It is calculated as

$$1.25\% \times (100\% - 1.25\%)^3 = 1.2037\%.$$

The probability of survival after Date 3 is $(100\% - 1.25\%)^3$, and the probability of default on Date 4 is 1.25%. The product of the LGD and the POD is the expected loss. The present value of the expected loss, 0.6734, is the contribution to total CVA for Date 4. The sum of the CVAs for each year is the overall CVA.

Exhibit 13 Credit Valuation Adjustment for the 3.50% Annual Payment Corporate Bond

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	103.2862	61.9717	1.2500%	1.002506	0.7766
2	101.5481	60.9289	1.2344%	0.985093	0.7409
3	101.0433	60.6260	1.2189%	0.955848	0.7064
4	102.0931	61.2559	1.2037%	0.913225	0.6734
5	103.5000	62.1000	1.1887%	0.870016	0.6422
			6.0957%	CVA =	3.5394

The fixed-income analyst concludes that the fair value of the corporate bond is 100.0056 per 100 of par value: 103.5450 – 3.5394 = 100.0056. Depending on the current market price for the bond, the analyst might recommend a buy or sell decision.

The yield to maturity (YTM) for the corporate bond given a fair value of 100.0056 is 3.4988%:

$$100.0056 = \frac{3.50}{(1+\text{YTM})^1} + \frac{3.50}{(1+\text{YTM})^2} + \frac{3.50}{(1+\text{YTM})^3} + \frac{3.50}{(1+\text{YTM})^4} + \frac{103.50}{(1+\text{YTM})^5}.$$

$$\text{YTM} = 0.034988.$$

The five-year par yield for the government bond in Exhibit 9 is 2.75%. Therefore, the credit spread over the benchmark bond is 0.7488% (= 3.4988% – 2.75%). In practice, the credit spread is typically measured against the actual yield on the comparable-maturity government bond, which might be trading at a premium or a discount.

We can say that the credit risk on this corporate bond is captured by a CVA of 3.5394 per 100 in par value as of Date 0 or as an annual spread of 74.88 bps per year for five years. This conclusion, however, assumes that the observed credit spread is based entirely on credit risk. In fact, there usually are liquidity and tax differences between government and corporate bonds. Those differences are neglected in this analysis to focus on credit risk. Stated differently, the liquidity and tax differences are represented in the credit spread.

EXAMPLE 6

Using Credit Analysis in Decision Making

Lori Boller is a fixed-income money manager specializing in taking long positions on high-yield corporate bonds that she deems to be undervalued. In particular, she looks for bonds for which the credit spread over government securities appears to indicate too high a probability of default or too low a recovery rate if default were to occur. Currently, she is looking at a three-year, 4.00% annual payment bond that is priced at 104 (per 100 of par value). In her opinion, this bond should be priced to reflect an annual default probability of 2.25% given a recovery rate of 40%. Ms. Boller is comfortable with an assumption of 10% volatility in government bond yields over the next few years. Should she consider buying this bond for her portfolio? Use the government par curve in Exhibit 9 and the binomial interest rate tree in Exhibit 10 in the solution.

Solution:

Ms. Boller needs to calculate the fair value of the three-year, 4% annual payment corporate bond given her assumptions about the credit risk parameters. The results are shown in Exhibit 14.

Exhibit 14 Fair Value of the Three-Year, 4% Annual Payment Corporate Bond

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	107.0902	64.2541	2.2500%	1.002506	1.4493
2	104.9120	62.9472	2.1994%	0.985093	1.3638
3	104.0000	62.4000	2.1499%	0.955848	1.2823
			6.5993%	CVA =	4.0954

The VND for the bond is 107.3586. The calculations for the bond values in the binomial interest rate tree are as follows:

104/1.037026 = 100.2868.

104/1.030315 = 100.9400.

104/1.024820 = 101.4812.

$$\frac{(0.5 \times 100.2868) + (0.5 \times 100.9400) + 4}{1.019442} = 102.6183.$$

$$\frac{(0.5 \times 100.9400) + (0.5 \times 101.4812) + 4}{1.015918} = 103.5621.$$

$$\frac{(0.5 \times 102.6183) + (0.5 \times 103.5621) + 4}{0.997500} = 107.3586.$$

The CVA for the bond is 4.0954 given the assumption of an annual default probability of 2.25% and a recovery rate of 40% of the expected exposure. The following are calculations for the Date 1 and Date 2 expected exposures:

(0.50 × 102.6183) + (0.50 × 103.5621) + 4 = 107.0902.

$$(0.25 \times 100.2868) + (0.50 \times 100.9400) + (0.25 \times 101.4812) + 4 = 104.9120.$$

The calculations for the LGD are as follows:

$$107.0902 \times (1 - 0.40) = 64.2541.$$

$$104.9120 \times (1 - 0.40) = 62.9472.$$

$$104 \times (1 - 0.40) = 62.4000.$$

The following are calculations for the POD for Date 2 and Date 3:

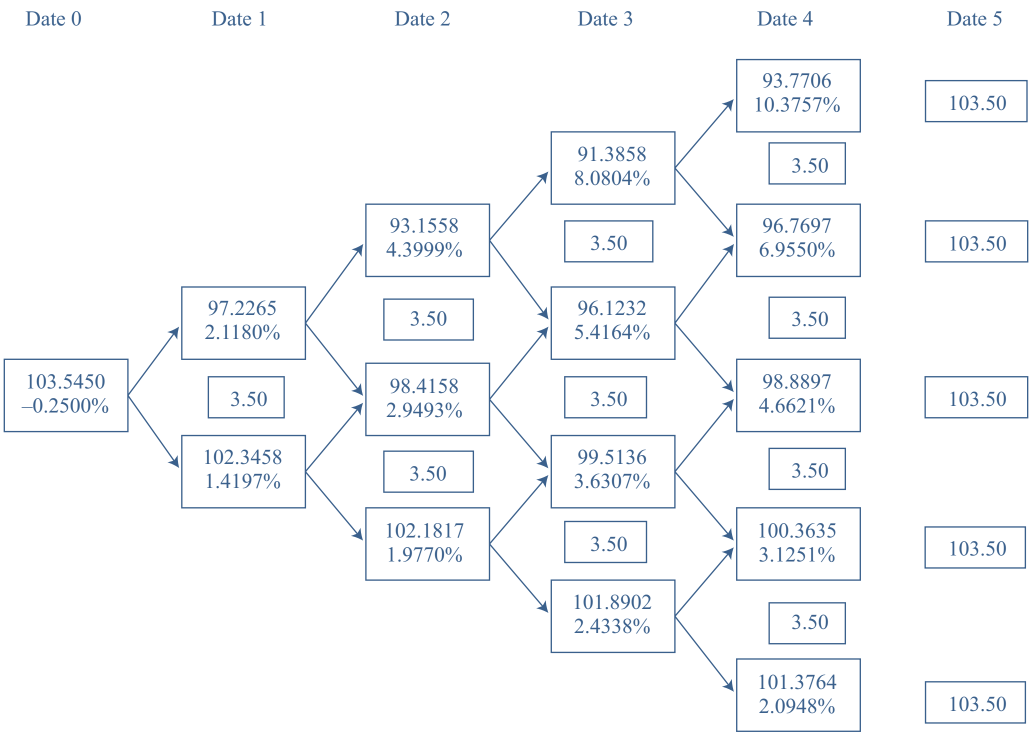
$$2.25\% \times (100\% - 2.25\%) = 2.1994\%.$$

$$2.25\% \times (100\% - 2.25\%)^2 = 2.1499\%.$$

Ms. Boller determines, on the basis of her assumed credit risk parameters, that the fair value for the high-yield corporate bond is 103.2632 (= 107.3586 – 4.0954). Given that the bond is trading at 104, she would likely decline to purchase because in her opinion the bond is overvalued.

A change in the assumed level of interest rate volatility can be shown to have a small impact on the fair value of the corporate bond. Usually the effect of a change in volatility is demonstrated with a bond having an embedded option, such as a callable or putable bond. Here we see an impact of the calculation of CVA on a bond having no embedded options. This is illustrated with Exhibits 15 and 16, which use a no-arbitrage binomial interest rate tree for 20% volatility to value the five-year, 3.50% annual payment corporate bond using the same credit risk parameters as in the previous calculations.

Exhibit 15 VND Calculation for the 3.50% Corporate Bond Assuming No Default and 20% Volatility



Notice in Exhibit 15 that with 20% volatility, the range in forward rates for each date is now wider. With 10% volatility, the Date 4 rates go from a low of 3.2764% to a high of 7.2918%. Now, with 20% volatility, the range is from 2.0948% to 10.3757%. The key point is that changing all the bond values still results in a VND of 103.5450. This confirms that the tree has been correctly calibrated and that the assumed level of future interest rate volatility has no impact on the value of a default-risk-free government bond. Changes in the fair value of a corporate bond arising from a change in the assumed rate volatility occur only when there are embedded options and, as demonstrated in Exhibit 16, when there is credit risk.

Exhibit 16 CVA Calculation for the 3.50% Corporate Bond Assuming 20% Volatility

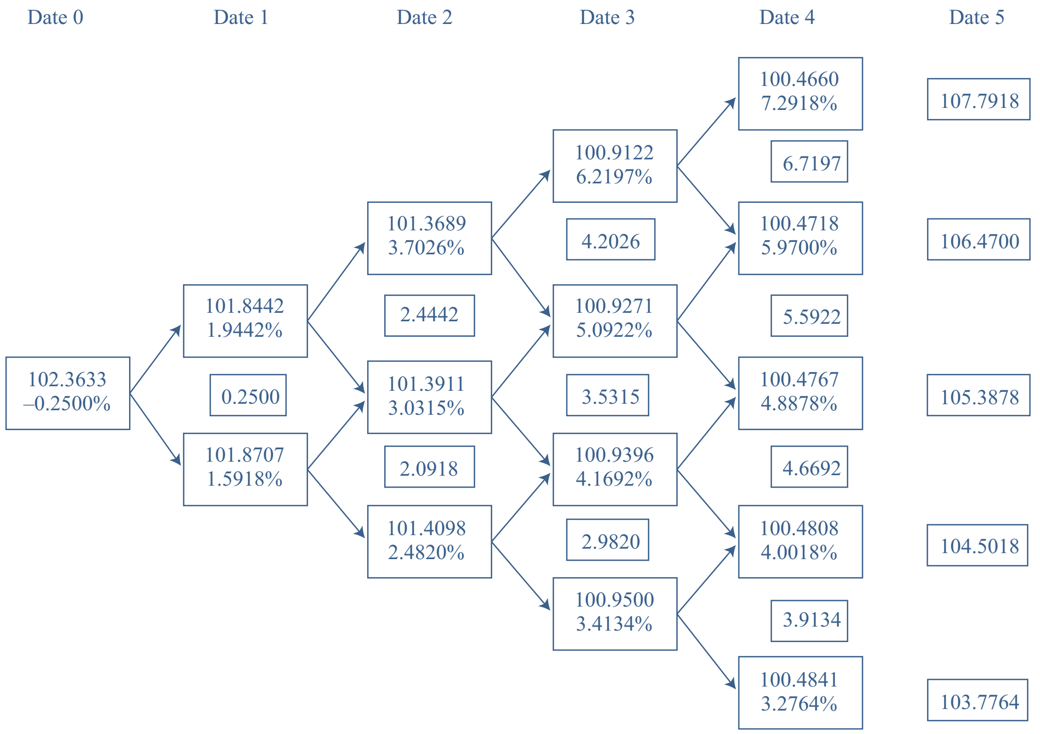
Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	103.2862	61.9717	1.2500%	1.002506	0.7766
2	101.5423	60.9254	1.2344%	0.985093	0.7408
3	101.0233	60.6140	1.2189%	0.955848	0.7062
4	102.0636	61.2382	1.2037%	0.913225	0.6732
5	103.5000	62.1000	1.1887%	0.870016	0.6422
			6.0957%	CVA =	3.5390

Exhibit 16 presents the table to calculate the CVA for 20% volatility. The expected exposures to default loss are slightly lower for Dates 2, 3, and 4 compared with Exhibit 13 for 10% volatility. These small changes feed through the table, reducing the loss given default and the contribution to total CVA for those dates. Overall, the CVA is 3.5390 per 100 of par value. The fair value of the bond is now slightly higher at 100.0060 (= 103.5450 – 3.5390), compared with the value for 10% volatility of 100.0056 (= 103.5450 – 3.5394).

The reason for the small volatility impact on the fair value is the asymmetry in the forward rates produced by the lognormality assumption in the interest rate model. In building the tree, rates are spread out around the implied forward rate for each date—more so the greater the given level of volatility. However, the range is not symmetric about the implied forward rate. For example, the one-year forward rate four years into the future is 4.9665% in Exhibit 9. With 20% volatility, the Date 4 rate at the top of the tree is higher by 5.4092% (= 10.3757% – 4.9665%), while the rate at the bottom of the tree is lower by 2.8717% (= 4.9665% – 2.0948%). The net effect is to reduce the expected exposure to default loss. The top of the tree shows less potential loss because the current value of the bond is lower, which more than offsets the greater exposure to loss at the bottom of the tree.

The arbitrage-free framework can be adapted to value a risky floating-rate note. Consider a five-year "floater" that pays annually the one-year benchmark rate plus 0.50%. This 50 bp addition to the index rate is called the *quoted margin* and typically is fixed over the lifetime of the security. Exhibit 17 demonstrates that the VND for the floater is 102.3633 per 100 of par value, using the binomial interest rate tree for 10% interest rate volatility. Notice that the interest payment is "in arrears," meaning that the rate is set at the beginning of the period and paid at the end of the period. That is why the interest payments set to the right of each rate vary depending on the realized rate in the tree. The interest payment for Date 1 is 0.25 because the Date 0 reference rate is –0.25%: (–0.25% + 0.50%) × 100 = 0.25. The final payment on Date 5 when the floater matures is 105.3878 if the one-year rate is 4.8878% on Date 4: (4.8878% + 0.50%) × 100 + 100 = 105.3878.

Exhibit 17 Value of a Floating-Rate Note Paying the Benchmark Rate Plus 0.50% Assuming No Default and 10% Volatility



Notice that the bond values for each date are very similar for the various forward rates. That, of course, is the intent of a floating-rate note. The bond values would all be exactly 100.0000 if the note paid the benchmark rate "flat," meaning a quoted margin of zero. The VND of 102.3633 is obtained via backward induction (i.e., beginning at maturity and working backward in time). The following are the calculations for the bond values for Date 4:

107.7918/1.072918 = 100.4660.

106.4700/1.059700 = 100.4718.

105.3878/1.048878 = 100.4767.

104.5018/1.040018 = 100.4808.

103.7764/1.032764 = 100.4841.

These are the calculations for Date 3:

$$\frac{(0.5 \times 100.4660) + (0.5 \times 100.4718) + 6.7197}{1.062197} = 100.9122.$$

$$\frac{(0.50 \times 100.4718) + (0.5 \times 100.4767) + 5.5922}{1.050922} = 100.9271.$$

$$\frac{(0.5 \times 100.4767) + (0.5 \times 100.4808) + 4.6692}{1.041692} = 100.9396.$$

$$\frac{(0.5 \times 100.4808) + (0.5 \times 100.4841) + 3.9134}{1.034134} = 100.9500.$$

These are the calculations for the bond values for Date 2:

$$\frac{(0.5 \times 100.9122) + (0.5 \times 100.9271) + 4.2026}{1.037026} = 101.3689.$$

$$\frac{(0.5 \times 100.9271) + (0.5 \times 100.9396) + 3.5315}{1.030315} = 101.3911.$$

$$\frac{(0.5 \times 100.9396) + (0.5 \times 100.9500) + 2.9820}{1.024820} = 101.4098.$$

These are the calculations for the bond values for Date 1 and Date 0:

$$\frac{(0.5 \times 101.3689) + (0.5 \times 101.3911) + 2.4442}{1.019442} = 101.8442.$$

$$\frac{(0.5 \times 101.3911) + (0.5 \times 101.4098) + 2.0918}{1.015918} = 101.8707.$$

$$\frac{(0.5 \times 101.8442) + (0.5 \times 101.8707) + 0.2500}{0.997500} = 102.3633.$$

Exhibit 18 shows the credit risk table for the floating-rate note. For this example, we assume that for the first three years, the annual default probability is 0.50% and the recovery rate is 20%. The credit risk of the issuer then worsens: For the final two years, the annual probability of default goes up to 0.75% and the recovery rate goes down to 10%. This is an example in which the assumed annual default probability changes over the lifetime of the bond.

Exhibit 18 CVA Calculation for the Value of a Floating-Rate Note Paying the Benchmark Rate Plus 0.50%

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	102.1074	81.6859	0.5000%	1.002506	0.4095
2	103.6583	82.9266	0.4975%	0.985093	0.4064
3	104.4947	83.5957	0.4950%	0.955848	0.3955
4	105.6535	95.0881	0.7388%	0.913225	0.6416
5	105.4864	94.9377	0.7333%	0.870016	0.6057
			2.9646%	CVA =	2.4586

Note: Credit risk parameter assumptions: for Dates 1–3, annual default probability = 0.50% and recovery rate = 20%; for Dates 4–5, annual default probability = 0.75% and recovery rate = 10%.

The calculation for the expected exposure recognizes that the bond values for each date follow the probabilities of attaining those rates, whereas possible interest payments use the probabilities for the prior date. For example, the expected exposure to default loss for Date 4 is 105.6535:

$$\begin{bmatrix}(0.0625 \times 100.4660) + (0.25 \times 100.4718) + (0.375 \times 100.4767) \\ + (0.25 \times 100.4808) + (0.0625 \times 100.4841)\end{bmatrix}$$
$$+ \left[(0.125 \times 6.7197) + (0.375 \times 5.5922) + (0.375 \times 4.6692) + (0.125 \times 3.9134)\right]$$
$$= 105.6535.$$

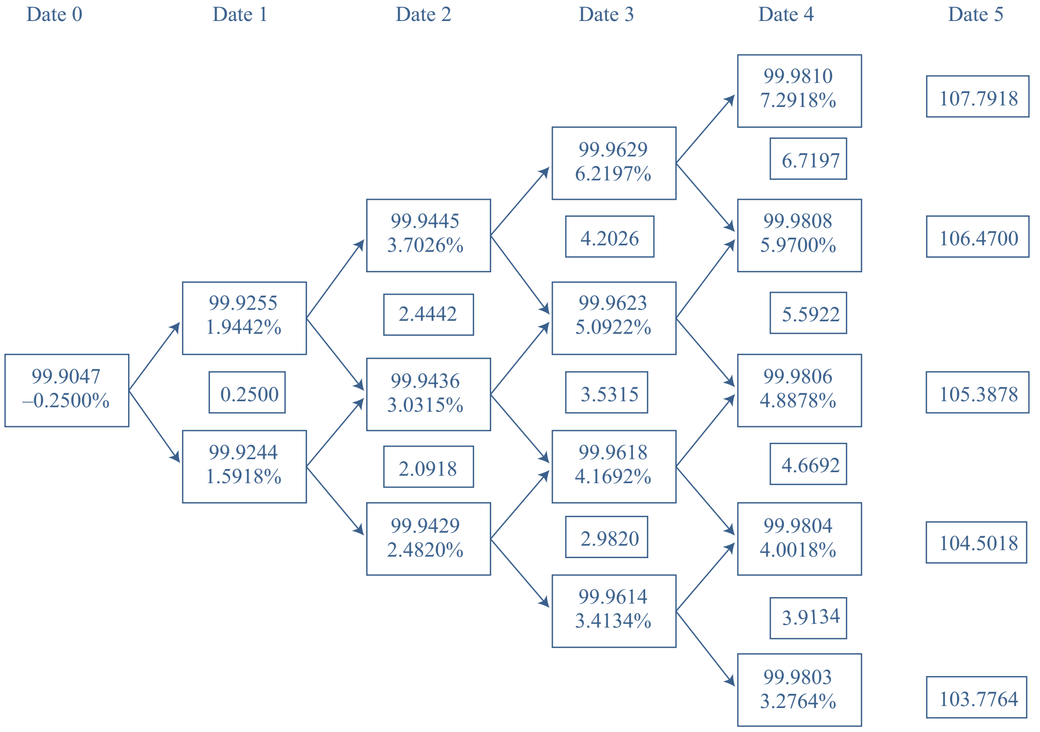
The first term in brackets is the expected bond value using the Date 4 probabilities for each of the five possible rates. The second term is the expected interest payment using the Date 3 probabilities for each of the four possible rates.

The expected LGD for Date 2 is 82.9266 [= 103.6583 × (1 – 0.20)]; for Date 4, it is 95.0881 [= 105.6535 × (1 – 0.10)]. The PODs in Exhibit 18 reflect the probability of default for each year. For Date 2, the POD is 0.4975%, conditional on no default on Date 1: 0.50% × (100% – 0.50%) = 0.4975%. For Date 3, the POD is 0.4950%: 0.50% × (100% – 0.50%)^2 = 0.4950%. The probability of survival into the fourth year is 98.5075%: (100% – 0.50%)^3 = 98.5075%. Therefore, the POD for Date 4 increases to 0.7388% because of the assumed worsening credit risk: 0.75% × 98.5075% = 0.7388%. The probability of survival into the fifth year is 97.7687% (= 98.5075% – 0.7388%). The POD for Date 5 is 0.7333% (= 0.75% × 97.7687%). The cumulative probability of default over the lifetime of the floater is 2.9646%.

Given these assumptions about credit risk, the CVA for the floater is 2.4586. The fair value is 99.9047, the VND of 102.3633 minus the CVA. Because the security is priced below par value, its *discount margin* (DM) must be higher than the quoted margin of 0.50%. The discount margin for a floating-rate note is a yield measure commonly used on floating-rate notes in the same manner that the credit spread is used with fixed-rate bonds.

The arbitrage-free framework can be used to determine the DM for this floater by trial-and-error search (or GoalSeek or Solver in Excel). We add a trial DM to benchmark rates that are used to get the bond values at each node in the tree. Then the trial DM is changed until the Date 0 value matches the fair value of 99.9047. Exhibit 19 shows that the DM for this floater is 0.52046%, slightly above the quoted margin because the security is priced at a small discount below par value.

Exhibit 19 The Discount Margin for the Floating-Rate Note Paying the Benchmark Rate Plus 0.50%, Assuming 10% Volatility



These are the calculations for the bond values for Date 2:

$$\frac{(0.5 \times 99.9629) + (0.5 \times 99.9623) + 4.2026}{1 + 0.037026 + 0.0052046} = 99.9445.$$

$$\frac{(0.5 \times 99.9623) + (0.5 \times 99.9618) + 3.5315}{1 + 0.030315 + 0.0052046} = 99.9436.$$

$$\frac{(0.5 \times 99.9618) + (0.5 \times 99.9614) + 2.9820}{1 + 0.024820 + 0.0052046} = 99.9429.$$

Throughout the binomial interest rate tree, the assumed DM is added to the benchmark rate to factor in credit risk. After a trial-and-error search, a DM of 0.52046% gives the same Date 0 value for the floating-rate note of 99.9047 as is obtained with the VND and CVA models.

EXAMPLE 7

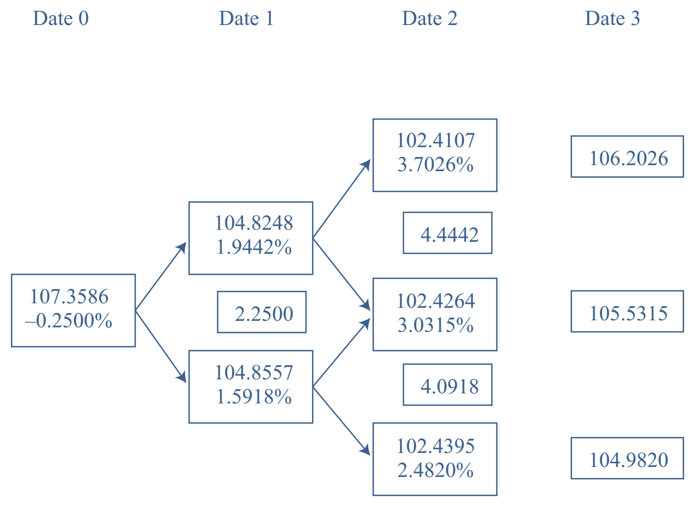
Evaluating a Floating-Rate Note

Omar Yassin is an experienced credit analyst at a fixed-income investment firm. His current assignment is to assess potential purchases of distressed high-yield corporate bonds. One intriguing prospect is a three-year, annual payment floating-rate note paying the one-year benchmark rate plus 2.50%. The floater is rated CCC and is priced at 84 per 100 of par value. Based on various research reports on and prices of the issuer's credit default swaps, Mr. Yassin believes the probability of default in the next year is about 30%. If the issuer goes into bankruptcy at any time, he expects the recovery rate to be at least 50%; it could be as high as 60% because of some valuable real estate holdings. He further believes that if the issuer is able to survive this next year, the default probability for the remaining two years will be only about 10% for each year. Based on these assumptions about the credit risk parameters and an expectation of 10% volatility for interest rates, should Mr. Yassin recommend purchasing the floating-rate note?

Solution:

Mr. Yassin calculates the fair value of the three-year, annual payment floating-rate note given his assumptions about the default probabilities and the recovery rate ranging between 50% and 60%. The results are shown in Exhibit 20.

Exhibit 20 Fair Value of the Three-Year, Annual Payment Floating-Rate Note Paying the One-Year Rate Plus 2.50%



Assumed 50% Recovery Rate

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	107.0902	53.5451	30.0000%	1.002506	16.1038
2	106.6938	53.3469	7.0000%	0.985093	3.6786
3	105.5619	52.7810	6.3000%	0.955848	3.1784
			43.3000%	CVA =	22.9608

Fair value = 107.3586 – 22.9608 = 84.3978.

Assumed 60% Recovery Rate

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	107.0902	42.8361	30.0000%	1.002506	12.8830

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
2	106.6938	42.6775	7.0000%	0.985093	2.9429
3	105.5619	42.2248	6.3000%	0.955848	2.5427
			43.3000%	CVA =	18.3686

Fair value = 107.3586 – 18.3686 = 88.9900.

Each projected interest payment in the tree is the benchmark rate at the beginning of the year plus 2.50% times 100. The rate is –0.25% on Date 0; the "in-arrears" interest payment on Date 1 is 2.2500 [= (–0.25% + 2.50%) × 100]. If the rate is 2.4820% on Date 2, the payment at maturity on Date 3 is 104.9820 [= (2.4820% + 2.50%) × 100 + 100].

The VND for the floater is 107.3586. The calculations for the bond values in the binomial interest rate tree are as follows:

106.2026/1.037026 = 102.4107.

105.5315/1.030315 = 102.4264.

104.9820/1.024820 = 102.4395.

$$\frac{(0.5 \times 102.4107) + (0.5 \times 102.4264) + 4.4442}{1.019442} = 104.8248.$$

$$\frac{(0.5 \times 102.4264) + (0.5 \times 102.4395) + 4.0918}{1.015918} = 104.8557.$$

$$\frac{(0.5 \times 104.8248) + (0.5 \times 104.8557) + 2.2500}{0.997500} = 107.3586.$$

These are the calculations for the expected exposures to default loss:

(0.5 × 104.8248) + (0.5 × 104.8557) + 2.2500 = 107.0902.

(0.25 × 102.4107) + (0.5 × 102.4264) + (0.25 × 102.4395) + (0.5 × 4.4442) + (0.5 × 4.0918) =106.6938.

(0.25 × 106.2026) + (0.5 × 105.5315) + (0.25 × 104.9820) = 105.5619.

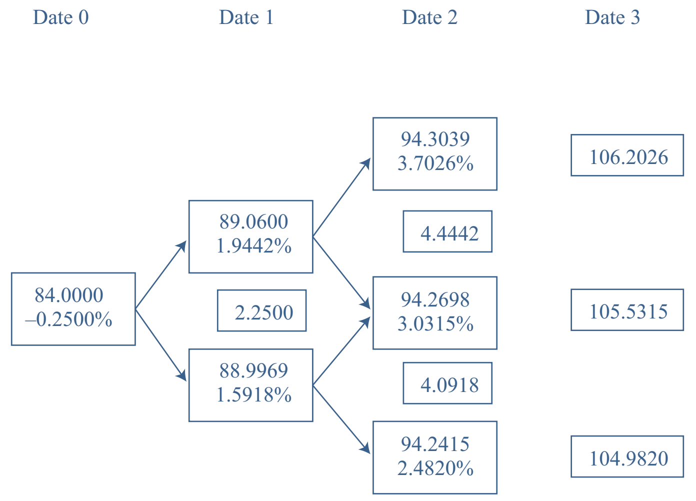
The assumed default probability for the first year is 30%. The POD for Date 2 is 7.00%, which is the probability of survival into the second year, 70%, times the 10% probability of default. The probability of survival into the third year is 63% (= 70% – 7%); the POD for Date 3 is 6.30% (= 10% × 63%).

The decision to consider purchase of the floating-rate note comes down to the assumption about recovery. Exhibit 20 first shows the results for 50% recovery of the expected exposure. The LGD on Date 2 is 53.3469 [= 106.6938 × (1 – 0.50)]. The overall CVA is 22.9608, giving a fair value of 84.3978 (= 107.3586 – 22.9608). Exhibit 20 next shows the results for 60% recovery. With this assumption, the LGD for Date 2 is just 42.6775 [= 106.6938 × (1 – 0.60)]. Stronger recovery reduces the overall CVA to 18.3686. The fair value for the floater is now 88.9900.

Mr. Yassin should recommend purchasing the distressed floating-rate note. Although there is a significant 43.3% probability of default at some point over the three years, the security appears to be fairly priced at 84 given a recovery rate of 50%. At 60% recovery, it is significantly undervalued.

In addition, there is still a 57.7% (= 100% – 43.3%) chance of no default. Exhibit 21 shows the calculation for the discount margin, which is a measure of the return to the investor assuming no default (like a yield to maturity on a fixed-rate bond). Found by a trial-and-error search, the DM is 8.9148%, considerably higher than the quoted margin because the floater is priced at a deep discount.

Exhibit 21 Discount Margin on the Three-Year, Annual Payment Floating-Rate Note Paying the One-Year Rate Plus 2.50%



These are the calculations for the bond values for Date 1 and Date 0:

$$\frac{(0.5 \times 94.3039) + (0.5 \times 94.2698) + 4.4442}{1 + 0.019442 + 0.089148} = 89.0600.$$

$$\frac{(0.5 \times 94.2698) + (0.5 \times 94.2415) + 4.0918}{1 + 0.015918 + 0.089148} = 88.9969.$$

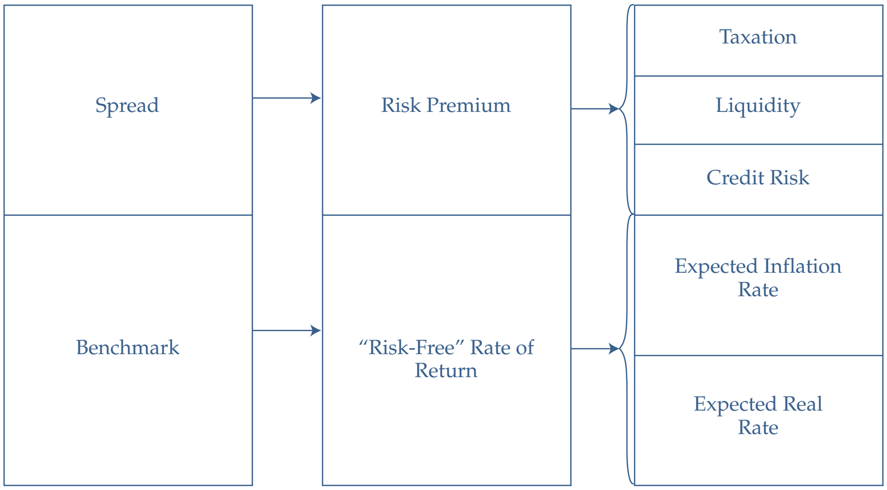
$$\frac{(0.5 \times 89.0600) + (0.5 \times 88.9969) + 2.2500}{1 - 0.0025 + 0.089148} = 84.0000.$$

6 INTERPRETING CHANGES IN CREDIT SPREADS

f interpret changes in a credit spread

Corporate and benchmark bond yields and the credit spread between them change from day to day. The challenge for a fixed-income analyst is to understand that and be able to explain *why* the yields and spreads change. Exhibit 22 offers a breakdown of the main components of bond yields. Benchmark bond yields, in general, capture the *macroeconomic* factors affecting all debt securities. These are the expected inflation rate and the expected real rate of return. Risk-averse investors in benchmark bonds also might require compensation for uncertainty regarding those variables.

Exhibit 22 Components of a Corporate Bond Yield



The spread over the benchmark bond yield captures the *microeconomic* factors that pertain to the corporate issuer and the specific issue itself. The chief microeconomic factor is the expected loss due to default. There also are liquidity and tax differences between the corporate and benchmark bonds. Moreover, it can be difficult to separate these factors. Securities for which it becomes more difficult for analysts to assess a probability of default and a recovery rate typically become less liquid. Similarly, an uncertain tax status on a bond's gains and losses will increase the time and cost to estimate value. That makes the bond less liquid. Another factor in the observed spread between the corporate and benchmark bond yields can be compensation to risk-averse investors for uncertainty regarding credit risk, as well as liquidity and tax factors.

Research groups at major banks and consultancies have been working on models to better include counterparty credit risk, funding costs, and liquidity and taxation effects in the valuations of derivatives. First, a value is obtained using benchmark discount factors, in practice, derived from rates on overnight indexed swaps (OIS). These are interest rate swaps that reference an average daily interest rate. For instance, in the United States this daily rate is the effective federal funds rate. Then this OIS value, which is comparable to the VND in the previous section, is adjusted for the other factors. These valuation adjustments collectively are known as the XVA. The credit valuation adjustment is the most developed and most used in practice. Others include a funding valuation adjustment (FVA), a liquidity valuation adjustment (LVA), and a taxation valuation adjustment (TVA). In principle, the same ideas apply to debt securities in that these XVA comprise the observed spread between corporate and benchmark bond yields. For the purposes of our coverage, we focus only on the credit risk component, the CVA.

We can use the arbitrage-free framework and the credit risk model to examine the connections between the default probability, the recovery rate, and the credit spread. To be sure, this is a simple model to illustrate the much more complex models used in practice. These (which are called *XVA engines*) typically use Monte Carlo simulations for thousands of possible paths for interest rates. Our binomial interest rate tree has only 16 paths for the five years; it's a model of the actual model.

Consider again the five-year, 3.50% annual payment corporate bond examined earlier. In Exhibit 12, the value assuming no default was determined to be 103.5450 per 100 of par value. Now let us use the credit risk model to find the probabilities of default that would be consistent with various credit spreads and a recovery rate of 40%. Suppose, as in Exhibit 7, the credit spread for an AAA rated bond is 0.60%.

Using trial-and-error search, we find that an annual probability of default of 1.01% produces a 60 bp credit spread. The credit risk table is presented in Exhibit 23. Notice that the expected exposure to default loss and the loss given default are the same as in Exhibit 13. Only the default probabilities and the contributions to total CVA for each year change.

Exhibit 23 CVA Calculation for the 3.50% Corporate Bond Given a Default Probability of 1.01% and a Recovery Rate of 40%

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	103.2862	61.9717	1.0100%	1.002506	0.6275
2	101.5481	60.9289	0.9998%	0.985093	0.6001
3	101.0433	60.6260	0.9897%	0.955848	0.5735
4	102.0931	61.2559	0.9797%	0.913225	0.5481
5	103.5000	62.1000	0.9698%	0.870016	0.5240
			4.9490%	CVA =	2.8731

The CVA for the bond is 2.8731 per 100 of par. The fair value is 100.6719 (= 103.5450 – 2.8731). This gives a yield to maturity of 3.35%.

$$100.6719 = \frac{3.50}{(1 + \text{YTM})^1} + \frac{3.50}{(1 + \text{YTM})^2} + \frac{3.50}{(1 + \text{YTM})^3} + \frac{3.50}{(1 + \text{YTM})^4} + \frac{103.50}{(1 + \text{YTM})^5}.$$

$$\text{YTM} = 0.0335.$$

Given that the yield on the five-year benchmark bond is 2.75%, the credit spread is 0.60% (= 3.35% – 2.75%).

We can repeat this exercise for the other credit spreads and ratings shown in Exhibit 7. In each case, trial-and-error search is used to get the initial POD that corresponds to the CVA, the fair value, and the yield to maturity for each assumed spread. The results for the annual and cumulative default probabilities over the five years are shown in Exhibit 24.

Exhibit 24 Default Probabilities Consistent with Given Credit Ratings and Spreads and 40% Recovery

Credit Rating	Credit Spread	Annual Default Probability	Cumulative Default Probability
AAA	0.60%	1.01%	4.95%
AA	0.90%	1.49%	7.23%
A	1.10%	1.83%	8.82%
BBB	1.50%	2.48%	11.80%
BB	3.40%	5.64%	25.19%
B	6.50%	10.97%	44.07%
CCC, CC, C	9.50%	16.50%	59.41%

The default probabilities illustrated in Exhibit 24 might seem high, especially given the historical experience presented in Exhibit 6. Since 1995, no AAA rated company has defaulted; still, we model the likelihood to be over 1% for the first year and almost 5% for the next five years. However, as discussed earlier, these are *risk-neutral* probabilities of default and are higher than the actual probabilities because market prices reflect uncertainty over the timing of possible default. Investors are concerned about credit spread widening, especially if they do not intend to hold the bond to maturity. Credit rating migration from year to year, as illustrated in Exhibit 7, is a concern even for a high-quality investment-grade corporate bond. This is captured in the risk-neutral probability of default. Also, we must remember that observed credit spreads reflect more than just credit risk—there also are liquidity and tax differences. That further explains the difference between risk-neutral and actual default probabilities.

The relationship between the assumed recovery rate and the credit spread can be examined in the context of the credit risk model. Suppose that the five-year, 3.50% annual payment corporate bond has an initial probability of default of 1.83%. In Exhibit 24, we see that for a 40% recovery rate, the credit spread is 1.10%. What if the recovery rate is expected to be only 30%? Exhibit 25 shows the credit risk table for that assumption.

Exhibit 25 CVA Calculation for the 3.50% Corporate Bond Given a Default Probability of 1.83% and a Recovery Rate of 30%

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	103.2862	72.3003	1.8300%	1.002506	1.3264
2	101.5481	71.0837	1.7965%	0.985093	1.2580
3	101.0433	70.7303	1.7636%	0.955848	1.1923
4	102.0931	71.4652	1.7314%	0.913225	1.1300
5	103.5000	72.4500	1.6997%	0.870016	1.0714
			8.8212%	CVA =	5.9781

The reduction in the recovery rate from 40% to 30% has an impact on LGD and CVA for each year. The overall CVA is 5.9781 per 100 of par value. The fair value for the bond is 97.5670 (= 103.5450 – 5.9781), and the yield to maturity is 4.05%, giving a credit spread of 1.30% (= 4.05% – 2.75%).

$$97.5670 = \frac{3.50}{(1+\text{YTM})^1} + \frac{3.50}{(1+\text{YTM})^2} + \frac{3.50}{(1+\text{YTM})^3} + \frac{3.50}{(1+\text{YTM})^4} + \frac{103.50}{(1+\text{YTM})^5}.$$

$$\text{YTM} = 0.0405.$$

This example illustrates how a credit rating agency might use "notching" to combine the expected loss given default and the probability of default in setting the rating for a corporate bond. If the issuer were rated single A, associated with a default probability of 1.83% and a recovery rate of 40% on the company's senior unsecured debt, that debt might have a credit spread of 1.10%, comparable to other A rated companies. This particular bond is subordinated, leading analysts at the rating agency to believe that a lower recovery rate assumption of 30% is applicable. That could justify assigning a lower rating of A– or BBB+ on the subordinated debt, along with its 20 bp higher spread.

EXAMPLE 8

Evaluating Changes in Credit Risk Parameters

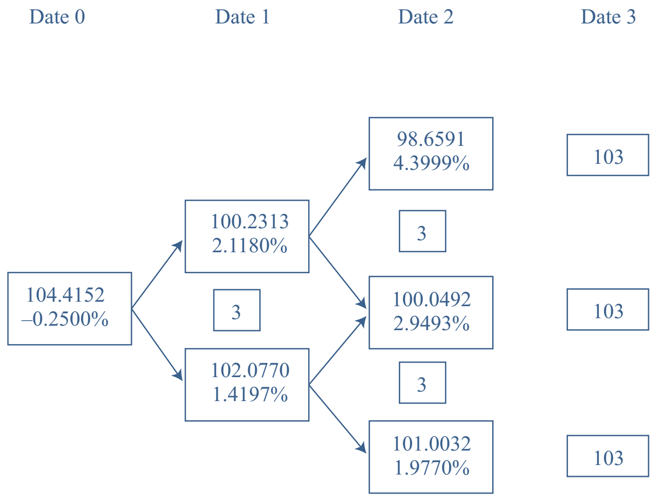
Edward Kapili is a summer intern working on a fixed-income trading desk at a major money-center bank. His supervisor asks him to value a three-year, 3% annual payment corporate bond using a binomial interest rate tree model for 20% volatility and the current par curve for benchmark government bonds. (This is the binomial tree in Exhibit 15.) The assumed annual probability of default is 1.50%, and the recovery rate is 40%.

The supervisor asks Mr. Kapili if the credit spread over the yield on the three-year benchmark bond, which is 1.50% in Exhibit 9, is likely to go up more if the default probability doubles to 3.00% or if the recovery rate halves to 20%. Mr. Kapili's intuition is that doubling the probability of default has a larger impact on the credit spread. Is his intuition correct?

Solution:

Mr. Kapili first determines the fair value of the three-year, 3% annual payment bond given the assumptions for the original credit risk parameters. The binomial interest rate tree and credit risk table are presented in Exhibit 26.

Exhibit 26 Fair Value of the Three-Year, 3% Annual Payment Corporate Bond Assuming 20% Volatility



Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	104.1541	62.4925	1.5000%	1.002506	0.9397
2	102.9402	61.7641	1.4775%	0.985093	0.8990
3	103.0000	61.8000	1.4553%	0.955848	0.8597
			4.4328%	CVA =	2.6984

Fair value = 104.4152 – 2.6984 = 101.7168.

The VND for the bond is 104.4152, the CVA is 2.6984, and the fair value is 101.7168 per 100 of par value. The yield to maturity is 2.40%, and the credit spread is 0.90% (= 2.40% – 1.50%).

$$101.7168 = \frac{3}{(1 + \text{YTM})^1} + \frac{3}{(1 + \text{YTM})^2} + \frac{103}{(1 + \text{YTM})^3}.$$

$$\text{YTM} = 0.0240.$$

Next, Mr. Kapili calculates the fair values under the new credit risk parameters, first for doubling the default probability and second for halving the recovery rate. These tables are shown in Exhibit 27.

Exhibit 27 Fair Value Calculations for Doubling the Default Probability and Halving the Recovery Rate

3.00% Default Probability, 40% Recovery Rate

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	104.1541	62.4925	3.0000%	1.002506	1.8795
2	102.9402	61.7641	2.9100%	0.985093	1.7705
3	103.0000	61.8000	2.8227%	0.955848	1.6674
			8.7327%	CVA =	5.3174

Fair value = 104.4152 – 5.3174 = 99.0978.

1.50% Default Probability, 20% Recovery Rate

Date	Expected Exposure	LGD	POD	Discount Factor	CVA per Year
0					
1	104.1541	83.3233	1.5000%	1.002506	1.2530
2	102.9402	82.3522	1.4775%	0.985093	1.1986
3	103.0000	82.4000	1.4553%	0.955848	1.1463
			4.4328%	CVA =	3.5978

Fair value = 104.4152 – 3.5978 = 100.8173.

The fair value of the corporate bond falls to 99.0978 when the default probability is raised to 3.00% and the recovery rate stays at 40%. The VND is the same, at 104.4152, and the CVA goes up to 5.3174. The yield to maturity increases to 3.32%, and the credit spread rises to 1.82% (= 3.32% – 1.50%).

$$99.0978 = \frac{3}{(1 + \text{YTM})^1} + \frac{3}{(1 + \text{YTM})^2} + \frac{103}{(1 + \text{YTM})^3}.$$

$$\text{YTM} = 0.0332.$$

The fair value of the corporate bond falls to 100.8173 when the recovery rate is reduced by half, to 20%, and the default probability is maintained at 1.50%. The VND is again the same, at 104.4152, and the CVA goes up to 3.5978. The yield to maturity increases to 2.71%, and the credit spread rises to 1.21% (= 2.71% – 1.50%).

$$100.8173 = \frac{3}{(1 + \text{YTM})^1} + \frac{3}{(1 + \text{YTM})^2} + \frac{103}{(1 + \text{YTM})^3}.$$

$$\text{YTM} = 0.0271.$$

Mr. Kapili's intuition is correct: Doubling the default probability has a greater impact on the credit spread than halving the recovery rate.

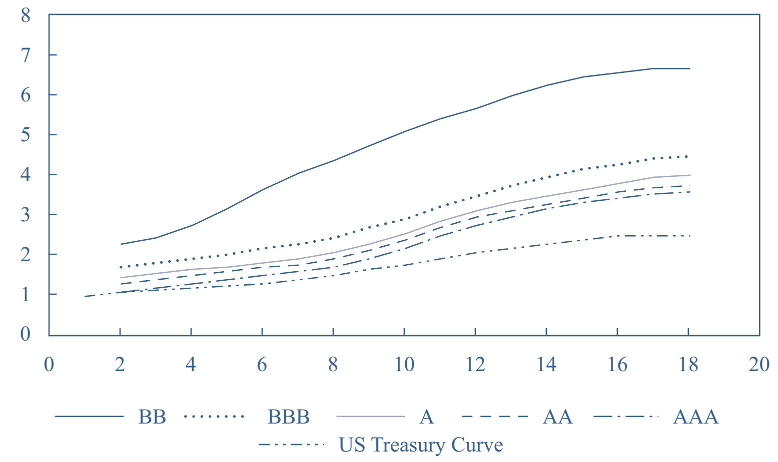
7 THE TERM STRUCTURE OF CREDIT SPREADS

g explain the determinants of the term structure of credit spreads and interpret a term structure of credit spreads

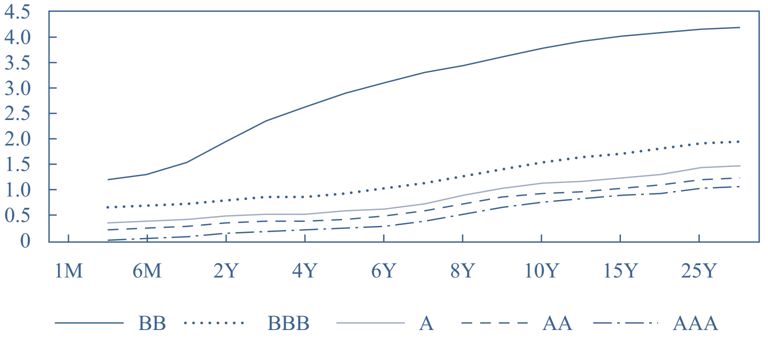
In the same way that the yield curve is composed of the interest rates on a single government issuer's debt across bond maturities, a credit curve shows the spread over a benchmark security for an issuer for outstanding fixed-income securities with shorter to longer maturities. For example, Exhibit 28 shows the relationship between US Treasury yields of a specific maturity and bonds rated AAA, AA, A, BBB, and BB. The total yields of the bonds are shown in Panel A, and spreads over the benchmark Treasury are shown in Panel B.

Exhibit 28 Composite Yield Graphs

A. Total Yields



B. Spreads



Source: Bloomberg.

The term structure of credit spreads is a useful gauge for issuers, underwriters, and investors in measuring the risk–return trade-off for a single issuer or a set of issuers across ratings and/or sectors across maturities. Issuers often work with their underwriter to consider the terms of a new issuance or a tender for existing debt based on relative credit spreads across maturities. For example, an investment-grade bond portfolio manager might use the existing credit curve for a particular issuer to determine a bid for a new primary debt issuance as well as to inform trading decisions for secondary debt positions. In some cases, investors, issuers, or underwriters might use the credit spread term structure for a particular rating or corporate sector either to derive prospective pricing for a new issuance or to determine fair value spreads for outstanding securities, which is an extension of matrix pricing. A high-yield debt investor might employ the term structure of credit spreads to gauge the risk/reward trade-offs between debt maturities. Given the impact of monetary and fiscal policies on risky debt markets, policymakers have extended their focus from default-risk-free yield curve dynamics to the term structure of credit spreads.

There are several key drivers of the term structure of credit spreads. First, credit quality is a key factor. For investment-grade securities with the highest credit ratings and extremely low spreads, credit spread migration is only possible in one direction given the implied lower bound of zero on credit spreads. As a result, the credit term structure for the most highly rated securities tends to be either flat or slightly upward sloping. Securities with lower credit quality, however, face greater sensitivity to the credit cycle. The greater likelihood of default associated with high-yield securities generally results in a steeper credit spread curve, both in cases where a weaker economy suggests credit spread widening and when an inverted credit spread curve suggests tighter spreads for longer maturities. As a high-yield bond moves further down the credit spectrum into a more distressed scenario, the contractual cash flows through maturity become less certain—with the value of distressed debt converging to a dollar price equal to the recovery rate as default becomes more certain, regardless of the remaining time to maturity. Such a scenario will result in a steeply inverted credit spread term structure. We now review the determinants of that term structure inversion and other implications of this scenario in more detail.

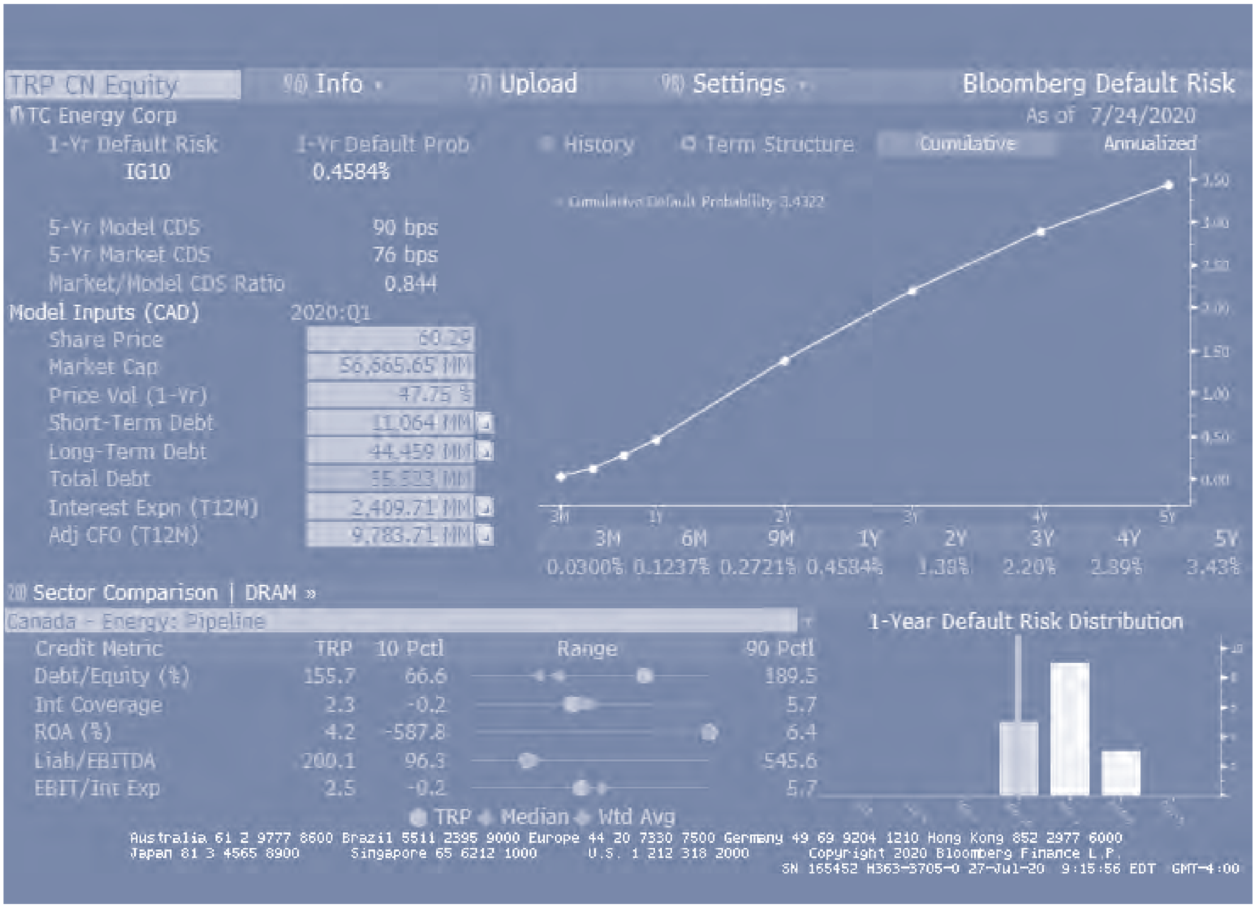
Financial conditions are another critical factor affecting the credit spread term structure. From a macroeconomic perspective, the credit risk of a bond is influenced by expectations for economic growth and inflation. A stronger economic climate is generally associated with higher benchmark yields but lower credit spreads for issuers whose default probability declines during periods of economic growth (cash flows tend to improve and profitability increases under such a scenario). The countercyclical relationship between spreads and benchmark rates is therefore commonly observed across the business cycle.

Market supply and demand dynamics are another critical factor influencing the credit curve term structure. Unlike default-risk-free government securities in developed markets, the relative liquidity of corporate bonds varies widely, with the vast majority of securities not trading on a daily basis. Given that new and most recently issued securities tend to represent the largest proportion of trading volume and are responsible for much of the volatility in credit spreads, the credit curve will be most heavily influenced by the most frequently traded securities. For example, although one might expect the credit curve to steepen for a borrower refinancing near-term maturities with long-term debt, this effect may be partially offset by a tighter bid–offer spread for longer credit maturities. This flattening may also occur within a specific rating or if market participants anticipate significant supply in a particular tenor. Infrequently traded bonds trading with wider bid–offer spreads can also impact the shape of the term structure, so it is important to gauge the size and frequency of trades in bonds across the maturity spectrum to ensure consistency.

Finally, from a microeconomic perspective, company-value model results discussed earlier are another key driver of the credit spread term structure. Under traditional credit analysis, the specific industry or industries within which an issuer operates are considered, as well as key financial ratios, such as cash flow, leverage, and profitability versus sector and ratings peers. This company-specific analysis based on fundamental data has been complemented by more probabilistic, forward-looking structural models for company valuation. These models take stock market valuation, equity volatility, and balance sheet information into account to derive the implied default probability for a company. Holding other factors constant, any microeconomic factor that increases the implied default probability, such as greater equity volatility, will tend to drive a steeper credit spread curve, and the reverse is true with a decline in equity volatility.

Practitioners will frequently employ these tools when analyzing the term structure of credit spreads to determine fair value. For example, the Bloomberg default risk screen (DRSK) shown in Exhibit 29 combines the company-value analysis with fundamental credit ratios for a composite analysis of TransCanada Corporation, a Canadian natural gas transmission and power services company.

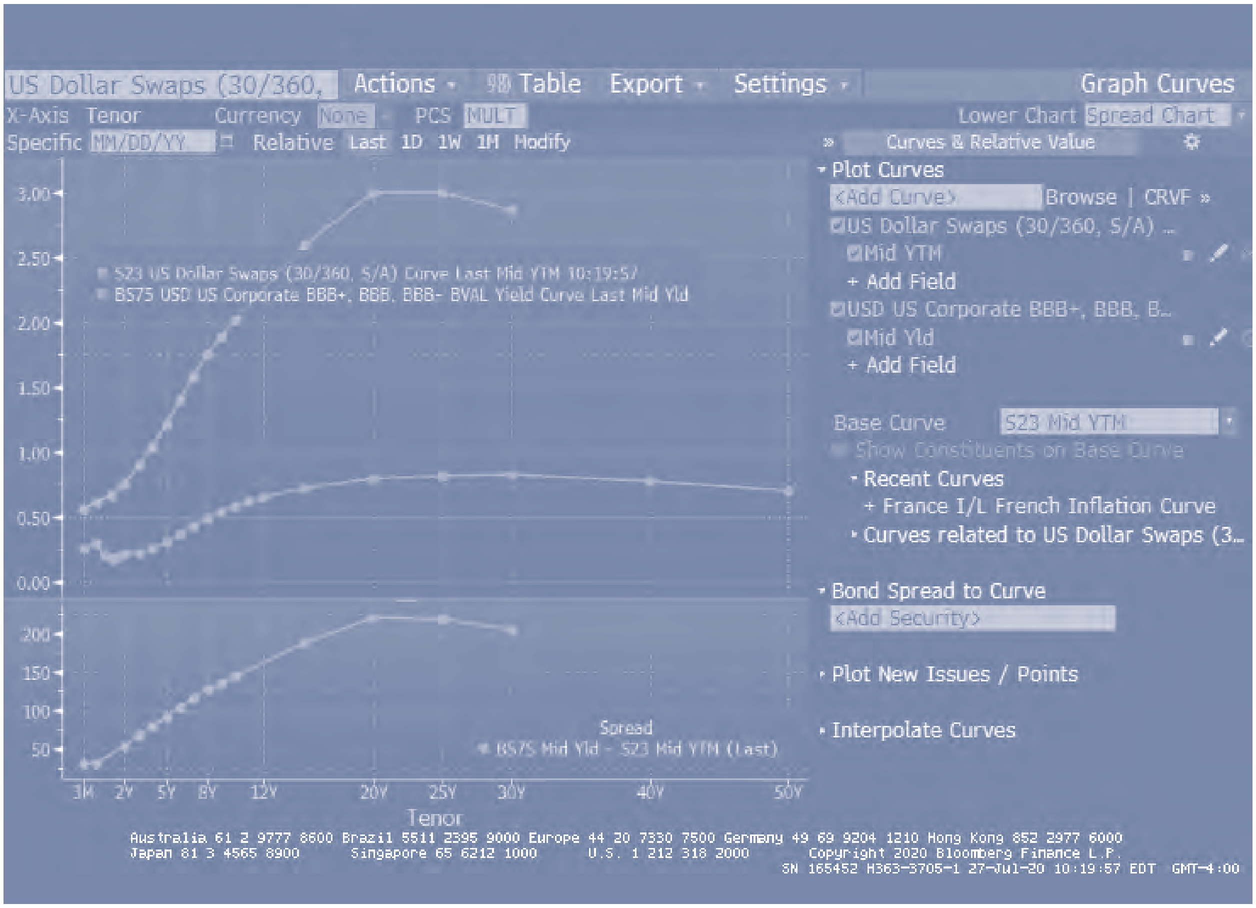
Exhibit 29 Default Risk Screen



Source: Bloomberg.

Two further considerations are important when analyzing the term structure of credit spreads. The first concerns the appropriate risk-free or benchmark rates used to determine spreads. A frequently traded government security with the nearest maturity to an outstanding corporate bond generally represents the lowest default risk for developed markets, so this is a logical benchmark choice. However, the duration and maturity of the most liquid or on-the-run government bonds rarely match those of corporate bonds trading in the secondary market, so it is often necessary to interpolate between yields of the two government securities with the closest maturity. Because the interpolation may impact the analysis for less liquid maturities, the benchmark swap curve based on interbank rates is often substituted for the government benchmark because of greater swap market liquidity for off-the-run maturities. For example, Exhibit 30 demonstrates the latter methodology on a Bloomberg screen for a composite of BBB rated US industrial corporate issuers versus the benchmark US dollar swap curve, showing a positive-sloped credit spread term structure across maturities.

Exhibit 30 Credit Spreads over Swap Rates



Source: Bloomberg.

The second consideration concerns the all-in spread over the benchmark itself. Term structure analysis should include only bonds with similar credit characteristics, which are typically senior unsecured general obligations of the issuer. Any bonds of the issuer with embedded options, first or second lien provisions, or other unique provisions should be excluded from the analysis. It is also important to note that such securities typically include cross-default provisions so that all securities across the maturity spectrum of a single issuer will be subject to recovery in the event of bankruptcy.

Using the models presented in prior sections, we can demonstrate that the *change* in market expectations of default over time is a key determinant of the shape of the credit curve term structure. This may be shown using a simple extension of the zero-coupon corporate bond example in Exhibit 2 by changing the probability of default. Using a recovery rate of 40% and changing the probability of default from 1.25% to 1.50% raises the credit spread from 77 bps in the original example to 92 bps. These calculations are shown in Exhibit 31.

Exhibit 31 Raising the Default Probability of the Five-Year, Zero-Coupon Corporate Bond

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0								
1	88.8487	35.5395	53.3092	1.5000%	98.5000%	0.7996	0.970874	0.7763
2	91.5142	36.6057	54.9085	1.4775%	97.0225%	0.8113	0.942596	0.7647
3	94.2596	37.7038	56.5558	1.4553%	95.5672%	0.8231	0.915142	0.7532
4	97.0874	38.8350	58.2524	1.4335%	94.1337%	0.8351	0.888487	0.7419
5	100.0000	40.0000	60.0000	1.4120%	92.7217%	0.8472	0.862609	0.7308
				7.2783%			CVA =	3.7670

Fair value = 86.2609 – 3.7670 = 82.4939.
Yield to maturity = 3.9240%.
Credit spread = 3.9240% – 3.00% = 0.9240%.

Flat credit spread curves imply a relatively stable expectation of default over time, whereas an upward-sloping credit curve implies that investors seek greater compensation for assuming issuer default risk over longer periods. For example, we can illustrate this in terms of a credit spread curve by holding the benchmark rate constant at 3.00% across 3-year, 5-year, and 10-year maturities while increasing the default probability over time. Although one could consider an increase in default probability each year, the following example in Exhibit 32 assumes a 1.00% default probability for Years 1, 2, and 3, a 2.00% probability of default in Years 4 and 5, and a 3.00% default probability in Years 6 through 10, with the recovery rate at a constant 40%. (Note that this is another example of the annual default probability changing over the lifetime of the bonds.) As shown in Exhibit 32, the credit spread rises from 62 bps to 86 bps to 132 bps.

Exhibit 32 Increasing the Default Probability for Longer Times to Maturity

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0								
1	94.2596	37.7038	56.5558	1.0000%	99.0000%	0.5656	0.970874	0.5491
2	97.0874	38.8350	58.2524	0.9900%	98.0100%	0.5767	0.942596	0.5436
3	100.0000	40.0000	60.0000	0.9801%	97.0299%	0.5881	0.915142	0.5382
				2.9701%			CVA =	1.6308

Fair value = 91.5142 – 1.6308 = 89.8833.
Yield to maturity = 3.6192%.
Credit spread = 3.6192% – 3.00% = 0.6192%.

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0								
1	88.8487	35.5395	53.3092	1.0000%	99.0000%	0.5331	0.970874	0.5176
2	91.5142	36.6057	54.9085	0.9900%	98.0100%	0.5436	0.942596	0.5124
3	94.2596	37.7038	56.5558	0.9801%	97.0299%	0.5543	0.915142	0.5073

(continued)

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
4	97.0874	38.8350	58.2524	1.9406%	95.0893%	1.1304	0.888487	1.0044
5	100.0000	40.0000	60.0000	1.9018%	93.1875%	1.1411	0.862609	0.9843
				6.8125%			CVA =	3.5259

Fair value = 86.2609 – 3.5259 = 82.7350.
Yield to maturity = 3.8633%.
Credit spread = 3.8633% – 3.00% = 0.8633%.

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0								
1	76.647	30.6567	45.9850	1.0000%	99.0000%	0.4599	0.970874	0.4465
2	78.9409	31.5764	47.3646	0.9900%	98.0100%	0.4689	0.942596	0.4420
3	81.3092	32.5237	48.7855	0.9801%	97.0299%	0.4781	0.915142	0.4376
4	83.7484	33.4994	50.2491	1.9406%	95.0893%	0.9751	0.888487	0.8664
5	86.2609	34.5044	51.7565	1.9018%	93.1875%	0.9843	0.862609	0.8491
6	88.8487	35.5395	53.3092	2.7956%	90.3919%	1.4903	0.837484	1.2481
7	91.5142	36.6057	54.9085	2.7118%	87.6801%	1.4890	0.813092	1.2107
8	94.2596	37.7038	56.5558	2.6304%	85.0497%	1.4876	0.789409	1.1744
9	97.0874	38.8350	58.2524	2.5515%	82.4982%	1.4863	0.766417	1.1391
10	100.0000	40.0000	60.0000	2.4749%	80.0233%	1.4850	0.744094	1.1050
				19.9767%			CVA =	8.9187

Fair value = 74.4094 – 8.9187 = 65.4907.
Yield to maturity = 4.3235%.
Credit spread = 4.3235% – 3.00% = 1.3235%.

Positive-sloped credit spread curves are likely when a high-quality issuer with a strong competitive position in a stable industry has low leverage, strong cash flow, and a high profit margin. This type of issuer tends to exhibit very low short-term credit spreads rising with increasing maturity given greater uncertainty due to the macroeconomic environment, potential adverse changes in the competitive landscape, technological change, or other factors that drive a higher implied probability of default over time. Empirical academic studies also tend to support the view that the credit spread term structure is upward-sloping for investment-grade bond portfolios (Bedendo, Cathcart, and El-Jahel 2007).

Alternatively, high-yield issuers in cyclical industries sometimes face a downward-sloping credit term structure because of issuer- or industry-specific reasons. For example, an ownership change resulting from a leveraged buyout or private equity acquisition may often be accompanied by a significant increase in leverage. In such a case, an inverted credit curve may indicate investor expectations that the new owners will create efficiencies in the restructured organization, leading to improved future cash flow and profitability that will benefit debt investors. Another example of an inverted credit term structure might result when issuers in a historically cyclical industry (such as oil and gas exploration or retail) find themselves at the bottom of an economic cycle, with investor expectations of a recovery in the industry tied to improving credit spreads over time.

That said, it is important to distinguish between scenarios where the contractual cash flows of a risky bond are likely to occur and distressed debt scenarios where investors expect to receive only the recovery rate in a likely bankruptcy scenario. Bonds with a very high likelihood of default tend to trade on a price basis that converges toward the recovery rate rather than on a spread to benchmark rates. This scenario leads to credit spread term structures that may be considered more of an "optical" phenomenon rather than a true reflection of the relative risks and rewards of long-term versus short-term bonds from a single issuer, as illustrated in the following discussion.

To demonstrate this using our zero-coupon bond example, let us shift to a scenario where bondholders with 5-year and 10-year bonds outstanding anticipate an imminent default scenario and both bonds trade at a recovery rate of 40%.

Note that if we solve for the fair value and resulting credit spread over the benchmark yield as in the instances where default probability was 1.25%, we end up with the same VNDs for the 5-year and 10-year bonds, respectively. However, when deriving a credit spread value for both securities assuming recovery in a bankruptcy scenario and cross-default provisions across maturities, the credit valuation adjustment representing the sum of expected losses is simply the difference between the VND and the recovery rate.

For the five-year example, we can thus calculate a VND of 86.2609, a CVA of 46.2609, and a fair value with recovery at 40. This results in a yield of 20.1124% and a credit spread over the government bond of 17.1124%. In the 10-year case, the VND may be shown as 74.4094, a CVA of 34.4094, and a fair value at 40. That gives a yield of 9.5958% and a credit spread of 6.5958%. We end up with a steep and inverted "credit spread" curve.

The interpretation of the credit spread term structure is important for investors seeking to capitalize on a market view that differs from that reflected in the credit curve. For example, if a portfolio manager disagrees with the market's expectation of a high near-term default probability that declines over time, she could sell short-term protection in the credit default swap market and buy longer-term protection. In a scenario where the issuer does not default, the investor retains the premium on protection sold and may either retain or choose to sell back the longer-term credit default swap to realize a gain.

CREDIT ANALYSIS FOR SECURITIZED DEBT

8

h compare the credit analysis required for securitized debt to the credit analysis of corporate debt

Unlike the general obligation nature of most private or sovereign fixed-income securities, securitized debt allows issuers to finance a specific set of assets or receivables (e.g., mortgages, automobile loans, or credit card receivables) rather than an entire balance sheet. Issuers in securitized debt markets are frequently motivated to undertake financing using these more structured securities given their ability to increase debt capacity and reduce the originator's need to maintain regulatory capital or retain residual risk. The isolation of securitized assets generally decreases the relative financing cost for these assets on a stand-alone basis as compared to a general obligation financing of the debt originator. By freeing up capital, an originator is also able to continue to generate income from further originations. Investors, however, seek to benefit from greater diversification, more stable and predictable underlying cash flows, and a return that is greater than that of securities with similar ratings, which provide a reward for accepting the greater complexity associated with collateralized debt. That said, the credit analysis of such structured finance instruments requires a fundamentally different approach compared with other risky bonds given the underlying collateral, the parties associated with the origination or servicing of the portfolio over the life of the security, and the issuing entity, as well as any structural and credit enhancement features typically present in these transactions.

It is important to distinguish first and foremost among the types of securitized debt issued globally, as well as the various forms. In its summary of structured finance asset types shown in Exhibit 33, the German-based rating agency Scope Ratings AG provides its general approach to credit assessment based not only on the underlying time horizon and collateral but also on asset characteristics referred to as granularity and homogeneity.

Exhibit 33 Summary of Asset Types and Characteristics of Core Structured Finance Asset Classes

Deal Type	Underlying Collateral	Risk Horizon	Granularity	Homogeneity	Credit Analysis Approach
Asset-backed CP	Commercial discount credits or credit advances	Short-term	Granular	Homogeneous	Book
Auto ABS	Auto loans or leases	Medium-term	Granular	Homogeneous	Portfolio
CMBS	Commercial mortgages	Typically long-term	Non-granular	Heterogeneous	Loan by loan
Consumer ABS	Consumer loans	Medium-term	Granular	Homogeneous	Portfolio
CRE loans	Commercial real estate loans	Long-term	Non-granular	Heterogeneous	Loan by loan
Credit cards	Credit card balances	Short-term	Granular	Homogeneous	Book
Credit-linked notes/ repackaging	Any financial assets	Typically medium-term	Typically single asset	NA	Pass-through rating/asset by asset
LL CLOs	Leveraged corporate loans	Medium-term	Non-granular	Heterogeneous	Loan by loan
PF CLOs	Project finance debt	Long-term	Non-granular	Heterogeneous	Loan by loan
RMBS	Residential mortgages	Long-term	Granular	Homogeneous	Loan by loan or portfolio
SME ABS	Loans to small- and medium-sized businesses	Typically medium-term	Granular	Mixed	Loan by loan or portfolio
Trade receivables	Commercial credit	Short-term	Typically granular	Homogeneous	Book

Source: Adapted from Scope Ratings AG (2016b, pp. 7–8).

The concept of homogeneity refers to the degree to which underlying debt characteristics within a structured finance instrument are similar across individual obligations. On the one hand, an investor or credit analyst might draw general conclusions about the nature of homogeneous credit card or auto loan obligations given that an individual obligation faces strict eligibility criteria to be included in a specific asset pool. On the other hand, heterogeneous leveraged loan, project finance, or real estate transactions require scrutiny on a loan-by-loan basis given their different characteristics. The granularity of the portfolio refers to the actual number of obligations that make up the overall structured finance instrument. A highly granular portfolio may have hundreds of underlying debtors, suggesting it is appropriate to draw conclusions about creditworthiness based on portfolio summary statistics rather than investigating each borrower. Alternatively, an asset pool with fewer more-discrete or non-granular investments would warrant analysis of each individual obligation.

The combination of asset type and tenor as well as the relative granularity and homogeneity of the underlying obligations drive the approach to credit analysis for a given instrument type. For example, short-term structured finance vehicles with granular, homogeneous assets tend to be evaluated using a statistics-based approach to the existing book of loans. This changes to a portfolio-based approach for medium-term granular and homogeneous obligations because the portfolio is not static but changes over time. For discrete or non-granular heterogeneous portfolios, a loan-by-loan approach to credit analysis is more appropriate. The following example of a credit card securitization will provide further insight into the process.

Exhibit 34 provides a summary from the prospectus of the Synchrony Credit Card Master Note Trust $750,000,000 Series 2016-1 Asset Backed Notes issued in March 2016. As is spelled out in the prospectus, the Synchrony transaction is backed by credit card receivables having the given credit score distribution presented in the exhibit.

Exhibit 34 A Structured Debt Example, Composition by FICO Credit Score Range

FICO Credit Score Range	Receivables Outstanding	Percentage of Outstanding
Less than or equal to 599	$995,522,016	6.6%
600 to 659	$2,825,520,245	18.7%
660 to 719	$6,037,695,923	39.9%
720 and above	$5,193,614,599	34.4%
No score	$64,390,707	0.4%
Total	**$15,116,743,490**	**100%**

Source: Synchrony Credit Card Master Note Trust $750,000,000 Series 2016-1 Asset Backed Notes Prospectus (p. 93; available at investors.synchronyfinancial.com).

Investors in this type of ABS will base their probability of default on the mean default probability, recovery rate, and variance of a portfolio of borrowers reflecting the distribution of FICO scores within the pool rather than conducting an analysis of individual borrowers. The prospectus provides a broad set of details beyond the FICO scores of borrowers for further in-depth portfolio analysis, including age of the receivables, average outstanding balances, and delinquency rates.

A heterogeneous portfolio of fewer loans, however, requires a fundamentally different approach. In this instance, each obligation within the asset pool may warrant its own analysis to determine whether an individual commercial property or leveraged company is able to meet its financial obligations under the ABS contract. Here the expected default probability and recovery rate on an asset-by-asset basis is the best gauge of how the investment will perform under various scenarios.

A second critical aspect of the credit exposure associated with ABS relates to the origination and servicing of assets over the life of the transaction. The prospectus and other related documents determine the roles and responsibilities of these related parties over the life of an ABS transaction. Upon inception of the transaction, investors rely on the originator/servicer to establish and enforce loan eligibility criteria, secure and maintain proper documentation and records, and maximize timely repayment and contract enforceability in cases of delinquency. Once the asset pool has been identified, investors are also exposed to operational and counterparty risk over the life of an ABS transaction. That is, they remain exposed to the ability of the servicer to effectively manage and service the portfolio over the life of the transaction. For an auto ABS transaction, this may involve the ability to repossess and sell a vehicle at a price close to the residual value in a timely manner in the event that a borrower is unable to pay, while in a commercial real estate transaction, it may involve identifying and replacing a non-performing tenant. Investors in an asset portfolio whose composition changes over time also face exposure to the replacement of obligors over time. In all such instances, not only is the creditworthiness of the servicer important but also of importance is its track record in meeting these servicing obligations, which are frequently gauged by analyzing the performance of more seasoned transactions handled by the same servicer over the credit cycle.

For example, in the case of the Synchrony Credit Card Master Note Trust transaction, Synchrony Financial acts as servicer of the trust and Synchrony Bank, as sub-servicer, is primarily responsible for receiving and processing collections on the receivables. A potential investor might therefore evaluate not only the performance of other debt backed by credit card receivables but also how outstanding notes serviced by Synchrony have performed over time versus its servicing competitors.

Finally, the structure of a collateralized or secured debt transaction is a critical factor in analyzing this type of investment. These structural aspects include both the nature of the obligor itself, which is often a special purpose entity (SPE) whose sole purpose is to acquire a specified pool of assets and issue ABS to finance the SPE, and any structural enhancements of the transaction, which may include overcollateralization, credit tranching (i.e., tiering the claim priorities of ownership or interest), or other characteristics.

A key question related to the issuer is its relationship to the originator—namely, the degree to which the bankruptcy of the obligor is related to that of the originator. The bankruptcy remoteness is typically determined by whether the transfer of the assets from the originator to the SPE may be deemed a true sale, which otherwise allows for the ability to separate risk between the originator and SPE at a later date.

Second, additional credit enhancements are a key structural element to be evaluated in the context of credit risk. Credit enhancements for ABS take on several forms beyond the bankruptcy remoteness of the SPE. For example, ABS transactions frequently have payout or performance triggers that protect investors in the case of adverse credit events. Certain events related to the servicer or seller—such as failure to make deposits or payments or other adverse events—may trigger early repayment ("amortization") of the security. For consumer transactions such as credit card or automotive ABS, the primary protection against a decline in asset quality for investors is additional return built into the transaction that is greater than the expected or historical loss of the asset pool. This additional return is often called the excess spread. Issuers create subordinated tranches of debt that provide added protection to those rated higher and benefit from a greater excess spread cushion over the life of the financing.

Covered bonds, which originated in Germany in the 18th century but have since been adopted by issuers across Europe, Asia, and Australia, have some similarities with these structured finance investments but also have fundamental differences that warrant special consideration. A covered bond is a senior debt obligation of a financial institution that gives recourse to both the originator/issuer and a predetermined underlying collateral pool. Each country or jurisdiction specifies the eligible collateral types and the specific structures permissible in its covered bond market. Covered bonds most frequently have either commercial or residential mortgages meeting specific criteria or public sector debt as underlying collateral.

The dual recourse to the issuing financial institution and the underlying asset pool has been a hallmark of covered bonds since their inception, but it was also reinforced under the European Union Bank Recovery and Resolution Directive (BRRD; see Scope Ratings AG 2016a). Under the BRRD, covered bonds enjoy unique protection among bank liabilities in the event of restructuring or regulatory intervention. Additionally, the financial institution has the ongoing obligation to maintain sufficient assets in the cover pool to satisfy the claims of covered bondholders at all times, and the obligations of the financial institution with respect to the cover pool are supervised by public or other independent bodies.

Another aspect of covered bonds that needs to be considered in credit analysis is the dynamic nature of the cover pool. In contrast to a static pool of mortgage loans (which expose investors to prepayment risk in the case of US mortgage-backed securities), cover pool sponsors must replace any prepaid or non-performing assets in the cover pool to ensure sufficient cash flows to the maturity of the covered bond.

Analysts should also be aware of various redemption regimes that exist to align the covered bond's cash flows as closely as possible to the original maturity schedule in the event of default of a covered bond's financial sponsor. These include hard-bullet covered bonds; if payments do not occur according to the original schedule, a bond default is triggered and bond payments are accelerated. Another type is soft-bullet covered bonds, which delay the bond default and payment acceleration of bond cash flows until a new final maturity date, which is usually up to a year after the original maturity date. Conditional pass-through covered bonds, in contrast, convert to pass-through securities after the original maturity date if all bond payments have not yet been made.

Credit analysis for covered bonds follows traditional credit analysis in evaluating both the issuer and the cover pool. Given the additional credit enhancements, recovery rates tend to be high and default probabilities low, making covered bonds a relatively safe credit asset. As a result, rating agencies often assign a credit rating to covered bonds that is several notches above that of the issuing financial institution.

SUMMARY

We have covered several important topics in credit analysis. Among the points made are the following:

- Three factors important to modeling credit risk are the expected exposure to default, the recovery rate, and the loss given default.
- These factors permit the calculation of a credit valuation adjustment that is subtracted from the (hypothetical) value of the bond, if it were default risk free, to get the bond's fair value given its credit risk. The credit valuation adjustment is calculated as the sum of the present values of the expected loss for each period in the remaining life of the bond. Expected values are computed using risk-neutral probabilities, and discounting is done at the risk-free rates for the relevant maturities.
- The CVA captures investors' compensation for bearing default risk. The compensation can also be expressed in terms of a credit spread.
- Credit scores and credit ratings are third-party evaluations of creditworthiness used in distinct markets.

- Analysts may use credit ratings and a transition matrix of probabilities to adjust a bond's yield to maturity to reflect the probabilities of credit migration. Credit spread migration typically reduces expected return.
- Credit analysis models fall into two broad categories: structural models and reduced-form models.
- Structural models are based on an option perspective of the positions of the stakeholders of the company. Bondholders are viewed as owning the assets of the company; shareholders have call options on those assets.
- Reduced-form models seek to predict *when* a default may occur, but they do not explain the *why* as structural models do. Reduced-form models, unlike structural models, are based only on observable variables.
- When interest rates are assumed to be volatile, the credit risk of a bond can be estimated in an arbitrage-free valuation framework.
- The discount margin for floating-rate notes is similar to the credit spread for fixed-coupon bonds. The discount margin can also be calculated using an arbitrage-free valuation framework.
- Arbitrage-free valuation can be applied to judge the sensitivity of the credit spread to changes in credit risk parameters.
- The term structure of credit spreads depends on macro and micro factors.
- As it concerns macro factors, the credit spread curve tends to become steeper and to widen in conditions of weak economic activity. Market supply and demand dynamics are important. The most frequently traded securities tend to determine the shape of this curve.
- Issuer- or industry-specific factors, such as the chance of a future leverage-decreasing event, can cause the credit spread curve to flatten or invert.
- When a bond is very likely to default, it often trades close to its recovery value at various maturities; moreover, the credit spread curve is less informative about the relationship between credit risk and maturity.
- For securitized debt, the characteristics of the asset portfolio themselves suggest the best approach for a credit analyst to take when deciding among investments. Important considerations include the relative concentration of assets and their similarity or heterogeneity as it concerns credit risk.

REFERENCES

Bedendo, Mascia, Lara Cathcart, and Lina El-Jahel. 2007. "The Slope of the Term Structure of Credit Spreads: An Empirical Investigation." *Journal of Financial Research* 30 (2): 237–57.

Berd, Arthur, Roy Mashal, and Peili Wang. 2004. "*Defining, Estimating and Using Credit Term Structures. Part 1: Consistent Valuation Measures.*" Lehman Brothers Working paper (November).

Black, Fisher, and Myron Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81:637–54.

Duffie, Darrell, and Kenneth J. Singleton. 1999. "Modeling the Term Structure of Defaultable Bonds." *Review of Financial Studies* 12:687–720.

Duffie, Darrell, and Kenneth J. Singleton. 2003. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press.

Fabozzi, Frank J. 2013. *Bond Markets, Analysis, and Strategies*. 8th ed., Pearson.

Jarrow, Robert, and Stuart Turnbull. 1995. "Pricing Derivatives on Financial Securities Subject to Default Risk." *Journal of Finance* 50:53–85.

Merton, Robert. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29:449–70.

Scope Ratings AG. 2016a. "*Covered Bond Rating Methodology*" (22 July): www.scoperatings.com.

Scope Ratings AG. 2016b. "*General Structured Finance Rating Methodology*" (31 August): www.scoperatings.com.

Smith, Donald J. 2011. "Hidden Debt: From Enron's Commodity Prepays to Lehman's Repo 105s." *Financial Analysts Journal* 67 (5): 15–22.

Smith, Donald J. 2017. *Valuation in a World of CVA, DVA, and FVA: A Tutorial on Debt Securities and Interest Rate Derivatives*. World Scientific Publishing Company.

PRACTICE PROBLEMS

The following information relates to Questions 1–15

Daniela Ibarra is a senior analyst in the fixed-income department of a large wealth management firm. Marten Koning is a junior analyst in the same department, and David Lok is a member of the credit research team.

The firm invests in a variety of bonds. Ibarra is presently analyzing a set of bonds with some similar characteristics, such as four years until maturity and a par value of €1,000. Exhibit 1 includes details of these bonds.

Exhibit 1 A Brief Description of the Bonds Being Analyzed

Bond	Description
B1	A zero-coupon, four-year corporate bond with a par value of €1,000. The wealth management firm's research team has estimated that the risk-neutral probability of default for each date for the bond is 1.50%, and the recovery rate is 30%.
B2	A bond similar to B1, except that it has a fixed annual coupon rate of 6% paid annually.
B3	A bond similar to B2 but rated AA.
B4	A bond similar to B2 but the coupon rate is the one-year benchmark rate plus 4%.

Ibarra asks Koning to assist her with analyzing the bonds. She wants him to perform the analysis with the assumptions that there is no interest rate volatility and that the government bond yield curve is flat at 3%.

Ibarra performs the analysis assuming an upward-sloping yield curve and volatile interest rates. Exhibit 2 provides the data on annual payment benchmark government bonds.

She uses these data to construct a binomial interest rate tree based on an assumption of future interest rate volatility of 20%.

Exhibit 2 Par Curve for Annual Payment Benchmark Government Bonds

Maturity	Coupon Rate	Price	Discount Factor	Spot Rate	Forward Rate
1	–0.25%	€100	1.002506	–0.2500%	
2	0.75%	€100	0.985093	0.7538%	1.7677%
3	1.50%	€100	0.955848	1.5166%	3.0596%
4	2.25%	€100	0.913225	2.2953%	4.6674%

© 2019 CFA Institute. All rights reserved.

Answer the first five questions (1–5) based on the assumptions made by Marten Koning, the junior analyst. Answer Questions 8–12 based on the assumptions made by Daniela Ibarra, the senior analyst.

Note: All calculations in this problem set are carried out on spreadsheets to preserve precision. The rounded results are reported in the solutions.

1 The market price of Bond B1 is €875. The bond is:

A fairly valued.

B overvalued.

C undervalued.

2 Koning realizes that an increase in the recovery rate would lead to an increase in the bond's fair value, whereas an increase in the probability of default would lead to a decrease in the bond's fair value. He is not sure, however, which effect would be greater. So, he increases both the recovery rate and the probability of default by 25% of their existing estimates and recomputes the bond's fair value. The recomputed fair value is closest to:

A €843.14.

B €848.00.

C €855.91.

3 The fair value of Bond B2 is closest to:

A €1,069.34.

B €1,111.51.

C €1,153.68.

4 The market price of Bond B2 is €1,090. If the bond is purchased at this price and there is a default on Date 3, the rate of return to the bond buyer would be closest to:

A –28.38%.

B –41.72%.

C –69.49%.

5 Bond B3 will have a modified duration of 2.75 at the end of the year. Based on the representative one-year corporate transition matrix in Exhibit 3 and assuming no default, how should the analyst adjust the bond's yield to maturity to assess the expected return on the bond over the next year?

Exhibit 3 Representative One-Year Corporate Transition Matrix (entries are in %)

From/To	AAA	AA	A	BBB	BB	B	CCC, CC, C	D
AAA	**90.00**	9.00	0.60	0.15	0.10	0.10	0.05	0.00
AA	1.50	**88.00**	9.50	0.75	0.15	0.05	0.03	0.02
A	0.05	2.50	**87.50**	8.40	0.75	0.60	0.12	0.08
BBB	0.02	0.30	4.80	**85.50**	6.95	1.75	0.45	0.23
BB	0.01	0.06	0.30	7.75	**79.50**	8.75	2.38	1.25
B	0.00	0.05	0.15	1.40	9.15	**76.60**	8.45	4.20
CCC, CC, C	0.00	0.01	0.12	0.87	1.65	18.50	**49.25**	29.60
Credit Spread	0.60%	0.90%	1.10%	1.50%	3.40%	6.50%	9.50%	

A Add 7.7 bps to YTM.

B Subtract 7.7 bps from YTM.

C Subtract 9.0 bps from YTM.

6 David Lok has estimated the probability of default of Bond B1 to be 1.50%. He is presenting the approach the research team used to estimate the probability of default. Which of the following statements is Lok likely to make in his presentation if the team used a reduced-form credit model?

A Option pricing methodologies were used, with the volatility of the underlying asset estimated based on historical data on the firm's stock price.

B Regression analysis was used, with the independent variables including both firm-specific variables, such as the debt ratio and return on assets, and macroeconomic variables, such as the rate of inflation and the unemployment rate.

C The default barrier was first estimated, followed by the estimation of the probability of default as the portion of the probability distribution that lies below the default barrier.

7 In the presentation, Lok is asked why the research team chose to use a reduced-form credit model instead of a structural model. Which statement is he likely to make in reply?

A Structural models are outdated, having been developed in the 1970s; reduced-form models are more modern, having been developed in the 1990s.

B Structural models are overly complex because they require the use of option pricing models, whereas reduced-form models use regression analysis.

C Structural models require "inside" information known to company management, whereas reduced-form models can use publicly available data on the firm.

8 As previously mentioned, Ibarra is considering a future interest rate volatility of 20% and an upward-sloping yield curve, as shown in Exhibit 2. Based on her analysis, the fair value of Bond B2 is closest to:

A €1,101.24.

B €1,141.76.

C €1,144.63.

9 Ibarra wants to know the credit spread of Bond B2 over a theoretical comparable-maturity government bond with the same coupon rate as this bond. The foregoing credit spread is closest to:

A 108 bps.

B 101 bps.

C 225 bps.

10 Ibarra is interested in analyzing how a simultaneous decrease in the recovery rate and the probability of default would affect the fair value of Bond B2. She decreases both the recovery rate and the probability of default by 25% of their existing estimates and recomputes the bond's fair value. The recomputed fair value is closest to:

A €1,096.59.

B €1,108.40.

C €1,111.91.

11 The wealth management firm has an existing position in Bond B4. The market price of B4, a floating-rate note, is €1,070. Senior management has asked Ibarra to make a recommendation regarding the existing position. Based on the assumptions used to calculate the estimated fair value only, her recommendation should be to:

A add to the existing position.

B hold the existing position.

C reduce the existing position.

12 The issuer of the floating-rate note, B4, is in the energy industry. Ibarra believes that oil prices are likely to increase significantly in the next year, which will lead to an improvement in the firm's financial health and a decline in the probability of default from 1.50% in Year 1 to 0.50% in Years 2, 3, and 4. Based on these expectations, which of the following statements is correct?

A The CVA will decrease to €22.99.

B The note's fair value will increase to €1,177.26.

C The value of the FRN, assuming no default, will increase to €1,173.55.

13 The floating-rate note, B4, is currently rated BBB by Standard & Poor's and Fitch Ratings (and Baa by Moody's Investors Service). Based on the research department assumption about the probability of default in Question 10 and her own assumption in Question 11, which action does Ibarra *most likely* expect from the credit rating agencies?

A Downgrade from BBB to BB.

B Upgrade from BBB to AAA.

C Place the issuer on watch with a positive outlook.

14 During the presentation about how the research team estimates the probability of default for a particular bond issuer, Lok is asked for his thoughts on the shape of the term structure of credit spreads. Which statement is he most likely to include in his response?

A The term structure of credit spreads typically is flat or slightly upward sloping for high-quality investment-grade bonds. High-yield bonds are more sensitive to the credit cycle, however, and can have a more upwardly sloped term structure of credit spreads than investment-grade bonds or even an inverted curve.

B The term structure of credit spreads for corporate bonds is always upward sloping—more so the weaker the credit quality because probabilities of default are positively correlated with the time to maturity.

C There is no consistent pattern for the term structure of credit spreads. The shape of the credit term structure depends entirely on industry factors.

15 The final question for Lok is about covered bonds. The person asking says, "I've heard about them but don't know what they are." Which statement is Lok most likely to make to describe a covered bond?

A A covered bond is issued in a non-domestic currency. The currency risk is then fully hedged using a currency swap or a package of foreign exchange forward contracts.

B A covered bond is issued with an attached credit default swap. It essentially is a "risk-free" government bond.

C A covered bond is a senior debt obligation giving recourse to the issuer as well as a predetermined underlying collateral pool, often commercial or residential mortgages.

The following information relates to Questions 16–22

Anna Lebedeva is a fixed-income portfolio manager. Paulina Kowalski, a junior analyst, and Lebedeva meet to review several positions in Lebedeva's portfolio.

Lebedeva begins the meeting by discussing credit rating migration. Kowalski asks Lebedeva about the typical impact of credit rating migration on the expected return on a bond. Lebedeva asks Kowalski to estimate the expected return over the next year on a bond issued by Entre Corp. The BBB rated bond has a yield to maturity of 5.50% and a modified duration of 7.54. Kowalski calculates the expected return on the bond over the next year given the partial credit transition and credit spread data in Exhibit 1. She assumes that market spreads and yields will remain stable over the year.

Exhibit 1 One-Year Transition Matrix for BBB Rated Bonds and Credit Spreads

	AAA	AA	A	BBB	BB	B	CCC, CC, C
Probability (%)	0.02	0.30	4.80	85.73	6.95	1.75	0.45
Credit spread	0.60%	0.90%	1.10%	1.50%	3.40%	6.50%	9.50%

Lebedeva next asks Kowalski to analyze a three-year bond, issued by VraiRive S.A., using an arbitrage-free framework. The bond's coupon rate is 5%, with interest paid annually and a par value of 100. In her analysis, she makes the following three assumptions:

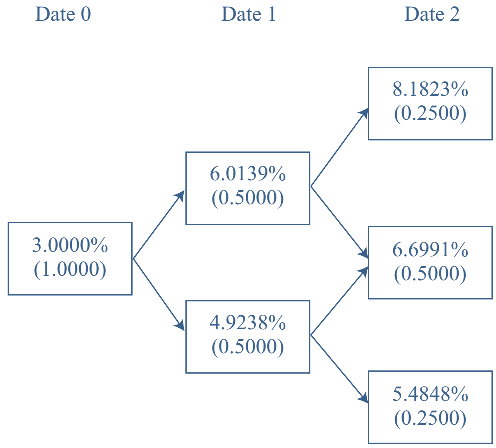
- The annual interest rate volatility is 10%.
- The recovery rate is one-third of the exposure each period.
- The annual probability of default each year is 2.00%.

Selected information on benchmark government bonds for the VraiRive bond is presented in Exhibit 2, and the relevant binomial interest rate tree is presented in Exhibit 3.

Exhibit 2 Par Curve Rates for Annual Payment Benchmark Government Bonds

Maturity	Coupon Rate	Price	Discount Factor	Spot Rate	Forward Rate
1	3.00%	100	0.970874	3.0000%	3.0000%
2	4.20%	100	0.920560	4.2255%	5.4656%
3	5.00%	100	0.862314	5.0618%	6.7547%

Exhibit 3 One-Year Binomial Interest Rate Tree for 10% Volatility (risk-neutral probabilities in parentheses)



Kowalski estimates the value of the VraiRive bond assuming no default(VND) as well as the fair value of the bond. She then estimates the bond's yield to maturity and the bond's credit spread over the benchmark in Exhibit 2. Kowalski asks Lebedeva, "What might cause the bond's credit spread to decrease?"

Lebedeva and Kowalski next discuss the drivers of the term structure of credit spreads. Kowalski tells Lebedeva the following:

Statement 1	The credit term structure for the most highly rated securities tends to be either flat or slightly upward sloping.
Statement 2	The credit term structure for lower-rated securities is often steeper, and credit spreads widen with expectations of strong economic growth.

Next, Kowalski analyzes the outstanding bonds of DLL Corporation, a high-quality issuer with a strong, competitive position. Her focus is to determine the rationale for a positive-sloped credit spread term structure.

Lebedeva ends the meeting by asking Kowalski to recommend a credit analysis approach for a securitized asset-backed security (ABS) held in the portfolio. This non-static asset pool is made up of many medium-term auto loans that are homogeneous, and each loan is small relative to the total value of the pool.

16 The *most appropriate* response to Kowalski's question regarding credit rating migration is that it has:

- **A** a negative impact.
- **B** no impact.
- **C** a positive impact.

17 Based on Exhibit 1, the one-year expected return on the Entre Corp. bond is *closest* to:

- **A** 3.73%.
- **B** 5.50%.
- **C** 7.27%.

18 Based on Kowalski's assumptions and Exhibits 2 and 3, the credit spread on the VraiRive bond is *closest* to:

- **A** 0.6949%.

B 0.9388%.

C 1.4082%.

19 The *most appropriate* response to Kowalski's question relating to the credit spread is:

A an increase in the probability of default.

B an increase in the loss given default.

C a decrease in the risk-neutral probability of default.

20 Which of Kowalski's statements regarding the term structure of credit spreads is correct?

A Only Statement 1

B Only Statement 2

C Both Statement 1 and Statement 2

21 DLL's credit spread term structure is *most* consistent with the firm having:

A low leverage.

B weak cash flow.

C a low profit margin.

22 Given the description of the asset pool of the ABS, Kowalski should recommend a:

A loan-by-loan approach.

B portfolio-based approach.

C statistics-based approach.

The following information relates to Questions 23–30

Lena Liecken is a senior bond analyst at Taurus Investment Management. Kristel Kreming, a junior analyst, works for Liecken in helping conduct fixed-income research for the firm's portfolio managers. Liecken and Kreming meet to discuss several bond positions held in the firm's portfolios.

Bonds I and II both have a maturity of one year, an annual coupon rate of 5%, and a market price equal to par value. The risk-free rate is 3%. Historical default experiences of bonds comparable to Bonds I and II are presented in Exhibit 1.

Exhibit 1 Credit Risk Information for Comparable Bonds

Bond	Recovery Rate	Percentage of Bonds That Survive and Make Full Payment
I	40%	98%
II	35%	99%

Bond III is a zero-coupon bond with three years to maturity. Liecken evaluates similar bonds and estimates a recovery rate of 38% and a risk-neutral default probability of 2%, assuming conditional probabilities of default. Kreming creates Exhibit 2 to compute Bond III's credit valuation adjustment. She assumes a flat yield curve at 3%, with exposure, recovery, and loss given default values expressed per 100 of par value.

Exhibit 2 Analysis of Bond III

Date	Exposure	Recovery	Loss Given Default	Probability of Default	Probability of Survival	Expected Loss	Present Value of Expected Loss
0							
1	94.2596	35.8186	58.4410	2.0000%	98.0000%	1.1688	1.1348
2	97.0874	36.8932	60.1942	1.9600%	96.0400%	1.1798	1.1121
3	100.0000	38.0000	62.0000	1.9208%	94.1192%	1.1909	1.0898
Sum				5.8808%		3.5395	3.3367

Bond IV is an AA rated bond that matures in five years, has a coupon rate of 6%, and a modified duration of 4.2. Liecken is concerned about whether this bond will be downgraded to an A rating, but she does not expect the bond to default during the next year. Kreming constructs a partial transition matrix, which is presented in Exhibit 3, and suggests using a model to predict the rating change of Bond IV using leverage ratios, return on assets, and macroeconomic variables.

Exhibit 3 Partial One-Year Corporate Transition Matrix (entries in %)

From/To	AAA	AA	A
AAA	92.00	6.00	1.00
AA	2.00	89.00	8.00
A	0.05	1.00	85.00
Credit Spread (%)	0.50	1.00	1.75

Kreming calculates the risk-neutral probabilities, compares them with the actual default probabilities of bonds evaluated over the past 10 years, and observes that the actual and risk-neutral probabilities differ. She makes two observations regarding the comparison of these probabilities:

Observation 1 Actual default probabilities include the default risk premium associated with the uncertainty in the timing of the possible default loss.

Observation 2 The observed spread over the yield on a risk-free bond in practice includes liquidity and tax considerations, in addition to credit risk.

23 The expected exposure to default loss for Bond I is:

A less than the expected exposure for Bond II.

B the same as the expected exposure for Bond II.

C greater than the expected exposure for Bond II.

24 Based on Exhibit 1, the loss given default for Bond II is:

A less than that for Bond I.

B the same as that for Bond I.

C greater than that for Bond I.

25 Based on Exhibit 1, the expected future value of Bond I at maturity is *closest* to:

A 98.80.

B 103.74.

C 105.00.

26 Based on Exhibit 1, the risk-neutral default probability for Bond I is *closest* to:

A 2.000%.

B 3.175%.

C 4.762%.

27 Based on Exhibit 2, the credit valuation adjustment for Bond III is *closest* to:

A 3.3367.

B 3.5395.

C 5.8808.

28 Based on Exhibit 3, if Bond IV's credit rating changes during the next year to an A rating, its expected price change would be *closest* to:

A –8.00%.

B –7.35%.

C –3.15%.

29 Kreming's suggested model for Bond IV is a:

A structural model.

B reduced-form model.

C term structure model.

30 Which of Kreming's observations regarding actual and risk-neutral default probabilities is correct?

A Only Observation 1

B Only Observation 2

C Both Observation 1 and Observation 2

SOLUTIONS

1 B is correct. The following table shows that the credit valuation adjustment (CVA) for the bond is €36.49, the sum of the present values of expected loss. The steps taken to complete the table are as follows.

Step 1 Exposure at date T is $\frac{€1,000}{(1+r)^{4-T}}$, where r is 3%. That is, exposure is computed by discounting the face value of the bond using the risk-free rate and the number of years until maturity.

Step 2 Recovery = Exposure × Recovery rate.

Step 3 Loss given default (LGD) = Exposure – Recovery.

Step 4 Probability of default (POD) on Date 1 is 1.50%. The probability of survival (POS) on Date 1 is 98.50%.

For subsequent dates, POD is calculated as the annual default probability multiplied by the previous date's POS.

For example, to determine the Date 2 POD (1.4775%), the annual default probability (1.50%) is multiplied by the Date 1 POS (98.50%).

Step 5 POS in Dates 2–4 = POS in the previous year – POD.

That is, POS in year T = POS in year (T – 1) – POD in year T.

POS can also be determined by subtracting the annual default probability from 100% and raising it to the power of the number of years:

$(100\% - 1.5000\%)^1 = 98.5000\%$.

$(100\% - 1.5000\%)^2 = 97.0225\%$.

$(100\% - 1.5000\%)^3 = 95.5672\%$.

$(100\% - 1.5000\%)^4 = 94.1337\%$.

Step 6 Expected loss = LGD × POD.

Step 7 Discount factor (DF) for date T is $\frac{1}{(1+r)^T}$, where r is 3%.

Step 8 PV of expected loss = Expected loss × DF.

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0								
1	€915.14	€274.54	€640.60	1.5000%	98.5000%	€9.61	0.970874	€9.33
2	€942.60	€282.78	€659.82	1.4775%	97.0225%	€9.75	0.942596	€9.19
3	€970.87	€291.26	€679.61	1.4553%	95.5672%	€9.89	0.915142	€9.05
4	€1,000.00	€300.00	€700.00	1.4335%	94.1337%	€10.03	0.888487	€8.92
							CVA =	€36.49

The value of the bond if it were default free would be 1,000 × DF for Date 4 = €888.49.

Fair value of the bond considering CVA = €888.49 – CVA = €888.49 – €36.49 = €852.00.

Because the market price of the bond (€875) is greater than the fair value of €852, B is correct.

A is incorrect because the market price of the bond differs from its fair value. C is incorrect because although the bond's value if the bond were default free is greater than the market price, the bond has a risk of default, and CVA lowers its fair value to below the market price.

2 B is correct. The recovery rate to be used now in the computation of fair value is 30% × 1.25 = 37.5%, whereas the default probability to be used is 1.50% × 1.25 = 1.875%.

Using the steps outlined in the solution to Question 1, the following table is prepared, which shows that the bond's CVA increases to 40.49. Thus, Koning concludes that a change in the probability of default has a greater effect on fair value than a similar change in the recovery rate. The steps taken to complete the table are the same as those in the previous problem. There are no changes in exposures and discount factors in this table.

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0								
1	€915.14	€343.18	€571.96	1.8750%	98.1250%	€10.72	0.970874	€10.41
2	€942.60	€353.47	€589.12	1.8398%	96.2852%	€10.84	0.942596	€10.22
3	€970.87	€364.08	€606.80	1.8053%	94.4798%	€10.95	0.915142	€10.03
4	€1,000.00	€375.00	€625.00	1.7715%	92.7083%	€11.07	0.888487	€9.84
							CVA =	€40.49

Changes in the default probability and recovery rates do not affect the value of the default-free bond. So, it is the same as in the previous question: €888.49.

Fair value of the bond considering CVA = €888.49 – CVA = €888.49 – €40.49 = €848.00

3 A is correct. The following table shows that the CVA for the bond is €42.17, the sum of the present values of expected loss. The steps taken to complete the table are as follows.

Step 1 Exposure at Date 4 is €1,000 + Coupon amount = €1,000 + €60 = €1,060. Exposure at a date T prior to that is the coupon on date T + PV at date T of subsequent coupons + PV of €1,000 to be received at Date 4. For example, exposure at Date 2 is

$$€60 + \frac{€60}{1 + 0.03} + \frac{€60}{(1 + 0.03)^2} + \frac{€1{,}000}{(1 + 0.03)^2} = €60 + \frac{€60}{1 + 0.03} + \frac{€1{,}060}{(1 + 0.03)^2}$$
$$= €1{,}117.40.$$

Steps 2 through 8 are the same as those in the solution to Question 1.

Date	Exposure	Recovery	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0								
1	€1,144.86	€343.46	€801.40	1.5000%	98.5000%	€12.02	0.970874	€11.67
2	€1,117.40	€335.22	€782.18	1.4775%	97.0225%	€11.56	0.942596	€10.89
3	€1,089.13	€326.74	€762.39	1.4553%	95.5672%	€11.10	0.915142	€10.15
4	€1,060.00	€318.00	€742.00	1.4335%	94.1337%	€10.64	0.888487	€9.45
							CVA =	€42.17

The value of the bond if it were default free would be €60 × DF_1 + €60 × DF_2 + €60 × DF_3 + €1,060 × DF_4 = €1,111.51.

Fair value of the bond considering CVA = €1,111.51 – €42.17 = €1,069.34.

4 A is correct. If default occurs on Date 3, the rate of return can be obtained by solving the following equation for internal rate of return (IRR):

$$€1{,}090 = \frac{€60}{1 + \text{IRR}} + \frac{€60}{(1 + \text{IRR})^2} + \frac{€326.74}{(1 + \text{IRR})^3}.$$

In this equation, €60 is the amount of coupon received at Dates 1 and 2 prior to default at Date 3. The amount €326.74 is the recovery at Time 3 (from the CVA table in the solution to the previous question). The solution to the foregoing equation can be obtained using the cash flow IRR function on your calculator.

5 B is correct. For each possible transition, the expected percentage price change, computed as the product of the modified duration and the change in the spread as shown in Exhibit 3 (relating to question 5), is calculated as follows:

From AA to AAA: –2.75 × (0.60% – 0.90%) = +0.83%.

From AA to A: –2.75 × (1.10% – 0.90%) = –0.55%.

From AA to BBB: –2.75 × (1.50% – 0.90%) = –1.65%.

From AA to BB: –2.75 × (3.40% – 0.90%) = –6.88%.

From AA to B: –2.75 × (6.50% – 0.90%) = –15.40%.

From AA to C: –2.75 × (9.50% – 0.90%) = –23.65%.

The expected percentage change in the value of the AA rated bond is computed by multiplying each expected percentage price change for a possible credit transition by its respective transition probability given in Exhibit 3 and summing the products:

(0.0150 × 0.83%) + (0.8800 × 0%) + (0.0950 × –0.55%) + (0.0075 × –1.65%) + (0.0015 × –6.88%) + (0.0005 × –15.40%) + (0.0003 × –23.65%)= –0.0774%.

Therefore, the expected return on the bond over the next year is its YTM minus 0.0774%, assuming no default.

6 B is correct. Statement B is correct because a reduced-form credit model involves regression analysis using information generally available in the financial markets, such as the measures mentioned in the statement.

Statement A is incorrect because it is consistent with the use of a structural model and not a reduced-form model. It is a structural model that is based on the premise that a firm defaults on its debt if the value of its assets falls below its liabilities and that the probability of that event has the characteristics of an option.

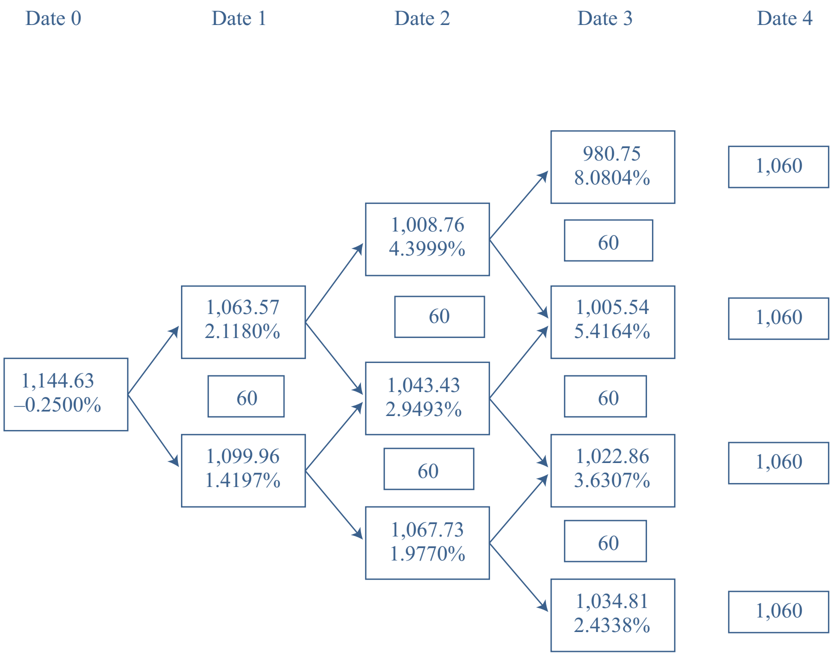
Statement C is incorrect because it is consistent with the use of a structural model and not a reduced-form model. A structural model involves the estimation of a default barrier, and default occurs if the value of firm's assets falls below the default barrier.

7 C is correct. Structural models require information best known to the managers of the company. Reduced-form models require information only generally available in financial markets.

A is incorrect because although it is literally true, when the models were developed is immaterial. Structural models are currently used in practice by commercial banks and credit rating agencies.

B is incorrect because computer technology facilitates valuation using option pricing models as well as regression analysis.

8 A is correct. The following tree shows the valuation assuming no default of Bond B2, which pays a 6% annual coupon.



The scheduled year-end coupon and principal payments are placed to the right of each forward rate in the tree. For example, the Date 4 values are the principal plus the coupon of 60. The following are the four Date 3 values for the bond, shown above the interest rate at each node:

$$€1,060/1.080804 = €980.75.$$

$$€1,060/1.054164 = €1,005.54.$$

$$€1,060/1.036307 = €1,022.86.$$

$$€1,060/1.024338 = €1,034.81.$$

These are the three Date 2 values:

$$\frac{(0.5 \times €980.75) + (0.5 \times €1,005.54) + €60}{1.043999} = €1,008.76.$$

$$\frac{(0.5 \times €1,005.54) + (0.5 \times €1,022.86) + €60}{1.029493} = €1,043.43.$$

$$\frac{(0.5 \times €1,022.86) + (0.5 \times €1,034.81) + €60}{1.019770} = €1,067.73.$$

These are the two Date 1 values:

$$\frac{(0.5 \times €1,008.76) + (0.5 \times €1,043.43) + €60}{1.021180} = €1,063.57.$$

$$\frac{(0.5 \times €1,043.43) + (0.5 \times €1,067.73) + €60}{1.014197} = €1,099.96.$$

This is the Date 0 value:

$$\frac{(0.5 \times €1{,}063.57) + (0.5 \times €1{,}099.96) + €60}{0.997500} = €1{,}144.63.$$

So, the value of the bond assuming no default is 1,144.63. This value could also have been obtained more directly using the benchmark discount factors from Exhibit 2:

€60 × 1.002506 + €60 × 0.985093 + €60 × 0.955848 + €1,060 × 0.913225 = €1,144.63.

The benefit of using the binomial interest rate tree to obtain the VND is that the same tree is used to calculate the expected exposure to default loss.

The credit valuation adjustment table is now prepared following these steps:

Step 1 Compute the expected exposures as described in the following, using the binomial interest rate tree prepared earlier.

The expected exposure for Date 4 is €1,060.

The expected exposure for Date 3 is

(0.1250 × €980.75) + (0.3750 × €1,005.54) + (0.3750 × €1,022.86) + (0.1250 × €1,034.81) + 60 = €1,072.60.

The expected exposure for Date 2 is

(0.25 × €1,008.76) + (0.50 × €1,043.43) + (0.25 × €1,067.73) + €60 = €1,100.84.

The expected exposure for Date 1 is

(0.50 × €1,063.57) + (0.50 × €1,099.96) + 60 = €1,141.76.

Step 2 LGD = Exposure × (1 – Recovery rate).

Step 3 The initial default probability is 1.50%. For subsequent dates, POD is calculated as the default probability multiplied by the previous date's POS.

For example, to determine the Date 2 POD (1.4775%), the default probability (1.5000%) is multiplied by the Date 1 POS (98.5000%).

Step 4 POS is determined by subtracting the default probability from 100% and raising it to the power of the number of years:

$(100\% - 1.5000\%)^1 = 98.5000\%$.

$(100\% - 1.5000\%)^2 = 97.0225\%$.

$(100\% - 1.5000\%)^3 = 95.5672\%$.

$(100\% - 1.5000\%)^4 = 94.1337\%$.

Step 5 Expected loss = LGD × POD.

Step 6 Discount factors in year T are obtained from Exhibit 2.

Step 7 PV of expected loss = Expected loss × DF.

Date	Exposure	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0							
1	€1,141.76	€799.23	1.5000%	98.5000%	€11.99	1.002506	€12.02

(continued)

Date	Exposure	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
2	€1,100.84	€770.58	1.4775%	97.0225%	€11.39	0.985093	€11.22
3	€1,072.60	€750.82	1.4553%	95.5672%	€10.93	0.955848	€10.44
4	€1,060.00	€742.00	1.4335%	94.1337%	€10.64	0.913225	€9.71
						CVA =	€43.39

Fair value of the bond considering CVA = €1,144.63 – CVA = €1,144.63 – €43.39 = €1,101.24.

9 A is correct. The corporate bond's fair value is computed in the solution to Question 8 as €1,101.24. The YTM can be obtained by solving the following equation for IRR:

$$€1{,}101.24 = \frac{€60}{1 + \text{IRR}} + \frac{€60}{(1 + \text{IRR})^2} + \frac{€60}{(1 + \text{IRR})^3} + \frac{€1{,}060}{(1 + \text{IRR})^4}.$$

The solution to this equation is 3.26%.

Valuation of a four-year, 6% coupon bond under no default is computed in the solution to Question 8 as 1,144.63. So, the YTM of a theoretical comparable-maturity government bond with the same coupon rate as the corporate bond, B2, can be obtained by solving the following equation for IRR:

$$€1{,}144.63 = \frac{€60}{1 + \text{IRR}} + \frac{€60}{(1 + \text{IRR})^2} + \frac{€60}{(1 + \text{IRR})^3} + \frac{€1{,}060}{(1 + \text{IRR})^4}.$$

The solution to this equation is 2.18%. So, the credit spread that the analyst wants to compute is 3.26% – 2.18% = 1.08%, or 108 bps.

B is incorrect because it is the spread over the four-year government par bond that has a YTM of 2.25% in Exhibit 2: 3.26% – 2.25% = 1.01%, or 101 bps. Although this spread is commonly used in practice, the analyst is interested in finding the spread over a theoretical 6% coupon government bond.

C is incorrect because it is the YTM of the coupon four-year government bond in Exhibit 2.

10 B is correct. The recovery rate to be used now in the computation of fair value is 30% × 0.75 = 22.500%, whereas the default probability to be used is 1.50% × 0.75 = 1.125%.

The tree that shows the valuation assuming no default of Bond B2 in the solution to Question 8 will not be affected by the foregoing changes. Accordingly, VND remains €1,144.63.

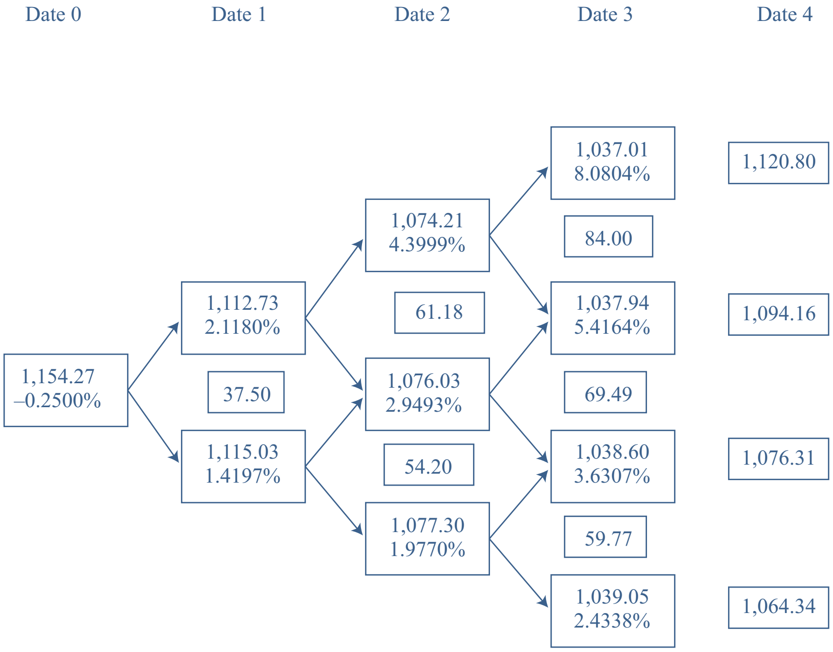
Following the steps outlined in the solution to Question 8, the following table is prepared, which shows that the CVA for the bond decreases to €36.23. Thus, Ibarra concludes that a decrease in the probability of default has a greater effect on fair value than a similar decrease in the recovery rate. The steps taken to complete the table are the same as those in Question 8. There are no changes in exposures or discount factors in this table.

Date	Exposure	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0							
1	€1,141.76	€884.87	1.1250%	98.8750%	€9.95	1.002506	€9.98
2	€1,100.84	€853.15	1.1123%	97.7627%	€9.49	0.985093	€9.35

Date	Exposure	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
3	€1,072.60	€831.26	1.0998%	96.6628%	€9.14	0.955848	€8.74
4	€1,060.00	€821.50	1.0875%	95.5754%	€8.93	0.913225	€8.16
						CVA =	€36.23

Fair value of the bond considering CVA = €1,144.63 – CVA = €1,144.63 – €36.23 = €1,108.40.

11 A is correct. The following tree shows the valuation assuming no default of the floating-rate note (FRN), B4, which has a quoted margin of 4%.



The scheduled year-end coupon and principal payments are placed to the right of each forward rate in the tree. For example, the four Date 4 values are the principal plus the coupon.

€1,000 × (1 + 0.080804 + 0.04) = €1,120.80.

€1,000 × (1 + 0.054164 + 0.04) = €1,094.16.

€1,000 × (1 + 0.036307 + 0.04) = €1,076.31.

€1,000 × (1 + 0.024338 + 0.04) = €1,064.34.

The following are the four Date 3 bond values for the note, shown above the interest rate at each node:

€1,120.80/1.080804 = €1,037.01.

€1,094.16/1.054164 = €1,037.94.

€1,076.31/1.036307 = €1,038.60.

€1,064.34/1.024338 = €1,039.05.

The three Date 3 coupon amounts are computed based on the interest rate at Date 2 plus the quoted margin of 4%:

€1,000 × (0.043999 + 0.04) = €84.00.

€1,000 × (0.029493 + 0.04) = €69.49.

€1,000 × (0.019770 + 0.04) = €59.77.

There are three Date 2 bond values:

$$\frac{(0.5 \times €1{,}037.01) + (0.5 \times €1{,}037.94) + €84.00}{1.043999} = €1{,}074.21.$$

$$\frac{(0.5 \times €1{,}037.94) + (0.5 \times €1{,}038.60) + €69.49}{1.029493} = €1{,}076.03.$$

$$\frac{(0.5 \times €1{,}038.60) + (0.5 \times €1{,}039.05) + €59.77}{1.019770} = €1{,}077.30.$$

The two Date 2 coupon amounts are computed based on the interest rate at Date 1 plus the quoted margin of 4%:

€1,000 × (0.021180 + 0.04) = €61.18.

€1,000 × (0.014197 + 0.04) = €54.20.

The Date 1 coupon amount is computed based on the interest rate at date 0 plus the quoted margin of 4%:

€1,000 × (–0.0025 + 0.04) = €37.50.

These are the calculations for the bond values for Date 1 and Date 0:

$$\frac{(0.5 \times €1{,}074.21) + (0.5 \times €1{,}076.03) + €61.18}{1.021180} = €1{,}112.73.$$

$$\frac{(0.5 \times €1{,}076.06) + (0.5 \times €1{,}077.30) + €54.20}{1.014197} = €1{,}115.0.$$

Then, the VND is calculated as follows:

$$\frac{(0.5 \times €1{,}112.73) + (0.5 \times €1{,}115.03) + €37.50}{0.9975} = €1{,}154.27.$$

The expected exposures are then computed using the binomial interest rate tree prepared earlier. For example, the expected exposure for Date 4 is computed as follows:

(0.125 × €1,120.80) + (0.375 × €1,094.16) + (0.375 × €1,076.31) + (0.125 × €1,064.34) = €1,087.07.

Similarly, the expected exposure for Date 3 is computed as follows:

(0.125 × €1,037.01) + (0.375 × €1,037.94) + (0.375 × €1,038.60) + (0.125 × €1,039.05) + (0.250 × €84) + (0.500 × €69.49) + (0.250 × €59.77) = €1,108.90.

The expected exposures for Dates 2 and 1 are computed similarly, and the credit valuation adjustment table is completed following Steps 2–7 outlined in the solution to Question 8.

Date	Exposure	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0							
1	€1,151.38	€805.97	1.5000%	98.5000%	€12.09	1.002506	€12.12

Date	Exposure	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
2	€1,133.58	€793.51	1.4775%	97.0225%	€11.72	0.985093	€11.55
3	€1,108.90	€776.23	1.4553%	95.5672%	€11.30	0.955848	€10.80
4	€1,087.07	€760.95	1.4335%	94.1337%	€10.91	0.913225	€9.96
						CVA =	€44.43

Fair value of the FRN considering CVA = €1,154.27 – CVA = €1,154.27 – €44.43 = €1,109.84.

Because the market price of €1,070 is less than the estimated fair value, the analyst should recommend adding to existing positions in the FRN.

B and C are incorrect because the FRN is perceived to be undervalued in the market.

12 A is correct. The changing probability of default will not affect the binomial tree prepared in the solution to Question 11. The Date 1 value remains €1,154.27, which is also the VND. The expected exposures, loss given default, and discount factors are also unaffected by the changing probability of default. The following is the completed credit valuation adjustment table.

Date	Exposure	LGD	POD	POS	Expected Loss	DF	PV of Expected Loss
0							
1	€1,151.38	€805.97	1.5000%	98.5000%	€12.09	1.002506	€12.12
2	€1,133.58	€793.51	0.4925%	98.0075%	€3.91	0.985093	€3.85
3	€1,108.90	€776.23	0.4900%	97.5175%	€3.80	0.955848	€3.64
4	€1,087.07	€760.95	0.4876%	97.0299%	€3.71	0.913225	€3.39
						CVA =	€22.99

Thus, CVA decreases to €22.99.

13 C is correct. The credit rating agencies typically make incremental changes, as seen in a transition matrix provided in Exhibit 3. Ibarra believes the bond is undervalued, because her assessment of the probability of default and the recovery rate is more optimistic than that of the agencies. Therefore, she most likely expects the credit rating agencies to put the issuer on a positive watch.

A is incorrect because the bond is perceived to be undervalued, not overvalued. Ibarra is not expecting a credit downgrade.

B is incorrect because it is not the *most likely* expectation. The rating agencies rarely change an issuer's rating from BBB all the way to AAA. In Exhibit 3 (relating to question 5) the probability of a BBB rated issuer going from BBB to AAA is 0.02%, whereas to go from BBB to A it is 4.80%.

14 A is correct.

B is incorrect because, although generally true for investment-grade bonds, the statement neglects the fact that high-yield issuers sometimes face a downward-sloping credit term structure. Credit term structures are not *always* upward sloping.

C is incorrect because there is a consistent pattern for the term structure of credit spreads: Typically, it is upwardly sloped because greater time to maturity is associated with higher projected probabilities of default and lower recovery rates.

15 C is correct. A covered bond is a senior debt obligation of a financial institution that gives recourse to the originator/issuer as well as a predetermined underlying collateral pool. Each country or jurisdiction specifies the eligible collateral types as well as the specific structures permissible in the covered bond market. Covered bonds usually have either commercial or residential mortgages meeting specific criteria or public sector exposures as underlying collateral.

A is incorrect. The term "covered" is used in foreign exchange analysis, for instance, "covered interest rate parity." In the context of securitized debt, a covered bond is secured by specific assets in addition to the overall balance sheet of the issuer.

B is incorrect because a covered bond does not involve a credit default swap. In addition, an issuer is not likely to sell a credit default swap on its own liability.

16 A is correct. Credit spread migration typically reduces the expected return for two reasons. First, the probabilities for rating changes are not symmetrically distributed around the current rating; they are skewed toward a downgrade rather than an upgrade. Second, the increase in the credit spread is much larger for downgrades than is the decrease in the spread for upgrades.

17 A is correct. The expected return on the Entre Corp. bond over the next year is its yield to maturity plus the expected percentage price change in the bond over the next year. In the following table, for each possible transition, the expected percentage price change is the product of the bond's modified duration of 7.54, multiplied by –1, and the change in the spread, weighted by the given probability:

Expected percentage price change = (0.0002 × 6.786%) + (0.0030 × 4.524%) + (0.0480 × 3.016%) + (0.8573 × 0.000%) + (0.0695 × –14.326%) + (0.0175 × –37.700%) + (0.0045 × –60.320%) = –1.76715%.

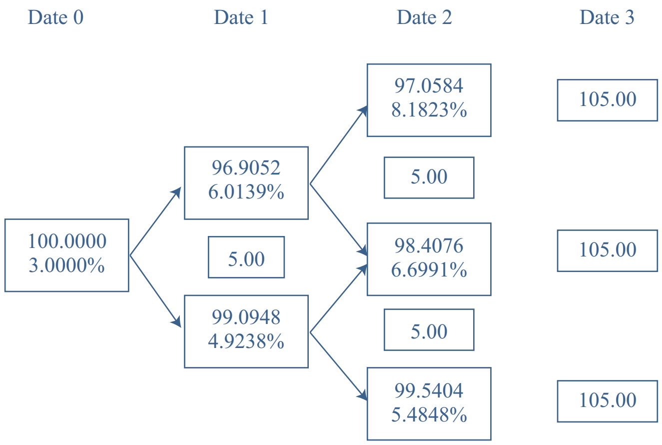
So, the expected return on the Entre Corp. bond is its yield to maturity plus the expected percentage price change due to credit migration:

Expected return = 5.50% – 1.77% = 3.73%.

	Expected % Price Change (1)	Probability (2)	Expected % Price Change × Probability (1 × 2)
From BBB to AAA	–7.54 × (0.60% – 1.50%) = 6.786%	0.0002	0.00136
From BBB to AA	–7.54 × (0.90% – 1.50%) = 4.524%	0.0030	0.01357
From BBB to A	–7.54 × (1.10% – 1.50%) = 3.016%	0.0480	0.14477
From BBB to BB	–7.54 × (3.40% – 1.50%) = –14.326%	0.0695	–0.99566
From BBB to B	–7.54 × (6.50% – 1.50%) = –37.700%	0.0175	–0.65975
From BBB to CCC, CC, or C	–7.54 × (9.50% – 1.50%) = –60.320%	0.0045	–0.27144
		Total:	–1.76715

18 C is correct. The credit spread can be calculated in three steps:

Step 1 Estimate the value of the three-year VraiRive bond assuming no default. Based on Kowalski's assumptions and Exhibits 2 and 3, the value of the three-year VraiRive bond assuming no default is 100.0000.



Supporting calculations:

The bond value in each node is the value of next period's cash flows discounted by the forward rate. For the three nodes on Date 2, the bond values are as follows:

105/1.081823 = 97.0584.

105/1.066991 = 98.4076.

105/1.054848 = 99.5404.

For the two nodes on Date 1, the two bond values are as follows:

[(0.5 × 97.0584) + (0.5 × 98.4076) + 5.00]/1.060139 = 96.9052.

[(0.5 × 98.4076) + (0.5 × 99.5404) + 5.00]/1.049238 = 99.0948.

Finally, for the node on Date 0, the bond value is

[(0.5 × 96.9052) + (0.5 × 99.0948) + 5.00]/1.030000 = 100.0000.

Therefore, the VND for the VraiRive bond is 100.0000.

Step 2 Calculate the credit valuation adjustment, and then subtract the CVA from the VND from Step 1 to establish the fair value of the bond. The CVA equals the sum of the present values of each year's expected loss and is calculated as follows:

Date	Expected Exposure	Loss Given Default	Probability of Default	Discount Factor	Present Value of Expected Loss
1	103.0000	68.6667	2.0000%	0.970874	1.3333
2	103.3535	68.9023	1.9600%	0.920560	1.2432
3	105.0000	70.0000	1.9208%	0.862314	1.1594
				CVA =	3.7360

Supporting calculations:

The expected exposures at each date are the bond values at each node, weighted by their risk-neutral probabilities, plus the coupon payment:

Date 1: (0.5 × 96.9052) + (0.5 × 99.0948) + 5.00 = 103.0000.

Date 2: (0.25 × 97.0584) + (0.5 × 98.4076) + (0.25 × 99.5404) + 5.00 = 103.3535.

Date 3: 105.0000

The loss given default on each date is 2/3 of the expected exposure.

The probability of default on each date is as follows:

Date 1: 2%

Date 2: 2% × (100% – 2%) = 1.96%.

Date 3: 2% × $(100\% - 2\%)^2$ = 1.9208%.

The discount factor on each date is 1/(1 + spot rate for the date) raised to the correct power.

Finally, the credit valuation adjustment each year is the product of the LGD times the POD times the discount factor, as shown in the last column of the table. The sum of the three annual CVAs is 3.7360.

So, the fair value of the VraiRive bond is the VND less the CVA, or VND – CVA = 100 – 3.7360 = 96.2640.

Step 3 Based on the fair value from Step 2, calculate the yield to maturity of the bond, and solve for the credit spread by subtracting the yield to maturity on the benchmark bond from the yield to maturity on the VraiRive bond. The credit spread is equal to the yield to maturity on the VraiRive bond minus the yield to maturity on the three-year benchmark bond (which is 5.0000%). Based on its fair value of 96.2640, the VraiRive bond's yield to maturity is

$$96.2640 = \frac{5}{(1 + \text{YTM})} + \frac{5}{(1 + \text{YTM})^2} + \frac{105}{(1 + \text{YTM})^3}$$

Solving for YTM, the yield to maturity is 6.4082%. Therefore, the credit spread on the VraiRive bond is 6.4082% – 5.0000% = 1.4082%.

19 C is correct. A decrease in the risk-neutral probability of default would decrease the credit valuation adjustment and decrease the credit spread. In contrast, increasing the bond's loss-given-default assumption and increasing the probability-of-default assumption would increase the credit valuation adjustment and decrease the fair value of the bond (and increase the yield to maturity and the credit spread over its benchmark).

20 A is correct. For investment-grade bonds with the highest credit ratings, credit spreads are extremely low, and credit migration is possible only in one direction given the implied lower bound of zero on credit spreads. As a result, the credit term structure for the most highly rated securities tends to be either flat or slightly upward sloping. Securities with lower credit quality, however, face greater sensitivity to the credit cycle. Credit spreads would decrease, not increase, with the expectation of economic growth. There is a countercyclical relationship between credit spreads and benchmark rates over the business cycle. A strong economic climate is associated with higher benchmark yields but lower credit spreads because the probability of issuers defaulting declines in such good times.

21 A is correct. Positive-sloped credit spread curves may arise when a high-quality issuer with a strong competitive position in a stable industry has low leverage, strong cash flow, and a high profit margin. This type of issuer tends to exhibit very low short-term credit spreads that rise with increasing maturity given greater uncertainty due to the macroeconomic environment, potential adverse changes in the competitive landscape, technological change, or other factors that drive a higher implied probability of default over time. Empirical academic studies also tend to support the view that the credit spread term structure is upward sloping for investment-grade bond portfolios.

22 B is correct. The auto ABS is granular, with many small loans relative to the size of the total portfolio. The auto loans are also homogeneous. These characteristics support using the portfolio-based approach. A loan-by-loan approach would be inefficient because of the large number of basically similar loans; this approach is best for a portfolio of discrete, large loans that are heterogeneous. A statistics-based approach would work for a static book of loans, whereas the auto loan portfolio would be dynamic and would change over time.

23 B is correct. The expected exposure is the projected amount of money that an investor could lose if an event of default occurs, before factoring in possible recovery. The expected exposure for both Bond I and Bond II is 100 + 5 = 105.

24 C is correct. The loss given default is a positive function of the expected exposure to default loss and a negative function of the recovery rate. Because Bond II has a lower recovery rate than Bond I and the same expected exposure to default loss (100 + 5 = 105), it will have a higher loss given default than Bond I will have. The loss given default for Bond I is 105 × (1 – 0.40) = 63.00. The loss given default for Bond II is 105 × (1 – 0.35) = 68.25.

25 B is correct. In the event of no default, the investor is expected to receive 105. In the event of a default, the investor is expected to receive 105 – [105 × (1 – 0.40)] = 42. The expected future value of the bond is, therefore, the weighted average of the no-default and default amounts, or (105 × 0.98) + (42 × 0.02) = 103.74.

26 B is correct. The risk-neutral default probability, P^*, is calculated using the current price, the expected receipt at maturity with no default (that is, 100 + 5 = 105), the expected receipt at maturity in the event of a default (that is, 0.40 × 105 = 42), and the risk-free rate of interest (0.03):

$$100 = \frac{\left[105 \times \left(1 - P^*\right)\right] + \left(42 \times P^*\right)}{1.03}.$$

Solving for P^* gives 0.031746, or 3.1746%.

27 A is correct. The CVA is the sum of the present value of expected losses on the bond, which from Exhibit 2 is 3.3367.

28 C is correct. The expected percentage price change is the product of the negative of the modified duration and the difference between the credit spread in the new rating and the old rating:

Expected percentage price change = –4.2 × (0.0175 – 0.01) = –0.0315, or –3.15%.

29 B is correct. A reduced-form model in credit risk analysis uses historical variables, such as financial ratios and macroeconomic variables, to estimate the default intensity. A structural model for credit risk analysis, in contrast, uses option pricing and relies on a traded market for the issuer's equity.

30 B is correct. Observation 1 is incorrect, but Observation 2 is correct. The actual default probabilities do not include the default risk premium associated with the uncertainty in the timing of the possible default loss. The observed spread over the yield on a risk-free bond in practice does include liquidity and tax considerations, in addition to credit risk.

READING

32

Credit Default Swaps

by Brian Rose and Don M. Chance, PhD, CFA

Brian Rose (USA). Don M. Chance, PhD, CFA, is at Louisiana State University (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe credit default swaps (CDS), single-name and index CDS, and the parameters that define a given CDS product;
☐	**b.** describe credit events and settlement protocols with respect to CDS;
☐	**c.** explain the principles underlying and factors that influence the market's pricing of CDS;
☐	**d.** describe the use of CDS to manage credit exposures and to express views regarding changes in the shape and/or level of the credit curve;
☐	**e.** describe the use of CDS to take advantage of valuation disparities among separate markets, such as bonds, loans, equities, and equity-linked instruments.

INTRODUCTION

1

Derivative instruments in which the underlying is a measure of a borrower's credit quality are widely used and well established in a number of countries. We explore basic definitions of such instruments, explain the main concepts, cover elements of valuation and pricing, and discuss applications.

BASIC DEFINITIONS AND CONCEPTS

2

a describe credit default swaps (CDS), single-name and index CDS, and the parameters that define a given CDS product

© 2021 CFA Institute. All rights reserved.

A **credit derivative** is a derivative instrument in which the underlying is a measure of a borrower's credit quality. Four types of credit derivatives are (1) total return swaps, (2) credit spread options, (3) credit-linked notes, and (4) credit default swaps, or CDS. CDS are the most liquid of the four and, as such, are the topic we focus on. In a CDS, one party makes payments to the other and receives in return the promise of compensation if a third party defaults.

In any derivative, the payoff is based on (derived from) the performance of an underlying instrument, rate, or asset that we call the "underlying." For a CDS, the underlying is the credit quality of a borrower. At its most fundamental level, a CDS provides compensation equal to expected recovery when a credit event occurs, but it also changes in value to reflect changes in the market's perception of a borrower's credit quality well in advance of default. The value of a CDS will rise and fall as opinions change about the likelihood and severity of a potential default. The actual event of default might never occur, but a decline in the price of a bond when investors perceive an increase in the likelihood of default is a mark-to-market loss to the bondholder. The most common credit events include bankruptcy, failure to pay, and restructuring. Another type of credit event which may be encountered in sovereign and municipal government bond markets is a moratorium or, more drastically, a repudiation of debt in which the governmental authority declares a moratorium on payments due under the terms of the obligation or challenges the validity of the entire debt obligation. (Other, less common credit events are also defined in the International Swaps and Derivatives Association's Credit Derivatives Definitions, but we will not consider them here.) Credit default swaps are designed to protect creditors against credit events such as these. The industry has expended great effort to provide clear guidance on what credit events are covered by a CDS contract. As with all efforts to write a perfect contract, however, no such device exists and disputes do occasionally arise. We will take a look at these issues later.

In addition to hedging credit risk, investors use CDS to

- leverage their portfolios,
- access maturity exposures not available in the cash market,
- access credit risk while limiting interest rate risk, and
- improve the liquidity of their portfolios given the illiquidity in the corporate bond market.

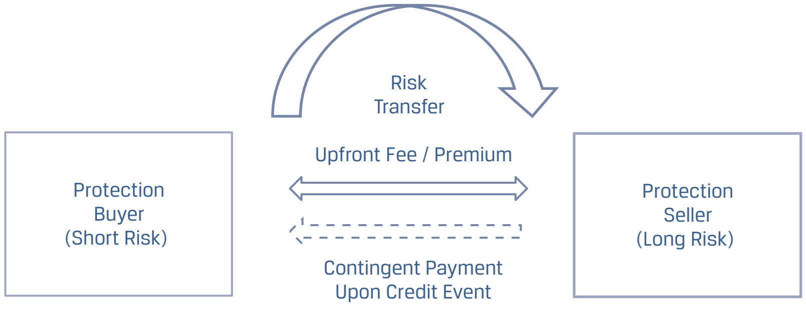
In addition, the CDS market has increased transparency and insight into the actual cost of credit risk. The higher relative liquidity and relative sophistication of CDS investors allow for more accurate price discovery and facilitate trading during liquidity events when the cash market for bonds becomes illiquid. While many of the applications listed above are beyond the scope of this reading, a basic understanding of this important fixed-income tool is necessary for all investment professionals.

Let's now define a **credit default swap**:

> *A credit default swap is a derivative contract between two parties, a credit protection buyer and credit protection seller, in which the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from a credit event in an underlying.*

In a CDS contract there are two counterparties, the **credit protection buyer** and the **credit protection seller**. The buyer agrees to make a series of periodic payments to the seller over the life of the contract (which are determined and fixed at contract initiation) and receives in return a promise that if default occurs, the protection seller will compensate the protection buyer. If default occurs, the periodic payments made by the protection buyer to the protection seller terminate. Exhibit 1 shows the structure of payment flows.

Exhibit 1 Payment Structure of a CDS



Credit default swaps are somewhat similar to put options. Put options effectively enable the option holder to sell (put) the underlying security to the option seller if the underlying performs poorly relative to the exercise price. Similarly, in the event of a credit event on the underlying security, the buyer of credit protection receives a payment from the credit protection seller equal to the par or notional value of the security less the expected recovery value. If the credit quality of the underlying deteriorates but there is no outright credit event, the credit protection buyer is compensated only if the contract is unwound. How that compensation occurs and how much protection it provides are some points we will discuss.

A CDS does not eliminate credit risk. The definition of a default in the swap contract may not perfectly align with a traditional default event, so the magnitude of the change in value of the contract may differ from the change in value of the underlying. In addition, the credit protection buyer assumes counterparty risk with respect to the credit protection seller. Although there are no guarantees that the credit protection seller will not default, as was seen with several large financial institutions in the financial crisis that started in 2007, most credit protection sellers are relatively high-quality borrowers. If they were not, they could not be active sellers of credit protection.

The majority of CDS are written on debt issued by corporate borrowers, which will be our focus in this reading. But note that CDS can also be written on the debt of sovereign governments and state and local governments. In addition, CDS can be written on portfolios of loans, mortgages, or debt securities.

2.1 Types of CDS

There are three types of CDS: single-name CDS, index CDS, and tranche CDS. Other CDS-related instruments, such as options on CDS (or CDS swaptions) are beyond the scope of this discussion. A CDS on one specific borrower is called a **single-name CDS**. The borrower is called the **reference entity**, and the contract specifies a **reference obligation**, a particular debt instrument issued by the borrower. Only a small subset of issuers, typically with large outstanding liquid debt, have single-name CDS. The designated instrument is usually a senior unsecured obligation, but the reference obligation is not the only instrument covered by the CDS. Any debt obligation issued by the borrower that is ranked equal to or higher than the reference obligation with respect to the priority of claims is covered. The payoff of the CDS is determined by the **cheapest-to-deliver** obligation, which is the debt instrument that can be purchased and delivered at the lowest cost but has the same seniority as the reference obligation.

EXAMPLE 1

Cheapest-to-Deliver Obligation

Assume that a company with several debt issues trading in the market files for bankruptcy (i.e., a credit event takes place). What is the cheapest-to-deliver obligation for a CDS contract where the reference bond is a five-year senior unsecured bond?

A A subordinated unsecured bond trading at 20% of par

B A five-year senior unsecured bond trading at 50% of par

C A two-year senior unsecured bond trading at 45% of par

Solution:

C is correct. The cheapest-to-deliver, or lowest-priced, instrument is the two-year senior unsecured bond trading at 45% of par. Although the bond in A trades at a lower dollar price, it is subordinated and, therefore, does not qualify for coverage under the CDS. Note that even though the CDS holder holds the five-year bonds, he will receive payment on the CDS based on the cheapest-to-deliver obligation, not the specific obligation he holds.

A second type of credit default swap, an **index CDS**, involves a portfolio of single-name CDS. This type of instrument allows participants to take positions on the credit risk of a combination of companies, in much the same way that investors can trade index or exchange-traded funds that are combinations of the equities of companies. The two most commonly traded CDS index products are the North American indexes (CDS) and the European, Asian, and Australian indexes (iTraxx). Correlation of defaults is a strong determinant of a portfolio's behavior. For index CDS, this concept takes the form of a factor called **credit correlation**, and it is a key determinant of the value of an index CDS. Analyzing the effects of those correlations is a highly specialized subject, but be aware that much effort is placed on modeling how defaults by certain companies are connected to defaults by other companies. The more correlated the defaults, the more costly it is to purchase protection for a combination of the companies. In contrast, for a diverse combination of companies whose defaults have low correlations, it will be much less expensive to purchase protection.

A third type of CDS is the **tranche CDS**, which covers a combination of borrowers but only up to pre-specified levels of losses—much in the same manner that asset-backed securities are divided into tranches, each covering particular levels of losses. Coverage of tranche CDS is beyond the scope of this reading.

3 IMPORTANT FEATURES OF CDS MARKETS AND INSTRUMENTS, CREDIT AND SUCCESSION EVENTS, AND SETTLEMENT PROPOSALS

b describe credit events and settlement protocols with respect to CDS

As we will describe in more detail later, the CDS market is large, global, and well organized. The unofficial industry governing body is the International Swaps and Derivatives Association (ISDA), which publishes industry-supported conventions that facilitate the functioning of the market. Parties to CDS contracts generally agree that their contracts will conform to ISDA specifications. These terms are specified in a document called the **ISDA Master Agreement**, which the parties to a CDS sign. In Europe, the standard CDS contract is called the Standard Europe Contract, and in the United States and Canada, it is called the Standard North American Contract. Other standardized contracts exist for Asia, Australia, Latin America, and a few other specific countries.

Each CDS contract specifies a **notional amount**, or "notional" for short, which is the amount of protection being purchased. The notional amount can be thought of as the *size* of the contract. It is important to understand that the total notional amount of CDS can exceed the amount of debt outstanding of the reference entity. As we will discuss later, the credit protection buyer does not have to be an actual creditor holding exposure (i.e., owning a loan, bond, or other debt instrument). It can be simply a party that believes that there will be a change in the credit quality of the reference entity.

As with all derivatives, the CDS contract has an expiration or maturity date, and coverage is provided up to that date. The typical maturity range is 1 to 10 years, with 5 years being the most common and actively traded maturity, but the two parties can negotiate any maturity. Maturity dates are typically the 20th day of March, June, September, or December. The March and September maturity dates are the most liquid, as these are when the index CDS contracts roll.

The buyer of a CDS pays a periodic premium to the seller, referred to as the **CDS spread**, which is a return over a market reference rate required to protect against credit risk. It is sometimes referred to as a credit spread. Conceptually, it is the same as the credit spread on a bond, the compensation for bearing credit risk.

An important advancement in the development of CDS has been in establishing standard annual coupon rates on CDS contracts. (Note that the reference bond will make payments that are referred to collectively as the coupon while a CDS on the reference bond will have its own coupon rate.) Formerly, the coupon rate on the CDS was set at the credit spread. If a CDS required a rate of 4% to compensate the protection seller for the assumption of credit risk, the protection buyer made quarterly payments amounting to 4% annually. Now CDS coupon rates are standardized, with the most common coupons being either 1% or 5%. The 1% rate typically is used for a CDS on an investment-grade company or index, and the 5% rate is used for a CDS on a high-yield company or index. Obviously, either standardized rate might not be the appropriate rate to compensate the seller. Clearly, not all investment-grade companies have equivalent credit risk, and not all high-yield companies have equivalent credit risk. In effect, the standard rate may be too high or too low. This discrepancy is accounted for by an **upfront payment**, commonly called the **upfront premium**. The differential between the credit spread and the standard rate is converted to a present value basis. Thus, a credit spread greater than the standard rate would result in a cash payment from the protection buyer to the protection seller. Similarly, a credit spread less than the standard rate would result in a cash payment from the protection seller to the protection buyer.

Regardless of whether either party makes an upfront payment, the reference entity's credit quality could change during the life of the contract, thereby resulting in changes in the value of the CDS. These changes are reflected in the price of the CDS in the market. Consider a high-yield company with a 5% credit spread and a CDS coupon of 5%. Therefore, there is no upfront payment. The protection buyer simply agrees to make payments equal to 5% of the notional over the life of the CDS. Now suppose that at some later date, the reference entity experiences a decrease in its credit quality. The credit protection buyer is thus paying 5% for risk that now merits a rate higher than 5%. The coverage and cost of protection are the same, but the risk being covered is greater. The value of the CDS to the credit protection buyer has, therefore, increased, and if desired, she could unwind the position to capture the gain. The credit protection seller has experienced a loss in value of the instrument because he is receiving 5% to cover a risk that is higher than it was when the contract was initiated. It should be apparent that absent any other exposure to the reference entity, if the credit quality of the reference entity decreases, the credit protection buyer gains and the credit protection seller loses. The market value of the CDS reflects these gains and losses.

The terminology in CDS markets can be confusing. In equity and fixed-income markets, we think of buyers as being long and sellers as being short. In the CDS market, however, that is not always true. In single-name CDS, the *buyer* of credit protection is *short credit exposure* and the *seller* of credit protection is *long credit exposure*. This is consistent with the fact that in the financial world, "shorts" are said to benefit when things go badly. When credit quality deteriorates, the credit protection buyer benefits, and when it improves, the credit protection seller benefits. To make things even more confusing, though, the opposite is true in CDS index positions: The *buyer* of a CDX is *long* credit exposure and the seller of a CDX is *short* credit exposure. To minimize the confusion, we use the terms *credit protection seller* and *credit protection buyer* throughout our discussion. .

3.1 Credit and Succession Events

The **credit event** is what defines default by the reference entity—that is, the event that triggers a payment from the credit protection seller to the credit protection buyer. This event must be unambiguous: Did it occur, or did it not? For the market to function well, the answer to this question must be clear.

As previously mentioned, the most common credit events include bankruptcy, failure to pay, and restructuring. **Bankruptcy** is a declaration provided for by a country's laws that typically involves the establishment of a legal procedure that forces creditors to defer their claims. Bankruptcy essentially creates a temporary fence around the company through which the creditors cannot pass. During the bankruptcy process, the defaulting party works with its creditors and the court to attempt to establish a plan for repaying the debt. If that plan fails, there is likely to be a full liquidation of the company, at which time the court determines the payouts to the various creditors. Until liquidation occurs, the company normally continues to operate. Many companies do not liquidate and are able to emerge from bankruptcy. A bankruptcy filing by the reference entity is universally regarded as a credit event in CDS contracts.

Another credit event recognized in standard CDS contracts is **failure to pay**, which occurs when a borrower does not make a scheduled payment of principal or interest on an outstanding obligation after a grace period, without a formal bankruptcy filing. (Failure to pay credit events are defined in the CDS contract. ISDA contracts define failure to pay events uniformly, but the same is not true for bespoke CDS.) The third type of event, **restructuring**, refers to a number of possible events, including reduction or deferral of principal or interest, change in seniority or priority of an obligation, or change in the currency in which principal or interest is scheduled to be paid. To qualify as a credit event, the restructuring must be either involuntary or coercive. An involuntary credit event is one that is forced on the borrower by the creditors. A coercive credit event is one that is forced on the creditors by the borrower. Debt restructuring is not a credit event in the United States; issuers generally restructure under *bankruptcy*, which *is* a credit event. Restructuring is a credit event in other countries where the use of bankruptcy court to reorganize is less common. The Greek debt crisis is a good example of a restructuring that triggered a credit event.

Determination of whether a credit event occurs is done by a 15-member group within the ISDA called the Determinations Committee (DC). Each region of the world has a Determinations Committee, which consists of 10 CDS dealer (sell-side) banks and 5 non-bank (buy-side) end users. To declare a credit event, there must be a supermajority vote of 12 members.

The Determinations Committees also play a role in determining whether a **succession event** occurred. A succession event arises when there is a change in the corporate structure of the reference entity, such as through a merger, a divestiture, a spinoff, or any similar action in which ultimate responsibility for the debt in question becomes unclear. For example, if a company acquires all of the shares of a target company, it ordinarily assumes the target company's debt as well. Many mergers, however, are more complicated and can involve only partial acquisition of shares. Spinoffs and divestitures can also involve some uncertainty about who is responsible for certain debts. When such a question arises, it becomes critical for CDS holders. The question is ordinarily submitted to a Determinations Committee, and its resolution often involves complex legal interpretations of contract provisions and country laws. If a succession event is declared, the CDS contract is modified to reflect the DC's interpretation of whoever it believes becomes the obligor for the original debt. Ultimately, the CDS contract could be split among multiple entities.

3.2 Settlement Protocols

If the DC declares that a credit event has occurred, the two parties to a CDS have the right, but not the obligation, to settle. **Settlement** typically occurs 30 days after declaration of the credit event by the DC. CDS can be settled by **physical settlement** or by **cash settlement**. The former is less common and involves actual delivery of the debt instrument in exchange for a payment by the credit protection seller of the notional amount of the contract. In cash settlement, the credit protection seller pays cash to the credit protection buyer. Determining the amount of that payment is a critical factor because opinions can differ about how much money has actually been lost. The payment should essentially be the loss that the credit protection buyer has incurred, but determining that amount is not straightforward. Default on a debt does not mean that the creditor will lose the entire amount owed. A portion of the loss could be recovered. The percentage of the loss recovered is called the **recovery rate** (RR). (In most models, the recovery rate applies only to the principal.) The complement is called the **loss given default** (LGD), which is essentially an estimate of the expected credit loss. The **payout amount** is determined as the loss given default multiplied by the notional.

Loss given default = 1 – Recovery rate (%).

Payout amount = LGD × Notional.

Actual recovery can be a very long process, however, and can occur much later than the payoff date of the CDS. To determine an appropriate LGD, the industry conducts an auction in which major banks and dealers submit bids and offers for the cheapest-to-deliver defaulted debt. This process identifies the market's expectation for the recovery rate and the complementary LGD, and the CDS parties agree to accept the outcome of the auction, even though the actual recovery rate can ultimately be quite different, which is an important point if the CDS protection buyer also holds the underlying debt.

EXAMPLE 2

Settlement Preference

A French company files for bankruptcy, triggering various CDS contracts. It has two series of senior bonds outstanding: Bond A trades at 30% of par, and Bond B trades at 40% of par. Investor X owns €10 million of Bond A and owns €10 million of CDS protection. Investor Y owns €10 million of Bond B and owns €10 million of CDS protection.

1 Determine the recovery rate for both CDS contracts.
2 Explain whether Investor X would prefer to cash settle or physically settle her CDS contract or whether she is indifferent.
3 Explain whether Investor Y would prefer to cash settle or physically settle his CDS contract or whether he is indifferent.

Solution to 1:

Bond A is the cheapest-to-deliver obligation, trading at 30% of par, so the recovery rate for both CDS contracts is 30%.

Solution to 2:

Investor X has no preference between settlement methods. She can cash settle for €7 million [(1 – 30%) × €10 million] and sell her bond for €3 million, for total proceeds of €10 million. Alternatively, she can physically deliver her entire €10 million face amount of bonds to the counterparty in exchange for €10 million in cash.

Solution to 3:

Investor Y would prefer a cash settlement because he owns Bond B, which is worth more than the cheapest-to-deliver obligation. He will receive the same €7 million payout on his CDS contract but can sell Bond B for €4 million, for total proceeds of €11 million. If he were to physically settle his contract, he would receive only €10 million, the face amount of his bond.

3.3 CDS Index Products

So far, we have mostly been focusing on single-name CDS. As noted, there are also index CDS products. A company called Markit has been instrumental in producing CDS indexes. Of course, a CDS index is not in itself a traded instrument any more than a stock index is a traded product. As with the major stock indexes, however, the industry has created traded instruments based on the Markit indexes. These instruments are CDS that generate a payoff based on any default that occurs on any entity covered by the index.

The Markit indexes are classified by region and further classified (or divided) by credit quality. The two most commonly traded regions are North America and Europe. North American indexes are identified by the symbol CDX, and European, Asian, and Australian indexes are identified as iTraxx. Within each geographic category are investment-grade and high-yield indexes. The former are identified as CDX IG and iTraxx Main, each comprising 125 entities. The latter are identified as CDX HY, consisting of 100 entities, and iTraxx Crossover, consisting of up to 75 high-yield entities. Investment-grade index CDS are typically quoted in terms of spreads, whereas high-yield index CDS are quoted in terms of prices. Both types of products use standardized coupons. All CDS indexes are equally weighted. Thus, if there are 125 entities, the settlement on one entity is 1/125 of the notional. (Note that some confusion might arise from quoting certain CDS as prices and some as spreads, but keep in mind that the bond market quotes bonds often as prices and sometimes as yields. For example, a Treasury bond can be described as having a price of 120 or a yield of 2.68%. Both terms, combined with the other characteristics of the bond, imply the same concept.)

Markit updates the components of each index every six months by creating new series while retaining the old series. The latest-created series is called the **on-the-run** series, whereas the older series are called **off-the-run** series. When an investor moves from one series to a new one, the move is called a **roll**. When an entity within an index defaults, that entity is removed from the index and settled as a single-name CDS based on its relative proportion in the index. The index then moves forward with a smaller notional.

Index CDS are typically used to take positions on the credit risk of the sectors covered by the indexes as well as to protect bond portfolios that consist of or are similar to the components of the indexes. (An important reminder: When you *buy* a CDS index position, you are *long the credit exposure*, but when you *buy* a single-name CDS position, you have *bought credit protection*. To avoid confusion, we do not talk about buying and selling CDS herein but focus on the desired exposure, using the terms *buy protection* and *sell protection*.)

Standardization is generally undertaken to increase trading volume, which is somewhat limited in the single-name market with so many highly diverse entities. With CDS indexes on standardized portfolios based on the credit risk of well-identified companies, market participants have responded by trading them in large volumes. Indeed, index CDS are typically more liquid than single-name CDS, with average daily trading volume several times that of single-name CDS.

EXAMPLE 3

Hedging and Exposure Using Index CDS

Assume that an investor sells $500 million of protection using the CDX IG index, which has 125 reference entities. Concerned about the creditworthiness of a few of the components, the investor hedges a portion of the credit risk in each. For Company A, he purchases $3 million of single-name CDS protection, and Company A subsequently defaults.

1 What is the investor's net notional exposure to Company A?
2 What proportion of his exposure to Company A has he hedged?
3 What is the remaining notional on his index CDS trade?

Solution to 1:

The investor is long $4 million notional credit exposure ($500 million/125) through the index CDS and is short $3 million notional credit exposure through the single-name CDS. His net notional credit exposure is $1 million.

Solution to 2:

He has hedged 75% of his exposure ($3 million out of $4 million).

Solution to 3:

His index CDS has $496 million remaining notional credit exposure ($500 million original notional minus the $4 million notional related to Company A, which is no longer in the index).

3.4 Market Characteristics

Credit default swaps trade in the over-the-counter market. To better understand this market, we will first review how credit derivatives and specifically CDS were started.

As financial intermediaries, banks draw funds from savings-surplus sectors, primarily consumers, and channel them to savings-deficit sectors, primarily businesses. Corporate lending is a core element of banking. When a bank makes a corporate loan, it assumes two primary risks. One is that the borrower will not repay principal and interest, and the other is that interest rates will change such that the return the bank is earning on its outstanding loans is less than the rate available on comparable instruments in the marketplace. The former is called **credit risk** or **default risk**, and the latter is called **interest rate risk**. There are many ways to manage interest rate risk. Until around the mid-1990s, credit risk was largely managed using traditional methods—such as analysis of the borrower, its industry, and the macroeconomy—as well as control methods, such as credit limits, monitoring, and collateral. In effect, the only defenses against credit risk were to not make a loan, to lend but require collateral (the value of which is also at risk), or to lend and closely monitor the borrower, hoping that any problems could be foreseen and dealt with before a default occurred.

Around 1995, credit derivatives were created to provide a new and potentially more effective method of managing credit risk. They allow credit risk to be transferred from the lender to another party. In so doing, they facilitate the separation of interest rate risk from credit risk. Banks can then provide their most important service—lending—knowing that the credit risk can be transferred to another party if so desired. This ability to easily transfer credit risk allows banks to greatly expand their loan business. Given that lending is such a large and vital component of any economy, credit derivatives facilitate economic growth and have expanded to cover, and indeed are primarily focused on, the short-, intermediate-, and long-term bond markets. In fact, credit derivatives are more effective in the bond market, in which terms and conditions are far more standard, than in the bank loan market. Of the four types of credit derivatives, credit default swaps have clearly established themselves as the most widely used instrument. Indeed, in today's markets CDS are nearly the only credit derivative used to any great extent. CDS transactions are executed in the over-the-counter market by phone, instant message, or the Bloomberg message service. Trade information is reported to the **Depository Trust and Clearinghouse Corporation**, which is a US-headquartered entity providing post-trade clearing, settlement, and information services for many kinds of securities. Regulations require the central clearing of many CDS contracts, meaning that parties will send their contracts through clearinghouses that collect and distribute payments and impose margin requirements, as well as mark positions to market. Central clearing of CDS has risen dramatically since 2010. Currently, slightly more than half of all CDS are centrally cleared, up from just 10% in 2010.

The CDS market today is considerably smaller than it was prior to the 2008 financial crisis. The Bank for International Settlements reported that as of December 2019, the gross notional amount of CDS was about $7.6 trillion with a market value of $199 billion. (For comparison, the notional amounts for interest rate contracts—forward rate agreements, swaps, options—as of December 2019 was about $449 trillion.) As of December 2007, CDS gross notional was $57.9 trillion, nearly 8 times larger.

More than 90% of all CDS market activity is now derived from trading in five major CDS indexes: iTraxx Europe, iTraxx Europe Crossover, iTraxx Europe Senior Financials, CDX IG, and CDS HY.

BASICS OF VALUATION AND PRICING

4

c explain the principles underlying and factors that influence the market's pricing of CDS

Derivatives are typically priced by solving for the cost of a position that fully offsets the underlying exposure and earns the risk-free rate. In the context of CDS, this "price" is the CDS spread or upfront payment for a particular coupon rate under the contract. Although CDS are referred to as "swaps," they in fact resemble options because of the contingent nature of the payment made by the protection seller to the protection buyer if a credit event occurs as established by the ISDA Determinations Committee as outlined above.

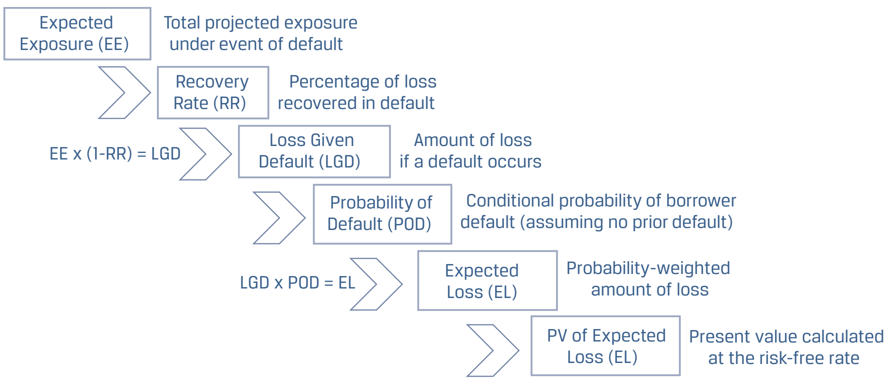
Unlike conventional derivative instruments, the CDS settlement amount under a credit event as declared by the ISDA Determinations Committee is far less clear than for derivatives whose underlying involves actively traded assets, such as equities, interest rates, or currencies. Credit does not "trade" in the traditional sense but, rather, exists implicitly within the bond and loan market. The unique debt structure and composition of each CDS reference entity adds to the complexity of establishing the basis between a CDS contract and a specific outstanding bond or loan.

The details of credit derivative models are beyond the scope of this reading, but it is important for investment industry analysts to have a thorough understanding of the factors that determine CDS pricing.

4.1 Basic Pricing Concepts

In our earlier coverage of credit strategies, we established that the credit valuation adjustment (CVA) may be thought of as the present value of credit risk for a loan, bond, or derivative obligation. In principle, the CVA should, therefore, be a reasonable approximation for the CDS hedge position outlined previously that would leave an investor with a risk-free rate of return. Exhibit 2 summarizes the CVA calculation for a financial exposure.

Exhibit 2 Credit Valuation Adjustment



CVA = Σ (PV of Expected Loss)

CVA is a function of expected exposure (EE), recovery rate, loss given default, the **probability of default** (POD) to arrive at an expected loss (EL), and a discount factor to arrive at the present value of expected loss.

Considering each of these CVA components in turn, the expected exposure reflects the notional value of the underlying CDS contract. Recall that the recovery rate is the percentage of loss recovered from a bond in default, whereas the loss given default is a function of the loss severity multiplied by the exposure amount.

The probability of default is a key element of CDS pricing that may be illustrated using a simple example. Consider a one-period CDS swap with no upfront payment where we ignore the time value of money and assume that default is possible only at maturity. The fair price of CDS protection for this period for a given borrower may be estimated as

$$\text{CDS spread} \approx (1 - \text{RR}) \times \text{POD}.$$

For example, if the probability of default is 2% and the recovery rate is 60%, the estimated CDS spread for the period would be 80 bps for the period. Assuming a \$100 notional contract value and a period of a year, the CDS contract fair value would be (the present value of) \$0.80.

It is important to note that the POD is a conditional probability over time. That is, assuming a two-period case, the probability of default in Period 2 is contingent on "surviving" to (i.e., not defaulting by) the end of Period 1. Note that we simplify the analysis by assuming discrete times of potential default versus the continuous time assumption common in CDS pricing models.

For example, consider a two-year, 5%, \$1,000 loan with one interest payment of \$50 due in one year and final interest and principal of \$1,050 due in two years. Assume further that we estimate a 2% chance of defaulting on the first payment and a 4% chance of defaulting on the second payment. To calculate the POD over the life of the loan, we first determine the **probability of survival** (POS) for Period 1. The POS is 0.98 (100% minus the 2% POD at T_1) multiplied by 0.96 (100% minus the 4% POD at T_2), approximately 94.08%. Thus, the POD over the life of the loan is 100% − 94.08% = 5.92%.

This conditional probability of default is also known as the **hazard rate**, as described in an earlier reading. The hazard rate is the probability that an event will occur *given that it has not already occurred.*

Now consider another possibility, a 10-year bond with an equivalent hazard rate of 2% each year. Suppose we want to know the probability that the borrower will not default during the entire 10-year period. The probability that a default will occur at some point during the 10 years is one minus the probability of no default in 10 years. The probability of no default in 10 years is $0.98 \times 0.98 \ldots 0.98 = 0.98^{10} = 0.817$. Thus, the probability of default is 1 − 0.817 = 0.183, or 18.3%. This somewhat simplified example illustrates how a low probability of default in any one period can turn into a surprisingly high probability of default over a longer period of time. Note that we have simplified the analysis by assuming a constant hazard rate, which may not be the case in practice.

EXAMPLE 4

Hazard Rate and Probability of Survival

Assume that a company's hazard rate is a constant 8% per year, or 2% per quarter. An investor sells five-year CDS protection on the company with the premiums paid quarterly over the next five years.

1 What is the probability of survival for the first quarter?

2 What is the conditional probability of survival for the second quarter?

3 What is the probability of survival through the second quarter?

Solution to 1:

The probability of survival for the first quarter is 98% (100% minus the 2% hazard rate).

Solution to 2:

The conditional probability of survival for the second quarter is also 98%, because the hazard rate is constant at 2%. In other words, *conditional on the company having survived the first quarter*, there is a 2% probability of default in the second quarter.

Solution to 3:

The probability of survival through the second quarter is 96.04%. The probability of survival through the first quarter is 98%, and the conditional probability of survival through the second quarter is also 98%. The probability of survival through the second quarter is thus 98% × 98% = 96.04%. Alternatively, 1 – 96.04% = 3.96% is the probability of default sometime during the first two quarters.

Understanding the concept of pricing a CDS is facilitated by recognizing that there are essentially two sides, or legs, of a contract. There is the **protection leg**, which is the contingent payment that the credit protection seller may have to make to the credit protection buyer, and the **premium leg**, which is the series of payments the credit protection buyer promises to make to the credit protection seller. Exhibit 3 provides an illustration of the process.

Exhibit 3 Determination of CDS Protection vs. Premium Legs

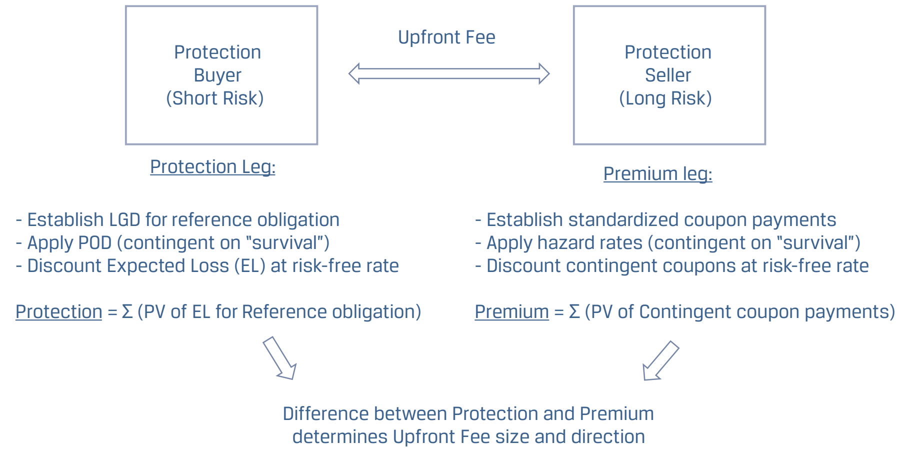


Exhibit 3 shows the upfront payment as the difference in value of the protection and premium legs. The party with a claim on the greater present value must pay the difference at the initiation date of the contract:

Upfront payment = PV (Protection leg) – PV (Premium leg).

If the result is greater (less) than zero, the protection buyer (seller) pays the protection seller (buyer). Actual CDS pricing and valuation models are more mathematically complex but are based on this conceptual framework.

4.2 The Credit Curve and CDS Pricing Conventions

The credit spread of a debt instrument is the rate in excess of a market reference rate (historically, Libor, although Libor is expected to be replaced by the end of 2021) that investors expect to receive to justify holding the instrument. The reference rate may itself contain some credit risk, as it reflects the rate at which commercial banks lend to one another. The credit spread can be expressed roughly as the probability of default multiplied by the loss given default, with LGD in terms of a percentage. The credit spreads for a range of maturities of a company's debt make up its **credit curve**. The credit curve is somewhat analogous to the term structure of interest rates, which is the set of rates on default-free debt over a range of maturities, but the credit curve applies to non-government borrowers and incorporates credit risk into each rate.

The CDS market for a given borrower is integrated with the credit curve of that borrower. In fact, given the evolution and high degree of efficiency of the CDS market, the credit curve is essentially determined by the CDS rates. The curve is affected by a number of factors, a key one of which is the set of aforementioned hazard rates. A constant hazard rate will tend to flatten the credit curve. Upward-sloping credit curves imply a greater likelihood of default in later years, whereas downward-sloping credit curves imply a greater probability of default in the earlier years. Downward-sloping curves are less common and often a result of severe near-term stress in the financial markets. The credit curve would not be completely flat even if the hazard rates are constant, because of discounting. For example, a company issuing 5- and 10-year zero-coupon bonds could have equally likely probabilities of default and hence equal expected payoffs. The present values of the payoffs are not the same, however, and so the discount rates that equate the present value to the expected payoffs will not be the same.

EXAMPLE 5

Change in Credit Curve

A company's 5-year CDS trades at a credit spread of 300 bps, and its 10-year CDS trades at a credit spread of 500 bps.

1 The company's 5-year spread is unchanged, but the 10-year spread widens by 100 bps. Describe the implication of this change in the credit curve.

2 The company's 10-year spread is unchanged, but the 5-year spread widens by 500 bps. Describe the implication of this change in the credit curve.

Solution to 1:

This change implies that although the company is not any riskier in the short term, its longer-term creditworthiness is less attractive. Perhaps the company has adequate liquidity for the time being, but after five years it must begin repaying debt or it will be expected to have cash flow difficulties.

Solution to 2:

This change implies that the company's near-term credit risk is now much greater. In fact, the probability of default will decrease if the company can survive for the next five years. Perhaps the company has run into liquidity issues that must be resolved soon, and if not resolved, the company will default.

4.3 CDS Pricing Conventions

With corporate bonds, we typically refer to their values in terms of prices or spreads. The spread is a more informative measure than price. A high-yield bond can be offered with a coupon equal to its yield and, therefore, a price of par value. An investment-grade bond with the same maturity can likewise be offered with a coupon equal to its yield, and therefore, its price is at par. These two bonds would have identical prices at the offering date, and their prices might even be close through much of their lives, but they are quite different bonds. Focusing on their prices would, therefore, provide little information. Their spreads are much more informative. With a market reference rate or the risk-free rate as a benchmark, investors can get a sense for the amount of credit risk implied by their prices, maturities, and coupons. The same is true for CDS. Although CDS have their own prices, their spreads are far more informative.

The reference entity will not necessarily have outstanding debt with credit spreads matching the 1% or 5% standardized coupons conventionally used in CDS contracts. Therefore, the present value of the promised payments from the credit protection buyer to the credit protection seller will most likely be different than the present value of the coupons on the reference entity's debt. The present value difference is the upfront premium paid from one party to the other.

Present value of credit spread = Upfront premium + Present value of fixed coupon.

A good rough approximation used in the industry is that the upfront premium is

Upfront premium ≈ (Credit spread – Fixed coupon) × Duration.

The upfront premium must ultimately be converted to a price, which is done by subtracting the percentage premium from 100.

Upfront premium % = 100 – Price of CDS in currency per 100 par.

Note that the duration used here is effective duration, since the cash flows arising from the coupon leg of the CDS are uncertain because they are contingent on the reference entity not defaulting.

EXAMPLE 6

Premiums and Credit Spreads

1. Assume a high-yield company's 10-year credit spread is 600 bps and the duration of the CDS is 8 years. What is the approximate upfront premium required to buy 10-year CDS protection? Assume high-yield companies have 5% coupons on their CDS.
2. Imagine an investor sold five-year protection on an investment-grade company and had to pay a 2% upfront premium to the buyer of protection. Assume the duration of the CDS to be four years. What are the company's credit spreads and the price of the CDS per 100 par?

Solution to 1:

To buy 10-year CDS protection, an investor would have to pay a 500 bp coupon plus the present value of the difference between that coupon and the current market spread (600 bps). In this case, the upfront premium would be approximately 100 bps × 8 (duration), or 8% of the notional.

Solution to 2:

The value of the upfront premium is equal to the premium (–2%) divided by the duration (4), or –50 bps. The sign of the upfront premium is negative because the seller is paying the premium rather than receiving it. The credit spread is equal to the fixed coupon (100 bps) plus the upfront premium, amortized over the duration of the CDS (–50 bps), or 50 bps. As a reminder, because the company's credit spread is less than the fixed coupon, the protection seller must pay the upfront premium to the protection buyer. The price in currency would be 100 minus the upfront premium, but the latter is negative, so the price is 100 – (–2) = 102.

4.4 Valuation Changes in CDS during Their Lives

As with any traded financial instrument, a CDS has a value that fluctuates during its lifetime. That value is determined in the competitive marketplace. Market participants constantly assess the current credit quality of the reference entity to determine its current value and (implied) credit spread. Clearly, many factors can change over the life of the CDS. By definition, the duration shortens through time. Likewise, the probability of default, the expected loss given default, and the shape of the credit curve will all change as new information is received. The exact valuation procedure of the CDS is precisely the same as it is when the CDS is first issued and simply incorporates the new inputs. The new market value of the CDS reflects gains and losses to the two parties.

Consider the following example of a five-year CDS with a fixed 1% coupon. The credit spread on the reference entity is 2.5%. In promising to pay 1% coupons to receive coverage on a company whose risk justifies 2.5% coupons, the present value of the protection leg exceeds the present value of the payment leg. The difference is the upfront premium, which will be paid by the credit protection buyer to the credit protection seller. During the life of the CDS, assume that the credit quality of the reference entity improves, such that the credit spread is now 2.1%. Now, consider a newly created CDS with the same remaining maturity and 1% coupon. The present value of the payment leg would still be less than the present value of the protection leg, but the difference would be less than it was when the original CDS was created because the risk is now less. Logically, it should be apparent that for the original transaction, the seller has gained and the buyer has lost. The difference between the original upfront premium and the new value is the seller's gain and buyer's loss. A rough approximation of the change in value of the CDS for a given change in spread is as follows:

Profit for the buyer of protection ≈ Change in spread in bps × Duration × Notional.

Alternatively, we might be interested in the CDS percentage price change, which is obtained as

% Change in CDS price = Change in spread in bps × Duration.

The percentage change in the price of a bond is approximately the change in its yield multiplied by its modified duration. For the CDS, the change in yield is analogous to the change in spread, measured in basis points. The duration of the CDS is analogous to the duration of the bond on which the CDS is written.

EXAMPLE 7

Profit and Loss from Change in Credit Spread

An investor buys $10 million of five-year protection, and the CDS contract has a duration of four years. The company's credit spread was originally 500 bps and widens to 800 bps.

1 Does the investor (credit protection buyer) benefit or lose from the change in credit spread?
2 Estimate the CDS price change and estimated profit to the investor.

Solution to 1:

The investor owns protection and is therefore short the credit exposure. As the credit spread widens (the credit quality of the underlying deteriorates), the value of the credit protection she owns increases.

Solution to 2:

The percentage price change is estimated as the change in spread (300 bps) multiplied by the duration (4), or 12%. The profit to the investor is 12% times the notional ($10 million), or $1.2 million.

4.5 Monetizing Gains and Losses

As with any financial instrument, changes in the price of a CDS give rise to opportunities to unwind the position and either capture a gain or realize a loss. This process is called **monetizing** a gain or loss. Keep in mind that the protection seller is effectively long the reference entity. He has entered into a contract to insure the debt of the reference entity, for which he receives a series of promised payments and possibly an upfront premium. He clearly benefits if the reference entity's credit quality improves because he continues to receive the same compensation but bears less risk. Using the opposite argument, the credit protection buyer benefits from a deterioration of the reference entity's credit quality. Thus, the credit protection seller is more or less long the company's bonds and the credit protection buyer is more or less short the company's bonds. As the company's credit quality changes through time, the market value of the CDS changes, giving rise to gains and losses for the CDS counterparties. The counterparties can realize those gains and losses by entering into new offsetting contracts, effectively selling their CDS positions to other parties.

Going back to the example in the previous section where the credit quality of the reference entity improved—the credit spread on the reference entity declined from 2.5% to 2.1%. The implied upfront premium on a new CDS that matches the terms of the original CDS with adjusted maturity is now the market value of the original CDS. The premium on the new CDS is smaller than that on the original CDS.

Now, suppose that the protection buyer in the original transaction wants to unwind her position. She would then enter into a new CDS as a protection seller and receive the newly calculated upfront premium. As we noted, this value is less than what he paid originally. Likewise, the protection seller in the original transaction could offset his position by entering into a new CDS as a protection buyer. He would pay an upfront premium that is less than what he originally received. The original protection buyer monetizes a loss, and the seller monetizes a gain. The transaction to unwind the CDS does not need to be done with the same original party, although doing so offers some advantages. Central clearing of CDS transactions facilitates the unwind transaction.

At this point, we have identified two ways of realizing a profit or loss on a CDS. One is to effectively exercise the CDS in response to a default. The other is to unwind the position by entering into a new offsetting CDS in the market. A third, less common method occurs if there is no default. A party can simply hold the position until expiration, at which time the credit protection seller has captured all of the premiums and has not been forced to make any payments, and the seller's obligation for any further payments is terminated. The spread of the CDS will go to zero, in much the same manner as a bond converges toward par as it approaches maturity.

The CDS seller clearly gains, having been paid to bear the risk of default that is becoming increasingly unlikely, and the CDS buyer loses. The buyer loses on the CDS because it paid premiums to receive protection in the event of a default, which did not occur. Although the CDS position itself is a loss, the buyer's overall position is not necessarily a loss. If the buyer is a creditor of the reference entity, the premium "loss" is no different than a homeowner's insurance premium payment on his house; he wouldn't consider that payment a loss simply because his house did not burn down.

5 APPLICATIONS OF CDS

d describe the use of CDS to manage credit exposures and to express views regarding changes in the shape and/or level of the credit curve

Credit default swaps, as demonstrated, facilitate the transfer of credit risk. As simple as that concept seems, there are many different circumstances under which CDS are used. In this section, we consider some applications of this instrument.

Any derivative instrument has two general uses. One is to exploit an expected movement in the underlying. The derivative typically requires less capital and is usually an easier instrument in which to create a short economic exposure as compared with the underlying. The derivatives market can also be more efficient, meaning that it can react to information more rapidly and have more liquidity than the market for the underlying. Thus, information or an expectation of movement in the underlying can often be exploited much more efficiently with the derivative than with the underlying directly.

The other trading opportunity facilitated by derivatives is in valuation differences between the derivative and the underlying. If the derivative is mispriced relative to the underlying, one can take the appropriate position in the derivative and an offsetting position in the underlying. If the valuation assessment is correct and other investors come to the same conclusion, the values of the derivative and underlying will converge, and the investor will earn a return that is essentially free of risk because the risk of the underlying has been hedged away by holding offsetting positions in the derivative and the underlying. Whether this happens as planned depends on both the efficiency of the market and the quality of the valuation model. Differences can also exist between the derivative and other derivatives on the same underlying.

These two general types of uses are also the major applications of CDS. We will refer to them as managing credit exposures, meaning the taking on or shedding of credit risk in light of changing expectations and/or valuation disparities. With valuation disparities, the focus is on differences in the pricing of credit risk in the CDS market relative to that of the underlying bonds.

5.1 Managing Credit Exposures

The most basic application of a CDS is to increase or decrease credit exposure. The most obvious such application is for a lender to buy protection to reduce its credit exposure to a borrower. For the seller of protection, the trade adds credit exposure. A lender's justification for using a CDS seems obvious. The lender may have assumed too much credit risk but does not want to sell the bond or loan because there can be significant transaction costs, because later it may want the bond or loan back, or because the market for the bond or loan is relatively illiquid. If the risk is temporary, it is almost always easier to temporarily reduce risk by using a CDS. Beyond financial institutions, any organization exposed to credit risk is potentially a candidate for using CDS.

The justification for selling credit protection is somewhat less obvious. The seller can be a CDS dealer, whose objective is to profit from making markets in CDS. A dealer typically attempts to manage its exposure by either diversifying its credit risks or hedging the risk by entering into a transaction with yet another party, such as by shorting the debt or equity of the reference entity, often accompanied by investment of the funds in a repurchase agreement, or repo. If the dealer manages the risk effectively, the risk assumed in selling the CDS is essentially offset when the payment for assuming the risk exceeds the cost of removing the risk. Achieving this outcome successfully requires sophisticated credit risk modeling.

Although dealers make up a large percentage of protection sellers, not all sellers are dealers. Consider that any bondholder is a buyer of credit and interest rate risk. If the bondholder wants only credit risk, it can obtain it by selling protection, which would require far less capital and incur potentially lower overall transaction costs than buying the bond. Moreover, the CDS can be more liquid than the bond, so the position can be unwound much more easily.

As noted, it is apparent why a party making a loan might want credit protection. Consider, however, that a party with no exposure to the reference entity might also purchase credit protection. Such a position is called a **naked credit default swap**, and it has resulted in some controversy in regulatory and political circles. In buying protection without owning the underlying, the investor is taking a position that the entity's credit quality will deteriorate, whereas the seller of protection without owning the underlying is taking the position that the entity's credit quality will improve or that the CDS was overpriced.

Some regulators and politicians believe it is inappropriate for a party with no exposure to a borrower to speculate that the borrower's financial condition will deteriorate. This controversy accelerated during the financial crisis of 2008–2009 because many investors bought protection without owning the underlying and benefited from the crisis.

The counterargument, however, is that elsewhere in the financial markets, such bets are made all of the time in the form of long puts, short futures, and short sales of stocks and bonds. These instruments are generally accepted as a means of protecting oneself against poor performance in the financial markets. Credit protection is also a means of protecting oneself against poor performance. In addition, proponents of naked CDS argue that they bring liquidity to the credit market, potentially providing more stability, not less. Nonetheless, naked CDS trading is banned in Europe for sovereign debt, although it is generally permitted otherwise.

CDS trading strategies, with or without naked exposure, can take several forms. An investor can choose to be long or short the credit exposure, as we have previously discussed. Alternatively, the party can be a credit protection seller on one reference entity and a credit protection buyer on a different entity. This is called a **long/short credit trade**. This transaction is a bet that the credit position of one entity will improve relative to that of another. The two entities might be related in some way or might produce substitute goods. For example, one might take a position that because of competition and changes in the luxury car industry, the credit quality of Daimler will improve and that of BMW will weaken, so selling protection on Daimler and buying protection on BMW would be appropriate. Similarly, an investor may undertake a long/short trade based on other factors, such as environmental, social, and governance (ESG) considerations. For instance, an investor may be concerned about a company's poor ESG-related practices and policies relative to another company. In this case, the investor could buy protection using the CDS of a company with weak ESG practices and policies and sell protection using the CDS of a company with strong ESG practices and policies. Example 8 provides a case study of ESG considerations in a long/short ESG trade.

EXAMPLE 8

Long/Short Trade with ESG Considerations

Overview

An analyst is evaluating two US apparel companies: Atelier and Trapp. Atelier is a large company that focuses on high-end apparel brands. It is profitable despite a high cost structure. Trapp is smaller and less profitable than Atelier. Trapp focuses on less expensive brands and strives to keep costs low. Both companies purchase their merchandise from suppliers all over the world. The analyst recognizes that apparel companies must maintain adequate oversight over their suppliers to control the risks of reputational damage and inventory disruptions. Supplier issues are particularly relevant for Atelier and Trapp following a recent fire that occurred at the factory of Global Textiles, a major supplier to both companies. The fire resulted in multiple casualties and unfavorable news headlines.

The analyst notices a significant difference in the way Atelier and Trapp approach ESG considerations. After the fire at its supplier, Atelier signed an "Accord on Fire and Building Safety," which is a legally binding agreement between global apparel manufacturers, retailers, and trade unions in the country where the fire occurred. After signing the accord, Atelier made a concerted effort to fix and enhance machinery in factories of its suppliers. Its objective was to improve workplace safety—notably, to reduce lost employee time due to factory incidents and the rate of factory accidents and fatalities.

Investors view Atelier's corporate governance system favorably because management interests and stakeholder interests are strongly aligned. Atelier's board of directors includes a high percentage of independent directors and is notably diverse. In contrast, Trapp's founder is the majority owner of the company and serves as CEO and chairman of the board of directors. Furthermore, Trapp's board is composed mainly of individuals who have minimal industry expertise. As a consequence, Trapp's board was unprepared to adequately respond to the Global Textiles fire. Given the lack of independence and expertise of Trapp's board, investors consider Trapp's corporate governance system to be poor. Because of its emphasis on low costs and reflecting its less experienced board, Trapp chose not to sign the accord.

Implications for CDS

Single-name CDS on both Atelier and Trapp are actively traded in the market, although Trapp's CDS is less liquid. Before the Global Textiles fire, five-year CDS for Trapp traded at a spread of 250 bps, compared to a spread of 150 bps for the five-year CDS for Atelier. The difference in spreads reflects Trapp's lower trading liquidity, perceived lower creditworthiness (primarily reflecting its smaller size and lower profitability), and hence higher default risk relative to Atelier.

After the Global Textiles fire, spreads on the CDS for all companies in the apparel sector widened considerably. Credit spreads for the five-year CDS on Atelier widened by 60 bps (to 210 bps), and credit spreads for the five-year CDS on Trapp widened by 75 bps (to 325 bps). The analyst believes that over the longer term, the implications of the fire at Global Textiles will be even more adverse for Trapp relative to Atelier. The analyst's view largely reflects Trapp's higher ESG-related risks, especially the perceived weaker safety in its factories and its weaker corporate governance system. In particular, the analyst believes that spreads of Trapp's CDS will remain wider than their pre-fire level of 250 bps, but Atelier's CDS spreads will return to their pre-fire level of 150 bps.

Describe how the analyst can use CDS to exploit the potential opportunity.

Solution

The analyst can try to exploit the potential opportunity by buying protection (shorting the credit) on Trapp using five-year CDS and selling protection (going long the credit) on Atelier using five-year CDS. This trade would reflect both the anticipated continuing adverse spreads for Trapp relative to the pre-fire level and the return of spreads for Atelier to their lower pre-fire levels. For example, assume Atelier's five-year CDS spread returns to 150 bps from 210 bps, but Trapp's five-year CDS spread narrows to just 300 bps from 325 bps. The difference in spreads between the two companies' CDS would have widened from 115 bps (325 bps – 210 bps) right after the factory fire occurred to 150 bps (300 bps – 150 bps). This 35 bp difference in spread would represent profit (excluding trading costs) to the analyst from the long/short trade.

Similar to a long/short trade involving individual entities (companies), an investor might also create a long/short trade using CDS indexes. For example, if the investor anticipates a weakening economy, she could buy protection using a high-yield CDS index and sell protection using an investment-grade CDS index. As high-yield spreads widen relative to investment-grade spreads, the trade would realize a profit. As another example, a trader expecting a strengthening in the Asian economy relative to the European economy could buy protection using a European CDS index and sell protection using an Asian CDS index. As Asia spreads narrow relative to European spreads, the trade would realize a profit.

Another type of long/short trade, called a **curve trade**, involves buying single-name or index protection at one maturity and selling protection on the same reference entity at a different maturity. Consider two CDS maturities, which we will call the short-term and the long-term to keep things simple. We will assume the more common situation of an upward-sloping credit curve, meaning that long-term CDS rates (and credit spreads) are higher than short-term rates. If the curve changes shape, it becomes either steeper or flatter. A steeper (flatter) curve means that long-term credit risk increases (decreases) relative to short-term credit risk. An investor who believes that long-term credit risk will increase relative to short-term credit risk (credit curve steepening) can buy protection by buying a long-term single-name CDS or selling a long-term CDS index and sell protection by selling a short-term single-name CDS or buying a short-term CDS index. In the short run, a curve-steepening trade is bullish. It implies that the short-term outlook for the reference entity is better than the long-term outlook. In the short run, a curve-flattening trade is bearish. It implies that the short-run outlook for the reference entity looks worse than the long-run outlook and reflects the expectation of near-term problems for the reference entity.

EXAMPLE 9

Curve Trading

An investor owns some intermediate-term bonds issued by a company and has become concerned about the risk of a near-term default, although he is not very concerned about a default in the long term. The company's two-year CDS currently trades at 350 bps, and the four-year CDS is at 600 bps.

1 Describe a potential curve trade that the investor could use to hedge the default risk.
2 Explain why an investor may prefer to use a curve trade as a hedge against the company's default risk rather than simply buying protection on the reference entity.

Solution to 1:

The investor anticipates a flattening credit curve for the reference company, with spreads rising at the shorter end of the curve. Thus, he would buy credit protection on the two year (buy the two-year single-name CDS) while selling credit protection further out on the curve (sell the four-year single-name CDS).

Solution to 2:

The long/short trade reduces the cost of buying near-term credit protection, with the cost of the credit protection offset by the premium received from selling protection further out on the curve. This works only as long as the investor's expectations about the relative risk of near- and longer-term default hold true.

Of course, there can be changes to the credit curve that take the form of simple shifts in the general level of the curve, whereby all spreads go up or down by roughly equal amounts. As with long-duration bonds relative to short-duration bonds, the values of longer-term CDS will move more than those of shorter-term CDS. As an example, a trader who believes that all spreads will go up will want to be a buyer of credit protection but will realize that longer-term CDS will move more than short-term CDS. Thus, she might want to buy protection at the longer part of the curve and hedge by selling protection at the shorter part of the curve. She will balance the sizes of the positions so that the volatility of the position she believes will gain in value will be more than that of the other position. If more risk is desired, she might choose to trade only the more volatile leg.

6 VALUATION DIFFERENCES AND BASIS TRADING

e describe the use of CDS to take advantage of valuation disparities among separate markets, such as bonds, loans, equities, and equity-linked instruments

Different investors will have different assessments of the price of credit risk. Such differences of opinion will lead to valuation disparities. Clearly, there can be only one appropriate price at which credit risk can be eliminated, but that price is not easy to determine. The party that has the best estimate of the appropriate price of credit risk can capitalize on its knowledge or ability at the expense of another party. Any such comparative advantage can be captured by trading the CDS against either the reference entity's debt or equity or derivatives on its debt or equity, but such trading is critically dependent on the accuracy of models that isolate the credit risk component of the return. The details of those models are left to CDS specialists, but it is important for candidates to understand the basic ideas.

The yield on the bond issued by the reference entity to a CDS contains a factor that reflects the credit risk. In principle, the amount of yield attributable to credit risk on the bond should be the same as the credit spread on a CDS. It is, after all, the compensation paid to the party assuming the credit risk, regardless of whether that risk is borne by a bondholder or a CDS seller. But there may be a difference in the credit risk compensation in the bond market and CDS market. This differential pricing can arise from mere differences of opinions, differences in models used by participants in the two markets, differences in liquidity in the two markets, and supply and demand conditions in the repo market, which is a primary source of financing for bond purchases. A difference in the credit spreads in these two markets is the foundation of a strategy known as a **basis trade**.

The general idea behind most basis trades is that any such mispricing is likely to be temporary and the spreads should return to equivalence when the market recognizes the disparity. For example, suppose the bond market implies a 5% credit risk premium whereas the CDS market implies a 4% credit risk premium. The trader does not know which is correct but believes these two rates will eventually converge. From the perspective of the CDS, its risk premium is too low relative to the bond credit risk premium. From the perspective of the bond, its risk premium is too high relative to the CDS market, which means its price is too low. So, the CDS market could be pricing in too little credit risk, and/or the bond market could be pricing in too much credit risk. Either market could be correct; it does not matter. The investor would buy the bond at a price that appears to overestimate its credit risk and, at the same time, buy credit protection at what appears to be an unjustifiably low premium, simultaneously hedging interest rate risk exposure with a duration strategy or interest rate derivatives. The risk is balanced because the default potential on the bond is protected by the CDS. If convergence occurs, the trade would capture the 1% differential in the two markets.

To determine the profit potential of such a trade, it is necessary to decompose the bond yield into the risk-free rate plus the funding spread plus the credit spread. The risk-free rate plus the funding spread is essentially the market reference rate. The credit spread is then the excess of the yield over the market reference rate and can be compared with the credit spread in the CDS market. If the spread is higher in the bond market than in the CDS market, it is said to be a negative basis. If the spread is higher in the CDS market than in the bond market, it is said to be a positive basis. Note that in practice, the above decomposition can be complicated by the existence of embedded options, such as with callable and convertible bonds or when the bond is not selling near par. Those factors would need to be accounted for in the calculations.

EXAMPLE 10

Bonds vs. Credit Default Swaps

An investor wants to be long the credit risk of a given company. The company's bond currently yields 6% and matures in five years. A comparable five-year CDS contract has a credit spread of 3.25%. The investor can borrow at Libor, which is currently 2.5%.

1 Calculate the bond's credit spread.
2 Identify a basis trade that would exploit the current situation.

Solution to 1:

The bond's credit spread is equal to the yield (6%) minus the market reference rate (2.5%). Therefore, the bond's credit spread is currently 3.5%.

Solution to 2:

The bond and CDS markets imply different credit spreads. Credit risk is cheap in the CDS market (3.25%) relative to the bond market (3.5%). The investor should buy protection in the CDS market at 3.25% and go long the bond, with its 3.5% credit spread, netting 25 bps.

Another type of trade using CDS can occur within the instruments issued by a single entity. Credit risk is an element of virtually every unsecured debt instrument or the capital leases issued by a company. Each of these instruments is priced to reflect the appropriate credit risk. Investors can use the CDS market to first determine whether any of these instruments is incorrectly priced relative to the CDS and then buy the cheaper one and sell the more expensive one. Again, there is the assumption that the market will adjust. This type of trading is much more complex, however, because priority of claims means that not all of the instruments pay off equally if default occurs.

EXAMPLE 11

Using CDS to Trade on a Leveraged Buyout

An investor believes that a company will undergo a leveraged buyout (LBO) transaction, whereby it will issue large amounts of debt and use the proceeds to repurchase all of the publicly traded equity, leaving the company owned by management and a few insiders.

1 Why might the CDS spread change?

2 What equity-versus-credit trade might an investor execute in anticipation of such a corporate action?

Solution to 1:

Taking on the additional debt will almost surely increase the probability of default, thereby increasing the CDS spread.

Solution to 2:

The investor might consider buying the stock and buying credit protection. Both legs will profit if the LBO occurs because the stock price will rise as the company repurchases all outstanding equity and the CDS price will rise as its spread widens to reflect the increased probability of default.

CDS indexes also create an opportunity for a type of arbitrage trade. If the cost of the index is not equivalent to the aggregate cost of the index components, an investor might go long the cheaper instrument and short the more expensive instrument. There is the implicit assumption that convergence will occur. If it does, the investor gains the benefit while basically having neutralized the risk. Transaction costs in this type of arbitrage trade can be quite significant and nullify the profit potential for all but the largest investors.

SUMMARY

- A credit default swap (CDS) is a contract between two parties in which one party purchases protection from another party against losses from the default of a borrower for a defined period of time.
- A CDS is written on the debt of a third party, called the reference entity, whose relevant debt is called the reference obligation, typically a senior unsecured bond.
- A CDS written on a particular reference obligation normally provides coverage for all obligations of the reference entity that have equal or higher seniority.
- The two parties to the CDS are the credit protection buyer, who is said to be short the reference entity's credit, and the credit protection seller, who is said to be long the reference entity's credit.
- The CDS pays off upon occurrence of a credit event, which includes bankruptcy, failure to pay, and, in some countries, involuntary restructuring.
- Settlement of a CDS can occur through a cash payment from the credit protection seller to the credit protection buyer as determined by the cheapest-to-deliver obligation of the reference entity or by physical delivery of the reference obligation from the protection buyer to the protection seller in exchange for the CDS notional.
- A cash settlement payoff is determined by an auction of the reference entity's debt, which gives the market's assessment of the likely recovery rate. The credit protection buyer must accept the outcome of the auction even though the ultimate recovery rate could differ.
- CDS can be constructed on a single entity or as indexes containing multiple entities. Bespoke CDS or baskets of CDS are also common.
- The fixed payments made from CDS buyer to CDS seller are customarily set at a fixed annual rate of 1% for investment-grade debt or 5% for high-yield debt.
- Valuation of a CDS is determined by estimating the present value of the payment leg, which is the series of payments made from the protection buyer to the protection seller, and the present value of the protection leg, which is the payment from the protection seller to the protection buyer in event of default. If the present value of the payment leg is greater than the present value of the protection leg, the protection buyer pays an upfront premium to the seller. If the present value of the protection leg is greater than the present value of the payment leg, the seller pays an upfront premium to the buyer.
- An important determinant of the value of the expected payments is the hazard rate, the probability of default given that default has not already occurred.
- CDS prices are often quoted in terms of credit spreads, the implied number of basis points that the credit protection seller receives from the credit protection buyer to justify providing the protection.
- Credit spreads are often expressed in terms of a credit curve, which expresses the relationship between the credit spreads on bonds of different maturities for the same borrower.
- CDS change in value over their lives as the credit quality of the reference entity changes, which leads to gains and losses for the counterparties, even though default may not have occurred or may never occur. CDS spreads approach zero as the CDS approaches maturity.

- Either party can monetize an accumulated gain or loss by entering into an offsetting position that matches the terms of the original CDS.
- CDS are used to increase or decrease credit exposures or to capitalize on different assessments of the cost of credit among different instruments tied to the reference entity, such as debt, equity, and derivatives of debt and equity.

PRACTICE PROBLEMS

The following information relates to Questions 1–6

UNAB Corporation

On 1 January 20X2, Deem Advisors purchased a $10 million six-year senior unsecured bond issued by UNAB Corporation. Six months later (1 July 20X2), concerned about the portfolio's credit exposure to UNAB, Doris Morrison, the chief investment officer at Deem Advisors, buys $10 million protection on UNAB with a standardized coupon rate of 5%. The reference obligation of the CDS is the UNAB bond owned by Deem Advisors. UNAB adheres to the ISDA CDS protocols.

On 1 January 20X3, Morrison asks Bill Watt, a derivatives analyst, to assess the current credit quality of UNAB bonds and the value of Deem Advisors' CDS on UNAB debt. Watt gathers the following information on UNAB's debt issues currently trading in the market:

Bond 1: A two-year senior unsecured bond trading at 40% of par

Bond 2: A five-year senior unsecured bond trading at 50% of par

Bond 3: A five-year subordinated unsecured bond trading at 20% of par

With respect to the credit quality of UNAB, Watt makes the following statement:

> "There is severe near-term stress in the financial markets, and UNAB's credit curve clearly reflects the difficult environment."

On 1 July 20X3, UNAB fails to make a scheduled interest payment on the outstanding subordinated unsecured obligation after a grace period; however, the company does not file for bankruptcy. Morrison asks Watt to determine if UNAB experienced a credit event and, if so, to recommend a settlement preference.

Kand Corporation

Morrison is considering purchasing protection on Kand Corporation debt to hedge the portfolio's position in Kand. She instructs Watt to determine if an upfront payment would be required and, if so, the amount of the premium. Watt presents the information for the CDS in Exhibit 1.

Exhibit 1	Summary Data for 10-year CDS on Kand Corporation
Credit spread	700 bps
Duration	7 years
Coupon rate	5%

© 2017 CFA Institute. All rights reserved.

Morrison purchases 10-year protection on Kand Corporation debt. Two months later the credit spread for Kand Corporation has increased by 200 bps. Morrison asks Watt to close out the firm's CDS position on Kand Corporation by entering into a new, offsetting contract.

Tollunt Corporation

Deem Advisors' chief credit analyst recently reported that Tollunt Corporation's five-year bond is currently yielding 7% and a comparable CDS contract has a credit spread of 4.25%. Since the current market reference rate is 2.5%, Watt has recommended executing a basis trade to take advantage of the pricing of Tollunt's bonds and CDS. The basis trade would consist of purchasing both the bond and the CDS contract.

1 If UNAB experienced a credit event on 1 July, Watt should recommend that Deem Advisors:
   - A prefer a cash settlement.
   - B prefer a physical settlement.
   - C be indifferent between a cash or a physical settlement.

2 According to Watt's statement, the shape of UNAB's credit curve is *most likely*:
   - A flat.
   - B upward-sloping.
   - C downward-sloping.

3 Should Watt conclude that UNAB experienced a credit event?
   - A Yes
   - B No, because UNAB did not file for bankruptcy
   - C No, because the failure to pay occurred on a subordinated unsecured bond

4 Based on Exhibit 1, the upfront premium as a percent of the notional for the CDS protection on Kand Corporation would be *closest* to:
   - A 2.0%.
   - B 9.8%.
   - C 14.0%.

5 If Deem Advisors enters into a new offsetting contract two months after purchasing protection on Kand Corporation, this action will *most likely* result in:
   - A a loss on the CDS position.
   - B a profit on the CDS position.
   - C neither a loss nor a profit on the CDS position.

6 If convergence occurs in the bond and CDS markets for Tollunt Corporation, a basis trade will capture a profit *closest* to:
   - A 0.25%.
   - B 1.75%.
   - C 2.75%.

The following information relates to Questions 7–14

John Smith, a fixed-income portfolio manager at a €10 billion sovereign wealth fund (the Fund), meets with Sofia Chan, a derivatives strategist with Shire Gate Securities (SGS), to discuss investment opportunities for the Fund. Chan notes that SGS adheres to ISDA (International Swaps and Derivatives Association) protocols for credit default swap (CDS) transactions and that any contract must conform to ISDA specifications. Before the Fund can engage in trading CDS products with SGS, the Fund must satisfy compliance requirements.

Smith explains to Chan that fixed-income derivatives strategies are being contemplated for both hedging and trading purposes. Given the size and diversified nature of the Fund, Smith asks Chan to recommend a type of CDS that would allow the Fund to simultaneously fully hedge multiple fixed-income exposures.

Smith and Chan discuss opportunities to add trading profits to the Fund. Smith asks Chan to determine the probability of default associated with a five-year investment-grade bond issued by Orion Industrial. Selected data on the Orion Industrial bond are presented in Exhibit 1.

Exhibit 1 Selected Data on Orion Industrial Five-Year Bond

Year	Hazard Rate
1	0.22%
2	0.35%
3	0.50%
4	0.65%
5	0.80%

Chan explains that a single-name CDS can also be used to add profit to the Fund over time. Chan describes a hypothetical trade in which the Fund sells £6 million of five-year CDS protection on Orion, where the CDS contract has a duration of 3.9 years. Chan assumes that the Fund closes the position six months later, after Orion's credit spread narrowed from 150 bps to 100 bps.

Chan discusses the mechanics of a long/short trade. In order to structure a number of potential trades, Chan and Smith exchange their respective views on individual companies and global economies. Chan and Smith agree on the following outlooks.

Outlook 1: The European economy will weaken.

Outlook 2: The US economy will strengthen relative to that of Canada.

Outlook 3: The credit quality of electric car manufacturers will improve relative to that of traditional car manufacturers.

Chan believes US macroeconomic data are improving and that the general economy will strengthen in the short term. Chan suggests that a curve trade could be used by the Fund to capitalize on her short-term view of a steepening of the US credit curve.

Another short-term trading opportunity that Smith and Chan discuss involves the merger and acquisition market. SGS believes that Delta Corporation may make an unsolicited bid at a premium to the market price for all of the publicly traded shares of Zega, Inc. Zega's market capitalization and capital structure are comparable to

Delta's; both firms are highly levered. It is anticipated that Delta will issue new equity along with 5- and 10-year senior unsecured debt to fund the acquisition, which will significantly increase its debt ratio.

7 To satisfy the compliance requirements referenced by Chan, the Fund is *most likely* required to:

A set a notional amount.

B post an upfront payment.

C sign an ISDA master agreement.

8 Which type of CDS should Chan recommend to Smith?

A CDS index

B Tranche CDS

C Single-name CDS

9 Based on Exhibit 1, the probability of Orion defaulting on the bond during the first three years is *closest* to:

A 1.07%.

B 2.50%.

C 3.85%.

10 To close the position on the hypothetical Orion trade, the Fund:

A sells protection at a higher premium than it paid at the start of the trade.

B buys protection at a lower premium than it received at the start of the trade.

C buys protection at a higher premium than it received at the start of the trade.

11 The hypothetical Orion trade generated an approximate:

A loss of £117,000.

B gain of £117,000.

C gain of £234,000.

12 Based on the three economic outlook statements, a profitable long/short trade would be to:

A sell protection using a Canadian CDX IG and buy protection using a US CDX IG.

B buy protection using an iTraxx Crossover and sell protection using an iTraxx Main.

C buy protection using an electric car CDS and sell protection using a traditional car CDS.

13 The curve trade that would *best* capitalize on Chan's view of the US credit curve is to:

A buy protection using a 20-year CDX and buy protection using a 2-year CDX.

B buy protection using a 20-year CDX and sell protection using a 2-year CDX.

C sell protection using a 20-year CDX and buy protection using a 2-year CDX.

14 A profitable equity-versus-credit trade involving Delta and Zega is to:

A short Zega shares and buy protection on Delta using the 10-year CDS.

B go long Zega shares and buy protection on Delta using 5-year CDS.

C go long Delta shares and buy protection on Delta using 5-year CDS.

SOLUTIONS

1 A is correct. Deem Advisors would prefer a cash settlement. Deem Advisors owns Bond 2 (trading at 50% of par), which is worth more than the cheapest-to-deliver obligation (Bond 1, also a senior secured bond, trading at 40% of par). Based on the price of this cheapest-to-deliver security, the estimated recovery rate is 40%. Thus, Deem Advisors can cash settle for $6 million [= (1 – 40%) × $10 million] on its CDS contract and sell the bond it owns, Bond 2, for $5 million, for total proceeds of $11 million. If Deem Advisors were to physically settle the contract, only $10 million would be received, the face amount of the bonds, and it would deliver Bond 2.

B is incorrect because if Deem Advisors were to physically settle the contract, it would receive only $10 million, which is less than the $11 million that could be obtained from a cash settlement. C is incorrect because Deem Advisors would not be indifferent between settlement protocols as the firm would receive $1 million more with a cash settlement in comparison to a physical settlement.

2 C is correct. A downward-sloping credit curve implies a greater probability of default in the earlier years than in the later years. Downward-sloping curves are less common and often are the result of severe near-term stress in the financial markets.

A is incorrect because a flat credit curve implies a constant hazard rate (conditional probability of default). B is incorrect because an upward-sloping credit curve implies a greater probability of default in later years.

3 A is correct. UNAB experienced a credit event when it failed to make the scheduled coupon payment on the outstanding subordinated unsecured obligation. Failure to pay, a credit event, occurs when a borrower does not make a scheduled payment of principal or interest on outstanding obligations after a grace period, even without a formal bankruptcy filing.

B is incorrect because a credit event can occur without filing for bankruptcy. The three most common credit events are bankruptcy, failure to pay, and restructuring.

C is incorrect because a credit event (failure to pay) occurs when a borrower does not make a scheduled payment of principal or interest on *any* outstanding obligations after a grace period, even without a formal bankruptcy filing.

4 C is correct. An approximation for the upfront premium is (Credit spread – Fixed coupon rate) × Duration of the CDS. To buy 10-year CDS protection, Deem Advisors would have to pay an approximate upfront premium of 1,400 bps [(700 – 500) × 7], or 14% of the notional.

A is incorrect because 200 bps, or 2%, is derived by taking the simple difference between the credit spread and the fixed coupon rate (700 – 500), ignoring the duration component of the calculation. B is incorrect because 980 bps, or 9.8%, is the result of dividing the credit spread by the fixed coupon rate and multiplying by the duration of the CDS [(700/500) × 7].

5 B is correct. Deem Advisors purchased protection and therefore is economically short and benefits from an increase in the company's spread. Since putting on the protection, the credit spread increased by 200 bps, and Deem Advisors realizes the profit by entering into a new, offsetting contract (sells protection to another party at a higher premium).

A is incorrect because a decrease (not increase) in the spread would result in a loss for the credit protection buyer. C is incorrect because Deem Advisors, the credit protection buyer, would profit from an increase in the company's credit spread, not break even.

6 A is correct. A difference in credit spreads in the bond market and CDS market is the foundation of the basis trade strategy. If the spread is higher in the bond market than in the CDS market, it is said to be a negative basis. In this case, the bond credit spread is currently 4.50% (bond yield minus Libor) and the comparable CDS contract has a credit spread of 4.25%. The credit risk is cheap in the CDS market relative to the bond market. Since the protection and the bond were both purchased, if convergence occurs, the trade will capture the 0.25% differential in the two markets (4.50% – 4.25%).

B is incorrect because the bond market implies a 4.50% credit risk premium (bond yield minus the market reference rate) and the CDS market implies a 4.25% credit risk premium. Convergence of the bond market credit risk premium and the CDS credit risk premium would result in capturing the differential, 0.25%. The 1.75% is derived by incorrectly subtracting Libor from the credit spread on the CDS (= 4.25% – 2.50%).

C is incorrect because convergence of the bond market credit risk premium and the CDS credit risk premium would result in capturing the differential, 0.25%. The 2.75% is derived incorrectly by subtracting the credit spread on the CDS from the current bond yield (= 7.00% – 4.25%).

7 C is correct. Parties to CDS contracts generally agree that their contracts will conform to ISDA specifications. These terms are specified in the ISDA master agreement, which the parties to a CDS sign before any transactions are made. Therefore, to satisfy the compliance requirements referenced by Chan, the sovereign wealth fund must sign an ISDA master agreement with SGS.

8 A is correct. A CDS index (e.g., CDX and iTraxx) would allow the Fund to simultaneously fully hedge multiple fixed-income exposures. A tranche CDS will also hedge multiple exposures, but it would only partially hedge those exposures.

9 A is correct. Based on Exhibit 1, the probability of survival for the first year is 99.78% (100% minus the 0.22% hazard rate). Similarly, the probability of survival for the second and third years is 99.65% (100% minus the 0.35% hazard rate) and 99.50% (100% minus the 0.50% hazard rate), respectively. Therefore, the probability of survival of the Orion bond through the first three years is equal to $0.9978 \times 0.9965 \times 0.9950 = 0.9893$, and the probability of default sometime during the first three years is $1 - 0.9893$, or 1.07%.

10 B is correct. The trade assumes that £6 million of five-year CDS protection on Orion is initially sold, so the Fund received the premium. Because the credit spread of the Orion CDS narrowed from 150 bps to 100 bps, the CDS position will realize a financial gain. This financial gain is equal to the difference between the upfront premium received on the original CDS position and the upfront premium to be paid on a new, offsetting CDS position. To close the position and monetize this gain, the Fund should unwind the position by buying protection for a lower premium (relative to the original premium collected).

11 B is correct. The gain on the hypothetical Orion trade is £117,000, calculated as follows.

Approximate profit = Change in credit spread (in bps) × Duration × Notional amount.

Approximate profit = (150 bps – 100 bps) × 3.9 × £6 million.

Approximate profit = 0.005 × 3.9 × £6 million.

= £117,000.

The Fund gains because it sold protection at a spread of 150 bps and closed out the position by buying protection at a lower spread of 100 bps.

12 B is correct. Based on Outlook 1, Chan and Smith anticipate that Europe's economy will weaken. In order to profit from this forecast, one would buy protection using a high-yield CDS index (e.g., iTraxx Crossover) and sell protection using an investment-grade CDS index (e.g., iTraxx Main).

13 B is correct. To take advantage of Chan's view of the US credit curve steepening in the short term, a curve trade will entail shorting (buying protection using) a long-term (20-year) CDX and going long (selling protection using) a short-term (2-year) CDX. A steeper curve means that long-term credit risk increases relative to short-term credit risk.

14 B is correct. The shares of Zega can be sold at a higher price as a result of the unsolicited bid in the market. If Delta Corporation issues significantly more debt, there is a higher probability that it may default. If the Fund sells protection on Delta now, the trade will realize a profit as credit spreads widen. An equity-versus-credit trade would be to go long (buy) the Zega shares and buy protection on Delta.

Glossary

Abandonment option The ability to terminate a project at some future time if the financial results are disappointing.

Abnormal earnings See *residual income*.

Abnormal return The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

Absolute convergence The idea that developing countries, regardless of their particular characteristics, will eventually catch up with the developed countries and match them in per capita output.

Absolute valuation model A model that specifies an asset's intrinsic value.

Absolute version of PPP An extension of the law of one price whereby the prices of goods and services will not differ internationally once exchange rates are considered.

Accounting estimates Estimates used in calculating the value of assets or liabilities and in the amount of revenue and expense to allocate to a period. Examples of accounting estimates include, among others, the useful lives of depreciable assets, the salvage value of depreciable assets, product returns, warranty costs, and the amount of uncollectible receivables.

Accumulated benefit obligation The actuarial present value of benefits (whether vested or non-vested) attributed, generally by the pension benefit formula, to employee service rendered before a specified date and based on employee service and compensation (if applicable) before that date. The accumulated benefit obligation differs from the projected benefit obligation in that it includes no assumption about future compensation levels.

Accuracy The percentage of correctly predicted classes out of total predictions. It is an overall performance metric in classification problems.

Acquirer The company in a merger or acquisition that is acquiring the target.

Acquiring company See *acquirer*.

Acquisition The purchase of some portion of one company by another; the purchase may be for assets, a definable segment of another entity, or the entire company.

Activation function A functional part of a neural network's node that transforms the total net input received into the final output of the node. The activation function operates like a light dimmer switch that decreases or increases the strength of the input.

Active factor risk The contribution to active risk squared resulting from the portfolio's different-than-benchmark exposures relative to factors specified in the risk model.

Active return The return on a portfolio minus the return on the portfolio's benchmark.

Active risk The standard deviation of active returns.

Active risk squared The variance of active returns; active risk raised to the second power.

Active share A measure of how similar a portfolio is to its benchmark. A manager who precisely replicates the benchmark will have an active share of zero; a manager with no holdings in common with the benchmark will have an active share of one.

Active specific risk The contribution to active risk squared resulting from the portfolio's active weights on individual assets as those weights interact with assets' residual risk.

Adjusted funds from operations (AFFO) Funds from operations adjusted to remove any non-cash rent reported under straight-line rent accounting and to subtract maintenance-type capital expenditures and leasing costs, including leasing agents' commissions and tenants' improvement allowances.

Adjusted present value As an approach to valuing a company, the sum of the value of the company, assuming no use of debt, and the net present value of any effects of debt on company value.

Adjusted R^2 A measure of goodness-of-fit of a regression that is adjusted for degrees of freedom and hence does not automatically increase when another independent variable is added to a regression.

Administrative regulations or administrative law Rules issued by government agencies or other regulators.

Advanced set An arrangement in which the reference interest rate is set at the time the money is deposited.

Advanced settled An arrangement in which a forward rate agreement (FRA) expires and settles at the same time, at the FRA expiration date.

Agency costs Costs associated with the conflict of interest present when a company is managed by non-owners. Agency costs result from the inherent conflicts of interest between managers and equity owners.

Agency costs of equity The smaller the stake managers have in the company, the less their share in bearing the cost of excessive perquisite consumption—consequently, the less their desire to give their best efforts in running the company.

Agency issues Conflicts of interest that arise when the agent in an agency relationship has goals and incentives that differ from the principal to whom the agent owes a fiduciary duty. Also called *agency problems* or *principal–agent problems*.

Agglomerative clustering A bottom-up hierarchical clustering method that begins with each observation being treated as its own cluster. The algorithm finds the two closest clusters, based on some measure of distance (similarity), and combines them into one new larger cluster. This process is repeated iteratively until all observations are clumped into a single large cluster.

Allowance for loan losses A balance sheet account; it is a contra asset account to loans.

Alpha The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

American Depositary Receipt A negotiable certificate issued by a depositary bank that represents ownership in a non-US company's deposited equity (i.e., equity held in custody by the depositary bank in the company's home market).

Analysis of variance (ANOVA) The analysis that breaks the total variability of a dataset (such as observations on the dependent variable in a regression) into components representing different sources of variation. With reference to regression, ANOVA provides the inputs for an *F*-test of

the significance of the regression as a whole, as well as the inputs for the coefficient of determination and the standard error of the estimate.

Application programming interface (API) A set of well-defined methods of communication between various software components and typically used for accessing external data.

Arbitrage (1) The simultaneous purchase of an undervalued asset or portfolio and sale of an overvalued but equivalent asset or portfolio in order to obtain a riskless profit on the price differential. Taking advantage of a market inefficiency in a risk-free manner. (2) The condition in a financial market in which equivalent assets or combinations of assets sell for two different prices, creating an opportunity to profit at no risk with no commitment of money. In a well-functioning financial market, few arbitrage opportunities are possible. (3) A risk-free operation that earns an expected positive net profit but requires no net investment of money.

Arbitrage-free models Term structure models that project future interest rate paths that emanate from the existing term structure. Resulting prices are based on a no-arbitrage condition.

Arbitrage-free valuation An approach to valuation that determines security values consistent with the absence of any opportunity to earn riskless profits without any net investment of money.

Arbitrage opportunity An opportunity to conduct an arbitrage; an opportunity to earn an expected positive net profit without risk and with no net investment of money.

Arbitrage portfolio The portfolio that exploits an arbitrage opportunity.

Ask price The price at which a trader will sell a specified quantity of a security. Also called *ask*, *offer price*, or *offer*.

Asset-based approach Approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities.

Asset-based valuation An approach to valuing natural resource companies that estimates company value on the basis of the market value of the natural resources the company controls.

Asset beta The unlevered beta; reflects the business risk of the assets; the asset's systematic risk.

Asset purchase An acquisition in which the acquirer purchases the target company's assets and payment is made directly to the target company.

Asymmetric information The differential of information between corporate insiders and outsiders regarding the company's performance and prospects. Managers typically have more information about the company's performance and prospects than owners and creditors.

At market contract When a forward contract is established, the forward price is negotiated so that the market value of the forward contract on the initiation date is zero.

Authorized participants (APs) A special group of institutional investors who are authorized by the ETF issuer to participate in the creation/redemption process. APs are large broker/dealers, often market makers.

Autocorrelations The correlations of a time series with its own past values.

Autoregressive model (AR) A time series regressed on its own past values in which the independent variable is a lagged value of the dependent variable.

Backtesting The process that approximates the real-life investment process, using historical data, to assess whether an investment strategy would have produced desirable results.

Backward integration A merger involving the purchase of a target ahead of the acquirer in the value or production chain; for example, to acquire a supplier.

Backward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving backward through the network's layers.

Backwardation A condition in futures markets in which the spot price exceeds the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is higher than the longer-term futures contract price.

Bag-of-words (BOW) A collection of a distinct set of tokens from all the texts in a sample dataset. BOW does not capture the position or sequence of words present in the text.

Bankruptcy A declaration provided for by a country's laws that typically involves the establishment of a legal procedure that forces creditors to defer their claims.

Barbell portfolio Fixed-income portfolio that combines short and long maturities.

Base error Model error due to randomness in the data.

Basic earnings per share (EPS) Net earnings available to common shareholders (i.e., net income minus preferred dividends) divided by the weighted average number of common shares outstanding during the period.

Basis The difference between the spot price and the futures price. As the maturity date of the futures contract nears, the basis converges toward zero.

Basis trade A trade based on the pricing of credit in the bond market versus the price of the same credit in the CDS market. To execute a basis trade, go long the "underpriced" credit and short the "overpriced" credit. A profit is realized as the implied credit prices converge.

Bear hug A tactic used by acquirers to circumvent target management's objections to a proposed merger by submitting the proposal directly to the target company's board of directors.

Bearish flattening Term structure shift in which short-term bond yields rise more than long-term bond yields, resulting in a flatter yield curve.

Benchmark value of the multiple In using the method of comparables, the value of a price multiple for the comparison asset; when we have comparison assets (a group), the mean or median value of the multiple for the group of assets.

Best ask The offer to sell with the lowest ask price. Also called *best offer* or *inside ask*.

Best bid The offer to buy with the highest bid price. Also called the *inside bid*.

Best offer See *best ask*.

Bias error Describes the degree to which a model fits the training data. Algorithms with erroneous assumptions produce high bias error with poor approximation, causing underfitting and high in-sample error.

Bid–ask spread The ask price minus the bid price.

Bid price The price at which a trader will buy a specified quantity of a security. Also called *bid*.

Bill-and-hold basis Sales on a bill-and-hold basis involve selling products but not delivering those products until a later date.

Blockage factor An illiquidity discount that occurs when an investor sells a large amount of stock relative to its trading volume (assuming it is not large enough to constitute a controlling ownership).

Bond indenture A legal contract specifying the terms of a bond issue.

Bond risk premium The expected excess return of a default-free long-term bond less that of an equivalent short-term bond.

Bond yield plus risk premium method An estimate of the cost of common equity that is produced by summing the before-tax cost of debt and a risk premium that captures the additional yield on a company's stock relative to its bonds. The additional yield is often estimated using historical spreads between bond yields and stock yields.

Bonding costs Costs borne by management to assure owners that they are working in the owners' best interest (e.g., implicit cost of non-compete agreements).

Bonus issue of shares *See stock dividend.*

Book value Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value of equity Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value per share The amount of book value (also called carrying value) of common equity per share of common stock, calculated by dividing the book value of shareholders' equity by the number of shares of common stock outstanding.

Bootstrap aggregating (or bagging) A technique whereby the original training dataset is used to generate *n* new training datasets or bags of data. Each new bag of data is generated by random sampling with replacement from the initial training set.

Bootstrapping The use of a forward substitution process to determine zero-coupon rates by using the par yields and solving for the zero-coupon rates one by one, from the shortest to longest maturities.

Bottom-up approach With respect to forecasting, an approach that usually begins at the level of the individual company or a unit within the company.

Breakup value The value derived using a sum-of-the-parts valuation.

Breusch–Pagan test A test for conditional heteroskedasticity in the error term of a regression.

Bullet portfolio A fixed-income portfolio concentrated in a single maturity.

Bullish flattening Term structure change in which the yield curve flattens in response to a greater decline in long-term rates than short-term rates.

Bullish steepening Term structure change in which short-term rates fall by more than long-term yields, resulting in a steeper term structure.

Buy-side analysts Analysts who work for investment management firms, trusts, bank trust departments, and similar institutions.

Buyback *See share repurchase.*

Callable bond Bond that includes an embedded call option that gives the issuer the right to redeem the bond issue prior to maturity, typically when interest rates have fallen or when the issuer's credit quality has improved.

Canceled shares Shares that were issued, subsequently repurchased by the company, and then retired (cannot be reissued).

Cannibalization Cannibalization occurs when an investment takes customers and sales away from another part of the company.

Capital charge The company's total cost of capital in money terms.

Capital deepening An increase in the capital-to-labor ratio.

Capital rationing A capital rationing environment assumes that the company has a fixed amount of funds to invest.

Capital structure The mix of debt and equity that a company uses to finance its business; a company's specific mixture of long-term financing.

Capitalization of earnings method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capitalization rate The divisor in the expression for the value of perpetuity. In the context of real estate, it is the divisor in the direct capitalization method of estimating value. The cap rate equals net operating income divided by value.

Capitalized cash flow method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity. Also called *capitalized cash flow model.*

Capitalized income method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capped floater Floating-rate bond with a cap provision that prevents the coupon rate from increasing above a specified maximum rate. It protects the issuer against rising interest rates.

Carry arbitrage model A no-arbitrage approach in which the underlying instrument is either bought or sold along with an opposite position in a forward contract.

Carry benefits Benefits that arise from owning certain underlyings; for example, dividends, foreign interest, and bond coupon payments.

Carry costs Costs that arise from owning certain underlyings. They are generally a function of the physical characteristics of the underlying asset and also the interest forgone on the funds tied up in the asset.

Cash available for distribution See *adjusted funds from operations.*

Cash-generating unit The smallest identifiable group of assets that generates cash inflows that are largely independent of the cash inflows of other assets or groups of assets.

Cash offering A merger or acquisition that is to be paid for with cash; the cash for the merger might come from the acquiring company's existing assets or from a debt issue.

Cash settlement A procedure used in certain derivative transactions that specifies that the long and short parties settle the derivative's difference in value between them by making a cash payment.

Catalyst An event or piece of information that causes the marketplace to re-evaluate the prospects of a company.

Categorical dependent variables An alternative term for qualitative dependent variables.

CDS spread A periodic premium paid by the buyer to the seller that serves as a return over a market reference rate required to protect against credit risk.

Ceiling analysis A systematic process of evaluating different components in the pipeline of model building. It helps to understand what part of the pipeline can potentially improve in performance by further tuning.

Centroid The center of a cluster formed using the *k*-means clustering algorithm.

Chain rule of forecasting A forecasting process in which the next period's value as predicted by the forecasting equation is substituted into the right-hand side of the equation to give a predicted value two periods ahead.

Cheapest-to-deliver The debt instrument that can be purchased and delivered at the lowest cost yet has the same seniority as the reference obligation.

Classification and regression tree A supervised machine learning technique that can be applied to predict either a categorical target variable, producing a classification tree, or a continuous target variable, producing a regression tree. CART is commonly applied to binary classification or regression.

Clean surplus relation The relationship between earnings, dividends, and book value in which ending book value is equal to the beginning book value plus earnings less dividends, apart from ownership transactions.

Club convergence The idea that only rich and middle-income countries sharing a set of favorable attributes (i.e., are members of the "club") will converge to the income level of the richest countries.

Cluster A subset of observations from a dataset such that all the observations within the same cluster are deemed "similar."

Clustering The sorting of observations into groups (clusters) such that observations in the same cluster are more similar to each other than they are to observations in other clusters.

Cobb–Douglas production function A function of the form $Y = K^{\alpha} L^{1-\alpha}$ relating output (Y) to labor (L) and capital (K) inputs.

Coefficient of determination The percentage of the variation of the dependent variable that is explained by the independent variable. Also referred to as the "*R*-squared" or "R^2."

Cointegrated Describes two time series that have a long-term financial or economic relationship such that they do not diverge from each other without bound in the long run.

Collateral return The component of the total return on a commodity futures position attributable to the yield for the bonds or cash used to maintain the futures position. Also called *collateral yield*.

Collection frequency (CF) The number of times a given word appears in the whole corpus (i.e., collection of sentences) divided by the total number of words in the corpus.

Commercial real estate properties Income-producing real estate properties; properties purchased with the intent to let, lease, or rent (in other words, produce income).

Commodity swap A type of swap involving the exchange of payments over multiple dates as determined by specified reference prices or indexes relating to commodities.

Common size statements Financial statements in which all elements (accounts) are stated as a percentage of a key figure, such as revenue for an income statement or total assets for a balance sheet.

Company fundamental factors Factors related to the company's internal performance, such as factors relating to earnings growth, earnings variability, earnings momentum, and financial leverage.

Company share-related factors Valuation measures and other factors related to share price or the trading characteristics of the shares, such as earnings yield, dividend yield, and book-to-market value.

Comparables Assets used as benchmarks when applying the method of comparables to value an asset. Also called *comps*, *guideline assets*, or *guideline companies*.

Competition laws A law that promotes or maintains market competition by regulating anti-competitive conduct. Known as "antitrust law" in the United States, "anti-monopoly law" in China and Russia, and often referred to as "trade practices law" in the United Kingdom and Australia.

Compiled financial statements Financial statements that are not accompanied by an auditor's opinion letter.

Complexity A term referring to the number of features, parameters, or branches in a model and to whether the model is linear or non-linear (non-linear is more complex).

Composite variable A variable that combines two or more variables that are statistically strongly related to each other.

Comprehensive income All changes in equity other than contributions by, and distributions to, owners; income under clean surplus accounting; includes all changes in equity during a period except those resulting from investments by owners and distributions to owners. Comprehensive income equals net income plus other comprehensive income.

Comps Assets used as benchmarks when applying the method of comparables to value an asset.

Concentrated ownership Ownership structure consisting of an individual shareholder or a group (controlling shareholders) with the ability to exercise control over the corporation.

Conditional convergence The idea that convergence of per capita income is conditional on the countries having the same savings rate, population growth rate, and production function.

Conditional heteroskedasticity Heteroskedasticity in the error variance that is correlated with the values of the independent variable(s) in the regression.

Conditional VaR (CVaR) The weighted average of all loss outcomes in the statistical (i.e., return) distribution that exceed the VaR loss. Thus, CVaR is a more comprehensive measure of tail loss than VaR is. Sometimes referred to as the *expected tail loss* or *expected shortfall*.

Confusion matrix A grid used for error analysis in classification problems, it presents values for four evaluation metrics including true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Conglomerate discount The discount possibly applied by the market to the stock of a company operating in multiple, unrelated businesses.

Conglomerate merger A merger involving companies that are in unrelated businesses.

Consolidation The combining of the results of operations of subsidiaries with the parent company to present financial statements as if they were a single economic unit. The assets, liabilities, revenues, and expenses of the subsidiaries are combined with those of the parent company, eliminating intercompany transactions.

Constant dividend payout ratio policy A policy in which a constant percentage of net income is paid out in dividends.

Constant returns to scale The condition that if all inputs into the production process are increased by a given percentage, then output rises by that same percentage.

Contango A condition in futures markets in which the spot price is lower than the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is lower than the longer-term futures contract price.

Contingent consideration Potential future payments to the seller that are contingent on the achievement of certain agreed-on occurrences.

Continuing earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *persistent earnings*, or *underlying earnings*.

Continuing residual income Residual income after the forecast horizon.

Continuing value The analyst's estimate of a stock's value at a particular point in the future.

Control premium An increment or premium to value associated with a controlling ownership interest in a company.

Convergence The property by which as expiration approaches, the price of a newly created forward or futures contract will approach the price of a spot transaction. At expiration, a forward or futures contract is equivalent to a spot transaction in the underlying.

Conversion period For a convertible bond, the period during which bondholders have the right to convert their bonds into shares.

Conversion price For a convertible bond, the price per share at which the bond can be converted into shares.

Conversion rate (or ratio) For a convertible bond, the number of shares of common stock that a bondholder receives from converting the bond into shares.

Conversion value For a convertible bond, the value of the bond if it is converted at the market price of the shares. Also called *parity value*.

Convertible bond Bond with an embedded conversion option that gives bondholders the right to convert their bonds into the issuer's common stock during a pre-determined period at a pre-determined price.

Convexity A measure of how interest rate sensitivity changes with a change in interest rates.

Core earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *persistent earnings*, or *underlying earnings*.

Core real estate investment style Investing in high-quality, well-leased, core property types with low leverage (no more than 30% of asset value) in the largest markets with strong, diversified economies. It is a conservative strategy designed to avoid real estate–specific risks, including leasing, development, and speculation in favor of steady returns. Hotel properties are excluded from the core categories because of the higher cash flow volatility resulting from single-night leases and the greater importance of property operations, brand, and marketing.

Corpus A collection of text data in any form, including list, matrix, or data table forms.

Cost approach An approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities. In the context of real estate, this approach estimates the value of a property based on what it would cost to buy the land and construct a new property on the site that has the same utility or functionality as the property being appraised.

Cost of carry model A model that relates the forward price of an asset to the spot price by considering the cost of carry (also referred to as future-spot parity model).

Cost of debt The cost of debt financing to a company, such as when it issues a bond or takes out a bank loan.

Cost of equity The required rate of return on common stock.

Covariance stationary Describes a time series when its expected value and variance are constant and finite in all periods and when its covariance with itself for a fixed number of periods in the past or future is constant and finite in all periods.

Covered bonds A senior debt obligation of a financial institution that gives recourse to the originator/issuer and a predetermined underlying collateral pool.

Covered interest rate parity The relationship among the spot exchange rate, the forward exchange rate, and the interest rates in two currencies that ensures that the return on a hedged (i.e., covered) foreign risk-free investment is the same as the return on a domestic risk-free investment. Also called *interest rate parity*.

Cox-Ingersoll-Ross model A general equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is directly related to the level of interest rates.

Creation basket The list of securities (and share amounts) the authorized participant (AP) must deliver to the ETF manager in exchange for ETF shares. The creation basket is published each business day.

Creation/redemption The process in which ETF shares are created or redeemed by authorized participants transacting with the ETF issuer.

Creation units Large blocks of ETF shares transacted between the authorized participant (AP) and the ETF manager that are usually but not always equal to 50,000 shares of the ETF.

Credit correlation The correlation of credit (or default) risks of the underlying single-name CDS contained in an index CDS.

Credit curve The credit spreads for a range of maturities of a company's debt.

Credit default swap A derivative contract between two parties in which the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit derivative A derivative instrument in which the underlying is a measure of the credit quality of a borrower.

Credit event The event that triggers a payment from the credit protection seller to the credit protection buyer.

Credit protection buyer One party to a credit default swap; the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit protection seller One party to a credit default swap; the seller makes a promise to pay compensation for credit losses resulting from the default.

Credit risk The risk that the borrower will not repay principal and interest. Also called *default risk*.

Credit valuation adjustment The value of the credit risk of a bond in present value terms.

Cross-validation A technique for estimating out-of-sample error directly by determining the error in validation samples.

Current exchange rate For accounting purposes, the spot exchange rate on the balance sheet date.

Current rate method Approach to translating foreign currency financial statements for consolidation in which all assets and liabilities are translated at the current exchange rate. The current rate method is the prevalent method of translation.

Curvature One of the three factors (the other two are level and steepness) that empirically explain most of the changes in the shape of the yield curve. A shock to the curvature factor affects mid-maturity interest rates, resulting in the term structure becoming either more or less hump-shaped.

Curve trade Buying a CDS of one maturity and selling a CDS on the same reference entity with a different maturity.

Cyclical businesses Businesses with high sensitivity to business- or industry-cycle influences.

Data mining The practice of determining a model by extensive searching through a dataset for statistically significant patterns.

Data preparation (cleansing) The process of examining, identifying, and mitigating (i.e., cleansing) errors in raw data.

Data snooping The subconscious or conscious manipulation of data in a way that produces a statistically significant result (i.e., the p-value is sufficiently small or the t-statistic sufficiently large to indicate statistical significance), such as by running multiple simulations and naively accepting the best result. Also known as p-hacking.

Data wrangling (preprocessing) This task performs transformations and critical processing steps on cleansed data to make the data ready for ML model training (i.e., preprocessing), and includes dealing with outliers, extracting useful variables from existing data points, and scaling the data.

"Dead-hand" provision A poison pill provision that allows for the redemption or cancellation of a poison pill provision only by a vote of continuing directors (generally directors who were on the target company's board prior to the takeover attempt).

Debt rating An objective measure of the quality and safety of a company's debt based upon an analysis of the company's ability to pay the promised cash flows. It includes an analysis of any indentures.

Deep learning Algorithms based on deep neural networks, ones with many hidden layers (more than two), that address highly complex tasks, such as image classification, face recognition, speech recognition, and natural language processing.

Deep neural networks Neural networks with many hidden layers—at least 2 but potentially more than 20—that have proven successful across a wide range of artificial intelligence applications.

Default risk See *credit risk*.

Defined benefit pension plans Plan in which the company promises to pay a certain annual amount (defined benefit) to the employee after retirement. The company bears the investment risk of the plan assets.

Defined contribution pension plans Individual accounts to which an employee and typically the employer makes contributions, generally on a tax-advantaged basis. The amounts of contributions are defined at the outset, but the future value of the benefit is unknown. The employee bears the investment risk of the plan assets.

Definitive merger agreement A contract signed by both parties to a merger that clarifies the details of the transaction, including the terms, warranties, conditions, termination details, and the rights of all parties.

Delay costs Implicit trading costs that arise from the inability to complete desired trades immediately. Also called *slippage.*

Delta The relationship between the option price and the underlying price, which reflects the sensitivity of the price of the option to changes in the price of the underlying. Delta is a good approximation of how an option price will change for a small change in the stock.

Dendrogram A type of tree diagram used for visualizing a hierarchical cluster analysis; it highlights the hierarchical relationships among the clusters.

Dependent variable The variable whose variation about its mean is to be explained by the regression; the left-side variable in a regression equation. Also referred to as the *explained variable*.

Depository Trust and Clearinghouse Corporation A US-headquartered entity providing post-trade clearing, settlement, and information services.

Descriptive statistics The study of how data can be summarized effectively.

Diluted earnings per share (Diluted EPS)Net income, minus preferred dividends, divided by the weighted average number of common shares outstanding considering all dilutive securities (e.g., convertible debt and options); the EPS that would result if all dilutive securities were converted into common shares.

Dilution A reduction in proportional ownership interest as a result of the issuance of new shares.

Dimension reduction A set of techniques for reducing the number of features in a dataset while retaining variation across observations to preserve the information contained in that variation.

Diminishing marginal productivity When each additional unit of an input, keeping the other inputs unchanged, increases output by a smaller increment.

Direct capitalization method In the context of real estate, this method estimates the value of an income-producing property based on the level and quality of its net operating income.

Discount To reduce the value of a future payment in allowance for how far away it is in time; to calculate the present value of some future amount. Also, the amount by which an instrument is priced below its face value.

Discount factor The present value or price of a risk-free single-unit payment when discounted using the appropriate spot rate.

Discount for lack of control An amount or percentage deducted from the pro rata share of 100% of the value of an equity interest in a business to reflect the absence of some or all of the powers of control.

Discount for lack of marketability An amount of percentage deducted from the value of an ownership interest to reflect the relative absence of marketability.

Discount function Discount factors for the range of all possible maturities. The spot curve can be derived from the discount function and vice versa.

Discount rate Any rate used in finding the present value of a future cash flow.

Discounted abnormal earnings model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Discounted cash flow (DCF) analysis In the context of merger analysis, an estimate of a target company's value found by discounting the company's expected future free cash flows to the present.

Discounted cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows. In the context of real estate, this method estimates the value of an income-producing property based on discounting future projected cash flows.

Discounted cash flow model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Discriminant analysis A multivariate classification technique used to discriminate between groups, such as companies that either will or will not become bankrupt during some time frame.

Dispersed ownership Ownership structure consisting of many shareholders, none of which has the ability to individually exercise control over the corporation.

Divestiture The sale, liquidation, or spin-off of a division or subsidiary.

Dividend A distribution paid to shareholders based on the number of shares owned.

Dividend coverage ratio The ratio of net income to dividends.

Dividend discount model (DDM) A present value model of stock value that views the intrinsic value of a stock as present value of the stock's expected future dividends.

Dividend displacement of earnings The concept that dividends paid now displace earnings in all future periods.

Dividend imputation tax system A taxation system that effectively assures corporate profits distributed as dividends are taxed just once and at the shareholder's tax rate.

Dividend index point A measure of the quantity of dividends attributable to a particular index.

Dividend payout ratio The ratio of cash dividends paid to earnings for a period.

Dividend policy The strategy a company follows with regard to the amount and timing of dividend payments.

Dividend rate The annualized amount of the most recent dividend.

Dividend yield Annual dividends per share divided by share price.

Divisive clustering A top-down hierarchical clustering method that starts with all observations belonging to a single large cluster. The observations are then divided into two clusters based on some measure of distance (similarity). The algorithm then progressively partitions the intermediate clusters into smaller ones until each cluster contains only one observation.

Document frequency (DF) The number of documents (texts) that contain a particular token divided by the total number of documents. It is the simplest feature selection method and often performs well when many thousands of tokens are present.

Document term matrix (DTM) A matrix where each row belongs to a document (or text file), and each column represents a token (or term). The number of rows is equal to the number of documents (or text files) in a sample text dataset. The number of columns is equal to the number of tokens from the BOW built using all the documents in the sample dataset. The cells typically contain the counts of the number of times a token is present in each document.

Dominance An arbitrage opportunity when a financial asset with a risk-free payoff in the future must have a positive price today.

Double taxation system Corporate earnings are taxed twice when paid out as dividends. First, corporate pretax earnings are taxed regardless of whether they will be distributed as dividends or retained at the corporate level. Second, dividends are taxed again at the individual shareholder level.

Downstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary) such that the investor company records a profit on its income statement. An example is a sale of inventory by the investor company to the associate or by a parent to a subsidiary company.

Dual-class shares Shares that grant one share class superior or even sole voting rights, whereas the other share class has inferior or no voting rights.

Due diligence Investigation and analysis in support of a recommendation; the failure to exercise due diligence may sometimes result in liability according to various securities laws.

Dummy variable A type of qualitative variable that takes on a value of 1 if a particular condition is true and 0 if that condition is false.

Duration A measure of the approximate sensitivity of a security to a change in interest rates (i.e., a measure of interest rate risk).

Dutch disease A situation in which currency appreciation driven by strong export demand for resources makes other segments of the economy (particularly manufacturing) globally uncompetitive.

Earnings surprise The difference between reported EPS and expected EPS. Also referred to as *unexpected earnings*.

Earnings yield EPS divided by price; the reciprocal of the P/E.

Economic profit See *residual income*.

Economic sectors Large industry groupings.

Economic value added (EVA®) A commercial implementation of the residual income concept; the computation of EVA® is the net operating profit after taxes minus the cost of capital, where these inputs are adjusted for a number of items.

Economies of scale A situation in which average costs per unit of good or service produced fall as volume rises. In reference to mergers, the savings achieved through the consolidation of operations and elimination of duplicate resources.

Edwards–Bell–Ohlson model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Effective convexity Sensitivity of duration to changes in interest rates.

Effective duration Sensitivity of the bond's price to a 100 bps parallel shift of the benchmark yield curve, assuming no change in the bond's credit spread.

Effective spread Two times the difference between the execution price and the midpoint of the market quote at the time an order is entered.

Eigenvalue A measure that gives the proportion of total variance in the initial dataset that is explained by each eigenvector.

Eigenvector A vector that defines new mutually uncorrelated composite variables that are linear combinations of the original features.

Embedded options Contingency provisions found in a bond's indenture or offering circular representing rights that enable their holders to take advantage of interest rate movements. They can be exercised by the issuer, by the bondholder, or automatically depending on the course of interest rates.

Ensemble learning A technique of combining the predictions from a collection of models to achieve a more accurate prediction.

Ensemble method The method of combining multiple learning algorithms, as in ensemble learning.

Enterprise value Total company value (the market value of debt, common equity, and preferred equity) minus the value of cash and investments.

Enterprise value multiple A valuation multiple that relates the total market value of all sources of a company's capital (net of cash) to a measure of fundamental value for the entire company (such as a pre-interest earnings measure).

Equilibrium The condition in which supply equals demand.

Equity carve-out A form of restructuring that involves the creation of a new legal entity and the sale of equity in it to outsiders.

Equity charge The estimated cost of equity capital in money terms.

Equity REITs REITs that own, operate, and/or selectively develop income-producing real estate.

Equity swap A swap transaction in which at least one cash flow is tied to the return on an equity portfolio position, often an equity index.

Error autocorrelations The autocorrelations of the error term.

Error term The difference between an observation and its expected value, where the expected value is based on the true underlying population relation between the dependent and independent variables. Also known simply as the *error*.

ESG integration An ESG investment approach that focuses on systematic consideration of material ESG factors in asset allocation, security selection, and portfolio construction decisions for the purpose of achieving the product's stated investment objectives.

Estimated parameters With reference to a regression analysis, the estimated values of the population intercept and population slope coefficients in a regression.

***Ex ante* tracking error** A measure of the degree to which the performance of a given investment portfolio might be expected to deviate from its benchmark; also known as *relative VaR*.

***Ex ante* version of PPP** The hypothesis that expected changes in the spot exchange rate are equal to expected differences in national inflation rates. An extension of relative purchasing power parity to expected future changes in the exchange rate.

Ex-dividend Trading ex-dividend refers to shares that no longer carry the right to the next dividend payment.

Ex-dividend date The first date that a share trades without (i.e., "ex") the right to receive the declared dividend for the period.

Excess earnings method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Exchange ratio The number of shares that target stockholders are to receive in exchange for each of their shares in the target company.

Exercise date The date when employees actually exercise stock options and convert them to stock.

Exercise value The value of an option if it were exercised. Also sometimes called *intrinsic value*.

Expanded CAPM An adaptation of the CAPM that adds to the CAPM a premium for small size and company-specific risk.

Expectations approach A procedure for obtaining the value of an option derived from discounting at the risk-free rate its expected future payoff based on risk neutral probabilities.

Expected exposure The projected amount of money an investor could lose if an event of default occurs, before factoring in possible recovery.

Expected holding-period return The expected total return on an asset over a stated holding period; for stocks, the sum of the expected dividend yield and the expected price appreciation over the holding period.

Expected shortfall See *conditional VaR*.

Expected tail loss See *conditional VaR*.

Exploratory data analysis (EDA) The preliminary step in data exploration, where graphs, charts, and other visualizations (heat maps and word clouds) as well as quantitative methods (descriptive statistics and central tendency measures) are used to observe and summarize data.

Exposure to foreign exchange risk The risk of a change in value of an asset or liability denominated in a foreign currency due to a change in exchange rates.

Extendible bond Bond with an embedded option that gives the bondholder the right to keep the bond for a number of years after maturity, possibly with a different coupon.

External growth Company growth in output or sales that is achieved by buying the necessary resources externally (i.e., achieved through mergers and acquisitions).

Extra dividend *See special dividend*.

F1 score The harmonic mean of precision and recall. F1 score is a more appropriate overall performance metric (than accuracy) when there is unequal class distribution in the dataset and it is necessary to measure the equilibrium of precision and recall.

Factor A common or underlying element with which several variables are correlated.

Factor betas An asset's sensitivity to a particular factor; a measure of the response of return to each unit of increase in a factor, holding all other factors constant.

Factor portfolio See *pure factor portfolio*.

Factor price The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors.

Factor risk premium The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors. Also called *factor price*.

Factor sensitivity See *factor betas*.

Failure to pay When a borrower does not make a scheduled payment of principal or interest on any outstanding obligations after a grace period.

Fair market value The market price of an asset or liability that trades regularly.

Fair value The amount at which an asset (or liability) could be bought (or incurred) or sold (or settled) in a current transaction between willing parties, that is, other than in a forced or liquidation sale. As defined in IFRS and US GAAP, it is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date.

Feature engineering A process of creating new features by changing or transforming existing features.

Feature selection A process whereby only pertinent features from the dataset are selected for model training. Selecting fewer features decreases model complexity and training time.

Features The independent variables (X's) in a labeled dataset.

Financial contagion A situation in which financial shocks spread from their place of origin to other locales. In essence, a faltering economy infects other, healthier economies.

Financial distress Heightened uncertainty regarding a company's ability to meet its various obligations because of lower or negative earnings.

Financial transaction A purchase involving a buyer having essentially no material synergies with the target (e.g., the purchase of a private company by a company in an unrelated industry or by a private equity firm would typically be a financial transaction).

First-differencing A transformation that subtracts the value of the time series in period $t-1$ from its value in period t.

First-order serial correlation Correlation between adjacent observations in a time series.

Fitting curve A curve which shows in- and out-of-sample error rates (E_{in} and E_{out}) on the y-axis plotted against model complexity on the x-axis.

Fixed price tender offer Offer made by a company to repurchase a specific number of shares at a fixed price that is typically at a premium to the current market price.

Fixed-rate perpetual preferred stock Non-convertible, non-callable preferred stock with a specified dividend rate that has a claim on earnings senior to the claim of common stock, and no maturity date.

Flight to quality During times of market stress, investors sell higher-risk asset classes such as stocks and commodities in favor of default-risk-free government bonds.

Flip-in pill A poison pill takeover defense that dilutes an acquirer's ownership in a target by giving other existing target company shareholders the right to buy additional target company shares at a discount.

Flip-over pill A poison pill takeover defense that gives target company shareholders the right to purchase shares of the acquirer at a significant discount to the market price, which has the effect of causing dilution to all existing acquiring company shareholders.

Float Amounts collected as premium and not yet paid out as benefits.

Floored floater Floating-rate bond with a floor provision that prevents the coupon rate from decreasing below a specified minimum rate. It protects the investor against declining interest rates.

Flotation cost Fees charged to companies by investment bankers and other costs associated with raising new capital.

Forced conversion For a convertible bond, when the issuer calls the bond and forces bondholders to convert their bonds into shares, which typically happens when the underlying share price increases above the conversion price.

Foreign currency transactions Transactions that are denominated in a currency other than a company's functional currency.

Forward curve The term structure of forward rates for loans made on a specific initiation date.

Forward dividend yield A dividend yield based on the anticipated dividend during the next 12 months.

Forward integration A merger involving the purchase of a target that is farther along the value or production chain; for example, to acquire a distributor.

Forward P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Forward price The fixed price or rate at which the transaction, scheduled to occur at the expiration of a forward contract, will take place. This price is agreed to at the initiation date of the forward contract.

Forward pricing model The model that describes the valuation of forward contracts.

Forward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving forward through the network's layers.

Forward rate An interest rate determined today for a loan that will be initiated in a future period.

Forward rate agreement An over-the-counter forward contract in which the underlying is an interest rate on a deposit. A forward rate agreement (FRA) calls for one party to make a fixed interest payment and the other to make an interest payment at a rate to be determined at contract expiration.

Forward rate model The forward pricing model expressed in terms of spot and forward interest rates.

Forward rate parity The proposition that the forward exchange rate is an unbiased predictor of the future spot exchange rate.

Forward value The monetary value of an existing forward contract.

Franking credit A tax credit received by shareholders for the taxes that a corporation paid on its distributed earnings.

Free cash flow The actual cash that would be available to the company's investors after making all investments necessary to maintain the company as an ongoing enterprise (also referred to as free cash flow to the firm); the internally generated funds that can be distributed to the company's investors (e.g., shareholders and bondholders) without impairing the value of the company.

Free cash flow hypothesis The hypothesis that higher debt levels discipline managers by forcing them to make fixed debt service payments and by reducing the company's free cash flow.

Free cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows.

Free cash flow to equity The cash flow available to a company's common shareholders after all operating expenses, interest, and principal payments have been made and necessary investments in working and fixed capital have been made.

Free cash flow to equity model A model of stock valuation that views a stock's intrinsic value as the present value of expected future free cash flows to equity.

Free cash flow to the firm The cash flow available to the company's suppliers of capital after all operating expenses (including taxes) have been paid and necessary investments in working and fixed capital have been made.

Free cash flow to the firm model A model of stock valuation that views the value of a firm as the present value of expected future free cash flows to the firm.

Frequency analysis The process of quantifying how important tokens are in a sentence and in the corpus as a whole. It helps in filtering unnecessary tokens (or features).

Friendly transaction A potential business combination that is endorsed by the managers of both companies.

Functional currency The currency of the primary economic environment in which an entity operates.

Fundamental factor models A multifactor model in which the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices.

Fundamentals Economic characteristics of a business, such as profitability, financial strength, and risk.

Funds available for distribution (FAD) See *adjusted funds from operations*.

Funds from operations (FFO) Net income (computed in accordance with generally accepted accounting principles) *plus* (1) gains and losses from sales of properties and (2) depreciation and amortization.

Futures price The price at which the parties to a futures contract agree to exchange the underlying (or cash). In commodity markets, the price agreed on to deliver or receive a defined quantity (and often quality) of a commodity at a future date.

Futures value The monetary value of an existing futures contract.

FX carry trade An investment strategy that involves taking long positions in high-yield currencies and short positions in low-yield currencies.

Gamma A measure of how sensitive an option's delta is to a change in the underlying. The change in a given instrument's delta for a given small change in the underlying's value, holding everything else constant.

Generalize When a model retains its explanatory power when predicting out-of-sample (i.e., using new data).

Generalized least squares A regression estimation technique that addresses heteroskedasticity of the error term.

Going-concern assumption The assumption that the business will maintain its business activities into the foreseeable future.

Going-concern value A business's value under a going-concern assumption.

Goodwill An intangible asset that represents the excess of the purchase price of an acquired company over the value of the net identifiable assets acquired.

Grant date The day that stock options are granted to employees.

Green bond Bonds in which the proceeds are designated by issuers to fund a specific project or portfolio of projects that have environmental or climate benefits.

Greenmail The purchase of the accumulated shares of a hostile investor by a company that is targeted for takeover by that investor, usually at a substantial premium over market price.

Greenwashing The risk that a green bond's proceeds are not actually used for a beneficial environmental or climate-related project.

Grid search A method of systematically training a model by using various combinations of hyperparameter values, cross validating each model, and determining which combination of hyperparameter values ensures the best model performance.

Gross domestic product A money measure of the goods and services produced within a country's borders over a stated period.

Gross lease A lease under which the tenant pays a gross rent to the landlord, who is responsible for all operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Ground truth The known outcome (i.e., target variable) of each observation in a labelled dataset.

Growth accounting equation The production function written in the form of growth rates. For the basic Cobb–Douglas production function, it states that the growth rate of output equals the rate of technological change plus α multiplied by the growth rate of capital plus $(1 - \alpha)$ multiplied by the growth rate of labor.

Growth capital expenditures Capital expenditures needed for expansion.

Growth option The ability to make additional investments in a project at some future time if the financial results are strong. Also called *expansion option*.

Guideline assets Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline companies Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline public companies Public-company comparables for the company being valued.

Guideline public company method A variation of the market approach; establishes a value estimate based on the observed multiples from trading activity in the shares of public companies viewed as reasonably comparable to the subject private company.

Guideline transactions method A variation of the market approach; establishes a value estimate based on pricing multiples derived from the acquisition of control of entire public or private companies that were acquired.

Harmonic mean A type of weighted mean computed by averaging the reciprocals of the observations and then taking the reciprocal of that average.

Hazard rate The probability that an event will occur, given that it has not already occurred.

Hedonic index Unlike a repeat-sales index, a hedonic index does not require repeat sales of the same property. It requires only one sale. The way it controls for the fact that different properties are selling each quarter is to include variables in the regression that control for differences in the characteristics of the property, such as size, age, quality of construction, and location.

Heteroskedastic With reference to the error term of regression, having a variance that differs across observations.

Heteroskedasticity The property of having a nonconstant variance; refers to an error term with the property that its variance differs across observations.

Heteroskedasticity-consistent standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Hierarchical clustering An iterative unsupervised learning procedure used for building a hierarchy of clusters.

Highest and best use The concept that the best use of a vacant site is the use that would result in the highest value for the land. Presumably, the developer that could earn the highest risk-adjusted profit based on time, effort, construction and development cost, leasing, and exit value would be the one to pay the highest price for the land.

Historical exchange rates For accounting purposes, the exchange rates that existed when the assets and liabilities were initially recorded.

Historical scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.
Historical simulation A simulation method that uses past return data and a random number generator that picks observations from the historical series to simulate an asset's future returns.
Historical simulation method The application of historical price changes to the current portfolio.
Historical stress testing The process that tests how investment strategies would perform under some of the most negative (i.e., adverse) combinations of events and scenarios.
Ho–Lee model The first arbitrage-free term structure model. The model is calibrated to market data and uses a binomial lattice approach to generate a distribution of possible future interest rates.
Holding period return The return that an investor earns during a specified holding period; a synonym for total return.
Holdout samples Data samples that are not used to train a model.
Homoskedasticity The property of having a constant variance; refers to an error term that is constant across observations.
Horizontal merger A merger involving companies in the same line of business, usually as competitors.
Horizontal ownership Companies with mutual business interests (e.g., key customers or suppliers) that have cross-holding share arrangements with each other.
Hostile transaction An attempt to acquire a company against the wishes of the target's managers.
Human capital The accumulated knowledge and skill that workers acquire from education, training, or life experience.
Hybrid approach With respect to forecasting, an approach that combines elements of both top-down and bottom-up analyses.
Hyperparameter A parameter whose value must be set by the researcher before learning begins.
I-spreads Shortened form of "interpolated spreads" and a reference to a linearly interpolated yield.
Illiquidity discount A reduction or discount to value that reflects the lack of depth of trading or liquidity in that asset's market.
Impairment Diminishment in value as a result of carrying (book) value exceeding fair value and/or recoverable value.
Impairment of capital rule A legal restriction that dividends cannot exceed retained earnings.
Implementation shortfall The difference between the money return (or value) on a notional or paper portfolio and the actual portfolio return (or value).
Implied volatility The standard deviation that causes an option pricing model to give the current option price.
In-sample forecast errors The residuals from a fitted time-series model within the sample period used to fit the model.
iNAVs "Indicated" net asset values are intraday "fair value" estimates of an ETF share based on its creation basket.
Income approach A valuation approach that values an asset as the present discounted value of the income expected from it. In the context of real estate, this approach estimates the value of a property based on an expected rate of return. The estimated value is the present value of the expected future income from the property, including proceeds from resale at the end of a typical investment holding period.
Incremental VaR (IVaR) A measure of the incremental effect of an asset on the VaR of a portfolio by measuring the difference between the portfolio's VaR while including a specified asset and the portfolio's VaR with that asset eliminated.
Indenture A written contract between a lender and borrower that specifies the terms of the loan, such as interest rate, interest payment schedule, or maturity.
Independent board directors Directors with no material relationship with the company with regard to employment, ownership, or remuneration.
Independent regulators Regulators recognized and granted authority by a government body or agency. They are not government agencies per se and typically do not rely on government funding.
Independent variable A variable used to explain the dependent variable in a regression; a right-side variable in a regression equation. Also referred to as the *explanatory variable*.
Index CDS A type of credit default swap that involves a combination of borrowers.
Indicator variable A variable that takes on only one of two values, 0 or 1, based on a condition. In simple linear regression, the slope is the difference in the dependent variable for the two conditions. Also referred to as a *dummy variable*.
Industry structure An industry's underlying economic and technical characteristics.
Information gain A metric which quantifies the amount of information that the feature holds about the response. Information gain can be regarded as a form of non-linear correlation between Y and X.
Information ratio (IR) Mean active return divided by active risk; or alpha divided by the standard deviation of diversifiable risk.
Informational frictions Forces that restrict availability, quality, and/or flow of information and its use.
Inside ask See *best ask*.
Inside bid See *best bid*.
Inside spread The spread between the best bid price and the best ask price. Also called the *market bid-ask spread*, *inside bid-ask spread*, or *market spread*.
Insiders Corporate managers and board directors who are also shareholders of a company.
Inter-temporal rate of substitution The ratio of the marginal utility of consumption *s* periods in the future (the numerator) to the marginal utility of consumption today (the denominator).
Intercept The expected value of the dependent variable when the independent variable in a simple linear regression is equal to zero.
Interest rate risk The risk that interest rates will rise and therefore the market value of current portfolio holdings will fall so that their current yields to maturity then match comparable instruments in the marketplace.
Interlocking directorates Corporate structure in which individuals serve on the board of directors of multiple corporations.
Internal rate of return Abbreviated as IRR. Rate of return that discounts future cash flows from an investment to the exact amount of the investment; the discount rate that makes the present value of an investment's costs (outflows) equal to the present value of the investment's benefits (inflows).

International Fisher effect The proposition that nominal interest rate differentials across currencies are determined by expected inflation differentials.

Intrinsic value The value of an asset given a hypothetically complete understanding of the asset's investment characteristics; the value obtained if an option is exercised based on current conditions. The difference between the spot exchange rate and the strike price of a currency.

Inverse price ratio The reciprocal of a price multiple—for example, in the case of a P/E, the "earnings yield" E/P (where P is share price and E is earnings per share).

Investment value The value to a specific buyer, taking account of potential synergies based on the investor's requirements and expectations.

ISDA Master Agreement A standard or "master" agreement published by the International Swaps and Derivatives Association. The master agreement establishes the terms for each party involved in the transaction.

Judicial law Interpretations of courts.

Justified (fundamental) P/E The price-to-earnings ratio that is fair, warranted, or justified on the basis of forecasted fundamentals.

Justified price multiple The estimated fair value of the price multiple, usually based on forecasted fundamentals or comparables.

***K*-fold cross-validation** A technique in which data (excluding test sample and fresh data) are shuffled randomly and then are divided into k equal sub-samples, with $k - 1$ samples used as training samples and one sample, the kth, used as a validation sample.

***K*-means** A clustering algorithm that repeatedly partitions observations into a fixed number, k, of non-overlapping clusters.

***K*-nearest neighbor** A supervised learning technique that classifies a new observation by finding similarities ("nearness") between this new observation and the existing data.

Kalotay–Williams–Fabozzi (KWF) model An arbitrage-free term structure model that describes the dynamics of the log of the short rate and assumes constant drift, no mean reversion, and constant volatility.

Key rate durations Sensitivity of a bond's price to changes in specific maturities on the benchmark yield curve. Also called *partial durations*.

***k*th-order autocorrelation** The correlation between observations in a time series separated by k periods.

Labeled dataset A dataset that contains matched sets of observed inputs or features (X's) and the associated output or target (Y).

Labor force Everyone of working age (ages 16 to 64) who either is employed or is available for work but not working.

Labor force participation rate The percentage of the working age population that is in the labor force.

Labor productivity The quantity of real GDP produced by an hour of labor. More generally, output per unit of labor input.

Labor productivity growth accounting equation States that potential GDP growth equals the growth rate of the labor input plus the growth rate of labor productivity.

Lack of marketability discount An extra return to investors to compensate for lack of a public market or lack of marketability.

LASSO Least absolute shrinkage and selection operator is a type of penalized regression which involves minimizing the sum of the absolute values of the regression coefficients. LASSO can also be used for regularization in neural networks.

Latency The elapsed time between the occurrence of an event and a subsequent action that depends on that event.

Law of one price A principle that states that if two investments have the same or equivalent future cash flows regardless of what will happen in the future, then these two investments should have the same current price.

Leading dividend yield Forecasted dividends per share over the next year divided by current stock price.

Leading P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Learning curve A curve that plots the accuracy rate (= 1 – error rate) in the validation or test samples (i.e., out-of-sample) against the amount of data in the training sample, which is thus useful for describing under- and overfitting as a function of bias and variance errors.

Learning rate A parameter that affects the magnitude of adjustments in the weights in a neural network.

Level One of the three factors (the other two are steepness and curvature) that empirically explain most yield curve shape changes. A shock to the level factor changes the yield for all maturities by an almost identical amount.

Leveraged buyout A transaction whereby the target company management team converts the target to a privately held company by using heavy borrowing to finance the purchase of the target company's outstanding shares.

Leveraged recapitalization A post-offer takeover defense mechanism that involves the assumption of a large amount of debt that is then used to finance share repurchases. The effect is to dramatically change the company's capital structure while attempting to deliver a value to target shareholders in excess of a hostile bid.

Libor–OIS spread The difference between Libor and the overnight indexed swap rate.

Limit order book The book or list of limit orders to buy and sell that pertains to a security.

Lin-log model A regression model in which the independent variable is in logarithmic form.

Linear classifier A binary classifier that makes its classification decision based on a linear combination of the features of each data point.

Linear regression Regression that models the straight-line relationship between the dependent and independent variables. Also known as *least squares regression* and *ordinary least squares regression*.

Linear trend A trend in which the dependent variable changes at a constant rate with time.

Liquidating dividend A dividend that is a return of capital rather than a distribution from earnings or retained earnings.

Liquidation To sell the assets of a company, division, or subsidiary piecemeal, typically because of bankruptcy; the form of bankruptcy that allows for the orderly satisfaction of creditors' claims after which the company ceases to exist.

Liquidation value The value of a company if the company were dissolved and its assets sold individually.

Liquidity preference theory A term structure theory that asserts liquidity premiums exist to compensate investors for the added interest rate risk they face when lending long term.

Liquidity premium The premium or incrementally higher yield that investors demand for lending long term.

Local currency The currency of the country where a company is located.

Local expectations theory A term structure theory that contends the return for all bonds over short periods is the risk-free rate.

Log-lin model A regression model in which the dependent variable is in logarithmic form.

Log-linear model With reference to time-series models, a model in which the growth rate of the time series as a function of time is constant.

Log-log model A regression model in which both the dependent and independent variables are in logarithmic form. Also known as the *double-log model*.

Log-log regression model A regression that expresses the dependent and independent variables as natural logarithms.

Logistic regression (logit model) A qualitative-dependent-variable multiple regression model based on the logistic probability distribution.

Long/short credit trade A credit protection seller with respect to one entity combined with a credit protection buyer with respect to another entity.

Look-ahead bias The bias created by using information that was unknown or unavailable in the time periods over which backtesting is conducted, such as company earnings and macroeconomic indicator values.

Lookback period The time period used to gather a historical data set.

Loss given default The amount that will be lost if a default occurs.

Macroeconomic factor model A multifactor model in which the factors are surprises in macroeconomic variables that significantly explain equity returns.

Macroeconomic factors Factors related to the economy, such as the inflation rate, industrial production, or economic sector membership.

Maintenance capital expenditures Capital expenditures needed to maintain operations at the current level.

Majority shareholders Shareholders that own more than 50% of a corporation's shares.

Majority-vote classifier A classifier that assigns to a new data point the predicted label with the most votes (i.e., occurrences).

Managerialism theories Theories that posit that corporate executives are motivated to engage in mergers to maximize the size of their company rather than shareholder value (a form of agency cost).

Marginal VaR (MVaR) A measure of the effect of a small change in a position size on portfolio VaR.

Market approach Valuation approach that values an asset based on pricing multiples from sales of assets viewed as similar to the subject asset.

Market conversion premium per share For a convertible bond, the difference between the market conversion price and the underlying share price, which allows investors to identify the premium or discount payable when buying a convertible bond rather than the underlying common stock.

Market conversion premium ratio For a convertible bond, the market conversion premium per share expressed as a percentage of the current market price of the shares.

Market efficiency A finance perspective on capital markets that deals with the relationship of price to intrinsic value. The **traditional efficient markets formulation** asserts that an asset's price is the best available estimate of its intrinsic value. The **rational efficient markets formulation** asserts that investors should expect to be rewarded for the costs of information gathering and analysis by higher gross returns.

Market fragmentation Trading the same instrument in multiple venues.

Market impact The effect of the trade on transaction prices. Also called *price impact.*

Market timing Asset allocation in which the investment in the market is increased if one forecasts that the market will outperform T-bills.

Market value of invested capital The market value of debt and equity.

Mature growth rate The earnings growth rate in a company's mature phase; an earnings growth rate that can be sustained long term.

Maximum drawdown The worst cumulative loss ever sustained by an asset or portfolio. More specifically, maximum drawdown is the difference between an asset's or a portfolio's maximum cumulative return and its subsequent lowest cumulative return.

Mean reversion The tendency of a time series to fall when its level is above its mean and rise when its level is below its mean; a mean-reverting time series tends to return to its long-term mean.

Mean square error (MSE) The sum of squares error divided by the degrees of freedom, $n - k - 1$; in a simple linear regression, $n - k - 1 = n - 2$.

Mean square regression (MSR) The sum of squares regression divided by the number of independent variables k; in a simple linear regression, $k = 1$.

Merger The absorption of one company by another; two companies become one entity and one or both of the pre-merger companies ceases to exist as a separate entity.

Metadata Data that describes and gives information about other data.

Method based on forecasted fundamentals An approach to using price multiples that relates a price multiple to forecasts of fundamentals through a discounted cash flow model.

Method of comparables An approach to valuation that involves using a price multiple to evaluate whether an asset is relatively fairly valued, relatively undervalued, or relatively overvalued when compared to a benchmark value of the multiple.

Midquote price The average, or midpoint, of the prevailing bid and ask prices.

Minority interest The proportion of the ownership of a subsidiary not held by the parent (controlling) company.

Minority shareholders Shareholders that own less than 50% of a corporation's shares.

Mispricing Any departure of the market price of an asset from the asset's estimated intrinsic value.

Mixed offering A merger or acquisition that is to be paid for with cash, securities, or some combination of the two.

Model specification With reference to regression, the set of variables included in the regression and the regression equation's functional form.

Molodovsky effect The observation that P/Es tend to be high on depressed EPS at the bottom of a business cycle and tend to be low on unusually high EPS at the top of a business cycle.

Momentum indicators Valuation indicators that relate either price or a fundamental (such as earnings) to the time series of their own past values (or in some cases to their expected value).

Monetary assets and liabilities Assets and liabilities with value equal to the amount of currency contracted for, a fixed amount of currency. Examples are cash, accounts receivable, accounts payable, bonds payable, and mortgages payable. Inventory is not a monetary asset. Most liabilities are monetary.

Monetary/non-monetary method Approach to translating foreign currency financial statements for consolidation in which monetary assets and liabilities are translated at the current exchange rate. Non-monetary assets and liabilities are translated at historical exchange rates (the exchange rates that existed when the assets and liabilities were acquired).

Monetizing Unwinding a position to either capture a gain or realize a loss.

Monitoring costs Costs borne by owners to monitor the management of the company (e.g., board of director expenses).

Monte Carlo simulation A technique that uses the inverse transformation method for converting a randomly generated uniformly distributed number into a simulated value of a random variable of a desired distribution. Each key decision variable in a Monte Carlo simulation requires an assumed statistical distribution; this assumption facilitates incorporating non-normality, fat tails, and tail dependence as well as solving high-dimensionality problems.

Mortgages Loans with real estate serving as collateral for the loans.

Multicollinearity A regression assumption violation that occurs when two or more independent variables (or combinations of independent variables) are highly but not perfectly correlated with each other.

Multiple linear regression Linear regression involving two or more independent variables.

Multiple linear regression model A linear regression model with two or more independent variables.

Mutual information Measures how much information is contributed by a token to a class of texts. MI will be 0 if the token's distribution in all text classes is the same. MI approaches 1 as the token in any one class tends to occur more often in only that particular class of text.

Mutually exclusive projects Mutually exclusive projects compete directly with each other. For example, if Projects A and B are mutually exclusive, you can choose A or B, but you cannot choose both.

N-grams A representation of word sequences. The length of a sequence varies from 1 to *n*. When one word is used, it is a unigram; a two-word sequence is a bigram; and a 3-word sequence is a trigram; and so on.

***n*-Period moving average** The average of the current and immediately prior $n - 1$ values of a time series.

Naked credit default swap A position where the owner of the CDS does not have a position in the underlying credit.

Name entity recognition An algorithm that analyzes individual tokens and their surrounding semantics while referring to its dictionary to tag an object class to the token.

Negative serial correlation Serial correlation in which a positive error for one observation increases the chance of a negative error for another observation, and vice versa.

Net asset balance sheet exposure When assets translated at the current exchange rate are greater in amount than liabilities translated at the current exchange rate. Assets exposed to translation gains or losses exceed the exposed liabilities.

Net asset value The difference between assets and liabilities, all taken at current market values instead of accounting book values.

Net asset value per share Net asset value divided by the number of shares outstanding.

Net lease A lease under which the tenant pays a net rent to the landlord and an additional amount based on the tenant's pro rata share of the operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Net liability balance sheet exposure When liabilities translated at the current exchange rate are greater assets translated at the current exchange rate. Liabilities exposed to translation gains or losses exceed the exposed assets.

Net operating income Gross rental revenue minus operating costs but before deducting depreciation, corporate overhead, and interest expense. In the context of real estate, a measure of the income from the property after deducting operating expenses for such items as property taxes, insurance, maintenance, utilities, repairs, and insurance but before deducting any costs associated with financing and before deducting federal income taxes. It is similar to EBITDA in a financial reporting context.

Net regulatory burden The private costs of regulation less the private benefits of regulation.

Network externalities The impact that users of a good, a service, or a technology have on other users of that product; it can be positive (e.g., a critical mass of users makes a product more useful) or negative (e.g., congestion makes the product less useful).

Neural networks Highly flexible machine learning algorithms that have been successfully applied to a variety of supervised and unsupervised tasks characterized by nonlinearities and interactions among features.

No-arbitrage approach A procedure for obtaining the value of an option based on the creation of a portfolio that replicates the payoffs of the option and deriving the option value from the value of the replicating portfolio.

No-growth company A company without positive expected net present value projects.

No-growth value per share The value per share of a no-growth company, equal to the expected level amount of earnings divided by the stock's required rate of return.

Non-cash rent An amount equal to the difference between the average contractual rent over a lease term (the straight-line rent) and the cash rent actually paid during a period. This figure is one of the deductions made from FFO to calculate AFFO.

Non-convergence trap A situation in which a country remains relatively poor, or even falls further behind, because it fails to implement necessary institutional reforms and/or adopt leading technologies.

Non-monetary assets and liabilities Assets and liabilities that are not monetary assets and liabilities. Non-monetary assets include inventory, fixed assets, and intangibles, and non-monetary liabilities include deferred revenue.

Non-renewable resources Finite resources that are depleted once they are consumed; oil and coal are examples.

Non-residential properties Commercial real estate properties other than multi-family properties, farmland, and timberland.

Nonearning assets Cash and investments (specifically cash, cash equivalents, and short-term investments).

Nonstationarity With reference to a random variable, the property of having characteristics, such as mean and variance, that are not constant through time.

Normal EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normalized EPS*.

Normalized earnings The expected level of mid-cycle earnings for a company in the absence of any unusual or temporary factors that affect profitability (either positively or negatively).

Normalized EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normal EPS*.

Normalized P/E P/E based on normalized EPS data.

Notional amount The amount of protection being purchased in a CDS.

NTM P/E Next 12-month P/E: current market price divided by an estimated next 12-month EPS.

Off-the-run A series of securities or indexes that were issued/created prior to the most recently issued/created series.

On-the-run The most recently issued/created series of securities or indexes.

One hot encoding The process by which categorical variables are converted into binary form (0 or 1) for machine reading. It is one of the most common methods for handling categorical features in text data.

One-sided durations Effective durations when interest rates go up or down, which are better at capturing the interest rate sensitivity of bonds with embedded options that do not react symmetrically to positive and negative changes in interest rates of the same magnitude.

One-tier board Board structure consisting of a single board of directors, composed of executive (internal) and non-executive (external) directors.

Opportunity cost The value that investors forgo by choosing a particular course of action; the value of something in its best alternative use.

Optimal capital structure The capital structure at which the value of the company is maximized.

Option-adjusted spread (OAS) Constant spread that, when added to all the one-period forward rates on the interest rate tree, makes the arbitrage-free value of the bond equal to its market price.

Orderly liquidation value The estimated gross amount of money that could be realized from the liquidation sale of an asset or assets, given a reasonable amount of time to find a purchaser or purchasers.

Organic growth Company growth in output or sales that is achieved by making investments internally (i.e., excludes growth achieved through mergers and acquisitions).

Other comprehensive income Changes to equity that bypass (are not reported in) the income statement; the difference between comprehensive income and net income.

Other post-employment benefits Promises by the company to pay benefits in the future, such as life insurance premiums and all or part of health care insurance for its retirees.

Out-of-sample forecast errors The differences between actual and predicted values of time series outside the sample period used to fit the model.

Overfitting When a model fits the training data too well and so does not generalize well to new data.

Overnight indexed swap (OIS) rate An interest rate swap in which the periodic floating rate of the swap equals the geometric average of a daily unsecured overnight rate (or overnight index rate).

Pairs trading An approach to trading that uses pairs of closely related stocks, buying the relatively undervalued stock and selling short the relatively overvalued stock.

Par curve A hypothetical yield curve for coupon-paying Treasury securities that assumes all securities are priced at par.

Par swap A swap in which the fixed rate is set so that no money is exchanged at contract initiation.

Parametric method A method of estimating VaR that uses the historical mean, standard deviation, and correlation of security price movements to estimate the portfolio VaR. Generally assumes a normal distribution but can be adapted to non-normal distributions with the addition of skewness and kurtosis. Sometimes called the *variance–covariance method* or the *analytical method*.

Partial regression coefficients The slope coefficients in a multiple regression. Also called *partial slope coefficients*.

Partial slope coefficients The slope coefficients in a multiple regression. Also called *partial regression coefficients*.

Parts of speech An algorithm that uses language structure and dictionaries to tag every token in the text with a corresponding part of speech (i.e., noun, verb, adjective, proper noun, etc.).

Payout amount The loss given default times the notional.

Payout policy The principles by which a company distributes cash to common shareholders by means of cash dividends and/or share repurchases.

Payouts Cash dividends and the value of shares repurchased in any given year.

Pecking order theory The theory that managers consider how their actions might be interpreted by outsiders and thus order their preferences for various forms of corporate financing. Forms of financing that are least visible to outsiders (e.g., internally generated funds) are most preferable to managers and those that are most visible (e.g., equity) are least preferable.

PEG ratio The P/E-to-growth ratio, calculated as the stock's P/E divided by the expected earnings growth rate.

Penalized regression A regression that includes a constraint such that the regression coefficients are chosen to minimize the sum of squared residuals *plus* a penalty term that increases in size with the number of included features.

Pension obligation The present value of future benefits earned by employees for service provided to date.

Perfect capital markets Markets in which, by assumption, there are no taxes, transaction costs, or bankruptcy costs and in which all investors have equal ("symmetric") information.

Perpetuity A perpetual annuity, or a set of never-ending level sequential cash flows, with the first cash flow occurring one period from now.

Persistent earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *continuing earnings*, or *underlying earnings*.

Pet projects Projects in which influential managers want the corporation to invest. Often, unfortunately, pet projects are selected without undergoing normal capital budgeting analysis.

Physical settlement Involves actual delivery of the debt instrument in exchange for a payment by the credit protection seller of the notional amount of the contract.

Point-in-time data Data consisting of the exact information available to market participants as of a given point in time. Point-in-time data is used to address look-ahead bias.

Poison pill A pre-offer takeover defense mechanism that makes it prohibitively costly for an acquirer to take control of a target without the prior approval of the target's board of directors.

Poison puts A pre-offer takeover defense mechanism that gives target company bondholders the right to sell their bonds back to the target at a pre-specified redemption price, typically at or above par value; this defense increases the need for cash and raises the cost of the acquisition.

Portfolio balance approach A theory of exchange rate determination that emphasizes the portfolio investment decisions of global investors and the requirement that global investors willingly hold all outstanding securities denominated in each currency at prevailing prices and exchange rates.

Positive serial correlation Serial correlation in which a positive error for one observation increases the chance of a positive error for another observation; a negative error for one observation increases the chance of a negative error for another observation.

Potential GDP The maximum amount of output an economy can sustainably produce without inducing an increase in the inflation rate. The output level that corresponds to full employment with consistent wage and price expectations.

Precision In error analysis for classification problems it is ratio of correctly predicted positive classes to all predicted positive classes. Precision is useful in situations where the cost of false positives (FP), or Type I error, is high.

Preferred habitat theory A term structure theory that contends that investors have maturity preferences and require yield incentives before they will buy bonds outside of their preferred maturities.

Premise of value The status of a company in the sense of whether it is assumed to be a going concern or not.

Premium leg The series of payments the credit protection buyer promises to make to the credit protection seller.

Premiums Amounts paid by the purchaser of insurance products.

Present value model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Present value of growth opportunities The difference between the actual value per share and the no-growth value per share. Also called *value of growth*.

Presentation currency The currency in which financial statement amounts are presented.

Price improvement When trade execution prices are better than quoted prices.

Price momentum A valuation indicator based on past price movement.

Price multiples The ratio of a stock's market price to some measure of value per share.

Price-setting option The operational flexibility to adjust prices when demand varies from what is forecast. For example, when demand exceeds capacity, the company could benefit from the excess demand by increasing prices.

Price-to-earnings ratio (P/E) The ratio of share price to earnings per share.

Priced risk Risk for which investors demand compensation for bearing (e.g., equity risk, company-specific factors, macroeconomic factors).

Principal components analysis (PCA) An unsupervised ML technique used to transform highly correlated features of data into a few main, uncorrelated composite variables.

Principle of no arbitrage In well-functioning markets, prices will adjust until there are no arbitrage opportunities.

Prior transaction method A variation of the market approach; considers actual transactions in the stock of the subject private company.

Private market value The value derived using a sum-of-the-parts valuation.

Probability of default The probability that a bond issuer will not meet its contractual obligations on schedule.

Probability of survival The probability that a bond issuer will meet its contractual obligations on schedule.

Procedural law The body of law that focuses on the protection and enforcement of the substantive laws.

Production-flexibility option The operational flexibility to alter production when demand varies from forecast. For example, if demand is strong, a company may profit from employees working overtime or from adding additional shifts.

Project sequencing To defer the decision to invest in a future project until the outcome of some or all of a current project is known. Projects are sequenced through time, so that investing in a project creates the option to invest in future projects.

Projection error The vertical (perpendicular) distance between a data point and a given principal component.

Prospective P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Protection leg The contingent payment that the credit protection seller may have to make to the credit protection buyer.

Protection period Period during which a bond's issuer cannot call the bond.

Provision for loan losses An income statement expense account that increases the amount of the allowance for loan losses.

Proxy fight An attempt to take control of a company through a shareholder vote.

Proxy statement A public document that provides the material facts concerning matters on which shareholders will vote.

Prudential supervision Regulation and monitoring of the safety and soundness of financial institutions to promote financial stability, reduce system-wide risks, and protect customers of financial institutions.

Pruning A regularization technique used in CART to reduce the size of the classification or regression tree—by pruning, or removing, sections of the tree that provide little classifying power.

Purchasing power gain A gain in value caused by changes in price levels. Monetary liabilities experience purchasing power gains during periods of inflation.

Purchasing power loss A loss in value caused by changes in price levels. Monetary assets experience purchasing power loss during periods of inflation.

Purchasing power parity (PPP) The idea that exchange rates move to equalize the purchasing power of different currencies.

Pure expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *unbiased expectations theory*.

Pure factor portfolio A portfolio with sensitivity of 1 to the factor in question and a sensitivity of 0 to all other factors.

Putable bond Bond that includes an embedded put option, which gives the bondholder the right to put back the bonds to the issuer prior to maturity, typically when interest rates have risen and higher-yielding bonds are available.

Qualitative dependent variables Dummy variables used as dependent variables rather than as independent variables.

Quality of earnings analysis The investigation of issues relating to the accuracy of reported accounting results as reflections of economic performance. Quality of earnings analysis is broadly understood to include not only earnings management but also balance sheet management.

Random forest classifier A collection of a large number of decision trees trained via a bagging method.

Random walk A time series in which the value of the series in one period is the value of the series in the previous period plus an unpredictable random error.

Rational efficient markets formulation See *market efficiency*.

Readme files Text files provided with raw data that contain information related to a data file. They are useful for understanding the data and how they can be interpreted correctly.

Real estate investment trusts Tax-advantaged entities (companies or trusts) that own, operate, and—to a limited extent—develop income-producing real estate property.

Real estate operating companies Regular taxable real estate ownership companies that operate in the real estate industry in countries that do not have a tax-advantaged REIT regime in place or that are engage in real estate activities of a kind and to an extent that do not fit in their country's REIT framework.

Real interest rate parity The proposition that real interest rates will converge to the same level across different markets.

Real options Options that relate to investment decisions such as the option to time the start of a project, the option to adjust its scale, or the option to abandon a project that has begun.

Rebalance return A return from rebalancing the component weights of an index.

Recall Also known as *sensitivity*, in error analysis for classification problems it is the ratio of correctly predicted positive classes to all actual positive classes. Recall is useful in situations where the cost of false negatives (FN), or Type II error, is high.

Reconstitution When dealers recombine appropriate individual zero-coupon securities and reproduce an underlying coupon Treasury.

Recovery rate The percentage of the loss recovered.

Redemption basket The list of securities (and share amounts) the authorized participant (AP) receives when it redeems ETF shares back to the ETF manager. The redemption basket is published each business day.

Reference entity The borrower (debt issuer) covered by a single-name CDS.

Reference obligation A particular debt instrument issued by the borrower that is the designated instrument being covered.

Regime With reference to a time series, the underlying model generating the times series.

Regression analysis A tool for examining whether a variable is useful for explaining another variable.

Regression coefficients The intercept and slope coefficient(s) of a regression.

Regular expression (regex) A series of texts that contains characters in a particular order. Regex is used to search for patterns of interest in a given text.

Regularization A term that describes methods for reducing statistical variability in high-dimensional data estimation problems.

Regulatory arbitrage Entities identify and use some aspect of regulations that allows them to exploit differences in economic substance and regulatory interpretation or in foreign and domestic regulatory regimes to their (the entities') advantage.

Regulatory burden The costs of regulation for the regulated entity.

Regulatory capture Theory that regulation often arises to enhance the interests of the regulated.

Regulatory competition Regulators may compete to provide a regulatory environment designed to attract certain entities.

Reinforcement learning Machine learning in which a computer learns from interacting with itself or data generated by the same algorithm.

Relative-strength indicators Valuation indicators that compare a stock's performance during a period either to its own past performance or to the performance of some group of stocks.

Relative valuation models A model that specifies an asset's value relative to the value of another asset.

Relative VaR See *ex ante tracking error*.

Relative version of PPP The hypothesis that changes in (nominal) exchange rates over time are equal to national inflation rate differentials.

Renewable resources Resources that can be replenished, such as a forest.

Rental price of capital The cost per unit of time to rent a unit of capital.

Repeat sales index As the name implies, this type of index relies on repeat sales of the same property. In general, the idea supporting this type of index is that because it is the same property that sold twice, the change in value between the two sale dates indicates how market conditions have changed over time.

Replacement cost In the context of real estate, the value of a building assuming it was built today using current construction costs and standards.

Reporting unit For financial reporting under US GAAP, an operating segment or one level below an operating segment (referred to as a component).

Required rate of return The minimum rate of return required by an investor to invest in an asset, given the asset's riskiness.

Residential properties Properties that provide housing for individuals or families. Single-family properties may be owner-occupied or rental properties, whereas multi-family properties are rental properties even if the owner or manager occupies one of the units.

Residual The difference between an observation and its predicted value, where the predicted value is based on the estimated linear relation between the dependent and independent variables using sample data.

Residual autocorrelations The sample autocorrelations of the residuals.

Residual income Earnings for a given period, minus a deduction for common shareholders' opportunity cost in generating the earnings. Also called *economic profit* or *abnormal earnings*.

Residual income method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Residual income model (RIM) A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share. Also called *discounted abnormal earnings model* or *Edwards–Bell–Ohlson model*.

Residual loss Agency costs that are incurred despite adequate monitoring and bonding of management.

Restructuring Reorganizing the capital structure of a firm.

Return on capital employed Operating profit divided by capital employed (debt and equity capital).

Return on invested capital A measure of the after-tax profitability of the capital invested by the company's shareholders and debtholders.

Reverse carry arbitrage A strategy involving the short sale of the underlying and an offsetting opposite position in the derivative.

Reverse stock split A reduction in the number of shares outstanding with a corresponding increase in share price but no change to the company's underlying fundamentals.

Reverse stress testing A risk management approach in which the user identifies key risk exposures in the portfolio and subjects those exposures to extreme market movements.

Reviewed financial statements A type of non-audited financial statements; typically provide an opinion letter with representations and assurances by the reviewing accountant that are less than those in audited financial statements.

Rho The change in a given derivative instrument for a given small change in the risk-free interest rate, holding everything else constant. Rho measures the sensitivity of the option to the risk-free interest rate.

Risk budgeting The allocation of an asset owner's total risk appetite among groups or divisions (in the case of a trading organization) or among strategies and managers (in the case of an institutional or individual investor).

Risk decomposition The process of converting a set of holdings in a portfolio into a set of exposures to risk factors.

Risk factors Variables or characteristics with which individual asset returns are correlated. Sometimes referred to simply as *factors*.

Risk parity A portfolio allocation scheme that weights stocks or factors based on an equal risk contribution.

Robust standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Roll When an investor moves its investment position from an older series to the most current series.

Roll return The component of the return on a commodity futures contract attributable to rolling long futures positions forward through time. Also called *roll yield*.

Rolling down the yield curve A maturity trading strategy that involves buying bonds with a maturity longer than the intended investment horizon. Also called *riding the yield curve*.

Rolling windows A backtesting method that uses a rolling-window (or walk-forward) framework, rebalances the portfolio after each period, and then tracks performance over time. As new information arrives each period, the investment manager optimizes (revises and tunes) the model and readjusts stock positions.

Root mean squared error (RMSE) The square root of the average squared forecast error; used to compare the out-of-sample forecasting performance of forecasting models.

Sale-leaseback A situation in which a company sells the building it owns and occupies to a real estate investor and the company then signs a long-term lease with the buyer to continue to occupy the building. At the end of the lease, use of the property reverts to the landlord.

Sales comparison approach In the context of real estate, this approach estimates value based on what similar or comparable properties (comparables) transacted for in the current market.

Scaled earnings surprise Unexpected earnings divided by the standard deviation of analysts' earnings forecasts.

Scaling The process of adjusting the range of a feature by shifting and changing the scale of the data. Two of the most common ways of scaling are normalization and standardization.

Scatter plot A chart in which two variables are plotted along the axis and points on the chart represent pairs of the two variables. In regression, the dependent variable is plotted on the vertical axis and the independent variable is plotted along the horizontal axis. Also known as a scattergram and a *scatter diagram*.

Scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Scree plots A plot that shows the proportion of total variance in the data explained by each principal component.

Screening The application of a set of criteria to reduce a set of potential investments to a smaller set having certain desired characteristics.

Seasonality A characteristic of a time series in which the data experience regular and predictable periodic changes; for example, fan sales are highest during the summer months.

Secured overnight financing rate (SOFR) A daily volume-weighted index of rates on qualified cash borrowings collateralized by US Treasuries that is expected to replace Libor as a floating reference rate for swaps.

Securities offering A merger or acquisition in which target shareholders are to receive shares of the acquirer's common stock as compensation.

Security selection risk See *active specific risk*.

Segmented markets theory A term structure theory that contends yields are solely a function of the supply and demand for funds of a particular maturity.

Self-regulating organizations (SROs) Self-regulating bodies that are given recognition and authority, including enforcement power, by a government body or agency.

Self-regulatory bodies Private, non-governmental organizations that both represent and regulate their members. Some self-regulating organizations are also independent regulators.

Sell-side analysts Analysts who work at brokerages.

Sensitivity analysis A technique for exploring how a target variable (e.g., portfolio returns) and risk profiles are affected by changes in input variables (e.g., the distribution of asset or factor returns).

Sentence length The number of characters, including spaces, in a sentence.

Serially correlated With reference to regression errors, errors that are correlated across observations.

Service period For employee stock options, usually the period between the grant date and the vesting date.

Settled in arrears An arrangement in which the interest payment is made (i.e., settlement occurs) at the maturity of the underlying instrument.

Settlement In the case of a credit event, the process by which the two parties to a CDS contract satisfy their respective obligations.

Shaping risk The sensitivity of a bond's price to the changing shape of the yield curve.

Share repurchase A transaction in which a company buys back its own shares. Unlike stock dividends and stock splits, share repurchases use corporate cash.

Shareholder activism Strategies used by shareholders to attempt to compel a company to act in a desired manner.

Shareholders' equity Total assets minus total liabilities.

Shark repellents A pre-offer takeover defense mechanism involving the corporate charter (e.g., staggered boards of directors and supermajority provisions).

Simple linear regression (SLR) A regression that summarizes the relation between the dependent variable and a single independent variable.

Simulation A technique for exploring how a target variable (e.g. portfolio returns) would perform in a hypothetical environment specified by the user, rather than a historical setting.

Single-name CDS Credit default swap on one specific borrower.

Sinking fund bond A bond that requires the issuer to set aside funds over time to retire the bond issue, thus reducing credit risk.

Slope coefficient The coefficient of an independent variable that represents the average change in the dependent variable for a one-unit change in the independent variable.

Soft margin classification An adaptation in the support vector machine algorithm which adds a penalty to the objective function for observations in the training set that are misclassified.

Special dividend A dividend paid by a company that does not pay dividends on a regular schedule or a dividend that supplements regular cash dividends with an extra payment.

Spin-off A form of restructuring in which shareholders of a parent company receive a proportional number of shares in a new, separate entity; shareholders end up owning stock in two different companies where there used to be one.

Split-off A form of restructuring in which shareholders of the parent company are given shares in a newly created entity in exchange for their shares of the parent company.

Split-rate tax system In reference to corporate taxes, a split-rate system taxes earnings to be distributed as dividends at a different rate than earnings to be retained. Corporate profits distributed as dividends are taxed at a lower rate than those retained in the business.

Spot curve The term structure of spot rates for loans made today.

Spot price The current price of an asset or security. For commodities, the current price to deliver a physical commodity to a specific location or purchase and transport it away from a designated location.

Spot rate The interest rate that is determined today for a risk-free, single-unit payment at a specified future date.

Spot yield curve The term structure of spot rates for loans made today.

Stabilized NOI In the context of real estate, the expected NOI when a renovation is complete.

Stable dividend policy A policy in which regular dividends are paid that reflect long-run expected earnings. In contrast to a constant dividend payout ratio policy, a stable dividend policy does not reflect short-term volatility in earnings.

Standard error of the estimate A measure of the fit of a regression line, calculated as the square root of the mean square error. Also known as the *standard error of the regression* and the *root mean square error*.

Standard error of the forecast A measure of the uncertainty associated with a forecasted value of the dependent variable that depends on the standard error of the estimate, the variability of the independent variable, the deviation of the forecasted independent variable from the mean in the regression, and the number of observations.

Standard error of the slope coefficient The standard error of the slope, which in a simple linear regression is the ratio of the model's standard error of the estimate (s_e) to the square root of the variation of the independent variable.

Standardized beta With reference to fundamental factor models, the value of the attribute for an asset minus the average value of the attribute across all stocks, divided by the standard deviation of the attribute across all stocks.

Standardized unexpected earnings Unexpected earnings per share divided by the standard deviation of unexpected earnings per share over a specified prior time period.

Static trade-off theory of capital structure A theory pertaining to a company's optimal capital structure. The optimal level of debt is found at the point where additional debt would cause the costs of financial distress to increase by a greater amount than the benefit of the additional tax shield.

Statistical factor model A multifactor model in which statistical methods are applied to a set of historical returns to determine portfolios that best explain either historical return covariances or variances.

Statutes Laws enacted by legislative bodies.

Statutory merger A merger in which one company ceases to exist as an identifiable entity and all its assets and liabilities become part of a purchasing company.

Steady-state rate of growth The constant growth rate of output (or output per capita) that can or will be sustained indefinitely once it is reached. Key ratios, such as the capital–output ratio, are constant on the steady-state growth path.

Steepness The difference between long-term and short-term yields that constitutes one of the three factors (the other two are level and curvature) that empirically explain most of the changes in the shape of the yield curve.

Stock dividend A type of dividend in which a company distributes additional shares of its common stock to shareholders instead of cash.

Stock purchase An acquisition in which the acquirer gives the target company's shareholders some combination of cash and securities in exchange for shares of the target company's stock.

Stop-loss limit Constraint used in risk management that requires a reduction in the size of a portfolio, or its complete liquidation, when a loss of a particular size occurs in a specified period.

Straight bond An underlying option-free bond with a specified issuer, issue date, maturity date, principal amount and repayment structure, coupon rate and payment structure, and currency denomination.

Straight-line rent The average annual rent under a multi-year lease agreement that contains contractual increases in rent during the life of the lease.

Straight-line rent adjustment See *non-cash rent.*

Straight voting Voting structure in which shareholders are granted the right of one vote for each share owned.

Stranded assets Assets that are obsolete or not economically viable.

Strategic transaction A purchase involving a buyer that would benefit from certain synergies associated with owning the target firm.

Stress tests A risk management technique that assesses the portfolio's response to extreme market movements.

Stripping A dealer's ability to separate a bond's individual cash flows and trade them as zero-coupon securities.

Subsidiary merger A merger in which the company being purchased becomes a subsidiary of the purchaser.

Substantive law The body of law that focuses on the rights and responsibilities of entities and relationships among entities.

Succession event A change of corporate structure of the reference entity, such as through a merger, a divestiture, a spinoff, or any similar action, in which ultimate responsibility for the debt in question is unclear.

Sum of squares error (SSE) The sum of the squared deviations of (1) the value of the dependent variable and (2) the value of the dependent variable based on the estimated regression line. Also referred to as the *residual sum of squares.*

Sum of squares regression (SSR) The sum of the squared deviations of (1) the value of the dependent variable based on the estimated regression line and (2) the mean of the dependent variable.

Sum of squares total (SST) The sum of the squared deviations of the dependent variable from its mean; the variation of the dependent variable. Also referred to as the *total sum of squares.*

Sum-of-the-parts valuation A valuation that sums the estimated values of each of a company's businesses as if each business were an independent going concern.

Summation operator A functional part of a neural network's node that multiplies each input value received by a weight and sums the weighted values to form the total net input, which is then passed to the activation function.

Supernormal growth Above-average or abnormally high growth rate in earnings per share.

Supervised learning Machine learning where algorithms infer patterns between a set of inputs (the *X*'s) and the desired output (*Y*). The inferred pattern is then used to map a given input set into a predicted output.

Support vector machine A linear classifier that determines the hyperplane that optimally separates the observations into two sets of data points.

Survivorship bias The bias that results when data as of a given date reflects only those entities that have survived to that date. Entities can include any element of an index or list that is constituted through time: stocks, investment funds, etc. Survivorship bias is a form of look-ahead bias.

Sustainable growth rate The rate of dividend (and earnings) growth that can be sustained over time for a given level of return on equity, keeping the capital structure constant and without issuing additional common stock.

Swap curve The term structure of swap rates.

Swap rate The "price" that swap traders quote among one another. It is the rate at which the present value of all the expected floating-rate payments received over the life of the floating-rate bond equal the present value of all the expected fixed-rate payments made over the life of the fixed-rate bond.

Swap rate curve The term structure of swap rates.

Swap spread The difference between the fixed rate on an interest rate swap and the rate on a Treasury note with equivalent maturity; it reflects the general level of credit risk in the market.

Systematic risk Risk that affects the entire market or economy; it cannot be avoided and is inherent in the overall market. Systematic risk is also known as non-diversifiable or market risk.

Systemic risk The risk of failure of the financial system.

Tail risk The risk that losses in extreme events could be greater than would be expected for a portfolio of assets with a normal distribution.

Takeover A merger; the term may be applied to any transaction but is often used in reference to hostile transactions.

Takeover premium The amount by which the takeover price for each share of stock must exceed the current stock price in order to entice shareholders to relinquish control of the company to an acquirer.

Tangible book value per share Common shareholders' equity minus intangible assets reported on the balance sheet, divided by the number of shares outstanding.

Target In machine learning, the dependent variable (*Y*) in a labeled dataset; the company in a merger or acquisition that is being acquired.

Target capital structure A company's chosen proportions of debt and equity.

Target company See *target.*

Target payout ratio A strategic corporate goal representing the long-term proportion of earnings that the company intends to distribute to shareholders as dividends.

Taxable REIT subsidiaries Subsidiaries that pay income taxes on earnings from non-REIT-qualifying activities like merchant development or third-party property management.

Technical indicators Momentum indicators based on price.

TED spread A measure of perceived credit risk determined as the difference between Libor and the T-bill yield of matching maturity.

Temporal method A variation of the monetary/non-monetary translation method that requires not only monetary assets and liabilities, but also non-monetary assets and liabilities that are measured at their current value on the balance sheet date to be translated at the current exchange rate. Assets and liabilities are translated at rates consistent with the timing of their measurement value. This method is typically used when the functional currency is other than the local currency.

Tender offer A public offer whereby the acquirer invites target shareholders to submit ("tender") their shares in return for the proposed payment.

Term frequency (TF) Ratio of the number of times a given token occurs in all the texts in the dataset to the total number of tokens in the dataset.

Term premium The additional return required by lenders to invest in a bond to maturity net of the expected return from continually reinvesting at the short-term rate over that same time horizon.

Terminal price multiples The price multiple for a stock assumed to hold at a stated future time.

Terminal share price The share price at a particular point in the future.

Terminal value of the stock The analyst's estimate of a stock's value at a particular point in the future. Also called *continuing value of the stock*.

Test sample A data sample that is used to test a model's ability to predict well on new data.

Theta The change in a derivative instrument for a given small change in calendar time, holding everything else constant. Specifically, the theta calculation assumes nothing changes except calendar time. Theta also reflects the rate at which an option's time value decays.

Time series A set of observations on a variable's outcomes in different time periods.

Tobin's *q* The ratio of the market value of debt and equity to the replacement cost of total assets.

Token The equivalent of a word (or sometimes a character).

Tokenization The process of splitting a given text into separate tokens. Tokenization can be performed at the word or character level but is most commonly performed at word level.

Top-down approach With respect to forecasting, an approach that usually begins at the level of the overall economy. Forecasts are then made at more narrowly defined levels, such as sector, industry, and market for a specific product.

Total factor productivity (TFP) A multiplicative scale factor that reflects the general level of productivity or technology in the economy. Changes in total factor productivity generate proportional changes in output for any input combination.

Total invested capital The sum of market value of common equity, book value of preferred equity, and face value of debt.

Tracking error The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking risk*.

Tracking risk The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking error*.

Trailing dividend yield The reciprocal of current market price divided by the most recent annualized dividend.

Trailing P/E A stock's current market price divided by the most recent four quarters of EPS (or the most recent two semi-annual periods for companies that report interim data semi-annually). Also called *current P/E*.

Training sample A data sample that is used to train a model.

Tranche CDS A type of credit default swap that covers a combination of borrowers but only up to pre-specified levels of losses.

Transaction exposure The risk of a change in value between the transaction date and the settlement date of an asset of liability denominated in a foreign currency.

Treasury shares/stock Shares that were issued and subsequently repurchased by the company.

Trend A long-term pattern of movement in a particular direction.

Triangular arbitrage An arbitrage transaction involving three currencies that attempts to exploit inconsistencies among pairwise exchange rates.

Trimming Also called truncation, it is the process of removing extreme values and outliers from a dataset.

Triple-net leases Common leases in the United States and Canada that require each tenant to pay its share of the following three operating expenses: common area maintenance and repair expenses; property taxes; and building insurance costs. Also known as *NNN leases*.

Two-tier board Board structure consisting of a supervisory board that oversees a management board.

Unbiased expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *pure expectations theory*.

Unconditional heteroskedasticity Heteroskedasticity of the error term that is not correlated with the values of the independent variable(s) in the regression.

Uncovered interest rate parity The proposition that the expected return on an uncovered (i.e., unhedged) foreign currency (risk-free) investment should equal the return on a comparable domestic currency investment.

Underlying earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *core earnings*, or *persistent earnings*.

Unexpected earnings The difference between reported EPS and expected EPS. Also referred to as an *earnings surprise*.

Unit root A time series that is not covariance stationary is said to have a unit root.

Unsupervised learning Machine learning that does not make use of labeled data.

Upfront payment The difference between the credit spread and the standard rate paid by the protection buyer if the standard rate is insufficient to compensate the protection seller. Also called *upfront premium*.

Upfront premium See *upfront payment*.

Upstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary company) such that the associate company records a profit on its income statement. An example is a sale of inventory by the associate to the investor company or by a subsidiary to a parent company.

Validation sample A data sample that is used to validate and tune a model.

Valuation The process of determining the value of an asset or service either on the basis of variables perceived to be related to future investment returns or on the basis of comparisons with closely similar assets.

Value additivity An arbitrage opportunity when the value of the whole equals the sum of the values of the parts.

Value at risk (VaR) The minimum loss that would be expected a certain percentage of the time over a certain period of time given the assumed market conditions.

Value of growth The difference between the actual value per share and the no-growth value per share.

Variance error Describes how much a model's results change in response to new data from validation and test samples. Unstable models pick up noise and produce high variance error, causing overfitting and high out-of-sample error.

Vasicek model A partial equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is constant.

Vega The change in a given derivative instrument for a given small change in volatility, holding everything else constant. A sensitivity measure for options that reflects the effect of volatility.

Venture capital investors Private equity investors in development-stage companies.

Vertical merger A merger involving companies at different positions of the same production chain; for example, a supplier or a distributor.

Vertical ownership Ownership structure in which a company or group that has a controlling interest in two or more holding companies, which in turn have controlling interests in various operating companies.

Vested benefit obligation The actuarial present value of vested benefits.

Vesting date The date that employees can first exercise stock options.

Visibility The extent to which a company's operations are predictable with substantial confidence.

Voting caps Legal restrictions on the voting rights of large share positions.

Web spidering (scraping or crawling) programs Programs that extract raw content from a source, typically web pages.

Weighted average cost of capital (WACC) A weighted average of the after-tax required rates of return on a company's common stock, preferred stock, and long-term debt, where the weights are the fraction of each source of financing in the company's target capital structure.

Weighted harmonic mean See *harmonic mean*.

White-corrected standard errors A synonym for robust standard errors.

White knight A third party that is sought out by the target company's board to purchase the target in lieu of a hostile bidder.

White squire A third party that is sought out by the target company's board to purchase a substantial minority stake in the target—enough to block a hostile takeover without selling the entire company.

Winner's curse The tendency for the winner in certain competitive bidding situations to overpay, whether because of overestimation of intrinsic value, emotion, or information asymmetries.

Winsorization The process of replacing extreme values and outliers in a dataset with the maximum (for large value outliers) and minimum (for small value outliers) values of data points that are not outliers.

Write-down A reduction in the value of an asset as stated in the balance sheet.

Yield curve factor model A model or a description of yield curve movements that can be considered realistic when compared with historical data.

Zero A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

Zero-coupon bond A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

CFA Institute®
CFA Program

DERIVATIVES, ALTERNATIVE INVESTMENTS, AND PORTFOLIO MANAGEMENT

CFA® Program Curriculum
2022 • LEVEL II • VOLUME 5

WILEY

CONTENTS

Alternative Investments

Portfolio Management

How to Use the CFA Program Curriculum

Congratulations on your decision to enter the Chartered Financial Analyst (CFA®) Program. This exciting and rewarding program of study reflects your desire to become a serious investment professional. You are embarking on a program noted for its high ethical standards and the breadth of knowledge, skills, and abilities (competencies) it develops. Your commitment should be educationally and professionally rewarding.

The credential you seek is respected around the world as a mark of accomplishment and dedication. Each level of the program represents a distinct achievement in professional development. Successful completion of the program is rewarded with membership in a prestigious global community of investment professionals. CFA charterholders are dedicated to life-long learning and maintaining currency with the ever-changing dynamics of a challenging profession. CFA Program enrollment represents the first step toward a career-long commitment to professional education.

The CFA exam measures your mastery of the core knowledge, skills, and abilities required to succeed as an investment professional. These core competencies are the basis for the Candidate Body of Knowledge (CBOK™). The CBOK consists of four components:

- A broad outline that lists the major CFA Program topic areas (www.cfainstitute.org/programs/cfa/curriculum/cbok);
- Topic area weights that indicate the relative exam weightings of the top-level topic areas (www.cfainstitute.org/programs/cfa/curriculum);
- Learning outcome statements (LOS) that advise candidates about the specific knowledge, skills, and abilities they should acquire from readings covering a topic area (LOS are provided in candidate study sessions and at the beginning of each reading); and
- CFA Program curriculum that candidates receive upon exam registration.

Therefore, the key to your success on the CFA exams is studying and understanding the CBOK. The following sections provide background on the CBOK, the organization of the curriculum, features of the curriculum, and tips for designing an effective personal study program.

BACKGROUND ON THE CBOK

CFA Program is grounded in the practice of the investment profession. CFA Institute performs a continuous practice analysis with investment professionals around the world to determine the competencies that are relevant to the profession, beginning with the Global Body of Investment Knowledge (GBIK®). Regional expert panels and targeted surveys are conducted annually to verify and reinforce the continuous feedback about the GBIK. The practice analysis process ultimately defines the CBOK. The CBOK reflects the competencies that are generally accepted and applied by investment professionals. These competencies are used in practice in a generalist context and are expected to be demonstrated by a recently qualified CFA charterholder.

© 2021 CFA Institute. All rights reserved.

The CFA Institute staff—in conjunction with the Education Advisory Committee and Curriculum Level Advisors, who consist of practicing CFA charterholders—designs the CFA Program curriculum in order to deliver the CBOK to candidates. The exams, also written by CFA charterholders, are designed to allow you to demonstrate your mastery of the CBOK as set forth in the CFA Program curriculum. As you structure your personal study program, you should emphasize mastery of the CBOK and the practical application of that knowledge. For more information on the practice analysis, CBOK, and development of the CFA Program curriculum, please visit www.cfainstitute.org.

ORGANIZATION OF THE CURRICULUM

The Level II CFA Program curriculum is organized into 10 topic areas. Each topic area begins with a brief statement of the material and the depth of knowledge expected. It is then divided into one or more study sessions. These study sessions should form the basic structure of your reading and preparation. Each study session includes a statement of its structure and objective and is further divided into assigned readings. An outline illustrating the organization of these study sessions can be found at the front of each volume of the curriculum.

The readings are commissioned by CFA Institute and written by content experts, including investment professionals and university professors. Each reading includes LOS and the core material to be studied, often a combination of text, exhibits, and in-text examples and questions. End of Reading Questions (EORQs) followed by solutions help you understand and master the material. The LOS indicate what you should be able to accomplish after studying the material. The LOS, the core material, and the EORQs are dependent on each other, with the core material and EORQs providing context for understanding the scope of the LOS and enabling you to apply a principle or concept in a variety of scenarios.

The entire readings, including the EORQs, are the basis for all exam questions and are selected or developed specifically to teach the knowledge, skills, and abilities reflected in the CBOK.

You should use the LOS to guide and focus your study because each exam question is based on one or more LOS and the core material and practice problems associated with the LOS. As a candidate, you are responsible for the entirety of the required material in a study session.

We encourage you to review the information about the LOS on our website (www.cfainstitute.org/programs/cfa/curriculum/study-sessions), including the descriptions of LOS "command words" on the candidate resources page at www.cfainstitute.org.

FEATURES OF THE CURRICULUM

End of Reading Questions/Solutions *All End of Reading Questions (EORQs) as well as their solutions are part of the curriculum and are required material for the exam.* In addition to the in-text examples and questions, these EORQs help demonstrate practical applications and reinforce your understanding of the concepts presented. Some of these EORQs are adapted from past CFA exams and/or may serve as a basis for exam questions.

Glossary For your convenience, each volume includes a comprehensive Glossary. Throughout the curriculum, a **bolded** word in a reading denotes a term defined in the Glossary.

Note that the digital curriculum that is included in your exam registration fee is searchable for key words, including Glossary terms.

LOS Self-Check We have inserted checkboxes next to each LOS that you can use to track your progress in mastering the concepts in each reading.

Source Material The CFA Institute curriculum cites textbooks, journal articles, and other publications that provide additional context or information about topics covered in the readings. As a candidate, you are not responsible for familiarity with the original source materials cited in the curriculum.

Note that some readings may contain a web address or URL. The referenced sites were live at the time the reading was written or updated but may have been deactivated since then.

Some readings in the curriculum cite articles published in the *Financial Analysts Journal*®, which is the flagship publication of CFA Institute. Since its launch in 1945, the *Financial Analysts Journal* has established itself as the leading practitioner-oriented journal in the investment management community. Over the years, it has advanced the knowledge and understanding of the practice of investment management through the publication of peer-reviewed practitioner-relevant research from leading academics and practitioners. It has also featured thought-provoking opinion pieces that advance the common level of discourse within the investment management profession. Some of the most influential research in the area of investment management has appeared in the pages of the *Financial Analysts Journal*, and several Nobel laureates have contributed articles.

Candidates are not responsible for familiarity with *Financial Analysts Journal* articles that are cited in the curriculum. But, as your time and studies allow, we strongly encourage you to begin supplementing your understanding of key investment management issues by reading this, and other, CFA Institute practice-oriented publications through the Research & Analysis webpage (www.cfainstitute.org/en/research).

Errata The curriculum development process is rigorous and includes multiple rounds of reviews by content experts. Despite our efforts to produce a curriculum that is free of errors, there are times when we must make corrections. Curriculum errata are periodically updated and posted by exam level and test date online (www.cfainstitute.org/en/programs/submit-errata). If you believe you have found an error in the curriculum, you can submit your concerns through our curriculum errata reporting process found at the bottom of the Curriculum Errata webpage.

DESIGNING YOUR PERSONAL STUDY PROGRAM

Create a Schedule An orderly, systematic approach to exam preparation is critical. You should dedicate a consistent block of time every week to reading and studying. Complete all assigned readings and the associated problems and solutions in each study session. Review the LOS both before and after you study each reading to ensure that

you have mastered the applicable content and can demonstrate the knowledge, skills, and abilities described by the LOS and the assigned reading. Use the LOS self-check to track your progress and highlight areas of weakness for later review.

Successful candidates report an average of more than 300 hours preparing for each exam. Your preparation time will vary based on your prior education and experience, and you will probably spend more time on some study sessions than on others.

You should allow ample time for both in-depth study of all topic areas and additional concentration on those topic areas for which you feel the least prepared.

CFA INSTITUTE LEARNING ECOSYSTEM (LES)

As you prepare for your exam, we will email you important exam updates, testing policies, and study tips. Be sure to read these carefully.

Your exam registration fee includes access to the CFA Program Learning Ecosystem (LES). This digital learning platform provides access, even offline, to all of the readings and End of Reading Questions found in the print curriculum organized as a series of shorter online lessons with associated EORQs. This tool is your one-stop location for all study materials, including practice questions and mock exams.

The LES provides the following supplemental study tools:

Structured and Adaptive Study Plans The LES offers two ways to plan your study through the curriculum. The first is a structured plan that allows you to move through the material in the way that you feel best suits your learning. The second is an adaptive study plan based on the results of an assessment test that uses actual practice questions.

Regardless of your chosen study path, the LES tracks your level of proficiency in each topic area and presents you with a dashboard of where you stand in terms of proficiency so that you can allocate your study time efficiently.

Flashcards and Game Center The LES offers all the Glossary terms as Flashcards and tracks correct and incorrect answers. Flashcards can be filtered both by curriculum topic area and by action taken—for example, answered correctly, unanswered, and so on. These Flashcards provide a flexible way to study Glossary item definitions.

The Game Center provides several engaging ways to interact with the Flashcards in a game context. Each game tests your knowledge of the Glossary terms a in different way. Your results are scored and presented, along with a summary of candidates with high scores on the game, on your Dashboard.

Discussion Board The Discussion Board within the LES provides a way for you to interact with other candidates as you pursue your study plan. Discussions can happen at the level of individual lessons to raise questions about material in those lessons that you or other candidates can clarify or comment on. Discussions can also be posted at the level of topics or in the initial Welcome section to connect with other candidates in your area.

Practice Question Bank The LES offers access to a question bank of hundreds of practice questions that are in addition to the End of Reading Questions. These practice questions, only available on the LES, are intended to help you assess your mastery of individual topic areas as you progress through your studies. After each practice question, you will receive immediate feedback noting the correct response and indicating the relevant assigned reading so you can identify areas of weakness for further study.

Mock Exams The LES also includes access to three-hour Mock Exams that simulate the morning and afternoon sessions of the actual CFA exam. These Mock Exams are intended to be taken after you complete your study of the full curriculum and take practice questions so you can test your understanding of the curriculum and your readiness for the exam. If you take these Mock Exams within the LES, you will receive feedback afterward that notes the correct responses and indicates the relevant assigned readings so you can assess areas of weakness for further study. We recommend that you take Mock Exams during the final stages of your preparation for the actual CFA exam. For more information on the Mock Exams, please visit www.cfainstitute.org.

PREP PROVIDERS

You may choose to seek study support outside CFA Institute in the form of exam prep providers. After your CFA Program enrollment, you may receive numerous solicitations for exam prep courses and review materials. When considering a prep course, make sure the provider is committed to following the CFA Institute guidelines and high standards in its offerings.

Remember, however, that there are no shortcuts to success on the CFA exams; reading and studying the CFA Program curriculum *is* the key to success on the exam. The CFA Program exams reference only the CFA Institute assigned curriculum; no prep course or review course materials are consulted or referenced.

SUMMARY

Every question on the CFA exam is based on the content contained in the required readings and on one or more LOS. Frequently, an exam question is based on a specific example highlighted within a reading or on a specific practice problem and its solution. To make effective use of the CFA Program curriculum, please remember these key points:

1 All pages of the curriculum are required reading for the exam.

2 All questions, problems, and their solutions are part of the curriculum and are required study material for the exam. These questions are found at the end of the readings in the print versions of the curriculum. In the LES, these questions appear directly after the lesson with which they are associated. The LES provides immediate feedback on your answers and tracks your performance on these questions throughout your study.

3 We strongly encourage you to use the CFA Program Learning Ecosystem. In addition to providing access to all the curriculum material, including EORQs, in the form of shorter, focused lessons, the LES offers structured and adaptive study planning, a Discussion Board to communicate with other candidates, Flashcards, a Game Center for study activities, a test bank of practice questions, and online Mock Exams. Other supplemental study tools, such as eBook and PDF versions of the print curriculum, and additional candidate resources are available at www.cfainstitute.org.

4 Using the study planner, create a schedule and commit sufficient study time to cover the study sessions. You should also plan to review the materials, answer practice questions, and take Mock Exams.

5 Some of the concepts in the study sessions may be superseded by updated rulings and/or pronouncements issued after a reading was published. Candidates are expected to be familiar with the overall analytical framework contained in the assigned readings. Candidates are not responsible for changes that occur after the material was written.

FEEDBACK

At CFA Institute, we are committed to delivering a comprehensive and rigorous curriculum for the development of competent, ethically grounded investment professionals. We rely on candidate and investment professional comments and feedback as we work to improve the curriculum, supplemental study tools, and candidate resources.

Please send any comments or feedback to info@cfainstitute.org. You can be assured that we will review your suggestions carefully. Ongoing improvements in the curriculum will help you prepare for success on the upcoming exams and for a lifetime of learning as a serious investment professional.

Derivatives

STUDY SESSIONS

Study Session 13	Derivatives

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to estimate the value of futures, forwards, options, and swaps and demonstrate how they may be used in various strategies.

Derivatives are used extensively to manage financial risk. Institutions and individuals use derivatives to transfer, modify, or eliminate unwanted interest rate, currency, cash flow, or market exposures. Besides their value in risk management, derivatives can also be effective tools for generating income, enhancing returns, and creating synthetic exposure. Efficiencies in cost, liquidity, ability to short, and limited capital outlay may make derivatives attractive alternatives to their underlying.

© 2021 CFA Institute . All rights reserved.

DERIVATIVES

STUDY SESSION 13

Derivatives

This study session introduces key valuation concepts and models for forward commitments (forwards, futures, swaps) and contingent claims (options). Option coverage includes the "Greeks," which measure the effects on value of small changes in underlying asset value, time, volatility, and the risk-free rate.

READING ASSIGNMENTS

© 2021 CFA Institute. All rights reserved.

READING

33

Pricing and Valuation of Forward Commitments

by Adam Schwartz, PhD, CFA

Adam Schwartz, PhD, CFA is at Bucknell University (USA).

CFA Institute would like to thank Robert Brooks, PhD, CFA and Barbara Valbuzzi, CFA for their contributions to earlier versions of this reading.

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe the carry arbitrage model without underlying cashflows and with underlying cashflows;
☐	**b.** describe how equity forwards and futures are priced, and calculate and interpret their no-arbitrage value;
☐	**c.** describe how interest rate forwards and futures are priced, and calculate and interpret their no-arbitrage value;
☐	**d.** describe how fixed-income forwards and futures are priced, and calculate and interpret their no-arbitrage value;
☐	**e.** describe how interest rate swaps are priced, and calculate and interpret their no-arbitrage value;
☐	**f.** describe how currency swaps are priced, and calculate and interpret their no-arbitrage value;
☐	**g.** describe how equity swaps are priced, and calculate and interpret their no-arbitrage value.

INTRODUCTION TO PRICING AND VALUATION OF FORWARD COMMITMENTS

1

b describe how equity forwards and futures are priced, and calculate and interpret their no-arbitrage value

© 2020 CFA Institute. All rights reserved.

Forward commitments include forwards, futures, and swaps. A forward contract is a promise to buy or sell an asset at a future date at a price agreed to at the contract's initiation. The forward contract has a linear payoff function, with both upside and downside risk.

A swap is essentially a promise to undertake a transaction at a set price or rate at several dates in the future. The technique we use to price and value swaps is to identify and construct a portfolio with cash flows equivalent to those of the swap. Then, we can use tools, such as the law of one price, to determine swap values from simpler financial instruments, such as a pair of bonds with a cash flow pattern similar to those of our swap.

Look out for the big picture: value additivity, arbitrage, and the law of one price are important valuation concepts.

Forwards and swaps are widely used in practice to manage a broad range of market risks. As well, more complex derivative instruments can sometimes be understood in terms of their basic building blocks: forwards and option-based components. Here are just some of the many and varied uses for forwards, futures, and swaps that you might encounter in your investment career:

- Use of equity index futures and swaps by a private wealth manager to hedge equity risk in a low tax basis, concentrated position in his high-net-worth client's portfolio.
- Use of interest rate swaps by a defined benefits plan manager to hedge interest rate risk and to manage the pension plan's duration gap.
- Use of derivatives (total return swaps, equity futures, bond futures, etc.) overlays by a university endowment for tactical asset allocation and portfolio rebalancing.
- Use of interest rate swaps by a corporate borrower to synthetically convert floating-rate debt securities to fixed-rate debt securities (or vice versa).
- Use of VIX futures and inflation swaps by a firm's market strategist to infer expectations about market volatility and inflation rates, respectively.

1.1 Principles of Arbitrage-Free Pricing and Valuation of Forward Commitments

a Describe the carry arbitrage model without underlying cashflows and with underlying cashflows

In this section, we examine arbitrage-free pricing and valuation of forward commitments—also known as the no-arbitrage approach to pricing and valuing such instruments. We introduce some guiding principles that heavily influence the activities of arbitrageurs, who are price setters in forward commitment markets.

There is a distinction between the pricing and the valuation of forward commitments. Forward commitment pricing involves determining the appropriate forward commitment price or rate when initiating the forward commitment contract. Forward commitment valuation involves determining the appropriate value of the forward commitment, typically after it has been initiated.

Our approach to pricing and valuation is based on the assumption that prices adjust to prevent arbitrage profits. Hence, the material will be covered from an arbitrageur's perspective. Key to understanding this material is to think like an arbitrageur. Specifically, the arbitrageur seeks to make a profit following two rules:

Rule #1: Do not use your own money; and

Rule #2: Do not take any price risk.

To make a profit, subject to these restrictions, the arbitrageur may need to borrow or lend money and buy or sell assets. The no-arbitrage approach considers the contract's cash flows from contract initiation (Time 0) to contract maturity (Time T). If an initial investment requires an outflow of 100 euros, then we will present it as a –100 euro cash flow. Cash inflows to the arbitrageur have a positive sign, and outflows are negative.

Pricing and valuation tasks based on the no-arbitrage approach imply an inability to create a portfolio that earns a risk-free profit without making a positive net investment of capital. In other words, if cash and forward markets are priced correctly with respect to each other, we cannot create such a portfolio. That is, we cannot create money today with no risk or future liability. This approach is built on the **law of one price**, which states that if two investments have equivalent future cash flows regardless of what will happen in the future, then these two investments should have the same current price. Alternatively, if the law of one price is violated, someone could buy the cheaper asset and sell the more expensive asset, resulting in a gain at no risk and with no commitment of capital. The law of one price can be used with the value additivity principle, which states that the value of a portfolio is simply the sum of the values of each instrument held in the portfolio.

Throughout this discussion of forward commitments, the following key assumptions are made: (1) replicating instruments are identifiable and investable; (2) market frictions are nil; (3) short selling is allowed with full use of proceeds; and (4) borrowing and lending are available at a known risk-free rate.

Our analyses will rely on the **carry arbitrage model**, a no-arbitrage approach in which the underlying instrument is either bought or sold along with establishing a forward position—hence the term "carry." Carry arbitrage models are also known as cost-of-carry arbitrage models or cash-and-carry arbitrage models. Carry arbitrage models account for costs to carry/hold the underlying asset. Carry costs include financing costs plus storage and insurance costs (for physical underlying, like gold). The carry arbitrage model must also adjust for any carry benefits (i.e., negative carry costs), including dividends and interest (such as bond coupons) received. Typically, each type of forward commitment will result in a different model, but common elements will be observed. Carry arbitrage models are a great first approximation to explaining observed forward commitment prices in many markets.

The central theme here is that forward commitments are generally priced so as to preclude arbitrage profits. Section 3 demonstrates how to price and value equity, interest rate, fixed-income, and currency forward contracts. We also explain how these results apply to futures contracts.

1.2 Pricing and Valuing Generic Forward and Futures Contracts

In this section, we examine the pricing of forward and futures contracts based on the no-arbitrage approach. The resulting carry arbitrage models are based on the replication of the forward contract payoff with a position in the underlying that is financed through an external source. Although the margin requirements, mark-to-market features, and centralized clearing in futures markets result in material differences between forward and futures markets in some cases, we focus mainly on cases in which the particular carry arbitrage model can be used in both markets.

1.2.1 *Forwards and Futures*

Forward and futures contracts are similar in that they are both agreements in which one party is legally obligated to sell and the other party is legally obligated to buy an asset (financial or otherwise) at an agreed price at some specific date in the future. The main difference is that a futures contract is an exchange-traded financial instrument. Contracts trading on an organized exchange, such as the Chicago Mercantile

Exchange (CME), incorporate standard features to facilitate trading and ensure both parties fulfill their obligations. For example, a gold futures contract traded on the CME (COMEX) features a standard contract size of 100 ounces, agreed upon deliverable assets (gold bars, perhaps), and a limited choice of maturity dates. To ensure performance of the long and the short parties, the futures exchange requires the posting and daily maintenance of a margin account.

A forward contract is an agreement to buy or sell a specific asset (financial or otherwise) at an agreed price at some specific date in the future. Forward contracts are bilateral non-exchange traded contracts, offering flexibility in terms of size, type of the underlying asset, expiration date, and settlement date. This customization comes at a price of potential credit risk and ability to unwind the position. Since the financial crisis, best practices for OTC contracts suggest daily settlement and margin requirements for forward contracts similar to those required by futures exchanges. Without daily settlement, a forward contract may accumulate (or may lose) value over time. Some of the differences and similarities between forwards and futures are summarized in Exhibit 1.

Exhibit 1 Characteristics of Futures and Forward Contracts

Futures	Forwards
Exchange-traded	Negotiated between the contract counterparties
Standardized dates and deliverables	Customized dates and deliverables
Trades guaranteed by a clearinghouse	Trading subject to counterparty risk
Initial value = 0	Initial value = 0 (Typically, but not required)
An initial margin deposit specified by the exchange is required. The margin account is adjusted for gains and losses daily. If daily losses cause the margin balance to drop below a limit set by the futures exchange (i.e., maintenance margin), additional funds must be deposited, or the position will be closed.	Margin requirements may be specified by the counterparties.
Daily settlement marks the contract price equal to the market price and contract value = 0.	Contract may outline a settlement schedule. The forward may accumulate (or lose) value between settlement periods or until maturity (if no early settlements are required).

Forward price (F) or **futures price** (f) refers to the price that is negotiated between the parties to the forward or futures contract, respectively.

Our notation will be as follows, let:

S_t represent spot price (cash price for immediate delivery) of the underlying instrument at any time t,

F_t represent forward price at any time t, and

f_t represent futures price at any time t.

Therefore, S_0, F_0, and f_0 denote, respectively, the spot, forward, and futures price, respectively, established at the initiation date, 0. The initial forward price is established to make the contract value zero for both the long and short parties. The forward (delivery) price does not change during the life of the contract. Time T represents the time at which the contract expires and the future transaction is scheduled to take place. Thus, S_T, F_T, and f_T are the spot, forward, and futures price, respectively, at expiration time T. Between initiation at time 0 and expiration at time T, the spot price of the underlying asset may fluctuate to a new value, S_t. The price of a newly created forward or futures contract at time t with the same underlying and expiration (at time T) may differ from the price agreed to at time 0. So, our forward or futures contract established at time 0 may have a positive or negative value at time t. V_t and v_t will later be used to describe, respectively, the value of a forward and futures contract at any time t.

As we approach expiration, the price of a newly created forward or futures contract will approach the price of a spot transaction. At expiration, a forward or futures contract is equivalent to a spot transaction in the underlying. This property is often called **convergence**, and it implies that at time T, both the forward price and the futures price are equivalent to the spot price—that is,

Convergence property: $F_T = f_T = S_T$.

The convergence property is intuitive. For example, the one-year forward price of gold (that is, the price set today to purchase gold one year from now) might be very different from the spot price of gold. However, the price to buy gold one hour in the future should be very close to the spot price. As the maturity of the forward or futures contract approaches, the forward or futures price will converge to the spot price. If the forward or futures price were higher than the spot at maturity, an arbitrageur would:

1. Sell the forward or futures contract.
2. Borrow funds using a loan to buy the asset.
3. Make delivery at expiration of the contract, repay the loan, and keep the profit.

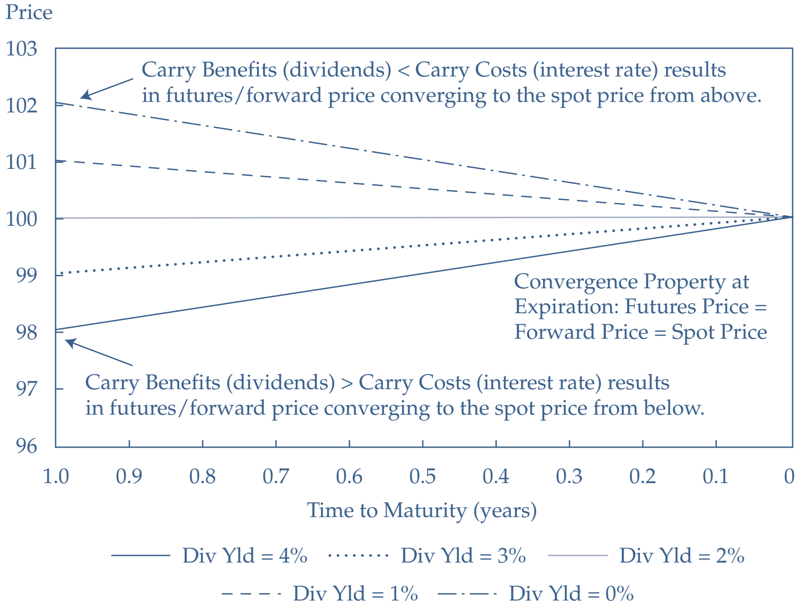
As market participants exploit this arbitrage opportunity, the forward or futures price will fall due to selling pressure.

If the futures price is below spot price, an arbitrageur would short sell the asset, invest the short-sale proceeds at the risk-free rate, and then enter into a long futures contract. He or she would take delivery of the asset at the futures contract expiration and use it to cover the short. The profit is simply the difference between the short-sale price and the futures price, after adjusting for carrying and financing costs. These actions on the part of arbitrageurs would act to enforce the convergence property.

Prior to expiration, the price of a newly created futures or forward contract will usually differ from the spot price. The forward and futures prices may even differ slightly from each other. For example, when the possibility of counterparty default exists or when the underlying asset price (such as a bond) is correlated with interest rates (which might impact the financing costs for daily settlement), the futures price might vary slightly from the forward price. For most cases, the generalist may assume the price of a futures contract and a forward contract will be same. That is $F_t = f_t$ before expiration.

Exhibit 2 shows the convergence property for a stock index futures/forward contact under continuous compounding and varying dividend yields. To carry a stock index, we must forego the interest rate that could be otherwise earned on our money, but we will collect dividend payments. As shown in Exhibit 2, the convergence path to the spot price at maturity depends on the costs and benefits of carrying the underlying asset. Here the stock index pays a dividend yield, which is a carry benefit. To hold the stock index, we must forego interest that could otherwise be earned on the investment. This financing rate (interest rate, r_c), assumed to be 2% in the following graph, is a cost to carry the index.

Exhibit 2 Convergence Property: Convergence of Forward Price to Spot Price (r_c = 2% and Index Level = 100)



As maturity of the contract approaches (at time = T), the price of a newly created forward or futures contract will approach the spot price so that at expiration $F_T = f_T = S_T$, according to the convergence property. Prior to expiration, the forward/futures prices may be above, below, or nearly equal to the current spot price S_t. For futures contracts, the difference between the spot price and the futures price is the **basis**. As the maturity date nears, the basis converges toward zero. According to the convergence property, the future price approaches the spot price as we move toward expiration. At expiration, the futures price is equal to the price today for delivery today (i.e., spot price). If the convergence property does not hold, arbitrage will force the prices to be equal at contract expiration. The nature of the pricing relationship between the spot and forward/futures prices shown here will be explained shortly using the carry arbitrage model. For example, carry arbitrage will help us understand why assets with carry benefits (dividends) greater than carry costs (costs to finance and store the underlying) will have forward prices that converge to the spot price from below.

As market prices change, the value of existing futures and forward positions will change also. The market value of the forward or futures contract, termed **forward value** or **futures value**, respectively, and sometimes just value, refers to the monetary value of an existing forward or futures contract. When the forward or futures contract is established, the price is negotiated so that the value of the contract on the initiation date is zero. Subsequent to the initiation date, the value can be significantly positive or negative.

For example, an industrial firm requires platinum to manufacture certain components used in automobile manufacturing. The firm enters a long forward contract on 10 March. Under the terms of the contract, the firm agrees to buy 4,500 ounces of platinum on 10 September for $900 per ounce from a metal producer. From the firm's point of view, this is effectively a six-month long forward contract at a price of F_0 = $900. If the price (technically, the September forward price) of platinum increases to $1,100 in May, the firm will be happy to have locked in a purchase price of $900 (long forward contract value is positive). If the price of platinum decreases to $800, the firm must still honor the forward agreement to buy platinum at $900 (long forward contract value is negative). To describe the value of a forward contract, let V_t be the value of the forward contract at any time t.

When the forward contract is established, the forward price is negotiated so that the market value of the forward contract on the initiation date is zero. Most forward contracts are structured this way and are referred to as **at market contracts**. Again, we assume no margin requirements. No money changes hands, meaning that the initial value is zero, so, $V_0 = 0$.

At expiration, the value of a forward contract V_T is realized and, as shown next, is straightforward to compute. Remember, the profit on any completed transaction is the sale price minus the purchase price. The profit or value of the forward contract at expiration is also the sale price minus the purchase price. At initiation, a forward or futures contract allows for either a future purchase price or a future sale price, F_0, to be known at time 0. In a long forward, a buyer can lock in a purchase price, $F_{0.}$ In a short forward, a seller can lock in a sale price, $F_{0.}$ Again, a forward contract allows a buyer or a seller to fix an initial price F_0, either the purchase price (long forward) or the sale price (short forward). The party long the forward effectively agrees to buy an asset in the future (at time T) at a price set today (at time 0), F_0.

At expiration, the asset can be sold in the spot market at a price $S_{T.}$ Therefore, a *long* position in a forward contract has a value at expiration of:

$$V_T = S_T - F_0.$$

A short position effectively locks in a sale price of $F_{0.}$ It is the negative of the long position. Therefore, the value of a short forward position at expiration is the sale price minus the purchase price of the asset:

$$-V_T = -(S_T - F_0) = F_0 - S_T.$$

For example, in January a fund manager agrees to sell a bond portfolio in May for F_0 = £10,000,000. The fund manager locks in the sale price, F_0. If the spot price of the bond portfolio at expiration (S_T) is £9,800,000, then the short forward contract will have an expiration value to the fund manager of:

$$-V_T = £10,000,000 - £9,800,000 = £200,000.$$

The fund manager makes a profit by selling at a higher price than the market price at expiration.

Value may accumulate or diminish with the passage of time in forward contracts, which is why forward contracts require the posting of collateral. Futures contract values, on the other hand, are settled by margining at the end of each trading day when the contract is marked-to-market. The gains and losses in the position over time accumulate in the futures traders' margin accounts. Prior to daily settlement, during the trading hours the market value of a long position in a futures contract is the current futures price less the future price at the last time the contract was marked-to-market times the multiplier, N_f (the multiplier is the standard contract size set by the futures exchange).

For a long futures contract, the value accumulated during the trading day (v_t) is:

$$v_t = \text{Multiplier} \times (\text{Current futures price} - \text{Previous settlement price}) \text{ or}$$

$$v_t = N_f \times (f_t - f_{t-1}).$$

Assume an investor is long one contract (N_f = 100 ounces/contract) of June gold, which settled at $1,300/ounce on the previous trading day. So, the investor is effectively agreeing to purchase 100 ounces of gold in June for $1,300 per ounce or $130,000 total. The trader need not pay the entire $130,000 today but must post a deposit in a margin account to guarantee his/her performance. During the current trading day, the price of June gold increases to $1,310. Before marking-to-market, the value of the long contract is 100 × (\$1,310 – \$1,300) = +\$1,000. After marking-to-market, the gain or loss is reflected in the trader's margin account and the new contract price is set equal to the settlement price. The futures contract value after daily settlement is 0 or $v_t = 0$.

2 CARRY ARBITRAGE

a describe the carry arbitrage model without underlying cashflows and with underlying cashflows

We first consider a generic forward contract, meaning that we do not specify the underlying as anything more than just an asset. As we move through this section, we will continue to address specific additional factors to bring each carry arbitrage model closer to real markets. Thus, we will develop several different carry arbitrage models, each one applicable to a specific forward commitment contract. We start with the simpler of the two base cases, carry of an asset without cash flows to the underlying, then move to the more complex case of forwards on assets with underlying cash flows, such as bonds with coupon payments or stocks that pay dividends.

2.1 Carry Arbitrage Model When There Are No Underlying Cash Flows

Carry arbitrage models receive their name from the literal interpretation of carrying the underlying asset over the life of the forward contract. If an arbitrageur enters a forward contract to sell an underlying instrument for delivery at time T, then to offset this exposure, one strategy is to buy the underlying instrument at time 0 with borrowed funds and carry it to the forward expiration date (time T). The asset can then be sold (or even delivered) under the terms of a forward contract. The risks of this scenario are illustrated in Exhibit 3.

Exhibit 3 Cash Flows from Carrying an Underlying Asset and Offsetting Short Forward Position

	Time 0	Time T
Borrowing Funds to Purchase and Carry an Underlying Asset		
Underlying	$-S_0$ (purchase)	$+S_T$ (sale)
Borrowed funds	$+S_0$ (inflow)	$-FV(S_0)$ (repayment)
Net Cash Flow	$+S_0 - S_0 = 0$	$+ST - FV(S_0)$
Short Forward Position		
Short Forward	$V_0 = 0$	$V_T = F_0 - S_T$
Overall Position: Long Asset + Borrowed Funds + Short Forward		
	$+S_0 - S_0 + V_0 = 0$	$+S_T - FV(S_0) + V_T = 0$ $+S_T - FV(S_0) + (F_0 - S_T) = 0$ $+F_0 - FV(S_0) = 0$
Net	**0**	$\mathbf{F_0 = FV(S_0)}$

The underlying asset is bought for S_0 with borrowed funds. The asset can be sold at time T for a price, S_T. At time T, the borrowed funds must be repaid at a cost of $FV(S_0)$; note that FV stands for the future value function. Clearly, when S_T is below (above) $FV(S_0)$, our underlying transaction will suffer a loss (earn a profit). A short forward position can be added to our long position in the underlying asset to offset any profit or loss in the underlying. Both positions have no initial (time 0) cash flow. To prevent arbitrage, the overall portfolio (Asset + Borrowed funds + Short forward) should have a value of zero at time T. If the cost to finance the purchase of the asset, $FV(S_0)$, is equal to the initially agreed upon forward price, F_0, then there is no arbitrage profit. So, we should have $F_0 = FV(S_0)$.

For now, we will keep the significant technical issues to a minimum. When possible, we will just use FV and PV to denote future value and present value, respectively. At this point we are not yet concerned about compounding conventions, day count conventions, or even the appropriate risk-free interest rate proxy.

Carry arbitrage models rest on the no-arbitrage assumptions. Therefore, the arbitrageur does not use his or her own money to acquire positions but borrows to purchase the underlying. Borrowing (if the underlying asset is purchased) and lending the proceeds (if the underlying asset is sold) are done at the risk-free interest rate. Furthermore, the arbitrageur offsets all transactions, meaning he/she does not take any price risk. We do not consider other risks, such as liquidity risk and counterparty credit risk, as they would unnecessarily complicate our basic presentation.

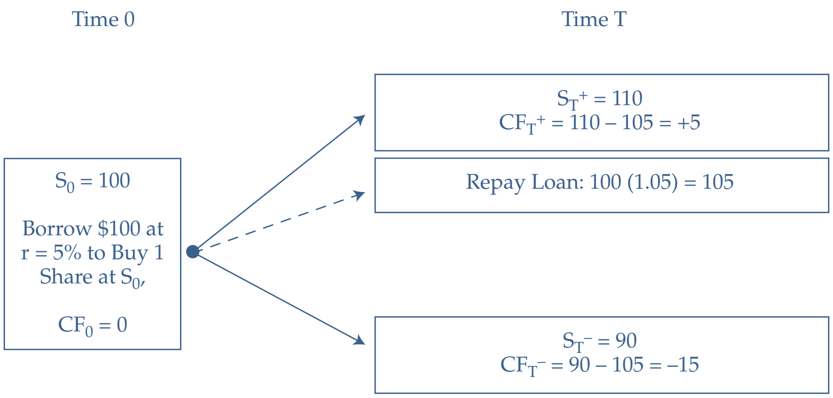
If we assume continuous compounding (r_c), then $FV(S_0) = S_0 exp^{r_c T}$. If we assume annual compounding (r), then $FV(S_0) = S_0(1 + r)^T$. Note that in practice, observed interest rates are derived from market prices. For example, a T-bill price implies the T-bill rate. Significant errors can occur if the quoted interest rate is used with the wrong compounding convention. When possible, we just use basic present value and future value representations to minimize confusion.

To help clarify, we first illustrate the price exposure solely from holding the underlying asset. Exhibit 4 shows the cash flows from carrying the underlying, a non-dividend-paying stock, assuming $S_0 = 100$, r = 5%, and T = 1. For illustration purposes, we allow the stock price at expiration to go down to $S_T^- = 90$ or up to $S_T^+ = 110$. The initial transactions will generate cash flows shown at times 0 and T. In practice, the set of transactions (market purchases, bank transactions) are executed simultaneously at each time period, not sequentially. Here are the two transactions at time 0 that produce a levered equity purchase.

Step 1 Purchase one unit of the underlying at time 0 (an outflow).

Step 2 Borrow the purchase price at the risk-free rate (an inflow).

Exhibit 4 Cash Flows for Long Financial Position

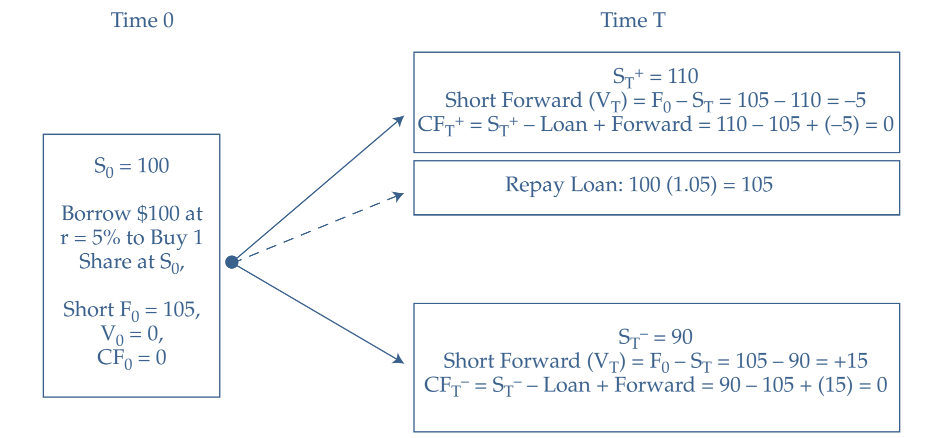


At time T (= 1), the stock price can jump up to $S_T^+ = 110$ or jump down to $S_T^- = 90$. Because the two outcomes are different, the strategy at this point has price risk. After the loan is repaid, the net cash flow will be +5 if the stock jumps up to 110 or −15 if the stock price jumps down to 90. To eliminate price risk, we must add another step to our list of simultaneous transactions. As suggested by Exhibit 3, we sell (go short) a forward contract to set a price today for the future sale of our underlying, and that price ($F_0 = FV(S_0)$) is 105.

Step 3 Sell a forward at $F_0 = 105$. For a short forward contract, F_0 is the price agreed to at time 0 to sell the asset at Time T.

The resulting portfolio with its offsetting transaction is illustrated in Exhibit 5.

Exhibit 5 Cash Flow for Long Financial Position with Short Forward Contract



Regardless of the value of the underlying at maturity, we owe 105 on the loan. Notice that at expiration the underlying is worth 90 or 110. Since we agreed to sell the asset at 105, the forward contract value is either 15 or −5, respectively. If the asset is selling for 90 at time T, the forward contract allows us to sell our underlying position for 15 more (105 – 90) than in the spot market. The combination of the proceeds from the sale of the underlying and the value of the short forward at maturity is always 105 (= 90 + 15 or 110 – 5), which is precisely the amount necessary to pay off the loan. So, there is zero net cash flow at expiration under any and all circumstances. Since this transaction has no risk (no uncertainty about value at time T), we require that the no-arbitrage forward price (F_0) is simply the future value of the underlying growing at the risk-free rate, or

$$F_0 = \text{Future value of underlying} = FV(S_0). \quad (1)$$

In our example, $F_0 = FV(S_0) = 105$. In fact, with annual compounding and T = 1, we have simply $F_0 = S_0(1 + r)^T = 100(1 + 0.05)^1$. The future value refers to the amount of money equal to the spot price invested at the compounded risk-free interest rate during the time period. It is not to be confused with or mistaken for the mathematical expectation of the spot price at time T.

Without market frictions, arbitrage may be possible when mispricing occurs. To better understand the arbitrage mechanics, suppose that $F_0 = 106$. Based on the prior information, we observe that the forward price is higher than the price suggested by the carry arbitrage model—recall $F_0 = FV(S_0) = 105$. Because the carry arbitrage model value is lower than the market's forward price, we conclude that the market's forward price is too high and should be sold. An arbitrage opportunity exists, and it will involve selling the forward contract at 106 (Step 1). Step 2 occurs when a second transaction is needed to borrow funds to undertake Step 3, purchase of the underlying instrument so that gains (or losses) in the underlying will offset losses (or gains) on the forward contract. Note, the second step ensures the arbitrageur does not use his or her own money. The third transaction, the purchase of the underlying security, guarantees the arbitrageur does not take any market price risk. Note that all three transactions are done simultaneously. To summarize, the arbitrage transactions for $F_0 > FV(S_0)$ can be represented in the following three steps:

Step 1 Sell the forward contract on the underlying.

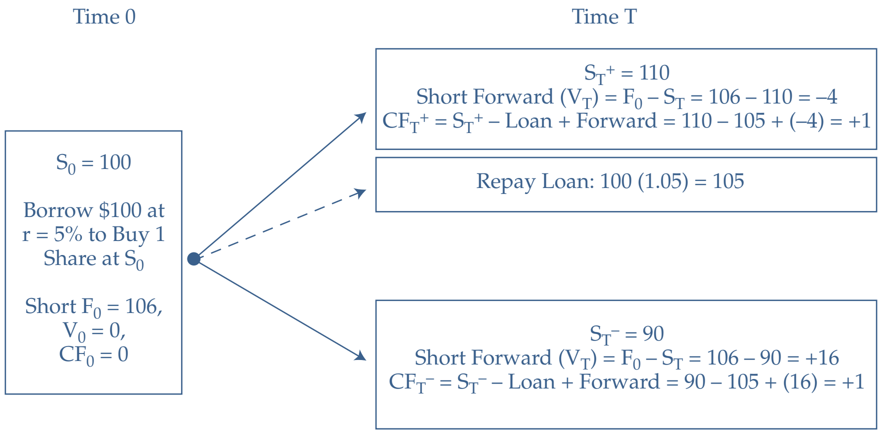
Step 2 Borrow the funds to purchase the underlying.

Step 3 Purchase the underlying.

Exhibit 6 shows the resulting cash flows from these transactions. This strategy is known as carry arbitrage because we are carrying—that is, we are long—the underlying instrument. At time T, we earn an arbitrage profit of +1. We do not use any of our own money and make a profit no matter the price of the underlying at maturity (i.e., 110, 90, or anything else). Since the profit of +1 at maturity occurs under every circumstance, it is considered risk-free. Any situation that allows a risk-free profit with no upfront cost will not be available for very long. It represents a clear arbitrage opportunity, one that will be pursued until forward prices fall and eliminate the arbitrage opportunity.

Note that if the forward price, $F_{0,}$ were 106, the value of the forward contract at time 0 would be the PV of the +1 cash flow at Time T. Thus, at time 0, the value of our short forward is $V_0 = PV[F_0 - FV(S_0)] = (106 - 105)/(1 + 0.05)^1 = 0.9524$.

Exhibit 6 Cash Flow with Forward Price Greater Than Carry Arbitrage Model Price



Suppose instead we observe a lower forward price, $F_0 = 104$. Based on the prior information, we conclude that the forward price is too low when compared to the forward price determined by the carry arbitrage model of $F_0 = FV(S_0) = 105$. Since the forward price is too low, Step 1 is to buy the forward contract, and the value at T is $S_T - F_0$. The arbitrageur does not want any price risk, so Step 2 is to sell short the underlying instrument. To accomplish Step 2, we must borrow the asset and sell it. Note that when an arbitrageur needs to sell the underlying, it must be assumed that he/she does not hold it in inventory and thus must sell it short. If the underlying were held in inventory, the investment in it would not be accounted for in the analysis. When the transaction calls for selling a derivative instrument, such as a forward contract, it is always just selling—technically, not short selling.

The long forward contract will allow us to cover our short later. The arbitrageur will then lend the short sale proceeds of 100 at the risk-free rate (Step 3). The deposit of 100 will grow to 105 at time T. Clearly, we will have a profit of +1 when we buy the asset at 104 and deliver it to clear the short. Again, to summarize, the arbitrage transactions when the forward price is too low—that is, $F_0 < FV(S_0)$—involve the following three steps:

Step 1 Buy the forward contract on the underlying.

Step 2 Sell the underlying short.

Step 3 Lend the short sale proceeds.

We must replace the asset at a price of S_T, but we have +105 from the loan and a long forward at 104. Remember, the value of a long forward at time T is $V_T = S_T - F_0$. So, using the prior information, the value of the forward at expiration will be 90 – 104 = –14 (if $S_T^- = 90$) or 110 – 104 = +6 (if $S_T^+ = 110$). Thus, the cash flows at maturity will be $CF^- = +105 - 14 - 90 = +1$ or $CF^+ = +105 + 6 - 110 = +1$. Again, we make a profit equal to the mispricing of +1 regardless of the stock value at time T. It is an arbitrage profit, since it was done with no money invested and with no risk.

Note that this set of transactions is the exact opposite of the prior case in Exhibit 6. This strategy is known as **reverse carry arbitrage** because we are doing the opposite of carrying the underlying instrument; that is, we are selling short the underlying instrument.

Therefore, unless $F_0 = FV(S_0)$, there is an arbitrage opportunity. Notice that if $F_0 > FV(S_0)$, then the forward contract is sold and the underlying is purchased. Thus, arbitrageurs drive down the forward price and drive up the underlying price until $F_0 = FV(S_0)$ and a risk-free positive cash flow today (i.e., in PV terms) no longer exists. Further, if $F_0 < FV(S_0)$, then the forward contract is purchased and the underlying is sold short. In this case, the forward price is driven up and the underlying price is driven down. Absent market frictions, arbitrageurs' market activities will drive forward prices to equal the future value of the underlying, bringing the law of one price into effect once again. Most importantly, if the forward contract is priced at its equilibrium price, there will be no arbitrage profit.

EXAMPLE 1

Forward Contract Price

An Australian stock paying no dividends is trading in Australian dollars for A$63.31, and the annual Australian interest rate is 2.75% with annual compounding.

1 Based on the current stock price and the no-arbitrage approach, which of the following values is *closest* to the equilibrium three-month forward price?

A A$63.31

B A$63.74

C A$65.05

2 If the interest rate immediately falls 50 bps to 2.25%, the three-month forward price will:

A decrease.

B increase.

C be unchanged.

Solution to 1:

B is correct. Based on the information given, S_0 = A\$63.31, r = 2.75% (annual compounding), and T = 0.25. Therefore,

$$F_0 = FV(S_0) = 63.31(1 + 0.0275)^{0.25} = A\$63.7408.$$

Solution to 2:

A is correct, because the forward price is directly related to the interest rate. Specifically,

$$F_0 = FV(S_0) = 63.31(1 + 0.0225)^{0.25} = A\$63.6632.$$

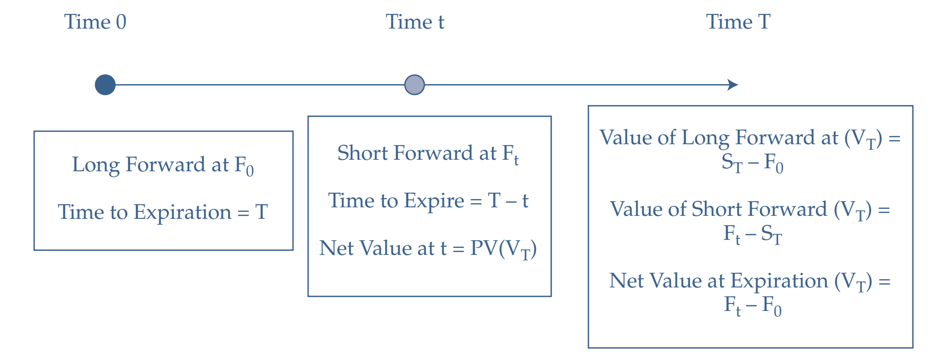
Therefore, we see in this case that a decrease in interest rates resulted in a decrease in the forward price. This relationship between forward prices and interest rates will generally hold so long as the underlying is not also influenced by interest rates.

As we see in Example 1, the quoted forward price does not directly reflect expectations of future underlying prices. The only factors that matter are the current price (S_0), the interest rate and time to expiration, and, of course, the absence of arbitrage. Other factors will be included later as we make the carry arbitrage model more realistic, but we will not be including expectations of future underlying prices. So, if we can carry the asset, an opinion that the underlying will increase in value, perhaps even substantially, has no bearing on the forward price.

We now turn to the task of understanding the value of an existing forward contract. There are many circumstances in which, once a forward contract has been entered, one wants to know the contract's fair value. The goal is to calculate the position's value at current market prices. The need may arise from market-based accounting, for example, in which the accounting statements need to reflect the current fair value of various instruments. Finally, it is simply important to know whether a position in a forward contract is making money or losing money (that is, the profit or loss from exiting the contract early).

The forward value prior to maturity is based on arbitrage. A timeline to help illustrate forward valuation is shown in Exhibit 7. Suppose the first transaction involves buying a forward contract with a price of F_0 at Time 0 with expiration at Time T. Now consider selling a new forward contract with price F_t at Time t, again with expiration at Time T. Exhibit 7 shows the potential cash flows. Remember the equivalence at expiration between the forward price, the futures price, and the underlying price will hold: $F_T = f_T = S_T$. Note that the middle of the timeline, "Time t" is the valuation date of the forward contract. Note also that we are seeking the forward value; therefore, this set of transactions would result in cash flows only if it is executed. We need not actually execute the transactions; we just need to see what they would produce if we did. This point is analogous to the fact that if we are holding a liquid asset, we need not sell it to determine its value; we can simply observe its market price, which gives us an estimate of the price at which we could sell the asset.

Exhibit 7 Long Forward Interim Value Timeline



Importantly, there are now three different points in time to consider: Time 0, Time t, and Time T. Note that once the offsetting forward is entered at time t, the net position is not subject to market risk. That is, the S_T terms cancel (in Exhibit 7), so the cash flow at Time T is not influenced by what happens to the spot price. The position is completely hedged. Therefore, the value observed at Time t of the original forward contract initiated at Time 0 and expiring at Time T is simply the present value of the difference in the forward prices, PV($F_t - F_0$), at Time t. To be clear, the PV discounts the time T cash flow at the risk-free rate, r, to Time t. Equation 2a shows the long forward value at time t under annual compounding.

Value of Long Forward Contract Prior to Maturity (Time t) =

V_t (long) = Present value of the difference in forward prices:

$$V_t = \mathrm{PV}\left[F_t - F_0\right] = \frac{\left[F_t - F_0\right]}{\left(1 + r\right)^{T-t}} \tag{2a}$$

where F_t is the current forward price and F_0 is the initial forward price.

Alternatively,

$$V_t = S_t - PV\left[F_0\right] = S_t - \frac{F_0}{\left(1 + r\right)^{T-t}}. \tag{2b}$$

Equation 2b can be derived from Equation 2a. Assuming annual compounding,

$$F_t = S_t(1 + r)^{(T-t)}, \text{ so } PV[F_t] = PV[S_t(1 + r)^{(T-t)}]$$
$$= S_t(1 + r)^{(T-t)} / (1 + r)^{(T-t)} = S_t.$$

While both are correct, Equation 2a may be useful in cases when market frictions may cause the observed forward price, F_t, to differ slightly from the correct arbitrage-free price. Equation 2b may be more intuitive and has the advantage that the spot price, S_t, may be more readily observed than the forward price, F_t.

As in Equation 2a, the long forward contract value can be viewed as the present value, determined using the given interest rate, of the difference in forward prices—the initial one and the new one that is priced at the point of valuation. If we know the underlying price at Time t, S_t, we can estimate the forward price, F_t = FV(S_t), and we can then solve for the forward value as in Equation 2a.

The interim valuation of a short forward contract is determined in a similar fashion. The short position value is also the present value of differences in forward prices and simply the negative of the long position value. So that,

Value of short forward contract prior to maturity (Time t) = $-V_t$

$$-V_t = \mathrm{PV}\left[F_0 - F_t\right] = \frac{\left[F_0 - F_t\right]}{\left(1 + r\right)^{T-t}},$$

or alternatively,

$$= \text{PV}\left[F_0\right] - S_t = \frac{F_0}{(1+r)^{T-t}} - S_t$$

EXAMPLE 2

Forward Contract Value

Assume that at Time 0 we entered into a one-year long forward contract with price F_0 = 105. Nine months later, at Time t = 0.75, the observed price of the underlying stock is $S_{0.75}$ = 110 and the interest rate is 5%. The value of the existing forward contract expiring in three months will be *closest* to:

A −6.34.

B 6.27.

C 6.34.

Solution:

B is correct. Note that, $S_{0.75}$ = 110, r = 5%, and T − t = 0.25.

Therefore, the three-month forward price at Time t is equal to $F_t = FV(S_t) = 110(1 + 0.05)^{0.25} = 111.3499$.

Based on F_0 = 105, we find that the value of the existing forward entered at Time 0 and valued at Time t using the difference method (Equation 2a) is:

$$V_t = PV[F_t - F_0] = (111.3499 - 105)/(1 + 0.05)^{0.25} = 6.2729.$$

Alternatively, using Equation 2b we have,

$$V_t = S_t - PV[F_0] = 110 - [105/(1 + 0.05)^{0.25}] = 6.2729.$$

Now that we have the basics of forward pricing and forward valuation, we introduce some other realistic carrying costs that influence pricing and valuation.

2.2 Carry Arbitrage Model When Underlying Has Cash Flows

We have seen that forward pricing and valuation are driven by arbitrageurs seeking to exploit mispricing by either carrying or reverse carrying the underlying instrument. Carry arbitrage, when $F_0 > FV(S_0)$, requires paying the interest cost from borrowing to fund purchase of the underlying, whereas reverse carry arbitrage, when $F_0 < FV(S_0)$, results in receiving the interest benefit from lending the proceeds from short-selling the underlying. For many instruments, there are other significant carry costs and benefits. We will now incorporate into forward pricing various costs and benefits related to the underlying instrument. For this reason, we need to introduce some notation.

Let CB denote the **carry benefits**: cash flows the owner might receive for holding the underlying assets (e.g., dividends, foreign currency interest, and bond coupon payments). Let CB_T denote the future value of underlying carry benefits at time T and CB_0 denote the present value at time 0 of underlying carry benefits. Let CC denote the **carry costs**. For financial instruments, carry costs are essentially zero. For commodities, however, carrying costs include such factors as waste, storage, and insurance. Let CC_T denote the future value of underlying carry costs at time T and CC_0 denote the present value of underlying carry costs at time 0. We do not cover commodities in this reading, but you should be aware of these costs. Moreover, you should note that carry costs are similar to financing costs. Holding a financial asset does not generate direct carry costs, but it does result in the opportunity cost of the interest that could be earned on the money tied up in carrying the spot asset. Remember, the financing costs at the risk-free rate are included in the calculation of $F_0 = FV[S_0]$. Other carrying costs that are common to physical assets (such as storage and insurance) are equivalent concepts. For example, to buy and hold gold, money is taken out of the bank (opportunity cost = r, the risk-free rate) to purchase the asset, and money must be paid to store and insure it. The cost to finance the spot asset purchase, the cost to store it, and any benefits that may result from holding the asset will all play a part in determination of the forward price.

The key forward pricing equation can be expressed as:

F_0 = Future value of the underlying adjusted for carry cash flows

$$= FV[S_0 + CC_0 - CB_0] \quad (3)$$

Equation 3 relates the forward price of an asset to the spot price by considering the cost of carry. It is sometimes referred to as the **cost of carry model** or future-spot parity. Carry costs and a positive rate of interest increase the burden of carrying the underlying instrument through time; hence, these costs are added in the forward pricing equation. Conversely, carry benefits decrease the burden of carrying the underlying instrument through time; hence, these benefits are subtracted in the forward pricing equation.

Based on Equation 3, $F_0 = FV(S_0 + CC_0 - CB_0)$, if there are no explicit carry costs ($CC_0 = 0$) as with many financial assets, then we have:

$$F_0 = FV(S_0) - FV(CB_0) = FV(S_0) - FV(\text{Benefits}).$$

For a stock paying a dividend (D), a benefit, prior to maturity of the forward contract, we have the forward contract price (F_0):

$$F_0 = FV(S_0 - PV(D)) = FV(S_0) - FV(D).$$

In words, the initial forward price (F_0) is equal to the future value of carrying the underlying (S_0) minus the future value of any ownership benefits, (FV(D)), for a dividend paying stock, prior to expiration. Note the FV computation for the stock price will likely use a different time period than the FV computation for the dividends. This is because the dividend FV is only compounded from the time the dividend is collected until the expiration of the forward contract. So, FV(PV(D)) for a dividend collected at time t and held to expiration at time T would be $FV(PV(D)) = FV(D/(1 + r)^t) = (1 + r)^T \times [(D/(1 + r)^t) = D(1 + r)^{T-t}$. The calculation of F_0 for a dividend paying stock is illustrated in Example 3.

EXAMPLE 3

Forward Contract Price with Underlying Cash Flows

A US stock paying a \$10 dividend in two months is trading at \$1,000. Assume the US interest rate is 5% with annual compounding.

1 Based on the current stock price and the no-arbitrage approach, which of the following values is *closest* to the equilibrium three-month forward price?

A \$1,002.23

B \$1,022.40

C \$1,025.31

2 If the dividend is instead paid in one month, the three-month forward price will:

A decrease.

B increase.

C be unchanged.

Solution to 1:

A is correct. Based on the information given, we know S_0 = \$1,000, r = 5% (annual compounding), and T = 0.25. After 2 months, we will receive the benefit of a \$10 dividend, which earns interest for 1 month. Therefore,

$$F_0 = FV(S_0) - FV(D) = 1{,}000(1 + 0.05)^{3/12} - 10(1 + 0.05)^{1/12}$$
$$= \$1{,}012.2722 - \$10.0407 = \$1{,}002.2315.$$

Using Equation 3, we could have arrived at the same result. Here $CC_0 = 0$, and CB_0 is the PV of the dividend at time 0 = $10/(1 + 0.05)^{2/12}$ = \$9.919. Then,

$$F_0 = FV(S_0 + CC_0 - CB_0) = FV(1{,}000 + 0 - 9.919)$$
$$= (990.081)\times(1 + 0.05)^{3/12} = \$1{,}002.23.$$

Solution to 2:

A is correct. The benefit of the dividend occurs one month earlier, so we can collect interest for one additional month. The future value of the dividend would be slightly higher. So, the forward price would decrease slightly,

$$F_0 = FV(S_0) - FV(D) = 1{,}000(1 + 0.05)^{3/12} - 10(1 + 0.05)^{2/12}$$
$$= \$1{,}012.2722 - \$10.0816 = \$1{,}002.1906.$$

The value for a long forward position when the underlying has carry benefits or carry costs is found in the same way as described previously except that the new forward price (F_t), as well as the initial one (F_0), are adjusted to account for these benefits and costs. Specifically,

V_t = Present value of the difference in forward prices adjusted for carry benefits and costs
$= PV[F_t - F_0]$.

This equation is Equation 2a. The forward value is equal to the present value of the difference in forward prices. The PV discounts the risk-free cash flow $[F_t - F_0]$ at time T to time t. The benefits and costs are reflected in this valuation equation because they are incorporated into the forward prices, where $F_t = FV(S_t + CC_t - CB_t)$ and $F_0 = FV(S_0 + CC_0 - CB_0)$. Again, the forward value is simply the present value of the difference in forward prices.

EXAMPLE 4

Forward Contract Price with Carry Costs and Benefits

A long one-year forward contract on a productive asset was entered at a forward price of ₡1,000. Now, seven months later, the underlying asset is selling for ₡1,050. The PV of the cost to store, insure, and maintain the asset for the next 5 months is ₡4.00, and the asset will generate income over the next 5 months with a PV of ₡28.00. Assume annual compounding for all costs and benefits and a risk-free rate of 2%.

Based on the current spot price and the no-arbitrage approach, which of the following values is *closest* to the equilibrium five-month forward value?

A ₡34.22

B ₡33.50

C ₡35.94

Solution to 1:

A is correct. Based on the information given, we know the following: F_0 = 1,000, S_t = 1,050, CC_t = 4, CB_t = 28, t = 7 months, T – t = 5 months, and r = 2%. The new forward price is $F_t = FV(S_t + CC_t - CB_t)$. So, with annual compounding, we have:

$F_t = (1{,}050 + 4 - 28)(1 + 0.02)^{5/12}$ = ₡1,034.50 and

$V_t = PV[F_t - F_0)$ = [₡1,034.50 – ₡1000]/$(1 + 0.02)^{5/12}$ = ₡34.22.

Now let us consider stock indexes, such as the EURO STOXX 50 or the US Russell 3000. With stock indexes, it is difficult to account for the numerous dividend payments paid by underlying stocks that vary in timing and amount. A **dividend index point** is a measure of the quantity of dividends attributable to a particular index. It is a useful measure of the amount of dividends paid, a very useful number for arbitrage trading. To simplify the problem, a continuous dividend yield is often assumed. This means it is assumed that dividends accrue continuously over the period in question rather than on specific discrete dates, which is not an unreasonable assumption for an index with a large number of component stocks.

The focus of the carry arbitrage model with continuous compounding is again the future value of the underlying adjusted for carry costs and benefits and can be expressed as:

$$F_0 = S_0 \exp^{(r_c + CC - CB)T} \quad (4)$$

(Future value of the underlying adjusted for carry).

Note that in this context, r_c, CC, and CB are continuously compounded rates.

The carry arbitrage model can also be used when the underlying asset requires storage costs, needs to be insured, and suffers from spoilage. In these cases, rather than lowering the carrying burden, these costs make it more expensive to carry and hence the forward price is higher. We now apply these results to equity forward and futures contracts.

3 PRICING EQUITY FORWARDS AND FUTURES

b describe how equity forwards and futures are priced, and calculate and interpret their no-arbitrage value

We now apply the concepts of arbitrage-free pricing and valuation to the specific types of forward and futures contracts typically used in investment management. We cover, in turn, equity, interest rate, fixed income, and currency forwards and futures. In doing so, we take account of the cash flows generated by the underlying (e.g., dividends, bond coupon payments, foreign currency interest) and the unique features of each of these contracts.

3.1 Equity Forward and Futures Contracts

Although we alluded to equity forward pricing and valuation in the last section, we will now illustrate with concrete examples the application of carry arbitrage models to equity forward and futures contracts. Remember that here we assume that forward contracts and futures contracts are priced in the same way. Additionally, remember that it is vital to treat the compounding convention of interest rates appropriately.

If the underlying is a stock, then the carry benefit is the dividend payments as illustrated in the next two examples.

EXAMPLE 5

Equity Futures Contract Price with Continuously Compounded Interest Rates

The continuously compounded dividend yield on the EURO STOXX 50 is 3%, and the current stock index level is 3,500. The continuously compounded annual interest rate is 0.15%. Based on the carry arbitrage model, the three-month futures price will be *closest* to:

A 3,473.85.

B 3,475.15.

C 3,525.03.

Solution:

B is correct. Based on the carry arbitrage model (see Equation 4), the futures price is

$$f_0 = S_0 \exp^{(r_c + CC - CB)T}$$

We assume the carry costs (CC) are 0 for a financial asset, such as a stock index. The carry benefit (CB), in this case a 3% continuous dividend yield, is greater than the financing cost r_c (0.15%), so the futures price will be below the spot price. The futures price, the future value of the underlying adjusted for carry (i.e., the dividend payments, over the next 3-months) is:

$$f_0 = 3{,}500\exp^{(0.0015 + 0 - 0.03)(3/12)} = 3{,}475.15.$$

EXAMPLE 6

Equity Forward Pricing and Forward Valuation with Discrete Dividends

Suppose Nestlé common stock is trading for CHF70 and pays a CHF2.20 dividend in one month. Further, assume the Swiss one-month risk-free rate is 1.0%, quoted on an annual compounding basis. Assume that the stock goes ex-dividend the same day the single stock forward contract expires. Thus, the single stock forward contract expires in one month.

The one-month forward price for Nestlé common stock will be *closest* to:

A CHF66.80.

B CHF67.86.

C CHF69.94.

Solution:

B is correct. In this case, we have $S_0 = 70$, $r = 1.0\%$, $T = 1/12$, and $FV(CB_0) = 2.20 = CB_T$. Therefore,

$$F_0 = FV(S_0 + CC_0 - CB_0) = FV(S_0) + FV\ (CC_0) - FV(CB_0)$$
$$= 70(1 + 0.01)^{1/12} + 0 - 2.20 = \text{CHF}67.86.$$

As shown in Equation 2a, the value of a forward contract is simply the present value (discounted from time T to time t) of the difference in the initial forward price and the current forward price, that is V_t (long) = $PV[F_t - F_0]$. We will employ this basic principal to value various forward and swap contracts. Here, we find the current value (at time t) of an equity forward contract initially entered at time 0. To reiterate, the value prior to expiration is the present value of the difference in the initial equity forward price and the current equity forward price as illustrated in the next example.

EXAMPLE 7

Equity Forward Valuation

Suppose we bought a one-year forward contract at 102, and there are now three months to expiration. The underlying is currently trading for 110, and interest rates are 5% on an annual compounding basis.

1 If there are no other carry cash flows, the forward value of the existing contract will be *closest* to:

   A –10.00.

   B 9.24.

   C 10.35.

2 If a dividend payment is announced between the forward's valuation and expiration dates, assuming the news announcement does not change the current underlying price, the forward value will *most likely*:

   A decrease.

   B increase.

   C be the same.

Suppose that instead of buying a forward contract, we buy a one-year *futures* contract at 102 and there are now three months to expiration. Today's futures price is 112.35. There are no other carry cash flows.

3 After marking to market, the futures value of the existing contract will be *closest* to:

   A –10.35.

   B 0.00.

   C 10.35.

Solution to 1:

B is correct. For this case, we have $F_0 = 102$, $S_{0.75} = 110$, $r = 5\%$, and $T - t = 0.25$. Note that the new forward price at t is simply $F_t = FV(S_t) = 110(1 + 0.05)^{0.25} = 111.3499$. Therefore, from Equation 2a we have:

$$V_t = PV[F_t - F_0] = (111.3499 - 102)/(1 + 0.05)^{0.25} = 9.2366, \text{ or}$$

alternatively, using Equation 2b,

$$V_t = S_t - PV[F_0] = 110 - 102/(1 + 0.05)^{0.25} = 9.2366.$$

Thus, we see that the current forward value is greater than the difference between the current underlying price of 110 and the initial forward price of 102 due to interest costs resulting in the new forward price being 111.35.

Solution to 2:

A is correct. The old forward price is fixed. The discounted difference in the new forward price and the old forward price is the value. If we impose a new dividend, it would lower the new forward price and thus lower the value of the old forward contract.

Solution to 3:

B is correct. Futures contracts are marked to market daily, which implies that the market value, resulting in profits and losses, is received or paid at each daily settlement. Hence, the equity futures value is zero each day after settlement has occurred.

We turn now to the widely used interest rate forward and futures contracts.

3.2 Interest Rate Forward and Futures Contracts

c describe how interest rate forwards and futures are priced, and calculate and interpret their no-arbitrage value

Libor, which stands for London Interbank Offered Rate, is a widely used interest rate that serves as the underlying for many derivative instruments. It represents the rate at which London banks can borrow from other London banks. When these loans are in dollars, they are known as Eurodollar time deposits, with the rate referred to as dollar Libor. There are, however, Libor rates for all major non-dollar currencies. Average Libor rates are derived and posted daily at 11:30 a.m. London time. Lenders and participants in the interest rate derivatives market use these posted Libor rates to determine the interest payments on loans and the payoffs of various derivatives. In 2008, financial regulators and many market participants began to suspect that the daily quoted Libor rates, which were compiled by the British Bankers Association (BBA), were being manipulated by certain banks that submitted their rates to the BBA for use in determining average Libor rates. The manipulation of Libor by some participants has resulted in the search for a new market reference rate (MRR), which would serve as a reliable reference for financial transactions. Candidates to replace Libor include SOFR (Secured Overnight Financing Rate), which would be determined by the Federal Reserve Bank of New York, and SONIA (Sterling Overnight Index Average), administered by the Bank of England. Libor is expected to be published until the end of 2021. Which MRR will replace Libor has yet to be decided, so for the examples in this section, we will continue to use Libor.

Currently, there are active forward and futures markets for derivatives based on Libor. We will use the symbol L_m to represent our spot MRR. Our focus will be on forward markets, as represented by forward rate agreements. In order to understand the forward market, however, let us first look at the MRR spot market.

Assume the following notation:

L_m = MRR spot rate (set at time t = 0) for an m-day deposit
NA = notional amount, quantity of funds initially deposited
NTD = number of total days in a year, used for interest calculations (360 in the Libor market)
t_m = accrual period, fraction of year for an m-day deposit—t_m = m/NTD = m/360 (for the Libor market)
TA = terminal amount, quantity of funds repaid when the Libor deposit is withdrawn

For example, suppose we are considering a 90-day Eurodollar deposit (m = 90). Dollar Libor is quoted at 2%; thus, $L_{90} = 0.02$. If \$50,000 is initially deposited, then NA = \$50,000. Libor is stated on an actual over 360-day count basis (often denoted ACT/360) with interest paid on an add-on basis. Add-on basis is the convention in the Libor market. The idea is that the interest is added on at the end—in contrast, for example, to the discount basis, in which the current price is discounted based on the amount paid at maturity. Hence, $t_m = 90/360 = 0.25$. Accordingly, the terminal amount can be expressed as:

$TA = NA \times [1 + L_m t_m]$, and the interest paid is $TA - NA = NA \times [L_m t_m]$.

In this example, TA = \$50,000 × [1 + 0.02(90/360)] = \$50,250 and the interest is \$50,250 – \$50,000 = \$250.

Now let us turn to the forward market for Libor (i.e., MRR). A **forward rate agreement** (FRA) is an over-the-counter (OTC) forward contract in which the underlying is an interest rate on a deposit. An FRA involves two counterparties: the fixed-rate payer (long), who is also the floating-rate receiver, and the fixed-rate receiver (short), who is also the floating-rate payer. Thus, a fixed-payer (long) FRA will profit when the MRR rises. If the floating rate is above the rate in the forward agreement, the long position can be viewed as having the benefit of borrowing at below market rates. The long will receive a payment. A long FRA would be well suited for a firm planning to borrow in the future and wishing to hedge against rising rates. A fixed-receiver (short) FRA might be a bank or financial institution hoping to lock in a fixed lending rate in the future. The fixed receiver, as the name implies, receives an interest payment based on a fixed rate and makes an interest payment based on a floating rate. If we are the fixed receiver, then it is understood without saying that we also are the floating payer, and vice versa. Because there is no initial exchange of cash flows, to eliminate arbitrage opportunities, the FRA price is the fixed interest rate such that the FRA value is zero on the initiation date.

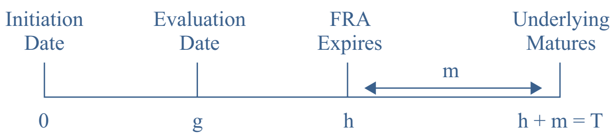
FRAs are identified in the form of "X × Y," where X and Y are months and the multiplication symbol, ×, is read as "by." To grasp this concept and the notion of exactly what is the underlying in an FRA, consider a 3 × 9 FRA, which is pronounced "3 by 9." The 3 indicates that the FRA expires in three months. After three months, we determine the FRA payoff based on an underlying rate. The underlying is implied by the difference in the 3 and the 9. That is, the payoff of the FRA is determined by a six-month (180-day) MRR (such as Libor) when the FRA expires in three months. The notation 3 × 9 is market convention, though it can seem confusing at first. If Libor is the MRR, the rate on the FRA will be determined by the relationship between the spot rate on a nine-month Libor deposit and the spot rate on a three-month Libor deposit when the FRA is initiated. A long FRA will effectively replicate going long a nine-month Libor deposit and short a three-month Libor deposit. Note that although market convention quotes the time periods as months, the calculations use days based on the assumption of 30 days in a month.

The contract established between the two counterparties settles in cash the difference between a fixed interest payment established on the initiation date and a floating interest payment established on the FRA expiration date. The underlying of an FRA is neither a financial asset nor even a financial instrument; it is just an interest payment. It is also important to understand that the parties to an FRA are not necessarily engaged in a Libor deposit in the spot market. The Libor spot market is simply the benchmark from which the payoff of the FRA is determined. Although a party may use an FRA in conjunction with a Libor deposit, it does not have to do so any more than a party that uses a forward or futures on a stock index has to have a position in the stock index.

In Exhibit 8, we illustrate the key time points in an FRA transaction. The FRA is created and priced at Time 0, the initiation date, and expires h days later. The underlying instrument has m days to maturity as of the FRA expiration date. Thus, the FRA payoff is based on the spot m-day MRR observed in h days from FRA initiation. We can only observe spot market reference rates, such as spot Libor. To price the FRA, we require two spot rates: L_h, which takes us to the expiration of the FRA, and L_T, which takes us to the underlying maturity.

The FRA helps hedge single period interest rate risk for an m-day period beginning h days in the future. After the initial FRA rate (FRA_0) is established, we may also wish to determine a value for our FRA at a later date (Time g). As the MRR changes, our interest rate agreement may take on a positive or negative value.

Exhibit 8 Important FRA Dates, Expressed in Days from Initiation



Using the notation in Exhibit 8, let FRA_0 denote the fixed forward rate set at Time 0 that expires at Time h wherein the underlying Libor deposit has m days to maturity at expiration of the FRA. Thus, the rate set at initiation of a contract expiring in 30 days in which the underlying is a 90-day MRR, denoted as a 1 x 4 FRA, will be such a number as 1% or 2.5%. Like all standard forward contracts, no money changes hands when an FRA is initiated, so our objective is to price the FRA, meaning to determine the fixed rate (FRA_0), such that the value of the FRA contract is zero on the initiation date.

When any interest rate derivative expires, there are technically two ways to settle at expiration: "advanced set, settled in arrears" and "advanced set, advanced settled." It is important to note that FRAs are typically settled based on "advanced set, advanced settled," whereas swaps and interest rate options are normally based on "advanced set, settled in arrears." Let us look at both approaches, because they are both used in the interest rate derivatives markets.

In the earlier example of a Libor deposit of $50,000 for 90 days at 2%, the rate was set when the money was deposited, and interest accrued over the life of the deposit. A payment of $50,250 (interest of $250 + principal of $50,000) was made at maturity, 90 days later. Here the term **advanced set** is used because the reference interest rate is set at the time the money is deposited. The advanced set convention is almost always used because most issuers and buyers of financial instruments want to know the rate on the instrument while they have a position in it.

In an FRA, the term "advanced" refers to the fact that the interest rate is set at Time h, the FRA expiration date, which is the time the underlying deposit starts. The term **settled in arrears** is used when the interest payment is made at Time h + m, the maturity of the underlying instrument. Thus, an FRA with advanced set, settled in arrears works the same way as a typical bank deposit as described in the previous example. At Time h, the interest rate is set at L_m, and the interest payment is made at Time T (h + m). Alternatively, when **advanced settled** is used, the settlement is made at Time h. Thus, in an FRA with the advanced set, advanced settled feature, the FRA expires and settles at the same time. Importantly, advanced set, advanced settled is almost always used in FRAs; although we will see advanced set, settled in arrears when we cover interest rate swaps, and it is also used in interest rate options. From this point forward in this discussion, all FRAs will be advanced set, advanced settled, as they are in practice.

The settlement amounts for advanced set, advanced settled are discounted in the following manner:

Settlement amount at h for receive-floating (Long):

$NA \times \{[L_m - FRA_0]\ t_m\}/[1 + D_m t_m]$.

Again, the FRA is a forward contract on interest rates; long FRA (floating receiver) wins when rates increase. Note the floating rate (Libor perhaps, L_m) is received and thus has a positive sign. Since floating is received, the fixed rate (FRA_0) is paid (outflow). The FRA rate (fixed at t = 0 for the period m, which runs from time h to time T) is an outflow for the long and has a negative sign. For receive fixed (short), the FRA rate is an inflow and the floating rate L_m is an outflow.

Settlement amount at h for receive-fixed (Short):

$NA \times \{[FRA_0 - L_m]\ t_m\}/[1 + D_m t_m]$.

The divisor, $1 + D_m t_m$, is a discount factor applied to the FRA payoff. It reflects the fact that the rate on which the payoff is determined, L_m, is obtained on day h from the spot market (advanced set), which uses settled in arrears. The discount factor is, therefore, appropriately applied to the FRA payment because the payment is received in advance, not in arrears. That is, the FRA payment is made early (advanced settled), but the interest on the loan is not due until later (settled in arrears). So, the settlement amount at time h is discounted to account for the fact that interest can be earned for m days on the advanced payment. Often it is assumed at time h that $D_m = L_m$, and we will commonly do so here, but it can be different.

Again, it is important to not be confused by the role played by an MRR, such as Libor spot market in an FRA. In the Libor spot market, deposits are made by various parties that are lending to banks. These rates are used as the benchmark for determining the payoffs of FRAs. The two parties to an FRA do not necessarily engage in any Libor spot transactions. Again, Libor spot deposits are settled in arrears, whereas FRA payoffs are settled in advance—hence the discounting.

EXAMPLE 8

Calculating Interest on Libor Spot and FRA Payments

In 30 days, a UK company expects to make a bank deposit of £10,000,000 for a period of 90 days at 90-day Libor set 30 days from today. The company is concerned about a possible decrease in interest rates. Its financial adviser suggests that it negotiate today a 1 × 4 FRA, an instrument that expires in 30 days and is based on 90-day Libor. The company enters a £10,000,000 notional amount 1 × 4 receive-fixed FRA that is advanced set, advanced settled (note the company is the short-side of this FRA contract). The appropriate discount rate for the FRA settlement cash flows is 2.40%. After 30 days, 90-day Libor in British pounds is 2.55%.

1 The interest actually paid at maturity on the UK company's bank deposit will be *closest* to:

A £60,000.

B £63,750.

C £67,500.

2 If the FRA was initially priced so that FRA_0 = 2.60%, the payment received to settle it will be *closest* to:

A –£2,485.08.

B £1,242.54.

C £1,250.00.

3 If the FRA was initially priced so that FRA_0 = 2.50%, the payment received to settle it will be *closest* to:

A –£1,242.54.

B £1,242.54.

C £1,250.00.

Solution to 1:

B is correct. This is a simple deposit of £10,000,000 for 90 days at the prevailing 90-day Libor. Since m = 90, we use L_{90} = 2.55%. Therefore, TA = 10,000,000 × [1 + 0.0255(0.25)] = £10,063,750. So, the interest paid at maturity is £63,750.

Solution to 2:

B is correct. In this example, m = 90 (number of days in the deposit), t_m = 90/360 (fraction of year until deposit matures observed at the FRA expiration date), and h = 30 (number of days initially in the FRA). The settlement amount of the 1 × 4 FRA at h for receive-fixed (the short) is:

$$NA \times \{[FRA_0 - L_m]t_m\}/[1 + D_m t_m]$$

$$= 10{,}000{,}000 \times \{[0.0260 - 0.0255](0.25)\}/[1 + 0.0240(0.25)]$$

= £1,242.54.Since the short FRA involves paying floating, the short benefited from a decline in rates. Note D_m does not equal L_m in this example.

Solution to 3:

A is correct. The data are similar to those in the previous question, but the initial FRA rate is now 2.50% and not 2.60%. Thus, the settlement amount of the 1 × 4 FRA at time h for receive-fixed (the short) is:

$$NA \times \{[FRA_0 - L_m]t_m\}/[1 + D_m t_m]$$

$$= 10{,}000{,}000 \times \{[0.0250 - 0.0255](0.25)\}/[1 + 0.0240(0.25)]$$

= –£1,242.54.

The short-side in the FRA suffered from a rise in rates because it is paying floating.

At this point, we highlight a few key concepts about FRAs and how to price and value them:

1 An FRA is a forward contract on interest rates. The long side of the FRA, fixed-rate payer (floating-rate receiver), incurs a gain when rates increase and incurs a loss when rates decrease. Conversely, the short side of the FRA, fixed-rate receiver (floating-rate payer), incurs a loss when rates increase and incurs a gain when rates decrease.

2 The FRA price, FRA_0, is the implied forward rate for the period beginning when the FRA expires to the underlying loan maturity. So, we require two spot rates to determine the initial forward rate. Therefore, pricing an FRA is like pricing a forward contract.

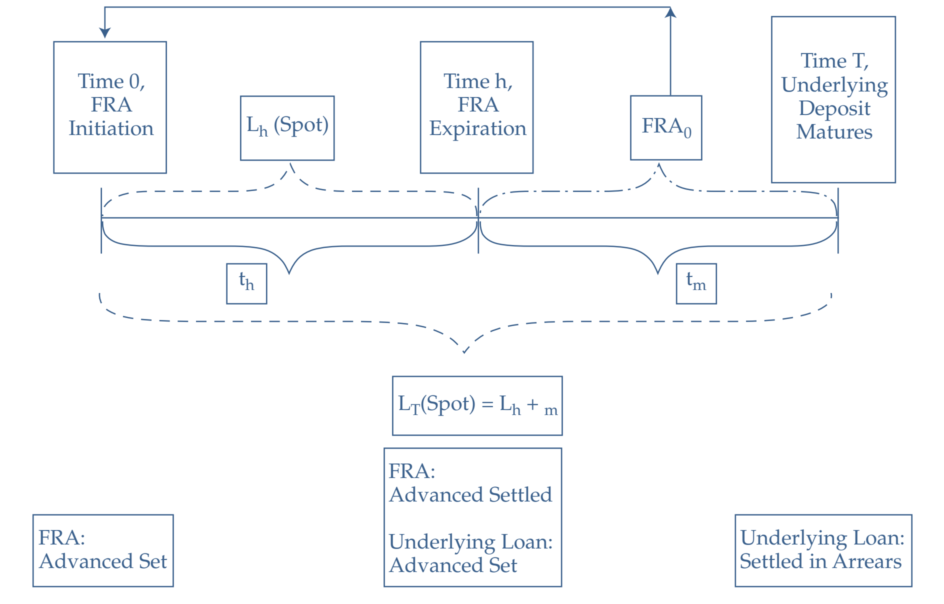
3 Although the interest on the underlying loan will not be paid until the end of the loan, the payoff on the FRA will occur at the expiration of the FRA (advanced settled). Therefore, the payoff of an FRA is discounted back to the expiration of the FRA.

As noted in point 2, the FRA price is the implied forward rate for the period beginning when the FRA expires at time h and running m days to the underlying loan maturity at time T. It is similar to any other forward contract. We wish to identify the appropriate FRA_0 rate that makes the value of the FRA equal to zero on the initiation date. The concept used to derive FRA_0 can be understood through a simple example.

Recall that with simple interest, a one-period forward rate is found by solving the expression $[1 + y(1)]\ [1 + F(1)] = [1 + y(2)]^2$, where y(1) denotes the one-period yield to maturity and y(2) the two-period yield to maturity. F denotes the forward rate in the next period. We can observe the spot rates y(1) and y(2). The forward rate is implied from those two rates. Borrowing or lending along the 2-year path must cost the same as borrowing or lending along the path using the 1-year spot and the 1-year forward. The solution for F(1) is simply $F(1) = ([1 + y(2)]^2/[1 + y(1)]) - 1$. Assume the one-year spot rate is 3% and the two-year spot rate is 4%. To prevent arbitrage, $F(1) = ([1 + 0.04]^2/[1 + 0.03])] - 1 = 0.0501$. If the forward rate was not 5.01%, an arbitrageur could make a risk-free profit through borrowing along one path and lending along another.

As depicted in Exhibit 9, the rate for an FRA is computed in the same manner. We derive the forward rate (or FRA rate, FRA_0) from two spot rates (such as Libor): the longer rate L_T and the shorter rate L_h. Borrowing or lending at L_T for T days should cost the same as borrowing or lending for h days at L_h and subsequently borrowing or lending for m days at FRA_0.

Exhibit 9 FRA Rates from Spot Market Reference Rate (MRR = LIBOR)



We can solve for the FRA rate by considering that the two paths must be equal to prevent arbitrage or:

$$[1 + L_h t_h][1 + FRA_0 t_m] = [1 + L_T t_T].$$

The solution in annualized form is shown in Equation 5:

$$FRA_0 = \{[1 + L_T t_T]/[1 + L_h t_h] - 1\}/t_m. \tag{5}$$

The result is the forward rate in the term structure.

So, if 180-day Libor is 2.0% and 90-day Libor is 1.5%, then the price of a 3 × 6 FRA would be:

$$\begin{aligned} FRA_0 &= \{[1 + L_T t_{180}]/[1 + L_h t_{90}] - 1\}/t_{90} \\ &= \{[1 + 0.02(180/360)]/[1 + 0.015(90/360)] - 1\}/(90/360) \\ &= 0.024907 \text{ or } 2.49\%. \end{aligned}$$

This result can be compared with the result from a simple approximation technique. Note that for this FRA, 90 is half of 180. Thus, we can use a simple arithmetic average equation—here, (1/2)1.5% + (1/2)X = 2.0%—and solve for the missing variable X: X = 2.5%. Knowing this approximation will always be biased slightly high, we know we are looking for an answer that is a little less than 2.5%. This is a helpful way to check your final answer.

EXAMPLE 9

FRA Fixed Rate

1 Based on market quotes on Canadian dollar (C$) Libor, the six-month C$ Libor and the nine-month C$ Libor rates are presently at 1.5% and 1.75%, respectively. Assume a 30/360-day count convention. The 6 × 9 FRA fixed rate (FRA_0) will be *closest* to:

A 2.00%.

B 2.23%.

C 2.25%.

Now consider the following information for problems 2 and 3.

Assume a 30/360-day count convention and the following spot rates:

1-Month USD Libor is 2.48%, 3-Month USD Libor is 2.58%, 6-Month USD Libor is 2.62%, and 1-Year USD Libor is 2.72%.

2 Given these four spot rates in the Libor term structure, how many FRA rates can be calculated?

A 4 FRA rates

B 6 FRA rates

C 12 FRA rates

3 The 1 × 3 FRA fixed rate will be *closest* to:

A 2.43%.

B 2.53%.

C 2.62%.

Solution to 1:

B is correct. Based on the information given, we know L_{180} = 1.50% and L_{270} = 1.75%. The 6 × 9 FRA rate is thus:

$FRA_0 = \{[1 + L_T t_T]/[1 + L_h t_h] - 1\}/t_m$
$FRA_0 = \{[1 + 0.0175(270/360)]/[1 + 0.015(180/360)] - 1\}/(90/360)$
$FRA_0 = [(1.013125/1.0075) - 1]/(0.25) = 0.022333$, or 2.23%.

Solution to 2:

B is correct. Based on the four Libor spot rates given, we can compute six separate FRA rates as follows: 1 × 3, 1 × 6, 1 × 12, 3 × 6, 3 × 12, and 6 × 12 FRA rates.

Solution to 3:

C is correct. Based on the information given, we know L_{30} = 2.48% and L_{90} = 2.58%. The 1 × 3 FRA rate is thus:

$FRA_0 = \{[1 + L_T t_T]/[1 + L_h t_h] - 1\}/t_m$
$FRA_0 = \{[1 + 0.0258(90/360)]/[1 + 0.0248(30/360)] - 1\}/(60/360)$
$FRA_0 = [(1.00645/1.00207) - 1]/(0.1667) = 0.026220$, or 2.62%.

We can now value an existing FRA (with rate FRA_0) using the same general approach as we did with the forward contracts previously covered. Specifically, we can enter into an offsetting transaction at the new rate that would be set on an FRA that expires at the same time as our original FRA. By taking the opposite position, the new FRA offsets the old one. That is, if we are long the old FRA, we will pay fixed and receive the floating rate L_m at h. We can go short a new FRA and receive fixed (with rate FRA_g) that will obligate us to pay L_m at h.

Consider the following strategy. Let us assume that we initiate an FRA that expires in 90 days and is based on 90-day Libor (so, a 3 x 6 FRA). The fixed rate at initiation FRA_0 = 2.49% and t_m = 90/360. We are long the FRA, so we will pay the fixed rate of 2.49% and receive floating Libor. Having entered the long FRA, we wish to value our position 30 days later, at Time g, when there are 60 days remaining in the life of the FRA (note that this is now a 2 x 5 FRA, as one month has passed since FRA initiation). Assume, at this point, the rate on an FRA based on 90-day Libor that expires in 60 days (FRA_g) is 2.59%. Remember, the original FRA has a fixed rate set at 2.49% when it was initiated. Now, 30 days later, a new offsetting FRA can be created at 2.59%. To value the original FRA (at Time g), we short a new FRA that will receive fixed at 2.59% and pay floating Libor at time h. Effectively, we are now receiving fixed at 2.59% and paying fixed at 2.49%. The value of the offset position is 10 bps times (90/360), as follows, times the notional amount, which is then discounted to back to Time g:

10 bps: $FRA_g - FRA_0 = 2.59\% - 2.49\% = 0.10\%$

90/360: $t_m = m/NTD$, as L_m is the 90-day Libor rate underlying both FRAs

Because the cash flows at T are now known with certainty at g, this offsetting transaction at Time g has eliminated any floating-rate risk at Time T. That is, we had a long FRA at time 0 and added a short FRA at time g. Since the notional amounts and times to maturity of the offsetting transaction are the same, the floating portion of the FRA cash flows (L_m) at time T will exactly cancel, $[L_m - FRA_0] + [FRA_g - L_m] = [FRA_g - FRA_0]$.

Our task, however, is to determine the fair value of the original FRA at Time g. Therefore, we need the present value of this Time T cash flow at Time g. That is, the value of the original FRA is the PV of the difference in the new FRA rate and the old FRA rate times the notional amount. Specifically, we let V_g be the value at Time g of the original FRA that was initiated at Time 0, expires at Time h, and is based on m-day Libor, L_m. Note that discounting will be over the period T − g. With D_{T-g} as the discount rate and NA as the notional amount. So,

Long FRA value at Time g: $V_g =$

$$NA \times \{[FRA_g - FRA_0]\ t_m\}/[1+ D_{(T-g)}t_{(T-g)}]. \tag{6}$$

Thus, the Time g value of the receive-floating FRA initiated at Time 0 (V_g) is just the present value of the difference in FRA rates, one entered at Time g and one entered at Time 0. Traditionally, it is assumed that the discount rate, D_m, is equal to the underlying floating rate, L_m, but that is not necessary. Note that here it is $D_{(T-g)}$.

The value of a receive-fixed or short FRA at time g is the negative of the long value ($-V_g$), so we have: $-V_g = -1 \times (NA \times \{[FRA_g - FRA_0]\ t_m\}/[1+ D_{(T-g)}\ t_{(T-g)}])$.

Short FRA value at Time g =

$$NA \times \{[FRA_0 - FRA_g]\ t_m\}/[1+ D_{(T-g)}\ t_{(T-g)}] \tag{6a}$$

EXAMPLE 10

FRA Valuation

Suppose we entered a receive-floating (long) 6 × 9 FRA with Canadian dollar notional amount of C$10,000,000 at Time 0. The six-month spot C$ Libor was 0.628%, and the nine-month C$ Libor was 0.712%. Also, assume the 6 × 9 FRA rate is quoted in the market at 0.877%. After 90 days have passed, the three-month C$ Libor is 1.25% and the six-month C$ Libor is 1.35%, which we will use as the discount rate to determine the value at g.

Assuming the appropriate discount rate is C$ Libor, the value of the original receive-floating 6 × 9 FRA will be *closest* to:

A C$14,105.

B C$14,200.

C C$14,625.

Solution:

A is correct. Initially, we have L_{180} = 0.628%, L_{270} = 0.712%, and FRA_0 = 0.877%.

After 90 days (g = 90), we have L_{90} = 1.25% and L_{180} = 1.35%. Interest rates rose during this period; hence, the FRA has gained value because the position is receive-floating. First, we compute the new FRA rate at Time g and then estimate the fair FRA value as the discounted difference in the new and old FRA rates. The new FRA rate at Time g, denoted FRA_g, is the rate on an FRA expiring in 90 days in which the underlying is 90-day C$ Libor (so, a 3 x 6 FRA). That rate is found using Equation 5. The shorter spot rate is now for h – g (180 – 90 = 90) days, which is the new time until both FRAs expire. The reference spot rate for the underlying maturity is now in T – g (270 – 90 = 180) days.

$$FRA_g = \{[1 + L_{180}\, t_{(T-g)}]/[1 + L_{90}\, t_{(h-g)}] - 1\}/t_m,$$

T – g = 180 days and h – g = 90 days, so we have:

$$FRA_g = \{[1 + L_{180}\,(180/360)]/[1 + L_{90}\,(90/360)] - 1\}/(90/360).$$

Substituting the values given in this problem, we find:

$$FRA_g = \{[1 + 0.0135\,(180/360)]/[1 + 0.0125\,(90/360)] - 1\}/(90/360) = [(1.006750/1.003125) - 1]/0.25 = 0.014455, \text{ or } 1.445\%.$$

Therefore, using Equation 6, we have:

$$V_g = 10{,}000{,}000 \times \{[0.01445 - 0.00877]\,(90/360)\}/[1 + 0.0135\,(180/360)] = 14{,}105.$$

We now turn to the specific features of various forward and futures markets. The same general principles will apply, but the specifics will be different.

4 PRICING FIXED-INCOME FORWARD AND FUTURES CONTRACTS

d describe how fixed-income forwards and futures are priced, and calculate and interpret their no-arbitrage value;

Fixed-income forward and futures contracts have several unique issues that influence the specifics of the carry arbitrage model. First, in some countries the prices of fixed-income securities (termed “bonds” here) are quoted without the interest that has accrued since the last coupon date. The quoted price is sometimes known as the clean price. Naturally when buying a bond, one must pay the full price, which is sometimes called the dirty price, so the accrued interest is included. Nonetheless, it is necessary to understand how the quoted bond price and accrued interest compose the true bond price and the effect this convention has on derivatives pricing. The quotation convention for futures contracts, whether based on clean or dirty prices, usually corresponds to the quotation convention in the respective bond market. In this section, we will largely treat forwards and futures the same, except in certain places where noted.

In general, accrued interest is computed based on the following linear interpolation formula:

Accrued interest = Accrual period × Periodic coupon amount, or

AI = (NAD/NTD) × (C/n),

where NAD denotes the number of accrued days since the last coupon payment, NTD denotes the number of total days during the coupon payment period, n denotes the number of coupon payments per year (commonly n = 2 for semi-annual), and C is the stated annual coupon amount. For example, after two months (60 days), a 3% semi-annual coupon bond with par of 1,000 would have accrued interest of AI = (60/180) × (30/2) = 5. Note that accrued interest is expressed in currency units (not percent), and the number of total days (NTD) depends on the coupon payment frequency. As in the example, semi-annual indicates coupons are paid twice per year, so with 360 days per year, NTD = 360/2= 180.

Second, fixed-income futures contracts often have more than one bond that can be delivered by the seller. Because bonds trade at different prices based on maturity and stated coupon, a mathematical adjustment to the amount required when settling a futures contract, known as the conversion factor (CF), is used to make all deliverable bonds approximately equal in price. According to the Chicago Mercantile Exchange, "A conversion factor is the approximate decimal price at which $1 par of a security would trade if it had a six percent yield-to-maturity." So, the CF adjusts each bond to an equivalent 6% coupon bond (i.e., benchmark bond). Other exchanges use different conversion factors, and these are illustrated later in the text and examples.

Third, when multiple bonds can be delivered for a particular futures contract, a cheapest-to-deliver bond typically emerges after adjusting for the conversion factor. The conversion factor adjustment, however, is not precise. Thus, if there are several candidates for delivery, the bond that will be delivered is the one that is least expensive for the seller to purchase in the open market to settle the obligation.

For bond markets in which the quoted price includes the accrued interest and in which futures or forward prices assume accrued interest is in the bond price quote, the futures or forward price simply conforms to the general formula we have previously discussed. Recall that the futures or forward price is simply the future value of the underlying in which finance costs, carry costs, and carry benefits are all incorporated, or

$$\begin{aligned} F_0 &= \text{Future value of underlying adjusted for carry cash flows} \\ &= FV(S_0 + CC_0 - CB_0). \end{aligned}$$

Let Time 0 be the forward contract trade initiation date and Time T be the forward contract expiration date, as shown in Exhibit 10. For the fixed-income bond, let Y denote the time to maturity of the bond at Time T, when the forward contract expires. Therefore, T + Y denotes the underlying instrument's current (Time 0) time to maturity. Let B_0 denote the quoted bond price observed at Time 0 of a fixed-rate bond that matures at Time T + Y and pays a fixed coupon rate.

Exhibit 10 Timeline for Bond Futures and Forwards

Time 0, Forward or Futures Contract Initiation

Time T, Forward or Futures Contract Expiration

Time Y, Time to Maturity of Bond at Time T

Time T + Y Underlying Bond Matures

Quoted Bond Price	B_0
Accrued Interest:	AI_0
Spot Bond price:	$S_0 = B_0 + AI_0$
Quoted Forward or Futures Price:	Q_0
Forward of Futures Adjusted Price:	$F_0 = Q_0 \times CF$

Quoted Bond Price	B_T
Accrued Interest:	AI_T
Spot Bond Price:	$S_T = B_T + AI_T$
Profit on Long Forward or Futures:	$V_T = B_T - F_0 = (S_T - AI_T) - F_0$
Profit on Short Forward or Futures:	$-V_T = F_0 - B_T = F_0 - (S_T - AI_T)$

For bonds quoted without accrued interest, let AI_0 denote the accrued interest at Time 0. The carry benefits are the bond's fixed coupon payments, so CB_0 = PVCI, meaning the present value of all coupon interest (CI) paid over the forward contract horizon from Time 0 to Time T. The corresponding future value of these coupons paid over the contract horizon to time T is CB_T = FVCI. Finally, there are no carry costs, and thus CC = 0. To be consistent with prior notation, we have:

S_0 = Quoted bond price + Accrued interest = $B_0 + AI_0$.

We could just insert this price (S_0) into the previous equation, letting CB_0 = PVCI, and thereby obtain the futures price the straightforward and traditional way. But fixed-income futures contracts often permit delivery of more than one bond and use the conversion factor system to provide this flexibility. In these markets, the futures price, F_0, is defined as the quoted futures price, Q_0, times the conversion factor, CF. Note that in this section, we will use the letter F to denote either the quoted forward price or the futures price times the conversion factor. In fact, the futures contract settles against the quoted bond price *without* accrued interest. Thus, as shown in Exhibit 10, the total profit or loss on a long position in fixed-income futures at expiration (Time T) is the quoted bond price minus the initial futures price or:

$v_T = B_T - F_0$. Moreover, based on our notation, we can also say,

$v_T = (S_T - AI_T) - F_0$.

The fixed-income forward or futures price including the conversion factor, termed the "adjusted price," can be expressed as:

$F_0 = Q_0 \times CF$

= FV of underlying adjusted for carry cash flows from Times 0 to T

$$= FV[S_0 + CC_0 - CB_0] = FV[S_0 + 0 - PVCI] = FV[B_0 + AI_0 - PVCI]. \quad (7)$$

In other words, the actual futures price is F_0, but in the market the availability of multiple deliverable bonds gives rise to the adjustment factor. Hence, the price you would see quoted is Q_0, where $Q_0 = F_0/CF$.

Recall that the bracketed term $B_0 + AI_0 - PVCI$ in Equation 7 is just the full spot price S_0 minus the present value of the coupons over the life of the forward or futures contract. The fixed-income forward or futures price (F_0) is thus the future value of the quoted bond price plus accrued interest less any coupon payments made during the life of the contract. Again, the quoted bond price plus the accrued interest is the spot price: It is in fact the price you would have to pay to buy the bond. Market conventions in some countries just happen to break this price out into the quoted price plus the accrued interest.

Why Equation 7 must hold is best understood by illustrating what happens when the futures price is not in equilibrium. In fact, in the following scenario, the futures are overpriced relative to the bond, giving rise to an arbitrage opportunity.

Assume we observe a 3-month forward contract, so T = 0.25, on a bond that expires at some time in the future, T + Y, and this bond is currently quoted (B_0) at 107% of par. There are no coupon payments for this bond over the life of the forward contract, so PVCI = 0.0. Other pertinent details of the bond and futures are presented in Exhibit 11.

Exhibit 11 Bond and Futures Information for Illustrating Disequilibrium and Arbitrage Opportunity

Bond		
Quoted Bond Price	B_0	107.00
PV of Coupon Interest	PVCI	0
Accrued Interest at Time 0	AI_0	0.07
Accrued Interest at Time T	AI_T	0.20
Futures		
Quoted Futures Price	Q_0	135.00
Conversion Factor	CF	0.80
Adjusted Futures Price	F_0 (= $Q_0 \times$ CF)	108.00
Interest Rate		
For Discounting/ Compounding	r	0.20%

We observe that the full spot price of the bond is:

$$S_0 = B_0 + AI_0 = 107 + 0.07 = 107.07.$$

The futures price (F_0), which is the future value adjusted for carry cash flows (using Equation 7), is:

$$F_0 = FV[B_0 + AI_0 - PVCI] = (107 + 0.07 - 0)(1.002)^{0.25} = 107.12.$$

Note that the adjusted futures price using the quoted futures price (Q_0 = 135) and the conversion factor (CF = 0.80) is F_0 = 108. Adding the accrued interest at expiration (AI_T = 0.20) to the adjusted futures price gives 108.20. Remember, if you are selling a bond you receive the accrued interest; if you are buying a bond you pay the accrued interest. The adjusted futures price plus accrued interest should equal the future value of the full bond price adjusted for any carry cash flows given by Equation 7. Here, the adjusted futures price (including accrued interest) is 108.20, while the cost to buy and carry the bonds is 107.12. This implies that the futures contract is overpriced by (108.2 – 107.12) = 1.08, thus there is an arbitrage opportunity. In this case, we would simultaneously: 1) sell the overpriced futures contract; 2) borrow funds to purchase the bonds; and 3) buy the underpriced deliverable bonds.

So, to capture the 1.08 with no risk, an arbitrageur might wish to buy this bond and carry it and short the futures contract at 108. At maturity, the arbitrageur simply delivers the bond to cover the futures contract and repays the loan. Arbitrage should allow for the capture of any over (or under) pricing. Selling the futures contract at 108 involves no initial cash flow. The short futures locks in a sale price of 108 + 0.2 =

108.20 for the bond just purchased for 107.07. Since there are no carry benefits, it costs the arbitrageur 107.12, = FV(107.07) = (107.07)(1+0.002)$^{0.25}$, to carry the bond to expiration. The result is a risk-free profit at expiration of 1.08, = 108.00 + 0.2 – 107.12, for which the Time 0 PV is 1.0795, = 1.08(1.002)$^{-0.25}$.

The value of the Time 0 cash flows should be zero to prevent an arbitrage opportunity. This example shows the arbitrage profit as a 1.0795 cash flow at Time 0 or 1.08 at time T per bond. If the value had been negative—meaning the full bond price exceeded the adjusted future price plus accrued interest—then the arbitrageur would conduct the reverse carry arbitrage of short selling the bond, lending the proceeds, and buying the futures (termed reverse carry arbitrage because the underlying is not carried but is sold short).

In equilibrium, the adjusted futures price of the bond plus any accrued interest must equal the cost of buying and holding the spot bond until time T. That is, to eliminate an arbitrage opportunity:

$$F_0 + AI_T = FV[B_0 + AI_0 - PVCI], \text{ which implies, } F_0 = FV(S_0) - AI_T - FVCI.$$

In this example, equilibrium is not met. The adjusted futures price, F_0 = 108, promises a profit of (108 – 106.92) = 1.08 at expiration, since

$$FV(S_0) - AI_T - FVCI = 107.12 - 0.2 - 0 = 106.92.$$

For clarity, substituting for F_0 and S_0 and solving for the quoted futures price (Q_0) results in Equation 8, the conversion factor adjusted futures price (i.e., quoted futures price):

$$Q_0 = [1/CF]\ \{FV\ [B_0 + AI_0\ ] - AI_T - FVCI\} \tag{8}$$

In this example we have,

$$\begin{aligned} Q_0 &= [1/CF]\ \{FV[B_0 + AI_0] - AI_T - FVCI\} \\ &= (1/0.8)\ \{(1 + 0.002)^{0.25}(107 + 0.07) - 0.20 - 0.0\} = 133.65. \end{aligned}$$

Recall, a futures price of 135 was used as the quoted price, Q_0 (108 was the adjusted futures price). Any quoted futures price higher than the equilibrium futures price of 133.65 (106.92 adjusted) will present arbitrage opportunities; hence, the arbitrage transaction of selling the futures contract resulted in a riskless positive cash flow.

EXAMPLE 11

Estimating the Euro-Bund Futures Price

Euro-bund futures have a contract value of €100,000, and the underlying consists of long-term German debt instruments with 8.5 to 10.5 years to maturity. They are traded on the Eurex. Suppose the underlying 2% coupon (semi-annual payment) German bund is quoted at €108 and has accrued interest of €0.083 (15 days since last coupon paid). The euro-bund futures contract matures in one month (30 days). At contract expiration, the underlying bund will have accrued interest of €0.25; there are no coupon payments due until after the futures contract expires; and the current one-month risk-free rate is 0.1%. The conversion factor is 0.729535.

In this case, we have the following:

T = 1/12, CF = 0.729535, B_0 = 108, FVCI = 0, AI_0 = (15/180 × 2%/2) = €0.083, AI_T = (45/180 × 2%/2) = €0.25, and r = 0.1%.

The equilibrium euro-bund quoted futures price (Q_0) based on the carry arbitrage model will be *closest* to:

A €147.57.

B €147.82.

C €148.15.

Solution:

B is correct. The carry arbitrage model for forwards and futures is simply the future value of the underlying with adjustments for unique carry features. With bond futures, the unique features include the conversion factor, accrued interest, and any coupon payments. Thus, the equilibrium euro-bund futures price can be found using the carry arbitrage model (Equation 8):

$$Q_0 = [1/CF]\{FV[B_0 + AI_0] - AI_T - FVCI\}.$$

Thus, we have:

$$Q_0 = [1/0.729535][(1 + 0.001)^{1/12}(108 + 0.083) - 0.25 - 0] = 147.82.$$

Note that the same result can be found by $Q_0 = F_0/CF$, where:

$$F_0 = FV(S_0) - AI_T - FVCI = (1 + 0.001)^{1/12}(108 + 0.083) - 0.25 - 0 = 107.84.$$

In equilibrium, the quoted euro-bund futures price should be approximately €147.82 based on the carry arbitrage model.

Because of the mark-to-market settlement procedure, the value of a bond future is essentially the price change since the previous day's settlement. That value is captured at the settlement at the end of the day, at which time the value of the bond futures contract, like other futures contracts, resets to zero.

We now turn to the task of estimating the fair value of the bond forward contract at a point in time during its life. Without daily settlement, the value of a forward is not formally realized until expiration. Suppose the first transaction is buying (at Time 0) an at-market bond forward contract priced at F_0 with expiration of Time T. Later (at Time t) consider selling a new bond forward contract priced at F_t, again with expiration of Time T. At the maturity of the forward contracts, we take delivery of the bond under the long forward and use it to make delivery under the short forward. Assuming the same underlying, there is no price risk. The net cash flow at maturity is the difference in the price at which we sold, F_t, and the price we agreed to pay, F_0, or $(F_t - F_0)$. To confirm the price risk on the underlying bond is zero, we could also add the values of the long and the short forward positions at expiration $V_{Long} + V_{Short} = (B_T - F_0) + (F_t - B_T) = F_t - F_0$. Since the position is riskless, the value to the long at time t should be:

$$V_t = \text{Present value of difference in forward prices at time t} = PV\,[F_t - F_0].$$

As a simple example of bond forward contract valuation, assume that two forward contracts have been entered as follows: long forward at $F_0 = 119.12$ and short forward at $F_t = 119.92$. Time t is one month before expiration, and both forward contracts expire at Time T. Therefore, time to expiration in one-month is $T - t = 1/12$. Finally, assume the appropriate interest rate for discounting is $r = 0.5\%$.

The forward value observed at Time t for the Time T maturity bond forward contracts is simply the present value of the difference in their forward prices —denoted $PV_{t,T}\,(F_t - F_0)$. That is, we have:

$$V_t = (119.92 - 119.12)/(1 + 0.005)^{1/12} = 0.7997.$$

EXAMPLE 12

Estimating the Value of a Euro-Bund Forward Position

Suppose that one month ago, we purchased *five* euro-bund forward contracts with two months to expiration and a contract notional value of €100,000 each at a price of 145 (quoted as a percentage of par). The euro-bund forward contract now has one month to expiration. The current annualized one-month risk-free rate is 0.1%. Based on the current forward price of 148, the value of the euro-bund forward position will be *closest* to:

A €2,190.

B €14,998.

C €15,012.

Solution:

B is correct. Because we are given both forward prices, the solution is simply the present value of the difference in forward prices at expiration.

$V_t = PV[F_t - F_0] = (148 - 145)/(1 + 0.001)^{1/12} = 2.99975.$

This is 2.9997 per €100 par value because this forward price was quoted as a percentage of par. Because five contracts each with €100,000 par were entered, we have 0.029997(€100,000)5 = €14,998.75. Note that when interest rates are low and the forward contract has a short maturity, then the present value effect is minimal (about €1.25 in this example).

We conclude this section with some observations on the similarities and differences between forward and futures contracts.

4.1 Comparing Forward and Futures Contracts

For every market considered here, the carry arbitrage model provides an approach for both pricing and valuing forward contracts. Recall the two generic expressions:

$F_0 = FV(S_0 + CC_0 - CB_0)$ (Forward pricing)

$V_t = PV[F_t - F_0]$ (Forward valuation)

Carry costs (CC) and financing costs increase the forward price, and carry benefits (CB) decrease the forward price. The arbitrageur is carrying the underlying, and costs increase the burden whereas benefits decrease the burden. The forward value can be expressed as either the present value of the difference in forward prices or as a function of the current underlying price adjusted for carry cash flows and the present value of the initial forward price.

Futures prices are generally found using the same model, but futures values are different because of the daily marking to market. Recall that the futures values are zero at the end of each trading day because profits and losses are taken daily.

In summary, the carry arbitrage model provides a compelling way to price and value forward and futures contracts. Stated concisely, the forward or futures price is simply the future value of the underlying adjusted for any carry cash flows. The forward value is simply the present value of the difference in forward prices at an intermediate time in the contract. The futures value is zero after marking to market. We turn now to pricing and valuing swaps.

PRICING AND VALUING SWAP CONTRACTS

5

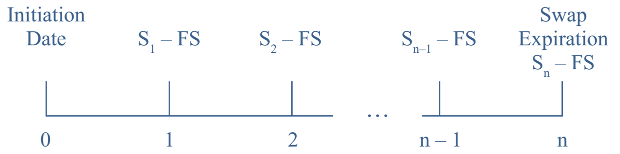
e describe how interest rate swaps are priced, and calculate and interpret their no-arbitrage value

Based on the foundational concepts we have studied on using the carry arbitrage model for pricing and valuing forward and futures contracts, we now apply this approach to pricing and valuing swap contracts.

A swap contract is an agreement to exchange (or swap) a series of cash flows at certain periodic dates. For example, an interest rate swap might exchange quarterly cash flows based on a floating rate for those based on a fixed rate. An interest rate swap is like an FRA except that it hedges multiperiod interest-rate risk, whereas an FRA only hedges single-period interest-rate risk. Similarly, in a currency swap the counterparties agree to exchange two series of interest payments, each denominated in a different currency, with the exchange of principal payments at inception and at maturity. Swap contracts can be synthetically created as either a portfolio of underlying instruments (such as bonds) or a portfolio of forward contracts (such as FRAs). Swaps are most easily understood as a portfolio of underlying bonds, so we will follow that approach.

Cash flows from a generic receive-floating and pay-fixed interest rate swap are shown in Exhibit 12. The cash flows are determined by multiplying a specified notional amount by a (fixed or floating) reference rate. In a fixed-for-floating interest rate swap (i.e., pay-fixed, receive-floating, also known as a plain vanilla swap), the fixed-rate payer in the swap would make a series of payments based on a fixed rate of interest applied to the notional amount. The counterparty would receive their fixed payments in return for making payments based on a floating rate applied to the same notional amount. The floating rate used as a reference will be referred to as the market reference rate (MRR). In our examples, we will use Libor as the MRR.

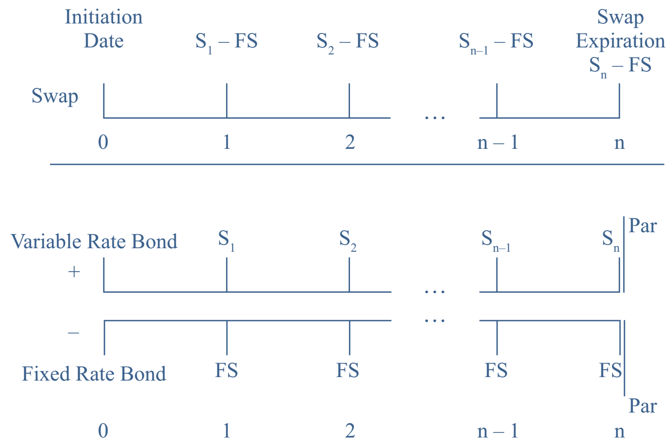
Exhibit 12 Generic Swap Cash Flows: Pay-Fixed, Receive-Floating



Our generic swap involves a series of n future cash flows at points in time represented simply here as 1, 2, ..., n. Let S_i denote the floating interest rate cash flow based on some underlying, and let FS denote the cash flow based on some fixed interest rate. Notice how the cash flows are netted. If the floating rate S_i increases above the agreed fixed rate FS, so $S_i > FS$, the fixed-rate payer (i.e., floating-rate receiver) will receive positive cash flow. If rates fall, so $S_i < FS$, the fixed-rate receiver (i.e., floating-rate payer) will receive the positive cash flow. We assume that the last cash flow occurs at the swap expiration. Later we will let S_i denote the floating cash flows tied to currency movements or equity movements.

We again will rely on the arbitrage approach for determining the pricing of a swap. This procedure involves finding the fixed rate such that the value of the swap at initiation is zero. Recall that the goal of the arbitrageur is to generate positive cash flows with no risk and no investment of one's own capital. To understand swap valuation, we match the swap cash flows by synthetically creating a replicating portfolio from other instruments. The swap must have the same value as the synthetic portfolio, or arbitrage will result. A pay-fixed, receive-floating swap is equivalent to a short position (i.e., issuer) in a fixed-rate bond and a long position (i.e., investor) in a floating-rate bond. Assuming both bonds were initially priced at par, the initial cash flows are zero and the par payments at maturity offset each other. In other words, the **swap rate** is the rate at which the present value of all the expected floating-rate payments received over the life of the floating-rate bond equal the present value of all the expected fixed-rate payments made over the life of the fixed-rate bond. Thus, the fixed bond payment should be equivalent to the fixed swap payment. Exhibit 13 shows the view of a swap as a pair of bonds. Note that the coupon dates on the bonds match the settlement dates on the swap, and the maturity date matches the expiration date of the swap. As with all derivative instruments, numerous technical details have been simplified here. We will explore some of these details shortly.

Exhibit 13 Receive-Floating, Pay-Fixed as a Portfolio of Bonds



It is worth noting that our replicating portfolio did not need to use a pair of bonds. Swaps can also be viewed as a portfolio of forward or futures contracts. However, in practice futures have standardized characteristics, so there is rarely a set of futures contracts that can perfectly replicate a swap. In addition, because a single forward contract can be viewed as a portfolio of a call and a put (a long call and a short put at the same strike price equal to the swap's fixed rate would replicate the payoffs on a pay-fixed swap), a swap can also be viewed as a portfolio of options. The procedure is fairly straightforward in all cases. Just match the swap cash flows with the cash flows from a portfolio of marketable underlying instruments and rely on the law of one price and the absence of arbitrage to provide a value. Again, bonds are perhaps the best instruments to replicate a swap because they are easy to value.

Market participants often use swaps to transform one series of cash flows into another. For example, suppose that because of the relative ease of issuance, REB, Inc. sells a fixed-rate bond to investors. Based on careful analysis of the interest rate sensitivity of the company's assets, REB's leadership deems a Libor-based variable rate bond to be a more appropriate liability. By entering a receive-fixed, pay-floating interest rate swap, REB can create a synthetic floating-rate bond, as illustrated in Exhibit 15. REB issues fixed-rate bonds and thus must make periodic fixed-rate-based payments to the bond investors, denoted FIX. REB then enters a receive-fixed (FIX) and pay-floating (FLT) interest rate swap. The two fixed-rate payments cancel, leaving on net the floating-rate payments. Thus, we say that REB has created a synthetic floating-rate bond.

Exhibit 14 REB's Synthetic Floating-Rate Bond Based on Fixed-Rate Bond Issuance with Receive-Fixed Swap

The example in Exhibit 14 is for a swap in which the underlying is an interest rate.

There are also currency swaps and equity swaps. Currency swaps can be used in a similar fashion, but the risks being addressed are both interest rate and currency exposures. Equity swaps can also be used in a similar fashion, but the risk being addressed is equity exposure.

Swaps have several technical nuances that can have a significant influence on pricing and valuation. Differences in payment frequency and day count methods often have a material impact on pricing and valuation. Another issue is identifying the appropriate discount rate to apply to the future cash flows. We turn now to examining three types of swap contracts—interest rate, currency, and equity—with a focus on pricing and valuation.

5.1 Interest Rate Swap Contracts

In this section we will focus on the pricing and valuing of interest rate swap contracts. Our approach will view a swap as a pair of bonds, a long position in one bond and a short position in another bond. At inception of a fixed-for-floating swap, a fixed rate is selected so that the present value of the floating-rate payments is equal to the present value of the fixed-rate payments, meaning the swap value is zero for both parties at inception. The fixed rate (FS) is the swap rate. Determining the swap rate is equivalent to pricing the swap. As the market rates change and time passes over the term of the swap, the value of the swap changes. The swap value (the value of the two constituent bonds) can be positive (an asset) or negative (a liability) to the pay-fixed or receive-fixed swap holders.

Swaps are OTC products with many variations. For example, a plain vanilla Libor-based interest rate swap can involve different frequencies of cash flow settlements and day count conventions. In fact, a swap can have both semi-annual payments and quarterly payments as well as actual day counts and day counts based on 30 days per month. Unless stated otherwise, we will assume for simplicity that the notional amounts are all equal to one (NA = 1). Swap values per 1 notional amount can be simply multiplied by the actual notional amount to arrive at the swap's fair market value.

Interest rate swaps have two legs, typically a floating leg (FLT) and a fixed leg (FIX). The floating leg cash flow—denoted S_i because the rate ($r_{FLT,i}$) may change (or float) during each period i—can be expressed as:

$$S_i = AP_{FLT} \times r_{FLT,i} = (NAD_{FLT}/NTD_{FLT}) \times r_{FLT,i}$$

and the fixed leg cash flow (denoted FS) can be expressed as:

$$FS = AP_{FIX} \times r_{FIX} = (NAD_{FIX}/NTD_{FIX}) \times r_{FIX}.$$

AP denotes the accrual period, $r_{FLT,i}$ denotes the observed floating rate appropriate for Time i, NAD denotes the number of accrued days during the payment period, NTD denotes the total number of days during the year applicable to each cash flow, and r_{FIX} denotes the fixed swap rate. The accrual period accounts for the payment frequency and day count methods. The two most popular day count methods are known as 30/360 and ACT/ACT. As the name suggests, 30/360 treats each month as having 30 days; thus, a year has 360 days. ACT/ACT treats the accrual period as having the actual number of days divided by the actual number of days in the year (365 or 366). Finally, the convention in the swap market is that the floating interest rate is assumed to be advanced set and settled in arrears; thus, $r_{FLT,i}$ is set at the beginning of the period and paid at the end. If we assume constant and equal accrual periods (so, $AP_{FLT} = AP_{FIX}$), the receive-fixed, pay-floating *net* cash flow can be expressed as:

$$FS - S_i = AP \times (r_{FIX} - r_{FLT,i}),$$

and the pay-fixed, receive-floating *net* cash flow can be expressed as:

$$S_i - FS = AP \times (r_{FLT,i} - r_{FIX}).$$

As a simple example, if the fixed rate is 5%, the floating rate is 5.2%, and the accrual period is 30 days based on a 360-day year, the payment of a receive-fixed, pay-floating swap is calculated as:

$$FS - S_i = (30/360) \times (0.05 - 0.052) = -0.000167 \text{ per notional of 1.}$$

Because the floating rate exceeds the fixed rate, the receive-fixed (pay-floating) party would pay this amount (0.000167 per notional of 1) to the pay-fixed (receive-floating) party. In other words, only a single net payment is made by the receive-fixed party to the counterparty. The sign of the net payment is negative as it is an outflow (i.e., negative cash flow) for the receive-fixed (pay-floating) party. Moreover, assuming the notional amount (NA) is £100 million, the net payment made by the receive-fixed party is £16,700 (= –0.000167 x £100,000,000). Finally, if, instead, the fixed rate exceeds the floating rate, the sign of the net payment would be positive as it would be an inflow (i.e., positive cash flow) to the receive-fixed party from the pay-fixed counterparty.

We now turn to swap pricing. Exhibit 15 shows the cash flows for an interest rate swap along with a pair of bonds of equal par value. Suppose (at Step 1) the arbitrageur enters a receive-fixed, pay-floating interest rate swap with some initial value, V_{swap}. Replicating this swap with bonds would entail being long a fixed-rate bond (as the arbitrageur is receiving the fixed-rate coupon) and short a floating-rate bond (as she is paying the floating rate). Therefore, to price this swap, the arbitrageur creates the *opposite* of the replicating portfolio. So, at Step 2 she purchases a floating-rate bond whose value is denoted V_{FLT}. Note that the terms of the variable rate bond are selected to match exactly the floating payments of the swap. Next, a fixed-rate bond is sold short (Step 3)—equivalent to borrowing funds—with terms to match exactly the fixed payments of the swap.

Exhibit 15 Cash Flows for Receive-Fixed, Pay-Floating Swap Offset with Bonds

Position	Step	Time 0	Time 1	Time 2	…	Time n
Swap	Receive-fixed, pay-floating swap	V_{swap}	$+FS - S_1$	$+FS - S_2$	…	$+FS - S_n$
Offsetting Portfolio	Buy floating-rate bond	$-V_{FLT}$	$+S_1$	$+S_2$	…	$+S_n$ + Par
	Short-sell fixed-rate bond	$+V_{FIX}$	–FS	–FS	…	–(FS + Par)
	Net Cash Flows	$\mathbf{V_{swap} = -V_{FLT} + V_{FIX} = 0}$	**0**	**0**	**0**	**0**

This portfolio offsets the cash flows from the swap, so the net cash flows from Time 1 to Time n will all be equal to zero. So, in equilibrium we must have $V_{swap} = -V_{FLT} + V_{FIX} = 0$ to prevent an arbitrage opportunity. The value of a receive-fixed, pay-floating swap is:

$$V_{swap} = \text{Value of fixed bond} - \text{Value of floating bond} = V_{FIX} - V_{FLT}. \quad (9)$$

The value of a receive-fixed, pay-floating interest rate swap is simply the value of buying a fixed-rate bond and issuing (i.e., selling) a floating-rate bond. Remember, the fixed-rate and floating-rate bond values are just the PVs of all the expected interest and par payments. Pricing the swap means to determine the fixed rate (r_{FIX}) such that the value of the swap at initiation is zero. Said differently, to price the swap, the value of the fixed bond must equal the value of the floating bond in Equation 9.

As stated earlier, the value of a fixed bond (V_{FIX}) is the sum of the PV(All coupons) + PV(Par). If C is the coupon amount and par is 1, the value of a fixed-rate bond is, V_{FIX} = sum of PV of all coupons (C) + PV of par value, or:

$$\text{Value fixed bond rate: } V_{FIX} = C\sum_{i=1}^{n}\text{PV}_i(1) + \text{PV}_n(1). \quad (10)$$

Notice the coupon amount in Equation 10 is multiplied by a summation term. This term includes the present value discount factors, PV(1), for each cash flow (or coupon payment). These PV factors are derived from the term structure of interest rates at the time of valuation. The summation adds up the PV factor for each coupon as it sequentially occurs. The sum of the PV of all the coupons is added to the PV of par at maturity (Time n). The present value expression is based on spot rates and is computed using the formula, $\text{PV}_i(1) = \dfrac{1}{1 + Rspot_i\left(\dfrac{\text{NAD}_i}{\text{NTD}}\right)}$. Spot interest rates ($Rspot_i$) will help us value each individual cash flow. As an illustration, consider the following term structure of rates for USD cash flows and the computation of their associated PV factors, as shown in Exhibit 16:

Exhibit 16 Present Value Factors Using the Term Structure

Days to Maturity	US$ Spot Interest Rates (%)	Present Value (US$1)
90	2.10	0.994777
180	2.25	0.988875
270	2.40	0.982318
360	2.54	0.975229
	Sum:	3.941199

The PV factors are computed for each rate in the term structure as:

$$\text{PV}_i(1) = \frac{1}{1 + Rspot_i\left(\dfrac{\text{NAD}_i}{\text{NTD}}\right)}$$

Using this formula, we compute the PV factor for a unit cash flow of 1. For example, at 90 days, we have a spot rate of 2.10%, which implies a discount (PV) factor of 0.994777 = $1/[1 + 0.0210 (90/360)]. Similarly, for 360 days, we have a spot rate of 2.54%, which implies a PV factor of 0.975229 = 1/[1 + 0.0254(360/360)].

The present value factors make it straightforward to value a fixed-rate bond under a given term structure. For example, the value of a fixed 4% bond with quarterly interest payments and Par = 1 under the term structure in Exhibit 16 can be computed using Equation 10. The quarterly coupon payment, C, is 4%/4 on par of 1 or 0.01/quarter.

$$V_{FIX} = C\sum_{i=1}^{n}\text{PV}_i(1) + \text{PV}_n(1) = 0.01(3.941199) + 0.975229(1) = 1.014641.$$

So, using Equation 10 and the PV factors and their sum from Exhibit 16, we can quickly value the bond at 101.464% of par.

To find the fixed rate needed to price a swap, we first make a slight modification to the notation in Equation 10. Since the coupon C is just the fixed interest rate multiplied by Par (and Par is assumed to be 1), we can substitute r_{FIX} = C, so that:

$$V_{FIX} = r_{FIX}\sum_{i=1}^{n}\text{PV}_i(1) + \text{PV}_n(1).$$

The value of a floating-rate bond, V_{FLT}, at the reset date is 1 (par) because the interest payment is set to match the discount rate. Recall that when the YTM (discount rate) of a bond is equal to the coupon rate, the bond sells at par. Here, we assume par is 1. Because the floating rate and the discount rate are initially the same for our floating bond, at the reset date we have V_{FLT} = par = 1.

Setting the value of the fixed bond in Equation 10 equal to 1 (the value of the floating bond at swap initiation, so $V_{FIX} = 1 = V_{FLT}$), we obtain:

$$V_{FIX} = r_{FIX}\sum_{i=1}^{n}\text{PV}_i(1) + \text{PV}_n(1) = 1.$$

This expression leads to the swap pricing equation, which sets r_{FIX} for the fixed bond:

$$r_{FIX} = \frac{1 - \text{PV}_n(1)}{\sum_{i=1}^{n}\text{PV}_i(1)} \text{ (Swap Pricing Equation).} \tag{11}$$

The fixed swap rate, the "price" that swap traders quote among one another, is simply one minus the *final* present value term divided by the sum of present values. The fixed swap leg cash flow (FS) for a unit of notional amount (NA) is simply the fixed swap rate adjusted for the accrual period, or:

FS = $AP_{FIX} \times r_{FIX}$ (Fixed swap cash flow per unit of NA).

We can multiply FS times the notional amount later to find the cash flow for a swap in practice.

EXAMPLE 13

Solving for the Fixed Swap Rate Based on Present Value Factors

Suppose we are pricing a five-year Libor-based interest rate swap with annual resets (30/360 day count). The estimated present value factors, PV_i (1), are given in the following table.

Maturity (years)	Present Value Factors
1	0.990099
2	0.977876
3	0.965136
4	0.951529
5	0.937467

The fixed rate of the swap (r_{FIX}) will be *closest* to:

A 1.0%.

B 1.3%.

C 1.6%.

Solution:

B is correct. Note that the sum of present values is:

$$\sum_{i=1}^{5} PV_i(1) = 0.990099 + 0.977876 + 0.965136 + 0.951529 + 0.937467 = 4.822107.$$

Since the final cash flow for a bond consists of the n^{th} coupon plus par, we use the PV factor for the last cash flow, here cash flow 5, twice in Equation 11. We sum it with the other PV factors for the individual coupons in the denominator, and we apply it to Par in the numerator. Therefore, the solution for the fixed swap rate is:

$$r_{FIX} = \frac{1 - 0.937467}{4.822107} = 0.012968, \text{ or } 1.2968\%.$$

From pricing a swap in Example 13, we now turn to interest rate swap valuation for a receive fixed (pay floating) swap. As noted previously, the fixed-rate receiver is effectively long a fixed bond and short a floating-rate bond. After initiation, this position will have a positive value when the fixed bond is trading at a premium to par (i.e., interest rates have fallen).

At any time after initiation, the market value of an existing swap can be understood by pricing a new offsetting swap. Assume $r_{FIX,0}$ is the swap rate at initiation. After initiation, the term structure of interest rates will likely imply a different swap rate, $r_{FIX,t}$.

The approach to value a multi-period swap is like the approach to valuing a single period FRA (i.e., multiplying the PV of the difference between the old FRA and the new FRA rates by a notional amount; Equation 6). Valuation is based on arbitrage transactions. Our initial swap position at Time 0 as a floating-rate payer would be offset by a position at Time t as a floating-rate receiver. The floating cash flows from paying and receiving will offset at each date (i), but the fixed payments will be different. We still receive the fixed rate, $r_{FIX,0}$, initially agreed to, but for the purposes of valuation we additionally assume the role as a fixed-rate payer at the new rate, $r_{FIX,t}$. The cash flows per unit of NA at each future date will always be based on the difference between the rate we initially received at Time 0 and the current rate paid at Time t, so $(FS_0 - FS_t) = AP(r_{FIX,0} - r_{FIX,t})$. Thus, the value of a *receive-fixed swap* at some future point in Time (t) is simply the sum of the present values of the difference in fixed swap rates times the stated notional amount (NA), or:

$$V_{SWAP,t} = NA \times (FS_0 - FS_t) \times \sum_{i=1}^{n} PV_i \text{ (Value of receive-fixed swap).} \quad (12)$$

In our valuation equation, n is the number of remaining cash flows from time t. Although this n may be different than the number of cash flows initially used to price the swap at time 0, we use the same notation. It is also important to be clear on which side of the swap this value applies. Notice the cash flow FS_0 in Equation 12 is positive. This is because the swap was initially set up (at Time 0) as a receive-fixed (FS_0), pay-floating swap. To establish a value, the swap is offset with a pay-fixed, receive-floating swap at Time t. Thus, when FS_0 has a positive sign, Equation 12 provides the value to the party initially receiving fixed. The negative of this amount is the value to the fixed-rate payer.

Now, since the *fixed-rate payer is effectively long a floating bond and short a fixed bond,* the position will have positive value when the fixed bond is trading at a discount to par (i.e., interest rates have risen). The fixed-rate payer is also the floating receiver and thus benefits as interest rates rise. At any date, the market value of a swap to the *fixed-rate payer* is based on the present value of the difference between the new offsetting fixed cash flow FS_t to be received and the fixed cash flow FS_0 he or she originally agreed to pay. It will be the negative of the receive-fixed swap value ($V_{SWAP,t}$) given by Equation 12, and we can compute it as follows:

$$-V_{SWAP,t} = -1\left[\text{NA} \times \left(\text{FS}_0 - \text{FS}_t\right) \times \sum_{i=1}^{n} \text{PV}_i\right]$$

$$= \text{NA} \times \left(\text{FS}_t - \text{FS}_0\right) \times \sum_{i=1}^{n} \text{PV}_i (\text{Value of pay-fixed swap}). \quad \text{(12a)}$$

Exhibit 17 provides a summary of the swap legs and the associated replicating and offsetting portfolios for each swap leg. The replicating portfolio (at time 0) provides the same cash flows as our swap. The offsetting portfolio (at time t) will offset the cash flows from our replication of the swap and help us determine a value. Note that the floating cash flows at Time 0 and Time t cancel each other out. For valuation purposes, this allows us to focus on the difference in fixed swap rates. So, the value of a receive-fixed swap at time t is based on the difference between the initial fixed swap rate and the fixed swap rate at time t, or $r_{FIX,0} - r_{FIX,t}$, as shown in the last row of Exhibit 17.

Exhibit 17 Swaps and Related Replicating and Offsetting Portfolios

		Receive-Fixed, Pay-Floating			Pay-Fixed, Receive-Floating		
Swap		**Portfolio Position**		**Rates**	**Portfolio Position**		**Rates**
Replicating Portfolio	Initiation t = 0	Long Fixed-Rate Bond	Short Floating-Rate Bond	$r_{FIX,0} - r_{FLT,0}$	Long Floating-Rate Bond	Short Fixed-Rate Bond	$r_{FLT,0} - r_{FIX,0}$
Offsetting Portfolio	Time = t	Short Fixed-Rate Bond	Long Floating-Rate Bond	$r_{FLT,t} - r_{FIX,t}$	Short Floating-Rate Bond	Long Fixed-Rate Bond	$r_{FIX,t} - r_{FLT,t}$
Rates for Swap Valuation	Time = t			$r_{FIX,0} - r_{FIX,t}$			$r_{FIX,t} - r_{FIX,0}$

The examples illustrated here show swap valuation only on a payment date. If a swap is being valued between payment dates, some adjustments are necessary. We do not pursue this topic here.

EXAMPLE 14

Solving for Receive-Fixed Swap Value Based on Present Value Factors

Suppose two years ago we entered a €100,000,000 seven-year receive-fixed Libor-based interest rate swap with annual resets. The fixed rate in the swap contract entered two years ago was 2.0%. The estimated present value factors, $PV_i(1)$, are repeated from the previous example.

Maturity (years)	Present Value Factors
1	0.990099
2	0.977876
3	0.965136
4	0.951529
5	0.937467
Sum	4.822107

We know from the previous example that the current equilibrium fixed swap rate is close to1.30% (two years after the swap was originally entered).

1 The value for the swap party receiving the fixed rate will be *closest* to:

A –€5,000,000.

B €3,375,000.

C €4,822,000.

2 The value for the swap party paying the fixed rate will be *closest* to:

A –€4,822,000.

B –€3,375,000.

C €5,000,000.

Solution to 1:

B is correct. $r_{FIX,0} = 2.0\%$, and $r_{FIX,t} = 1.3\%$. We assume annual resets (AP = 360/360 = 1), so the cash flow per unit notional is $\text{FS}_0 = 2.0\%$ and $FS_t = 1.3\%$.

The swap value to the fixed-rate receiver is:

$$\begin{aligned} V_{SWAP,t} &= \text{NA} \times \left(\text{FS}_0 - \text{FS}_t\right) \times \sum_{i=1}^{5} \text{PV}_i \\ &= €100{,}000{,}000 \times \left(0.02 - 0.013\right) \times 4.822107 = €3{,}375{,}000. \end{aligned}$$

Solution to 2:

B is correct. The equivalent pay-fixed swap value is simply the negative of the receive-fixed swap value:

$$\begin{aligned} -V_{SWAP,t} &= \text{NA} \times \left(\text{FS}_t - \text{FS}_0\right) \times \sum_{i=1}^{5'} \text{PV}_i \\ &= €100{,}000{,}000 \times \left(0.013 - 0.02\right) \times 4.822107 \\ &= -€3{,}375{,}000. \end{aligned}$$

6 PRICING AND VALUING CURRENCY SWAP CONTRACTS

f describe how currency swaps are priced, and calculate and interpret their no-arbitrage value;

A currency swap is a contract in which two counterparties agree to exchange future interest payments in different currencies. In a currency swap, one party is long a bond (fixed or floating) denominated in one currency and short a bond (fixed or floating) in another currency. The procedure for pricing and valuing currency swaps is like the pricing and valuation of interest rate swaps. Currency swaps come in a wide array of types and structures. We review a few key features:

1. Currency swaps often involve an exchange of notional amounts at both the initiation of the swap and at the expiration of the swap.
2. The payment on each leg of the swap is in a different currency unit, such as euros and Japanese yen, and the payments are not netted.
3. Each leg of the swap can be either fixed or floating.

Pricing a currency swap involves solving for three key variables: two fixed interest rates (each in a different currency) and one notional amount. We must determine the appropriate notional amount in one currency, given the notional amount in the other currency, as well as two fixed-interest rates such that the currency swap value is zero at initiation.

We will focus on fixed-for-fixed currency swaps, so we essentially trade cash flows on a bond in one currency for cash flows on a bond in another currency. Let k be the currency units, such as euros and yen. Letters are used here rather than numbers to avoid confusion with calendar time. The value of a fixed-rate bond in currency k with par of 1 can be expressed generically as the present value of the coupons plus the present value of par, or:

$$V_k = C_k \sum_{i=1}^{n} \mathrm{PV}_i(1) + \mathrm{PV}_n(\mathrm{Par}_k).$$

C_k is the coupon in currency k, and Par_k is the Par value paid at maturity in currency k. The value of a fixed-for-fixed currency swap, $V_{CS,}$ is the difference in the price of two bonds. That is, the value of a currency swap is simply the value of a bond in currency a (V_a) less the value of a bond in currency b (V_b), expressed in terms of currency a, as follows:

$$V_{CS} = V_a - S_0 V_b.$$

Here, S_0 is the spot exchange rate at time 0. To make each party indifferent between the two bonds, the par or principal notional amounts are set to reflect the current spot exchange rate. This will lead to the swap having zero value ($V_{CS} = 0$) at inception (to prevent any arbitrage opportunity), so

$$V_a = S_0 V_b.$$

The swap value may change after initiation as the exchange rate and interest rates on the two currencies fluctuate. Currency swap valuation is best understood by considering an example. Exhibit 18 provides an illustration of an at-market 10-year receive-fixed US$ and pay-fixed € swap. The US$ bond has an annual coupon of US$30 and par of

US$1,150. The annual coupon amount of the euro-denominated bond is €9 with par of €1,000. Both bonds are assumed to be trading at par (note, this is $1,150 for the US$ bond, not the usual $1,000) and have a 10-year maturity. We proceed as follows:

- Step 1: We enter the receive-fixed US$ and pay-fixed € swap.

  In Steps 2 and 3, we create a portfolio to offset the swap cash flows.
- Step 2 involves short-selling a US bond (so, paying the fixed US$ coupon on the bond) to offset the US dollar inflows from the swap.
- Step 3 involves purchasing a euro bond (so, receiving the fixed € coupon on the bond), which provides offsetting cash flows for the pay-fixed € portion of the swap.

Exhibit 18 Numerical Example of Currency Swap Offset with Bonds

Position	Step	Time 0	Time 1	Time 2	…	Time 10
Swap	1. Receive-fixed US$, pay-fixed euro swap	**0**	+$30 – ($1.5/€) x €9 = **+$16.5**	+$30 – ($1.1/€) x €9 = **+$20.1**	…	+($30 + $1,150) – ($1.2/€) x (€9 + €1,000) = **–$30.8**
Offsetting Bond Portfolio	2. Short-sell US$ bond	+$1,150	–$30	–$30	…	–($30 + $1,150)
	3. Buy euro bond	–($1.15/€) x €1,000 = –$1,150	+($1.5/€) x €9 = +$13.5	+($1.1/€) x €9 = +$9.9	…	+($1.2/€) x (€9 + €1,000) = $1,210.8
Offsetting Portfolio Cash Flows		**0**	**–$16.5**	**–$20.1**	…	**+$30.8**
Overall Net Cash Flows		**0**	**0**	**0**	**0**	**0**

The cash flows from the bond portfolio will exactly offset the cash flows from the swap. This illustration assumes a current spot exchange rate (S_0) at which €1 trades for US$1.15, so S_0 = $1.15/€1. Selected future spot exchange rates are S_1 = $1.50/€1, S_2 = $1.10/€1, and S_{10} = $1.20/€1. These future spot exchange rates are used to show the conversion of future euro cash flows into US dollars, but notice that the overall net cash flows are all zero regardless of the future spot exchange rates. In other words, we could have used any numbers for S_1, S_2, and S_{10}. Regardless of exchange rates in the future, the bond portfolio and the swap always have offsetting cash flows. Since the portfolio and swap produce identical (although opposite) cash flows, the law of one price will allow us to determine a value for our swap in terms of a pair of bonds.

Since the net cash flows are 0 at every time t, the portfolio must be worth 0 initially. Exhibit 18 provides the intuition for solving for the notional amount (NA). For a zero cash flow at initiation, the NA (or par value) of the bond denominated in currency a (NA_a) must equal the spot exchange S_0 rate times the notional amount (or par value) of the bond denominated in currency b (NA_b). That is,

$$NA_a = S_0 \times NA_b.$$

The exchange rate is stated as number of units of currency a to buy one unit of currency b. The spot exchange rate in Exhibit 18 is \$1.15/€1, so currency a (in the numerator) is US\$. At the prevailing exchange rate S_0, it takes \$1.15 to buy one euro. NA_a = \$1,150 and S_0 = \$1.15/€1, so NA_b = \$1,150/(\$1.15/€1) = €1,000. Therefore, the swap value at initiation is equal to zero, as it should be:

$$V_{CS} = V_a - S_0 V_b = \$1{,}150 - (\$1.15/€1) \times €1{,}000 = 0.$$

At any time during the life (tenor) of the swap shown in Exhibit 18, the opposite cash flows from the offsetting bond transactions result in a zero net cash flow. If the initial swap value is not at market or zero, then there are arbitrage opportunities. If the initial swap value is positive, then a set of arbitrage transactions would be implemented to capture the initial value with no net cash outflow. If the initial swap value is negative, then the opposite set of transactions would be implemented. The goal is to determine the fixed rates of the swap such that the current swap value is zero.

Because the fixed swap rate does not depend on the notional amounts, the fixed swap rates are found in the same manner as the fixed swap rate in an interest rate swap. For emphasis, we repeat the equilibrium fixed swap rate equations for each currency:

$$r_a = \frac{1 - PV_{n,a}(1)}{\sum_{i=1}^{n} PV_{i,a}(1)} \text{ and } r_b = \frac{1 - PV_{n,b}(1)}{\sum_{i=1}^{n} PV_{i,b}(1)} \tag{13}$$

We now have a solution for each of the three swap variables: one notional amount ($NA_a = S_0 \times NA_b$) and two fixed interest rates from Equation 13. Again, the fixed swap rate in each currency is simply one minus the final present value term divided by the sum of present values. We need to be sure that the present value terms are expressed in the appropriate currency. We illustrate currency swap pricing with spot rates by way of an example.

BEGIN BOX

EXAMPLE 15

Currency Swap Pricing with Spot Rates

A US company needs to borrow 100 million Australian dollars (A\$) for one year for its Australian subsidiary. The company decides to issue US\$-denominated bonds in an amount equivalent to A\$100 million. Then, the company enters into a one-year currency swap with quarterly reset (30/360 day count) and the exchange of notional amounts at initiation and at maturity. At the swap's initiation, the US company receives the notional amount in Australian dollars and pays to the counterparty, a swap dealer, the notional amount in US dollars. At the swap's expiration, the US company pays the notional amount in Australian dollars and receives from the counterparty the notional amount in US dollars. Based on interbank rates, we observe the following spot rates today, at Time 0, and compute their PV factors and sums:

Days to Maturity	A\$ Spot Interest Rates (%)	Present Value (A\$1)	US\$ Spot Interest Rates (%)	Present Value (US\$1)
90	2.50	0.993789[a]	0.10	0.999750
180	2.60	0.987167	0.15	0.999251[b]
270	2.70	0.980152	0.20	0.998502

360	2.80	0.972763	0.25	0.997506
	Sum:	*3.933870*	*Sum:*	*3.995009*

[a] A$0.993789 = 1/[1 + 0.0250(90/360)].
[b] US$0.999251 = 1/[1 + 0.00150(180/360)].

Assume that the counterparties in the currency swap agree to an A$/US$ spot exchange rate of 1.140 (expressed as number of Australian dollars for US$1).

1 The annual fixed swap rates for Australian dollars and US dollars, respectively, will be *closest* to:

A 2.80% and 0.10%.

B 2.77% and 0.25%.

C 2.65% and 0.175%.

2 The notional amount (in US$ millions) will be *closest* to:

A 88.

B 100.

C 114.

3 The fixed swap quarterly payments in the currency swap will be *closest* to:

A A$692,000 and US$55,000.

B A$220,000 and US$173,000.

C A$720,000 and US$220,000.

Solution to 1:

B is correct. Since the PV factors are given, we do not need to compute them from the spot rates. Using Equation 13, the Australian dollar periodic fixed swap rate is:

$$r_{AUD} = \frac{1 - PV_{n,AUD}(1)}{\sum_{i=1}^{4} PV_{i,AUD}(1)} = \frac{1 - 0.972763}{3.933870}$$

$$= 0.00692381 \text{ or } 0.692381\%.$$

The US dollar periodic fixed swap rate is:

$$r_{USD} = \frac{1 - PV_{n,USD}(1)}{\sum_{i=1}^{4} PV_{i,USD}(1)} = \frac{1 - 0.997506}{3.995009}$$

$$= 0.00062422 \text{ or } 0.062422\%.$$

The annualized rate is simply (360/90) times the period results: 2.7695% for Australian dollars and 0.2497% for US dollars.

Solution to 2:

A is correct. The US dollar notional amount is calculated as A$100 million divided by the current spot exchange rate, A$1.140/US$1. From $NA_a = S_0 \times NA_b$, we have A$100,000,000 = A$1.14/US$1 × N_b. Solving for N_b we have US$87,719,298 = A$100,000,000/(A$1.14/US$1).

Solution to 3:

A is correct. The fixed swap quarterly payments in currency units equal the *periodic* swap rate times the appropriate notional amounts. From the answers to 1 and 2, we have

$$FS_{A\$} = NA_{A\$} \text{ x (AP) x } r_{A\$}$$

$$= A\$100{,}000{,}000 \text{ x } (90/360) \text{ x } (0.027695)$$
$$= A\$692{,}375$$

and

$$FS_{US\$} = NA_{US\$} \text{ x (AP) x } r_{US\$}$$

$$= US\$87{,}719{,}298 \text{ x } (90/360) \text{ x } (0.002497)$$
$$= US\$54{,}759.$$

One approach to pricing currency swaps is to view the swap as a pair of fixed-rate bonds. The main advantage of this approach is that all foreign exchange considerations are moved to the initial exchange rate. We do not need to address future foreign currency transactions. Also, note that a fixed-for-floating currency swap (i.e., pay-fixed currency a, receive-floating currency b) is simply a fixed-for-fixed currency swap (i.e., pay-fixed currency a, receive-fixed currency b) paired with a fixed-for-floating interest rate swap (i.e., pay-fixed currency b, receive-floating currency b). Also, we do not technically "price" a floating-rate swap because we do not designate a single coupon rate and because the value of such a swap is par on any reset date. Thus, we have the capacity to price any variation of currency swaps.

We now turn to currency swap valuation. Recall that with currency swaps, there are two main sources of risk: interest rates associated with each currency and their exchange rate. The value of a fixed-for-fixed currency swap at some future point in time, say Time t, is simply the difference in a pair of fixed-rate bonds, one expressed in currency a and one expressed in currency b. To express the bonds in the same currency units, we convert the currency b bond into units of currency a through a spot foreign exchange transaction at a new rate, S_t. The value of a "receive currency a, pay currency b" (fixed-for-fixed) swap at any time t expressed in terms of currency a is the difference in bond values:

$$V_{CS} = V_a - S_t V_b.$$

Substituting the valuation equation for each of the bonds, we have:

$$V_{CS} = \left(FS_a \sum_{i=1}^{n} PV_i(1) + NA_a\, PV_n(1)\right) - S_t\left(FS_b \sum_{i=1}^{n} PV_i(1) + NA_b\, PV_n(1)\right)$$

Note that the fixed swap amount (FS) is the per-period fixed swap rate times the notional amount. Therefore, the currency swap valuation equation can be expressed as:

$$V_{CS} = NA_a\left(r_{Fix,a} \sum_{i=1}^{n} PV_i(1) + PV_n(1)\right) - S_t NA_b\left(r_{Fix,b} \sum_{i=1}^{n} PV_i(1) + PV_n(1)\right). \quad \textbf{(14)}$$

As mentioned, the terms in Equation 14 represent the difference in value of two fixed-rate bonds. The first term in braces is the value of a long position in a bond with face value of 1 unit of currency a, which is then multiplied by the notional amount of the swap in currency a (NA_a). This product represents the value of the cash inflows to the counterparty receiving interest payments in currency a. The second term (after the minus sign) implies outflows and represents the value of a short bond position with face value of 1 unit of currency b, which is multiplied by the product of the swap notional amount in currency b (NA_b) and the current (Time t) exchange rate,

S_t (stated in units of currency a per unit of currency b). This gives the value of the payments, in currency a terms, made by the party receiving interest in currency a and paying interest in currency b. V_{CS} is then the value of the swap to the party receiving currency a, while the value of the swap to the party receiving currency b is simply the negative of that amount, $-V_{CS}$.

Equation 14 seems formidable, but it is a straightforward idea. We hold a bond in currency a, and we are short a bond in currency b (which we must express in terms of currency a). It is best understood by an example of a firm that has entered a currency swap and needs to determine the current value.

Example 16 continues the case of the company using a currency swap to effectively convert a bond issued in US dollars into a bond issued in Australian dollars. In studying the problem, take care to identify currency a (implied by how the exchange rate, S_t, is given) and the party receiving interest payments in currency a in the swap.

EXAMPLE 16

Currency Swap Valuation with Spot Rates

This example builds on the previous example addressing currency swap pricing. Recall that a US company needed to borrow 100 million Australian dollars (A$) for one year for its Australian subsidiary. The company decided to borrow in US dollars (US$) an amount equivalent to A$100 million by issuing US-denominated bonds. The company then entered a one-year currency swap with a swap dealer. The swap uses quarterly reset (30/360 day count) and exchange of notional amounts at initiation and at maturity. At the swap's expiration, the US company pays the notional amount in Australian dollars and receives from the dealer the notional amount in US dollars. The fixed rates were found to be 2.7695% for Australian dollars and 0.2497% for US dollars. Initially, the notional amount in US dollars was determined to be US$87,719,298 with a spot exchange rate of A$1.14 for US$1.

Assume 60 days have passed since swap initiation and we now observe the following updated market information:

Days to Maturity	A$ Spot Interest Rates (%)	Present Value (A$1)	US$ Spot Interest Rates (%)	Present Value (US$1)
30	2.00	0.998336	0.50	0.999584
120	1.90	0.993707	0.40	0.998668
210	1.80	0.989609	0.30	0.998253
300	1.70	0.986031	0.20	0.998336
	Sum:	3.967683	*Sum:*	3.994841

The currency spot exchange rate (S_t) is now A$1.13 for US$1.

1 The current value to the swap dealer in A$ of the currency swap entered 60 days ago will be *closest* to:

A –A$13,557,000.

B A$637,620.

C A$2,145,200.

2 The current value to the US firm in US$ of the currency swap entered 60 days ago will be *closest* to:

A –$2,673,705.

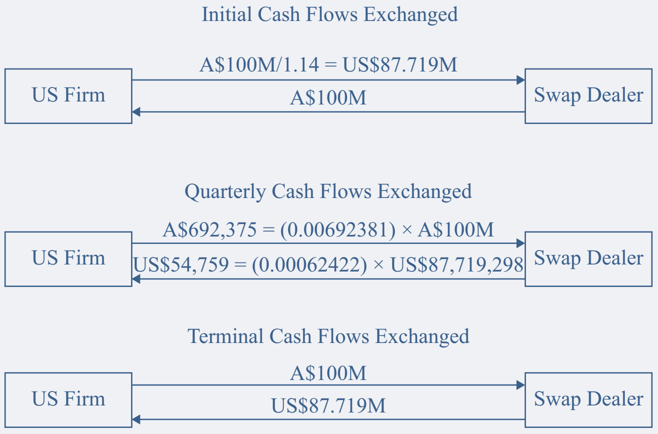
B −$1,898,400.

C $334,730.

Solution to 1:

C is correct. The US firm issues $87.7 million of bonds and enters a swap with the swap dealer. The initial exchange rate is given as 1.14A$/1US$, so currency a is A$. The swap dealer is receiving quarterly interest payments in currency a (A$). The swap is diagrammed for Example 15 and 16 as shown below:

Swap Cash Flows:



After 60 days, the new exchange rate is 1.13A$/1US$ and the term structure of interest rates has changed in both markets. Equation 14 gives the value of the swap at Time t, $V_{CS.}$ This is the value of the swap to the party receiving interest payments in Australian dollars, which is the swap dealer. Thus, using Equation 14, the value to the swap dealer receiving A$ is:

$$V_{CS} = NA_a\left(r_{Fix,a}\sum_{i=1}^{n}PV_i(1) + PV_n(1)\right) - S_t NA_b\left(r_{Fix,b}\sum_{i=1}^{n}PV_i(1) + PV_n(1)\right)$$

$$\begin{aligned} V_{CS} &= \text{A\$}100{,}000{,}000 \times [0.00692381\ (3.967683) + 0.986031] - 1.13 \\ &\quad (\text{A\$}/1\text{US\$}) \times (\text{US\$}87{,}719{,}298) \times [0.00062422\ (3.994841) + 0.998336] \\ &= \text{A\$}2{,}145{,}203. \end{aligned}$$

The first term in Equation 14 represents the PV of the dealer's incoming cash flows in A$, effectively a long position in an A$ bond. Remember, the dealer is receiving quarterly interest payments in A$ and will receive the A$100M terminal payment at swap maturity. To compute the PV of the A$ cash flows, the notional amount is multiplied by a term inside the braces, which represents the periodic interest rate multiplied by the sum of the PV factors for the four payments plus the PV factor for the terminal cash flow (where the PV factors reflect the new term structure). The second term is the PV of the dealer's US$ outflows (effectively a short bond in currency b, here US$). The PV of the quarterly interest payments and terminal payment are calculated using the new term structure and converted into A$ at S_t. Thus, we have the value of the long A$ bond minus the value of short US$ bond (stated in A$ terms). This gives V_{CS}, which is the value of the swap to the party receiving currency a and is the value from the perspective of the swap dealer.

Solution to 2:

B is correct. In terms of Solution 1, the current value of the swap to the US firm is $-V_{CS}$. This represents the value to the firm making interest payments in currency a (A$).

$-V_{CS}$ = −A$2,145,203, which when converted to US$ at S_t is:
$-V_{CS}$ = −A$2,145,203 × (1US$/1.13A$) = −US$1,898,410.

Note that the US company initially issues a bond in US$ in their home market and uses the swap to effectively convert to an A$ bond issue. Understanding the swap as two bonds, the US firm is long a US$ bond (US$ is currency b in this example, which the US firm is receiving) and short a bond in A$ (currency a, which the US firm is paying). The swap offsets the US firm's US$ bond issue. The swap allows the US firm to make A$ interest payments to the swap dealer, or to effectively issue a bond in A$ (currency a).

Alternatively, if the exchange rate had been stated as S_t = 1US$/1.13A$ or equivalently as S_t = $0.885/A$, then currency a would be US$. In that case, the swap value, V_{CS}, can be understood in terms of the firm receiving US$ since the swap gives the US firm the equivalent of a long position in a US$ bond. The first term in the following equation represents the value of the US$ bond to the US firm in the swap. The second term is the value of the A$ bond position (short for the US firm) expressed in US$ terms.

$$V_{CS} = NA_a\left(r_{Fix,a}\sum_{i=1}^{n}PV_i(1) + PV_n(1)\right) - S_t NA_b\left(r_{Fix,b}\sum_{i=1}^{n}PV_i(1) + PV_n(1)\right)$$

V_{CS} = $87,719,298 x [0.00062422 (3.994841) + 0.998336] − (1US$/A$1.13) x (A$100,000,000) x [0.00692381 (3.967683) + 0.986031]
= −US$1,898,410.

The swap value is negative to the US firm due to changes in the term structure and exchange rate. The A$ has strengthened against the US$, so now the US firm must pay periodic interest and principal cash flows in A$ at a rate of 1.13A$/1US$. That is, for each US$ the US firm gets fewer A$ for making payments to the dealer. The new term structure now offers lower interest rates to A$ borrowers, and this also contributes to the negative swap value for the US firm. The firm had agreed to pay higher periodic A$ rates in the swap, but now the present value of those outflows has increased.

7 PRICING AND VALUING EQUITY SWAP CONTRACTS

g describe how equity swaps are priced, and calculate and interpret their no-arbitrage value.

Drawing on our prior definition of a swap, we define an equity swap in the following manner: An **equity swap** is an OTC derivatives contract in which two parties agree to exchange a series of cash flows whereby one party pays a variable series that will be determined by an equity and the other party pays either (1) a variable series determined by a different equity or rate or (2) a fixed series. An equity swap is used to convert the returns from an equity investment into another series of returns, which, as noted, either can be derived from another equity series or can be a fixed rate. Equity swaps are widely used in equity portfolio investment management to modify returns and risks. Equity swaps allow parties to benefit from returns of an equity or index without owning any shares of the underlying equity. An equity swap may also be used to hedge risk exposure to an equity or index for a certain period.

We examine three types of equity swaps: 1) *receive-equity return, pay-fixed*; 2) *receive-equity return, pay-floating*; and 3) *receive-equity return, pay-another equity return*. Like interest rate swaps and currency swaps, equity swaps have several unique nuances. We highlight just a few. First, the underlying reference instrument for the equity leg of an equity swap can be an individual stock, a published stock index, or a custom portfolio. Second, the equity leg cash flow(s) can be with or without dividends. Third, all the interest rate swap nuances exist with equity swaps that have a fixed or floating interest rate leg.

We focus here on viewing an equity swap as a portfolio of an equity position and a bond. The equity swap cash flows can be expressed as follows:

NA(Equity return – Fixed rate) (for receive-equity, pay-fixed),

NA(Equity return – Floating rate) (for receive-equity, pay-floating), and

NA(Equity return$_a$ – Equity return$_b$) (for receive-equity, pay-equity),

where a and b denote different equities. Note that an equity-for-equity swap can be viewed simply as a receive-equity a, pay-fixed swap combined with a pay-equity b, receive-fixed swap. The fixed payments cancel out, and we have synthetically created an equity-for-equity swap.

The cash flows for an equity leg (S_i) of an equity swap can be expressed as:

$S_i = NA_E\ R_E$,

where R_E denotes the periodic return of the equity either with or without dividends as specified in the swap contract, and NA_E denotes the notional amount. The cash flows for a fixed-interest rate leg (FS) of an equity swap are the same as those of an interest rate swap, or:

$FS = NA_E \times AP_{FIX} \times r_{FIX}$,

where AP_{FIX} denotes the accrual period for the fixed leg (for which we assume the accrual period is constant) and r_{FIX} here denotes the fixed rate on the equity swap.

EXAMPLE 17

Equity Swap Cash Flows

Suppose we entered a receive-equity index and pay-fixed swap. It is quarterly reset, 30/360 day count, €5,000,000 notional amount, pay-fixed (1.6% annualized, quarterly pay, or 0.4% per quarter).

1 If the equity index return was 4.0% for the quarter (not annualized), the equity swap cash flow will be *closest* to:

A –€220,000.

B –€180,000.

C €180,000.

2 If the equity index return was –6.0% for the quarter (not annualized), the equity swap cash flow will be closest to:

A –€320,000.

B –€180,000.

C €180,000.

Solution to 1:

C is correct. Note that the equity index return is reported on a quarterly basis. It is not an annualized number. The fixed leg is often reported on an annual basis. Thus, one must carefully interpret the different return conventions. In this case, receive-equity index counterparty cash flows (S_i – FS = NA_E x (R_E – r_{FIX})) are as follows:

€5,000,000 x (0.040 – 0.004) = €180,000 (Receive 4%, pay 0.4% for the quarter).

Solution to 2:

A is correct. Similar to 1, we have (S_i – FS = NA_E x (R_E – r_{FIX})):

€5,000,000 x (–0.060 – 0.004) = –€320,000 (Receive –6%, pay 0.4% for the quarter).

When the equity leg of the swap is negative, then the receive-equity counterparty must pay both the equity return as well as the fixed rate (or whatever are the payment terms). Note also that equity swaps may cause liquidity problems. As seen here, if the equity return is negative, then the receive-equity return, pay-floating or pay-fixed swap may result in a large negative cash flow for the receive-equity return party.

For equity swaps, the equity position could be a wide variety of claims, including the return on a stock index with or without dividends and the return on an individual stock with or without dividends. For our objectives here, we ignore the influence of dividends with the understanding that the equity swap leg assumes all dividends are reinvested in the equity position. The arbitrage transactions for an equity swap when dividends are not included are extremely complex and beyond our objectives. The equity leg of the swap is produced by selling the equity position on a reset date and reinvesting the original equity notional amount (NA_E), leaving a remaining balance that is the cash flow required of the equity swap leg (S_i). Technically, we just sell off any equity value in excess of NA_E or purchase additional shares to return the equity value to NA_E, effectively generating S_i. Exhibit 19 shows the cash flows from an equity swap offset with an equity and bond portfolio.

Exhibit 19 Cash Flows for Receive-Fixed, Pay-Equity Swap Offset with Equity and Bond Portfolio

Position	Steps	Time 0	Time 1	Time 2	...	Time n
Equity Swap	1. Receive-fixed, pay-equity swap	$-V_{EQ}$	+FS – S_1	+FS – S_2	...	+FS – S_n
Offset Portfolio	2. Buy NA_E of equity	– NA_E	$+S_1$	$+S_2$	...	$+S_n$ + NA_E
	3. Short sell fixed-rate bond	$+V_{FIX}$, (C = FS)	–FS	–FS	...	–(FS + Par)
	Net cash flows	$-V_{EQ}$ – NA_E + V_{FIX}	0	0	0	NA_E – Par

Assume a portfolio manager has a large position in a stock that he/she expects to underperform in the future. Perhaps for liquidity or tax reasons, the manager prefers not to sell the stock but considers a receive-fixed, pay equity swap. Exhibit 19 shows the cash flows from such a swap as well as the offsetting portfolio (to eliminate arbitrage), which will assist us in valuing the swap. In Step 1, we enter a receive-fixed, pay equity swap. Steps 2 and 3 provide the offsetting cash flows to those of the swap, which are buy NA_E worth of equity and short sell a fixed-rate bond (with coupon equal to the fixed interest rate leg cash flows), respectively. Notice that from Time 1 to n – 1 the sum of these three transactions is always zero. Note also that the final (Time n) cash flow for the long position in the equity includes the periodic return (S_n) plus the sale proceeds of the underlying equity position (NA_E). For the terminal cash flows to equal zero, we must either set the bond par value to equal the initial equity position (NA_E = Par) or finance this difference. In this latter case, the bond par value could be different from the notional amount of equity.

As shown, the swap and pair of offsetting transactions produce 0 net cash flow from period 1 to period n – 1. In equilibrium, we require $-V_{EQ} - NA_E + V_{FIX} - PV(Par - NA_E) = 0$. That is, if the portfolio has initial value with no required cash outflow, then arbitrage will be possible. Hence, the equity swap value is:

$$V_{EQ} = V_{FIX} - NA_E - PV(Par - NA_E).$$

Assuming equilibrium ($V_{EQ} = 0$), the fixed swap rate can be expressed as the r_{FIX} rate such that $V_{FIX} = NA_E + PV(Par - NA_E)$. Note that assuming NA_E = Par = 1 and using our fixed bond pricing (Equation 10), we have the pricing equation for an equity swap:

$$r_{FIX} = \frac{1 - PV_n(1)}{\sum_{i=1}^{n} PV_i(1)}$$

You should recognize that the pricing of an equity swap is identical to Equation 11 for the pricing of a comparable interest rate swap, even though the future cash flows are dramatically different. If the swap required a floating payment, there would be no need to price the swap; the floating side effectively prices itself at par automatically at the start. If the swap involves paying one equity return against another, there would also be no need to price it. You could effectively view this arrangement as *paying equity "a" and receiving the fixed rate* as specified and *receiving equity "b" and paying the same fixed rate*. The fixed rates would cancel.

Finding the value of an equity swap after the swap is initiated, say at Time t (so, $V_{EQ,t}$), is similar to valuing an interest rate swap except that rather than adjusting the floating-rate bond for the last floating rate observed (remember, advanced set), we adjust the value of the notional amount of equity, as shown in Equation 15:

$$V_{EQ,t} = V_{FIX}(C_0) - (S_t/S_{t-1})NA_E - PV(Par - NA_E), \tag{15}$$

where $V_{FIX}(C_0)$ denotes the value at Time t of a fixed-rate bond initiated with coupon C_0 at Time 0, S_t denotes the current equity price, S_{t-1} denotes the equity price observed at the last reset date, and PV() denotes the present value function from the swap maturity date to Time t.

EXAMPLE 18

Equity Swap Pricing

In Examples 13 and 14 related to interest rate swaps, we considered a five-year, annual reset, 30/360 day count, Libor-based swap. The following table provides the present values per €1, PV_i (1).

Maturity (years)	Present Value Factors
1	0.990099
2	0.977876
3	0.965136

Maturity (years)	Present Value Factors
4	0.951529
5	0.937467

Assume an annual reset Libor floating-rate bond trading at par. The fixed rate was previously found to be 1.2968% (see Example 13). Given these same data (just shown), the fixed interest rate in the EURO STOXX 50 equity swap is *closest* to:

A 0.0%.

B 1.1%.

C 1.3%.

Solution:

C is correct. The fixed rate on an equity swap is the same as that on an interest rate swap, or 1.2968% as in Example 13. That is, the fixed rate on an equity swap is simply the fixed rate on a comparable interest rate swap.

$$\sum_{i=1}^{5} \mathrm{PV}_i(1) = 0.990099 + 0.977876 + 0.965136 + 0.951529 + 0.937467$$
$$= 4.822107.$$

Using Equation 11, the solution for the fixed swap rate is:

$$r_{FIX} = \frac{1 - 0.937467}{4.822107} = 0.012968, \ \textit{or} \ 1.2968\%$$

EXAMPLE 19

Equity Swap Valuation

Suppose six months ago we entered a receive-fixed, pay-equity five-year annual reset swap in which the fixed leg is based on a 30/360 day count. At the time the swap was entered, the fixed swap rate was 1.5%, the equity was trading at 100, and the notional amount was 10,000,000. Now all spot interest rates have fallen to 1.2% (a flat term structure), and the equity is trading for 105. Assume the Par value of the bond is equal to $\mathrm{NA_E}$.

1 The current fair value of this equity swap is *closest* to:

A –€300,000.

B –€500,000.

C €500,000.

2 The value of the equity swap will be *closest* to zero if the stock price is:

A 100.

B 102.

C 105.

Solution to 1:

A is correct. Because we have not yet passed the first reset date, there are five remaining cash flows for this equity swap. The fair value of this swap is found by solving for the fair value of the implied fixed-rate bond. We then adjust for the equity value. The fixed rate of 1.5% results in fixed cash flows of 150,000 at each settlement. Applying the respective present value factors, which are based on the new spot rates of 1.2% (i.e., new term structure is flat), gives us the following:

Date (Years)	Present Value Factors (PV)	Fixed Cash Flow	PV (Fixed Cash Flow)
0.5	0.994036	150,000	149,105
1.5*	0.982318	150,000	147,348
2.5	0.970874	150,000	145,631
3.5	0.959693	150,000	143,954
4.5	0.948767	10,150,000	9,629,981
		Total:	10,216,019

* Answers may differ due to rounding: PV(1.5) = 1/(1 + 3 × (0.012/2)) = 0.982318.

Using Equation 15, we have,

$V_{EQ,t} = V_{FIX}(C_0) - (S_t/S_{t-1})NA_E - PV(Par - NA_E)$.

Therefore, the fair value of this equity swap is:

$V_{EQ,t} = 10{,}216{,}019 - [(105/100) \times 10{,}000{,}000] - 0 = -283{,}981$.

Solution to 2:

B is correct. The value of the fixed leg of the swap is 102.16% of par, = (10,216,019/10,000,000) × 100]. Therefore, a stock price (S_t) of 102.1602 will result in a value of zero for the swap, as follows:

$V_{EQ,t} = 10{,}216{,}019 - [(102.1602/100) \times 10{,}000{,}000] - 0 = 0$.

SUMMARY

This reading on forward commitment pricing and valuation provides a foundation for understanding how forwards, futures, and swaps are both priced and valued.

Key points include the following:

- The arbitrageur would rather have more money than less and abides by two fundamental rules: Do not use your own money, and do not take any price risk.
- The no-arbitrage approach is used for the pricing and valuation of forward commitments and is built on the key concept of the law of one price, which states that if two investments have the same future cash flows, regardless of what happens in the future, these two investments should have the same current price.
- Throughout this reading, the following key assumptions are made:
  - Replicating and offsetting instruments are identifiable and investable.
  - Market frictions are nil.
  - Short selling is allowed with full use of proceeds.
  - Borrowing and lending are available at a known risk-free rate.

- Carry arbitrage models used for forward commitment pricing and valuation are based on the no-arbitrage approach.
- With forward commitments, there is a distinct difference between pricing and valuation. Pricing involves the determination of the appropriate fixed price or rate, and valuation involves the determination of the contract's current value expressed in currency units.
- Forward commitment pricing results in determining a price or rate such that the forward contract value is equal to zero.
- Using the carry arbitrage model, the forward contract price (F_0) is:

$$F_0 = FV(S_0) = S_0(1 + r)^T \text{ (assuming annual compounding, r)}$$

$$F_0 = FV(S_0) = S_0 \exp^{r_c T} \text{ (assuming continuous compounding, } r_c\text{)}$$

- The key forward commitment pricing equations with carry costs (CC) and carry benefits (CB) are:

$$F_0 = FV[S_0 + CC_0 - CB_0] \text{ (with discrete compounding)}$$

$$F_0 = S_0 \exp^{(r_c + CC - CB)T} \text{ (with continuous compounding)}$$

Futures contract pricing in this reading can essentially be treated the same as forward contract pricing.

- The value of a forward commitment is a function of the price of the underlying instrument, financing costs, and other carry costs and benefits.
- The key forward commitment valuation equations are:

$$\text{Long Forward: } V_t = PV[F_t - F_0] = \frac{[F_t - F_0]}{(1 + r)^{T-t}}$$

and

$$\text{Short Forward: } -V_t = PV[F_0 - F_t] = \frac{[F_0 - F_t]}{(1 + r)^{T-t}}$$

With the PV of the difference in forward prices adjusted for carry costs and benefits. Alternatively,

$$\text{Long Forward: } V_t = S_t - PV[F_0] = S_t - \frac{F_0}{(1 + r)^{T-t}}$$

and

$$\text{Short Forward: } -V_t = PV[F_0] - S_t = \frac{F_0}{(1 + r)^{T-t}} - S_t$$

- With equities and fixed-income securities, the forward price is determined such that the initial forward value is zero.
- A forward rate agreement (FRA) is a forward contract on interest rates. The FRA's fixed interest rate is determined such that the initial value of the FRA is zero.
- FRA settlements amounts at Time h are:

Pay-fixed (Long): NA × {[L_m − FRA_0] t_m}/[1 + $D_m t_m$] and

Receive-fixed (Short): NA × {FRA_0 − L_m] t_m}/[1 + $D_m t_m$].

- The FRA's fixed interest rate (annualized) at contract initiation is:

  $FRA_0 = \{[1 + L_T t_T]/[1 + L_h t_h] - 1\}/t_m$.

- The Time g value of an FRA initiated at Time 0 is:

  Long FRA: $V_g = NA \times \{[FRA_g - FRA_0]\ t_m\}/[1+ D_{(T-g)}\ t_{(T-g)}]$ and

  Short FRA: $-V_g = NA \times \{[FRA_0 - FRA_g]\ t_m\}/[1+ D_{(T-g)}\ t_{(T-g)}]$.

- The fixed-income forward (or futures) price including conversion factor (i.e., adjusted price) is:

  $F_0 = Q_0 \times CF = FV[S_0 + CC_0 - CB_0] = FV[B_0 + AI_0 - PVCI]$,

  and the conversion factor adjusted futures price (i.e., quoted futures price) is:

  $Q_0 = [1/CF]\ \{FV\ [B_0 + AI_0] - AI_T - FVCI\}$.

- The general approach to pricing and valuing swaps as covered here is using a replicating portfolio or offsetting portfolio of comparable instruments, typically bonds for interest rate and currency swaps and equities plus bonds for equity swaps.
- The swap pricing equation, which sets r_{FIX} for the implied fixed bond in an interest rate swap, is:

$$r_{FIX} = \frac{1 - \mathrm{PV}_n(1)}{\sum_{i=1}^{n} \mathrm{PV}_i(1)}$$

- The value of an interest rate swap at a point in Time t after initiation is the sum of the present values of the difference in fixed swap rates times the stated notional amount, or:

$$V_{SWAP,t} = \mathrm{NA} \times (\mathrm{FS}_0 - \mathrm{FS}_t) \times \sum_{i=1}^{n} \mathrm{PV}_i \text{(Value of receive-fixed swap)}$$

  and

$$-V_{SWAP,t} = \mathrm{NA} \times (\mathrm{FS}_t - \mathrm{FS}_0) \times \sum_{i=1}^{n} \mathrm{PV}_i \text{(Value of pay-fixed swap).}$$

- With a basic understanding of pricing and valuing a simple interest rate swap, it is a straightforward extension to pricing and valuing currency swaps and equity swaps.
- The solution for each of the three variables, one notional amount (NA_a) and two fixed rates (one for each currency, a and b), needed to price a fixed-for-fixed currency swap are :

$$\mathrm{NA_a} = \mathrm{S_0} \times \mathrm{NA_b};\quad r_a = \frac{1 - \mathrm{PV}_{n,a}(1)}{\sum_{i=1}^{n} \mathrm{PV}_{i,a}(1)} \quad \text{and} \quad r_b = \frac{1 - \mathrm{PV}_{n,b}(1)}{\sum_{i=1}^{n} \mathrm{PV}_{i,b}(1)}.$$

- The currency swap valuation equation, for valuing the swap at time *t* (after initiation), can be expressed as:

$$V_{CS} = \mathrm{NA}_a\left(r_{Fix,a} \sum_{i=1}^{n} \mathrm{PV}_i(1) + \mathrm{PV}_n(1)\right) - S_t \mathrm{NA}_b\left(r_{Fix,b} \sum_{i=1}^{n} \mathrm{PV}_i(1) + \mathrm{PV}_n(1)\right)$$

- For a receive-fixed, pay equity swap, the fixed rate (r_{FIX}) for the implied fixed bond that makes the swap's value (V_{EQ}) equal to "0" at initiation is:

$$r_{FIX} = \frac{1 - PV_n(1)}{\sum_{i=1}^{n} PV_i(1)}$$

- The value of an equity swap at Time t ($V_{EQ,t}$), after initiation, is:

$$V_{EQ,t} = V_{FIX}(C_0) - (S_t/S_{t-1})NA_E - PV(Par - NA_E)$$

where V_{FIX} (C_0) is the Time t value of a fixed-rate bond initiated with coupon C_0 at Time 0, S_t is the current equity price, S_{t-1} is the equity price at the last reset date, and PV() is the PV function from the swap maturity date to Time t.

PRACTICE PROBLEMS

The following information relates to Questions 1–5

Donald Troubadour is a derivatives trader for Southern Shores Investments. The firm seeks arbitrage opportunities in the forward and futures markets using the carry arbitrage model.

Troubadour identifies an arbitrage opportunity relating to a fixed-income futures contract and its underlying bond. Current data on the futures contract and underlying bond are presented in Exhibit 1. The current annual compounded risk-free rate is 0.30%.

Exhibit 1 Current Data for Futures and Underlying Bond

Futures Contract		Underlying Bond	
Quoted futures price	125.00	Quoted bond price	112.00
Conversion factor	0.90	Accrued interest since last coupon payment	0.08
Time remaining to contract expiration	Three months	Accrued interest at futures contract expiration	0.20
Accrued interest over life of futures contract	0.00		

Troubadour next gathers information on a Japanese equity index futures contract, the **Nikkei 225 Futures Contract**:

> Troubadour holds a long position in a Nikkei 225 futures contract that has a remaining maturity of three months. The continuously compounded dividend yield on the Nikkei 225 Stock Index is 1.1%, and the current stock index level is 16,080. The continuously compounded annual interest rate is 0.2996%.

Troubadour next considers an equity forward contract for Texas Steel, Inc. (TSI). Information regarding TSI common shares and a TSI equity forward contract is presented in Exhibit 2.

Exhibit 2 Selected Information for TSI

- The price per share of TSI's common shares is $250.
- The forward price per share for a nine-month TSI equity forward contract is $250.562289.
- Assume annual compounding.

© 2020 CFA Institute. All rights reserved.

Troubadour takes a short position in the TSI equity forward contract. His supervisor asks, "Under which scenario would our position experience a loss?"

Three months after contract initiation, Troubadour gathers information on TSI and the risk-free rate, which is presented in Exhibit 3.

Exhibit 3 Selected Data on TSI and the Risk-Free Rate (Three Months Later)

- The price per share of TSI's common shares is $245.
- The risk-free rate is 0.325% (quoted on an annual compounding basis).
- TSI recently announced its regular semiannual dividend of $1.50 per share that will be paid exactly three months before contract expiration.
- The market price of the TSI equity forward contract is equal to the no-arbitrage forward price.

---

1 Based on Exhibit 1 and assuming annual compounding, the arbitrage profit on the bond futures contract is *closest* to:

A 0.4158.

B 0.5356.

C 0.6195.

2 The current no-arbitrage futures price of the Nikkei 225 futures contract is *closest* to:

A 15,951.81.

B 16,047.86.

C 16,112.21.

3 Based on Exhibit 2, Troubadour should find that an arbitrage opportunity relating to TSI shares is

A not available.

B available based on carry arbitrage.

C available based on reverse carry arbitrage.

4 The *most appropriate* response to Troubadour's supervisor's question regarding the TSI forward contract is:

A a decrease in TSI's share price, all else equal.

B an increase in the risk-free rate, all else equal

C a decrease in the market price of the forward contract, all else equal.

5 Based on Exhibits 2 and 3, and assuming annual compounding, the per share value of Troubadour's short position in the TSI forward contract three months after contract initiation is *closest* to:

A $1.6549.

B $5.1561.

C $6.6549.

The following information relates to Questions 6–14

Sonal Johnson is a risk manager for a bank. She manages the bank's risks using a combination of swaps and forward rate agreements (FRAs).

Johnson prices a three-year Libor-based interest rate swap with annual resets using the present value factors presented in Exhibit 1.

Exhibit 1 Present Value Factors

Maturity (years)	Present Value Factors
1	0.990099
2	0.977876
3	0.965136

Johnson also uses the present value factors in Exhibit 1 to value an interest rate swap that the bank entered into one year ago as the pay-fixed (receive-floating) party. Selected data for the swap are presented in Exhibit 2. Johnson notes that the current equilibrium two-year fixed swap rate is 1.12%.

Exhibit 2 Selected Data on Fixed for Floating Interest Rate Swap

Swap notional amount	$50,000,000
Original swap term	Three years, with annual resets
Fixed swap rate (since initiation)	3.00%

One of the bank's investments is exposed to movements in the Japanese yen, and Johnson desires to hedge the currency exposure. She prices a one-year fixed-for-fixed currency swap involving yen and US dollars, with a quarterly reset. Johnson uses the interest rate data presented in Exhibit 3 to price the currency swap.

Exhibit 3 Selected Japanese and US Interest Rate Data

Days to Maturity	Yen Spot Interest Rates	US Dollar Spot Interest Rates
90	0.05%	0.20%
180	0.10%	0.40%
270	0.15%	0.55%
360	0.25%	0.70%

Johnson next reviews an equity swap with an annual reset that the bank entered into six months ago as the receive-fixed, pay-equity party. Selected data regarding the equity swap, which is linked to an equity index, are presented in Exhibit 4. At the time of initiation, the underlying equity index was trading at 100.00.

Exhibit 4	Selected Data on Equity Swap
Swap notional amount	$20,000,000
Original swap term	Five years, with annual resets
Fixed swap rate	2.00%

The equity index is currently trading at 103.00, and relevant US spot rates, along with their associated present value factors, are presented in Exhibit 5.

Exhibit 5 Selected US Spot Rates and Present Value Factors

Maturity (years)	Spot Rate	Present Value Factors
0.5	0.40%	0.998004
1.5	1.00%	0.985222
2.5	1.20%	0.970874
3.5	2.00%	0.934579
4.5	2.60%	0.895255

Johnson reviews a 6 × 9 FRA that the bank entered into 90 days ago as the pay-fixed/receive-floating party. Selected data for the FRA are presented in Exhibit 6, and current Libor (i.e., MRR) data are presented in Exhibit 7. Based on her interest rate forecast, Johnson also considers whether the bank should enter into new positions in 1 × 4 and 2 × 5 FRAs.

Exhibit 6	6 × 9 FRA Data
FRA term	6 × 9
FRA rate	0.70%
FRA notional amount	US$20,000,000
FRA settlement terms	Advanced set, advanced settle

Exhibit 7	Current Libor (Market Reference Rate)
30-day Libor	0.75%
60-day Libor	0.82%
90-day Libor	0.90%
120-day Libor	0.92%
150-day Libor	0.94%
180-day Libor	0.95%
210-day Libor	0.97%
270-day Libor	1.00%

Three months later, the 6 × 9 FRA in Exhibit 6 reaches expiration, at which time the three-month US dollar Libor is 1.10% and the six-month US dollar Libor is 1.20%. Johnson determines that the appropriate discount rate for the FRA settlement cash flows is 1.10%.

6 Based on Exhibit 1, Johnson should price the three-year Libor-based interest rate swap at a fixed rate *closest* to:

A 0.34%.

B 1.16%.

C 1.19%.

7 From the bank's perspective, using data from Exhibit 1, the current value of the swap described in Exhibit 2 is *closest* to:

A –$2,951,963.

B –$1,849,897.

C –$1,943,000.

8 Based on Exhibit 3, Johnson should determine that the annualized equilibrium fixed swap rate for Japanese yen is *closest* to:

A 0.0624%.

B 0.1375%.

C 0.2496%.

9 From the bank's perspective, using data from Exhibits 4 and 5, the fair value of the equity swap is *closest* to:

A –$1,139,425.

B –$781,322.

C –$181,323.

10 Based on Exhibit 5, the current value of the equity swap described in Exhibit 4 would be zero if the equity index was currently trading the *closest* to:

A 97.30.

B 99.09.

C 100.00.

11 From the bank's perspective, based on Exhibits 6 and 7, the value of the 6 × 9 FRA 90 days after inception is *closest* to:

A $14,820.

B $19,647.

C $29,635.

12 Based on Exhibit 7, the no-arbitrage fixed rate on a new 1 × 4 FRA is *closest* to:

A 0.65%.

B 0.73%.

C 0.98%.

13 Based on Exhibit 7, the fixed rate on a new 2 × 5 FRA is *closest* to:

A 0.61%.

B 1.02%.

C 1.71%.

14 Based on Exhibit 6 and the three-month US dollar Libor at expiration, the payment amount that the bank will receive to settle the 6 × 9 FRA is *closest* to:

A $19,945.

B $24,925.

C $39,781.

The following information relates to Questions 15–20

Tim Doyle is a portfolio manager at BestFutures Group, a hedge fund that frequently enters into derivative contracts either to hedge the risk of investments it holds or to speculate outside of those investments. Doyle works alongside Diane Kemper, a junior analyst at the hedge fund. They meet to evaluate new investment ideas and to review several of the firm's existing investments.

Carry Arbitrage Model

Doyle and Kemper discuss the carry arbitrage model and how they can take advantage of mispricing in bond markets. Specifically, they would like to execute an arbitrage transaction on a Eurodollar futures contract in which the underlying Eurodollar bond is expected to make an interest payment in two months. Doyle makes the following statements:

Statement 1 If the Eurodollar futures price is less than the price suggested by the carry arbitrage model, the futures contract should be purchased.

Statement 2 Based on the cost of carry model, the futures price would be higher if the underlying Eurodollar bond's upcoming interest payment was expected in five months instead of two.

Three-Year Treasury Note Futures Contract

Kemper then presents two investment ideas to Doyle. Kemper's first investment idea is to purchase a three-year Treasury note futures contract. The underlying 1.5%, semi-annual three-year Treasury note is quoted at a clean price of 101. It has been 60 days since the three-year Treasury note's last coupon payment, and the next coupon payment is payable in 120 days. Doyle asks Kemper to calculate the full spot price of the underlying three-year Treasury note.

10-Year Treasury Note Futures Contract

Kemper's second investment idea is to purchase a 10-year Treasury note futures contract. The underlying 2%, semi-annual 10-year Treasury note has a dirty price of 104.17. It has been 30 days since the 10-year Treasury note's last coupon payment. The futures contract expires in 90 days. The quoted futures contract price is 129. The current annualized three-month risk-free rate is 1.65%. The conversion factor is 0.7025. Doyle asks Kemper to calculate the equilibrium quoted futures contract price based on the carry arbitrage model.

Japanese Government Bonds

After discussing Kemper's new investment ideas, Doyle and Kemper evaluate one of their existing forward contract positions. Three months ago, BestFutures took a long position in eight 10-year Japanese government bond (JGB) forward contracts, with each contract having a contract notional value of 100 million yen. The contracts had a price of JPY153 (quoted as a percentage of par) when the contracts were purchased.

Now, the contracts have six months left to expiration and have a price of JPY155. The annualized six-month interest rate is 0.12%. Doyle asks Kemper to value the JGB forward position.

Interest Rate Swaps

Additionally, Doyle asks Kemper to price a one-year plain vanilla swap. The spot rates and days to maturity at each payment date are presented in Exhibit 1.

Exhibit 1 Selected US Spot Rate Data

Days to Maturity	Spot Interest Rates (%)
90	1.90
180	2.00
270	2.10
360	2.20

Finally, Doyle and Kemper review one of BestFutures's pay-fixed interest rate swap positions. Two years ago, the firm entered into a JPY5 billion five-year interest rate swap, paying the fixed rate. The fixed rate when BestFutures entered into the swap two years ago was 0.10%. The current term structure of interest rates for JPY cash flows, which are relevant to the interest rate swap position, is presented in Exhibit 2.

Exhibit 2 Selected Japanese Interest Rate Data

Maturity (Years)	Yen Spot Interest Rates (%)	Present Value Factors
1	0.03	0.9997
2	0.06	0.9988
3	0.08	0.9976
Sum		2.9961

Doyle asks Kemper to calculate the value of the pay-fixed interest rate swap.

15 Which of Doyle's statements regarding the Eurodollar futures contract price is correct?

A Only Statement 1

B Only Statement 2

C Both Statement 1 and Statement 2

16 The full spot price of the three-year Treasury note is:

A 101.00.

B 101.25.

C 101.50.

17 The equilibrium 10-year Treasury note quoted futures contract price is *closest* to:

A 147.94.

B 148.89.

C 149.78.

18 The value of the JGB long forward position is *closest* to:

A JPY15,980,823.

B JPY15,990,409.

C JPY16,000,000.

19 Based on Exhibit 1, the fixed rate of the one-year plain vanilla swap is *closest* to:

A 0.12%.

B 0.55%.

C 0.72%.

20 Based on Exhibit 2, the value of the pay-fixed interest rate swap is *closest* to:

A –JPY6,491,550.

B –JPY2,980,500.

C –JPY994,793.

SOLUTIONS

1 B is correct.

The no-arbitrage futures price is equal to the following:

$$F_0 = \text{FV}[B_0 + \text{AI}_0 - \text{PVCI}]$$
$$F_0 = (1 + 0.003)^{0.25}(112.00 + 0.08 - 0) = 112.1640.$$

The adjusted price of the futures contract is equal to the conversion factor multiplied by the quoted futures price:

$$F_0 = \text{CF} \times \text{Q}_0$$
$$F_0 = (0.90)(125) = 112.50.$$

Adding the accrued interest of 0.20 in three months (futures contract expiration) to the adjusted price of the futures contract gives a total price of 112.70.

This difference means that the futures contract is overpriced by 112.70 – 112.1640 = 0.5360. The available arbitrage profit is the present value of this difference: $0.5360/(1.003)^{0.25} = 0.5356$.

2 B is correct. The no-arbitrage futures price is

$$\text{F}_0 = S_0\exp^{(r_c + \text{CC} - \text{CB})T}$$

$$F_0 = 16{,}080\exp^{(0.002996 + 0 - 0.011)(3/12)} = 16{,}047.86.$$

3 A is correct. The carry arbitrage model price of the forward contract is

$$\text{FV}(S_0) = S_0(1 + r)^T = \$250(1 + 0.003)^{0.75} = \$250.562289.$$

The market price of the TSI forward contract is \$250.562289. A carry or reverse carry arbitrage opportunity does not exist because the market price of the forward contract is equal to the carry arbitrage model price.

4 B is correct. From the perspective of the long position, the forward value is equal to the present value of the difference in forward prices:

$$V_t = \text{PV}[F_t - F_0],$$

where

$$F_t = \text{FV}(S_t + \text{CC}_t - \text{CB}_t).$$

All else equal, an increase in the risk-free rate before contract expiration would cause the forward price, F_t, to increase. This increase in the forward price would cause the value of the TSI forward contract, from the perspective of the short, to decrease. Therefore, an increase in the risk-free rate would lead to a loss on the short position in the TSI forward contract.

5 C is correct. The no-arbitrage price of the forward contract, three months after contract initiation, is

$$F_{0.25} = \text{FV}_{0.25}(S_{0.25} + \text{CC}_{0.25} - \text{CB}_{0.25})$$
$$F_{0.25} = [\$245 + 0 - \$1.50/(1 + 0.00325)^{(0.5 - 0.25)}](1 + 0.00325)^{(0.75 - 0.25)} = \$243.8966.$$

Therefore, from the perspective of the long, the value of the TSI forward contract is

$V_{0.25} = \mathrm{PV}_{0.25}[F_{0.25} - F_0]$
$V_{0.25} = (\$243.8966 - \$250.562289)/(1 + 0.00325)^{0.75 - 0.25} = -\$6.6549.$

Because Troubadour is short the TSI forward contract, the value of his position is a gain of \$6.6549.

6 C is correct. The swap pricing equation is

$$r_{FIX} = \frac{1 - PV_n(1)}{\sum_{i=1}^{n} PV_i(1)}$$

That is, the fixed swap rate is equal to 1 minus the final present value factor (in this case, Year 3) divided by the sum of the present values (in this case, the sum of Years 1, 2, and 3). The sum of present values for Years 1, 2, and 3 is calculated as

$$\sum_{i=1}^{n} PV_i(1) = 0.990099 + 0.977876 + 0.965136 = 2.933111$$

Thus, the fixed-swap rate is calculated as

$$r_{FIX} = \frac{1 - 0.965136}{2.933111} = 0.01189 \text{ or } 1.19\%$$

7 B is correct. The value of a swap from the perspective of the receive-fixed (pay-floating) party is calculated as

$$V = \mathrm{NA} \times (\mathrm{FS}_0 - \mathrm{FS}_t) \times \sum_{i=1}^{n} \mathrm{PV}_i$$

The swap has two years remaining until expiration. The sum of the present values for Years 1 and 2 is

$$\sum_{i=1}^{n} \mathrm{PV}_i = 0.990099 + 0.977876 = 1.967975$$

Given the current equilibrium two-year swap rate of 1.12% and the fixed swap rate at initiation of 3.00%, the swap value per dollar notional is calculated as

$V = 1 \times (0.03 - 0.0112) \text{ x } 1.967975 = 0.036998.$

The current value of the swap, from the perspective of the receive-fixed party, is \$50,000,000 × 0.036998 = \$1,849,897.

From the perspective of the bank, as the pay-fixed party, the value of the swap is –\$1,849,897.

8 C is correct. The equilibrium swap fixed rate for yen is calculated as

$$r_{JPY} = \frac{1 - PV_{n,JPY}(1)}{\sum_{i=1}^{4} PV_{i,JPY}(1)}$$

The yen present value factors are calculated as

$$\mathrm{PV}(1)_{i,JPY} = \frac{1}{1 + Rspot_{i,JPY}\left(\frac{NAD_i}{NTD}\right)}$$

where

$$\begin{aligned}
\text{90-day PV factor} &= 1/[1 + 0.0005(90/360)] = 0.999875\\
\text{180-day PV factor} &= 1/[1 + 0.0010(180/360)] = 0.999500\\
\text{270-day PV factor} &= 1/[1 + 0.0015(270/360)] = 0.998876\\
\text{360-day PV factor} &= 1/[1 + 0.0025(360/360)] = 0.997506
\end{aligned}$$

Sum of present value factors = 3.995757

Therefore, the yen periodic rate is calculated as

$$r_{JPY} = \frac{1 - \text{PV}_n(1)}{\sum_{i=1}^{4}\text{PV}_i(1)} = \frac{1 - 0.997506}{3.995757} = 0.000624 = 0.0624\%.$$

The annualized rate is (360/90) times the periodic rate of 0.0624%, or 0.2496%.

9 B is correct. The value of an equity swap at time t is calculated as

$$\text{V}_{\text{EQ,t}} = \text{V}_{\text{FIX}}(\text{C}_0) - (\text{S}_t/\text{S}_{t-1})\text{NA}_\text{E} - \text{PV}(\text{Par} - \text{NA}_\text{E}).$$

The swap was initiated six months ago, so the first reset has not yet passed; thus, there are five remaining cash flows for this equity swap. The fair value of the swap is determined by comparing the present value of the implied fixed-rate bond with the return on the equity index. The fixed swap rate of 2.00%, the swap notional amount of $20,000,000, and the present value factors in Exhibit 5 result in a present value of the implied fixed-rate bond's cash flows of $19,818,678:

Date (in years)	PV Factors	Fixed Cash Flow	PV (fixed cash flow)
0.5	0.998004 or 1/[1 + 0.0040(180/360)]	$400,000	$399,202
1.5	0.985222 or 1/[1 + 0.0100(540/360)]	$400,000	$394,089
2.5	0.970874 or 1/[1 + 0.0120(900/360)]	$400,000	$388,350
3.5	0.934579 or 1/[1 + 0.0200(1,260/360)]	$400,000	$373,832
4.5	0.895255 or 1/[1 + 0.0260(1,620/360)]	$20,400,000	$18,263,205
Total			$19,818,678

The value of the equity leg of the swap is calculated as (103/100)($20,000,000) = $20,600,000.

Note the swap's notional amount and the implied fixed-rate bond's par value are both $20,000,000; therefore, the term – PV(Par – NA_E) reduces to zero.

The swap was designed to profit if rates fell or equities declined. Neither happened, so the swap value will be negative for the bank. The fair value of the equity swap, from the perspective of the bank (receive-fixed, pay-equity party) is calculated as

$$V_{EQ} = \$19{,}818{,}678 - \$20{,}600{,}000 = -\$781{,}322.$$

10 B is correct. The equity index level at which the swap's fair value would be zero can be calculated by setting the swap valuation formula equal to zero and solving for S_t:

$$V_{EQ,t} = V_{FIX}(C_0) - (S_t/S_{t-1})NA_E = 0.$$

The value of the fixed leg of the swap has a present value of $19,818,678, or 99.0934% of par value:

Date (years)	PV Factors	Fixed Cash Flow	PV (fixed cash flow)
0.5	0.998004	$400,000	$399,202
1.5	0.985222	$400,000	$394,089
2.5	0.970874	$400,000	$388,350
3.5	0.934579	$400,000	$373,832
4.5	0.895255	$20,400,000	$18,263,205
Total			$19,818,678

Treating the swap notional value as par value and substituting the present value of the fixed leg and S_0 into the equation yields

$$0 = 99.0934 - \left(\frac{S_t}{100}\right)100$$

Solving for S_t yields

$$S_t = 99.0934.$$

11 A is correct. The current value of the 6 × 9 FRA is calculated as

$$V_g = NA \times \{[FRA_g - FRA_0]t_m\}/[1 + D_{(T-g)}\, t_{(T-g)}].$$

The 6 × 9 FRA expires six months after initiation. The bank entered into the FRA 90 days ago; thus, the FRA will expire in 90 days. To value the FRA, the first step is to compute the new FRA rate, which is the rate on Day 90 of an FRA that expires in 90 days in which the underlying is the 90-day Libor:

$$FRA_g = \{[1 + L_Tt_T]/[1+L_ht_h] - 1\}/t_m$$
$$FRA_g = \{[1 + L_{180}(180/360)]/[1 + L_{90}\ (90/360)] - 1\}/(90/360)$$

Exhibit 7 indicates that $L_{90} = 0.90\%$ and $L_{180} = 0.95\%$, so

$$FRA_g = \{[1 + 0.0095(180/360)]/[1 + 0.0090(90/360)] - 1\}/(90/360)$$
$$FRA_g = \{[1.00475/1.00225] - 1]\} \times 4 = 0.009978, \text{ or } 0.9978\%.$$

Therefore, given the FRA rate at initiation of 0.70% and notional principal of $20 million from Exhibit 1, the current value of the forward contract is calculated as

$$V_g = \$20{,}000{,}000 \times [(0.009978 - 0.0070)(90/360)]/[1 + 0.0095(180/360)].$$
$$= \$14{,}890.00/1.00475 = \$14{,}819.61.$$

12 C is correct. The no-arbitrage fixed rate on the 1 × 4 FRA is calculated as

$$FRA_0 = \{[1 + L_Tt_T]/[1 + L_ht_h] - 1\}/t_m\,.$$

For a 1 × 4 FRA, the two rates needed to compute the no-arbitrage FRA fixed rate are L_{30} = 0.75% and L_{120} = 0.92%. Therefore, the no-arbitrage fixed rate on the 1 × 4 FRA rate is calculated as

$$FRA_0 = \{[1 + 0.0092(120/360)]/[1 + 0.0075(30/360)] - 1\}/(90/360).$$
$$FRA_0 = \{[1.003066/1.000625] - 1\} \times 4 = 0.009761, \text{ or } 0.98\% \text{ rounded.}$$

13 B is correct. The fixed rate on the 2 × 5 FRA is calculated as

$$FRA_0 = \{[1 + L_T t_T]/[1 + L_h t_h] - 1\}/t_m.$$

For a 2 × 5 FRA, the two rates needed to compute the no-arbitrage FRA fixed rate are L_{60}= 0.82% and L_{150} = 0.94%. Therefore, the no-arbitrage fixed rate on the 2 × 5 FRA rate is calculated as

$$FRA_0 = \{[1 + 0.0094(150/360)]/[1 + 0.0082(60/360)] - 1\}/(90/360)$$
$$FRA_0 = \{[(1.003917/1.001367) - 1\} \times 4 = 0.010186, \text{ or } 1.02\% \text{ rounded.}$$

14 A is correct. Given a three-month US dollar Libor of 1.10% at expiration, the settlement amount for the bank as the pay-fixed (receive-floating) party is calculated as

Settlement amount pay-fixed (receive floating)
$$= NA \times \{[L_m - FRA_0]t_m\}/[1 + D_m t_m]\}.$$
Settlement amount pay-fixed (receive floating)
$$= \$20{,}000{,}000 \times \{[0.011 - 0.0070] \times (90/360)]/[1 + 0.011(90/360)]\}.$$
Settlement amount pay-fixed (receive floating)
$$= \$20{,}000{,}000 \times (0.001)/1.00275 = \$19{,}945.15.$$

Therefore, the bank will receive $19,945 (rounded) as the receive-floating party.

15 C is correct. Doyle's first statement is correct. Unless the Eurodollar futures contract's quoted price is equal to the no-arbitrage futures price, there is an arbitrage opportunity. Moreover, if the quoted futures price is less than the no-arbitrage futures price, then to take advantage of the arbitrage opportunity, the Eurodollar futures contract should be purchased and the underlying Eurodollar bond should be sold short. Doyle would then lend the short sale proceeds at the risk-free rate. The strategy that comprises those transactions is known as reverse carry arbitrage.

Doyle's second statement is also correct. Based on the cost of carry model, the futures price is calculated as the future value of the sum of the underlying plus the underlying carry costs minus the future value of any ownership benefits. If the Eurodollar bond's interest payment was expected in five months instead of two, the benefit of the cash flow would occur three months later, so the future value of the benefits term would be slightly lower. Therefore, the Eurodollar futures contract price would be slightly higher if the Eurodollar bond's interest payment was expected in five months instead of two months.

A is incorrect because Doyle's Statement 2 is correct (not incorrect). Based on the cost of carry model, the futures price would be higher if the underlying Eurodollar bond's interest payment took place in five months instead of two months.

B is incorrect because Doyle's Statement 1 is correct (not incorrect). If the Eurodollar's futures contract price is less than the price suggested by the carry arbitrage model, the futures contract should be purchased.

16 B is correct. The full spot price of the three-year Treasury note is calculated as

$$S_0 = \text{Quoted bond price} + \text{Accrued interest} = B_0 + AI_0.$$

$$\text{Accrued interest (AI)} = \text{Accrural period} \times \text{Periodic coupon amount} = \left(\frac{\text{NAD}}{\text{NTD}}\right) \times \left(\frac{C}{n}\right).$$

AI = (60/180) × (0.015/2) = 0.25.

S_0 = 101 + 0.25 = 101.25.

A is incorrect because 101 is the quoted clean (not the full spot) price of the three-year Treasury note. The clean price excludes accrued interest; the full price, also referred to as the dirty price, includes accrued interest.

C is incorrect because the number of days until the next coupon payment (instead of the accrual period) is incorrectly used to compute accrued interest:

AI = (120/180) × (0.015/2) = 0.50.

S_0 = 101 + 0.50 = 101.50.

17 A is correct. The equilibrium 10-year quoted futures contract price based on the carry arbitrage model is calculated as

$Q_0 = (1/\text{CF}) \times [\text{FV}(B_0 + \text{AI}_0) - \text{AI}_T - \text{FVCI}]$.

CF = 0.7025.

B_0 = 104.00.

AI_0 = 0.17.

AI_T = (120/180 × 0.02/2) = 0.67.

FVCI = 0.

$$Q_0 = (1 / 0.7025) \times \left[(1 + 0.0165)^{3/12}(104.17) - 0.67 - 0\right] = 147.94.$$

B is incorrect because accrued interest at expiration is not subtracted in the equilibrium quoted futures contract price formula:

$$Q_0 = (1 / 0.7025) \times \left[(1 + 0.0165)^{3/12}(104.17) - 0\right] = 148.89.$$

C is incorrect because the future value is incorrectly calculated (the exponent of 3/12 is omitted):

$$Q_0 = (1 / 0.7025) \times \left[(1 + 0.0165)(104.17) - 0.67 - 0\right] = 149.78.$$

18 B is correct. The value of the JGB forward position is calculated as

$$V_t = \text{PV}\left[F_t - F_0\right] = (155 - 153) / (1 + 0.0012)^{\frac{6}{12}} = 1.9988.$$

Therefore, the value of the long forward position is 1.9988 per JPY100 par value.

For the long position in eight contracts with each contract having a par value of 100 million yen, the value of the position is calculated as

0.019988 × (JPY100,000,000) × 8 = JPY15,990,409.

A is incorrect because the present value of the difference between the price when the contracts were purchased and the current price of the contracts was incorrectly computed (the exponent of 6/12 is omitted):

$V_t = F_t - F_0$ = (155 − 153)/(1 + 0.0012) = 1.9980.

0.019980 × (JPY100,000,000) × 8 = JPY15,980,823.

C is incorrect because the absolute difference (not the present value of the difference) between the price when the contracts were purchased and the current price of the contracts was computed:

$V_t = F_t - F_0 = (155 - 153) = 2.$

$0.02 \times (\text{JPY}100{,}000{,}000) \times 8 = \text{JPY}16{,}000{,}000.$

19 B is correct. The swap's fixed rate is calculated as

$$r_{FIX} = \left[1 - \text{PV}_n(1)\right] / \sum_{i=1}^{n} \text{PV}_i(1).$$

$\text{PV}_i\,(1) = 1/[1 + \text{Rspot}_i\,(\text{NAD}_i/\text{NTD})].$

$90 - \text{day PV factor} = 1/[1 + 0.019 \times (90/360)] = 0.9953.$

$180 - \text{day PV factor} = 1/[1 + 0.020 \times (180/360)] = 0.9901.$

$270 - \text{day PV factor} = 1/[1 + 0.021 \times (270/360)] = 0.9845.$

$360 - \text{day PV factor} = 1/[1 + 0.022 \times (360/360)] = 0.9785.$

$$\sum_{i=1}^{4} \text{PV}_i(1) = 0.9953 + 0.9901 + 0.9845 + 0.9785 = 3.9483.$$

$r_{FIX} = (1 - 0.9785)/3.9483 = 0.0055 = 0.55\%.$

A is incorrect because the 90-day PV factor is incorrectly used in the numerator of the swap pricing equation instead of the final present value term:

$$r_{FIX} = \left[1 - \text{PV}_n(1)\right] / \sum_{i=1}^{n} \text{PV}_i(1).$$

$r_{FIX} = (1 - 0.9953)/3.9483 = 0.0012 = 0.12\%.$

C is incorrect because the sum of the present value terms excludes the final present value term:

$$\sum_{i=1}^{3} \text{PV}_i(1) = 0.9953 + 0.9901 + 0.9845 = 2.9699.$$

$$r_{FIX} = \left[1 - \text{PV}_n(1)\right] / \sum_{i=1}^{n} \text{PV}_i(1).$$

$r_{FIX} = (1 - 0.9785)/2.9699 = 0.0072 = 0.72\%.$

20 B is correct. The value of the pay-fixed interest rate swap is calculated as

$$-V_{SWAP,t} = \text{NA} \times \left(\text{FS}_t - \text{FS}_0\right) \times \sum\nolimits_{i=1}^{n} \text{PV}_i.$$

$$\text{FS}_t = r_{FIX} = \left[1 - \text{PV}_n(1)\right] / \sum_{i=1}^{3} \text{PV}_i(1) = (1 - 0.9976) / 2.9961 = 0.000801 = 0.08\%.$$

$$\begin{aligned}-V_{SWAP,t} &= \text{NA}\times\left(\text{FS}_t - \text{FS}_0\right)\times\sum\nolimits_{\text{i}=1}^{3}\text{PV}_i\\ &= \text{JPY5billion}\times\left(0.000801 - 0.001\right)\times 2.9961\\ &= -\text{JPY2,980,500.}\end{aligned}$$

Given that rates have declined since the inception of the swap, the value of the pay-fixed, receive-floating position is currently a loss of JPY2,980,500.

A is incorrect because the arithmetic average of the yen spot rates (instead of the current fixed swap rate) was incorrectly used to calculate the value of the pay-fixed swap:

Arithmetic average of yen spot rates = (0.0003 + 0.0006 + 0.0008)/3 = 0.0006.

$$\begin{aligned}-V_{SWAP,t} &= \text{NA}\times\left(\text{FS}_t - \text{FS}_0\right)\times\sum\nolimits_{i=1}^{3}\text{PV}_i\\ &= \text{JPY5billion}\times\left(0.0006 - 0.001\right)\times 2.9961\\ &= -\text{JPY6,491,550.}\end{aligned}$$

C is incorrect because the product of the notional amount and the difference between the initial swap fixed rate and the current swap fixed rate was not multiplied by the sum of the present values:

$$-V_{SWAP,t} = \text{NA}\times\left(\text{FS}_t - \text{FS}_0\right) = \text{JPY5billion}\times\left(0.0008 - 0.001\right) = -\text{JPY994,793.}$$

READING

34

Valuation of Contingent Claims

by Robert E. Brooks, PhD, CFA, and David Maurice Gentle, MEc, BSc, CFA

Robert E. Brooks, PhD, CFA, is at the University of Alabama (USA). David Maurice Gentle, MEc, BSc, CFA, is at Omega Risk Consulting (Australia).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe and interpret the binomial option valuation model and its component terms;
☐	**b.** calculate the no-arbitrage values of European and American options using a two-period binomial model;
☐	**c.** identify an arbitrage opportunity involving options and describe the related arbitrage;
☐	**d.** calculate and interpret the value of an interest rate option using a two-period binomial model;
☐	**e.** describe how the value of a European option can be analyzed as the present value of the option's expected payoff at expiration;
☐	**f.** identify assumptions of the Black–Scholes–Merton option valuation model;
☐	**g.** interpret the components of the Black–Scholes–Merton model as applied to call options in terms of a leveraged position in the underlying;
☐	**h.** describe how the Black–Scholes–Merton model is used to value European options on equities and currencies;
☐	**i.** describe how the Black model is used to value European options on futures;
☐	**j.** describe how the Black model is used to value European interest rate options and European swaptions;
☐	**k.** interpret each of the option Greeks;
☐	**l.** describe how a delta hedge is executed;
☐	**m.** describe the role of gamma risk in options trading;
☐	**n.** define implied volatility and explain how it is used in options trading.

© 2020 CFA Institute. All rights reserved.

1 INTRODUCTION AND PRINCIPLES OF A NO-ARBITRAGE APPROACH TO VALUATION

A contingent claim is a derivative instrument that provides its owner a right but not an obligation to a payoff determined by an underlying asset, rate, or other derivative. Contingent claims include options, the valuation of which is the objective of this reading. Because many investments contain embedded options, understanding this material is vital for investment management.

Our primary purpose is to understand how the values of options are determined. Option values, as with the values of all financial instruments, are typically obtained using valuation models. Any financial valuation model takes certain inputs and turns them into an output that tells us the fair value or price. Option valuation models, like their counterparts in the forward, futures, and swaps markets, are based on the principle of no arbitrage, meaning that the appropriate price of an option is the one that makes it impossible for any party to earn an arbitrage profit at the expense of any other party. The price that precludes arbitrage profits is the value of the option. Using that concept, we then proceed to introduce option valuation models using two approaches. The first approach is the binomial model, which is based on discrete time, and the second is the Black–Scholes–Merton (BSM) model, which is based on continuous time.

The reading is organized as follows. Section 1 introduces the principles of the no-arbitrage approach to pricing and valuation of options. In Sections 2–7, the binomial option valuation model is explored, and in Sections 8–10, the BSM model is covered. In Sections 11–13, the Black model, being a variation of the BSM model, is applied to futures options, interest rate options, and swaptions. Finally, in Sections 14–19, the Greeks are reviewed along with implied volatility.

1.1 Principles of a No-Arbitrage Approach to Valuation

Our approach is based on the concept of arbitrage. Hence, the material will be covered from an arbitrageur's perspective. Key to understanding this material is to think like an arbitrageur. Specifically, like most people, the arbitrageur would rather have more money than less. The arbitrageur, as will be detailed later, follows two fundamental rules:

Rule #1 Do not use your own money.

Rule #2 Do not take any price risk.

Clearly, if we can generate positive cash flows today and abide by both rules, we have a great business—such is the life of an arbitrageur. If traders could create a portfolio with no future liabilities and positive cash flow today, then it would essentially be a money machine that would be attractive to anyone who prefers more cash to less. In the pursuit of these positive cash flows today, the arbitrageur often needs to borrow to satisfy Rule #1. In effect, the arbitrageur borrows the arbitrage profit to capture it today and, if necessary, may borrow to purchase the underlying. Specifically, the arbitrageur will build portfolios using the underlying instrument to synthetically replicate the cash flows of an option. The underlying instrument is the financial instrument whose later value will be referenced to determine the option value. Examples of underlying instruments include shares, indexes, currencies, and interest rates. As we will see, with options we will often rely on a specific trading strategy that changes over time based on the underlying price behavior.

Based on the concept of comparability, the no-arbitrage valuation approach taken here is built on the concept that if two investments have the same future cash flows regardless of what happens, then these two investments should have the same current price. This principle is known as the **law of one price**. In establishing these foundations of option valuation, the following key assumptions are made: (1) Replicating instruments are identifiable and investable. (2) There are no market frictions, such as transaction costs and taxes. (3) Short selling is allowed with full use of proceeds. (4) The underlying instrument follows a known statistical distribution. (5) Borrowing and lending at a risk-free interest rate is available. When we develop the models in this reading, we will be more specific about what these assumptions mean, in particular what we mean by a known statistical distribution.

In an effort to demonstrate various valuation results based on the absence of arbitrage, we will rely heavily on cash flow tables, which are a representation of the cash flows that occur during the life of an option. For example, if an initial investment requires €100, then from an arbitrageur's perspective, we will present it as a –€100 cash flow. If an option pays off ¥1,000, we will represent it as a +¥1,000 cash flow. That is, cash outflows are treated as negative and inflows as positive.

We first demonstrate how to value options based on a two-period binomial model. The option payoffs can be replicated with a dynamic portfolio of the underlying instrument and financing. A dynamic portfolio is one whose composition changes over time. These changes are important elements of the replicating procedure. Based on the binomial framework, we then turn to exploring interest rate options using a binomial tree. Although more complex, the general approach is shown to be the same.

The multiperiod binomial model is a natural transition to the BSM option valuation model. The BSM model is based on the key assumption that the value of the underlying instrument follows a statistical process called geometric Brownian motion. This characterization is a reasonable way to capture the randomness of financial instrument prices while incorporating a pre-specified expected return and volatility of return. Geometric Brownian motion implies a lognormal distribution of the return, which implies that the continuously compounded return on the underlying is normally distributed.

We also explore the role of carry benefits, meaning the reward or cost of holding the underlying itself instead of holding the derivative on the underlying.

Next we turn to Fischer Black's futures option valuation model (Black model) and note that the model difference, versus the BSM model, is related to the underlying futures contract having no carry costs or benefits. Interest rate options and swaptions are valued based on simple modifications of the Black model.

Finally, we explore the Greeks, otherwise known as delta, gamma, theta, vega, and rho. The Greeks are representations of the sensitivity of the option value to changes in the factors that determine the option value. They provide comparative information essential in managing portfolios containing options. The Greeks are calculated based on an option valuation model, such as the binomial model, BSM model, or the Black model. This information is model dependent, so managers need to carefully select the model best suited for their particular situation. In the last section, we cover implied volatility, which is a measure derived from a market option price and can be interpreted as reflecting what investors believe is the volatility of the underlying.

The models presented here are useful first approximations for explaining observed option prices in many markets. The central theme is that options are generally priced to preclude arbitrage profits, which is not only a reasonable theoretical assumption but is sufficiently accurate in practice.

We turn now to option valuation based on the binomial option valuation model.

2 BINOMIAL OPTION VALUATION MODEL

a describe and interpret the binomial option valuation model and its component terms;

The binomial model is a valuable tool for financial analysts. It is particularly useful as a heuristic device to understand the unique valuation approach used with options. This model is extensively used to value path-dependent options, which are options whose values depend not only on the value of the underlying at expiration but also how it got there. The path-dependency feature distinguishes this model from the Black–Scholes–Merton option valuation model (BSM model) presented in the next section. The BSM model values only path-independent options, such as European options, which depend on only the values of their respective underlyings at expiration. One particular type of path-dependent option that we are interested in is American options, which are those that can be exercised prior to expiration. In this section, we introduce the general framework for developing the binomial option valuation models for both European and American options.

The binomial option valuation model is based on the no-arbitrage approach to valuation. Hence, understanding the valuation of options improves if one can understand how an arbitrageur approaches financial markets. An arbitrageur engages in financial transactions in pursuit of an initial positive cash flow with no possibility of a negative cash flow in the future. As it appears, it is a great business if you can find it.[1]

To understand option valuation models, it is helpful to think like an arbitrageur. The arbitrageur seeks to exploit any pricing discrepancy between the option price and the underlying spot price. The arbitrageur is assumed to prefer more money compared with less money, assuming everything else is the same. As mentioned earlier, there are two fundamental rules for the arbitrageur.

Rule #1 Do not use your own money. Specifically, the arbitrageur does not use his or her own money to acquire positions. Also, the arbitrageur does not spend proceeds from short selling transactions on activities unrelated to the transaction at hand.

Rule #2 Do not take any price risk. The focus here is only on market price risk related to the underlying and the derivatives used. We do not consider other risks, such as liquidity risk and counterparty credit risk.

We will rely heavily on these two rules when developing option valuation models. Remember, these rules are general in nature, and as with many things in finance, there are nuances.

In Exhibit 1, the two key dates are the option contract initiation date (identified as Time 0) and the option contract expiration date (identified as Time T). Based on the no-arbitrage approach, the option value from the initiation date onward will be estimated with an option valuation model.

1 There is not a one-to-one correspondence between arbitrage and great investment opportunities. An arbitrage is certainly a great investment opportunity because it produces a risk-free profit with no investment of capital, but all great investment opportunities are not arbitrage. For example, an opportunity to invest €1 today in return for a 99% chance of receiving €1,000,000 tomorrow or a 1% chance of receiving €0 might appear to be a truly great investment opportunity, but it is not arbitrage because it is not risk free and requires the investment of capital.

Exhibit 1 Illustration of Option Contract Initiation and Expiration

Let S_t denote the underlying instrument price observed at Time t, where t is expressed as a fraction of a year. Similarly, S_T denotes the underlying instrument price observed at the option expiration date, T. For example, suppose a call option had 90 days to expiration when purchased (T = 90/365), but now only has 35 days to expiration (t = 55/365). Further, let c_t denote a European-style call price at Time t and with expiration on Date t = T, where both t and T are expressed in years. Similarly, let C_t denote an American-style call price. At the initiation date, the subscripts are omitted, thus $c = c_0$. We follow similar notation with a put, using the letter p, in place of c. Let X denote the exercise price.[2]

For example, suppose on 15 April a 90-day European-style call option contract with a 14 July expiration is initiated with a call price of c = €2.50 and T = 90/365 = 0.246575.

At expiration, the call and put values will be equal to their intrinsic value or exercise value. These **exercise values** can be expressed as

$$c_T = \text{Max}(0,S_T - X) \text{ and}$$
$$p_T = \text{Max}(0,X - S_T),$$

respectively. If the option values deviate from these expressions, then there will be arbitrage profits available. The option is expiring, there is no uncertainty remaining, and the price must equal the market value obtained from exercising it or letting it expire.

Technically, European options do not have exercise values prior to expiration because they cannot be exercised until expiration. Nonetheless, the notion of the value of the option if it could be exercised, $\text{Max}(0,S_t - X)$ for a call and $\text{Max}(0,X - S_t)$ for a put, forms a basis for understanding the notion that the value of an option declines with the passage of time. Specifically, option values contain an element known as time value, which is just the market valuation of the potential for higher exercise value relative to the potential for lower exercise value. The time value is always non-negative because of the asymmetry of option payoffs at expiration. For example, for a call, the upside is unlimited, whereas the downside is limited to zero. At expiration, time value is zero.

Although option prices are influenced by a variety of factors, the underlying instrument has a particularly significant influence. At this point, the underlying is assumed to be the only uncertain factor affecting the option price. We now look in detail at the one-period binomial option valuation model. The one-period binomial model is foundational for the material that follows.

2 In financial markets, the exercise price is also commonly called the strike price.

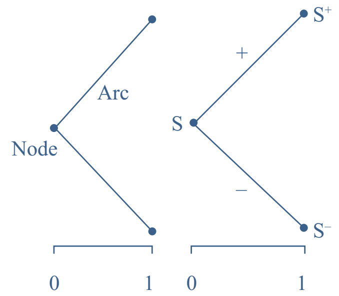
3 ONE-PERIOD BINOMIAL MODEL

a describe and interpret the binomial option valuation model and its component terms;

e describe how the value of a European option can be analyzed as the present value of the option's expected payoff at expiration;

Exhibit 2 illustrates the one-period binomial process for an asset priced at S. In the figure on the left, each dot represents a particular outcome at a particular point in time in the binomial lattice. The dots are termed nodes. At the Time 0 node, there are only two possible future paths in the binomial process, an up move and a down move, termed arcs. The figure on the right illustrates the underlying price at each node. At Time 1, there are only two possible outcomes: S^+ denotes the outcome when the underlying goes up, and S^- denotes the outcome when the underlying goes down.

Exhibit 2 One-Period Binomial Lattice with Underlying Distribution Illustrated



At Time 1, there are only two possible outcomes and two resulting values of the underlying, S^+ (up occurs) and S^- (down occurs). Although the one-period binomial model is clearly unrealistic, it will provide key insights into the more realistic multi-period binomial as well as the BSM model.

We further define the total returns implied by the underlying movements as

$$u = \frac{S^+}{S}\text{(up factor) and}$$

$$d = \frac{S^-}{S}\text{(down factor).}$$

The up factors and down factors are the total returns; that is, one plus the rate of return. The magnitudes of the up and down factors are based on the volatility of the underlying. In general, higher volatility will result in higher up values and lower down values.

We briefly review option valuation within a one-period binomial tree. With this review, we can move quickly to option valuation within a two-period binomial lattice by performing the one-period exercise three times.

We consider the fair value of a two-period call option value measured at Time 1 when an up move occurs, that is c^+. Based on arbitrage forces, we know this option value at expiration is either

$$c^{++} = \text{Max}(0,S^{++} - X) = \text{Max}(0,u^2S - X)\text{, or}$$
$$c^{+-} = \text{Max}(0,S^{+-} - X) = \text{Max}(0,udS - X).$$

At this point, we assume that there are no costs or benefits from owning the underlying instrument. Now consider the transactions illustrated in Exhibit 3. These transactions are presented as cash flows. Thus, if we write a call option, we receive money at Time Step 0 and may have to pay out money at Time Step 1. Suppose the first trade is to write or sell one call option within the single-period binomial model. The value of a call option is positively related to the value of the underlying. That is, they both move up or down together. Hence, by writing a call option, the trader will lose money if the underlying goes up and make money if the underlying falls. Therefore, to execute a hedge, the trader will need a position that will make money if the underlying goes up. Thus, the second trade needs to be a long position in the underlying. Specifically, the trader buys a certain number of units, h, of the underlying. The symbol h is used because it represents a hedge ratio.

Note that with these first two trades, neither arbitrage rule is satisfied. The future cash flow could be either $-c^- + hS^-$ or $-c^+ + hS^+$ and can be positive or negative. Thus, the cash flows at the Time Step 1 could result in the arbitrageur having to pay out money if one of these values is less than zero. To resolve both of these issues, we set the Time Step 1 cash flows equal to each other—that is, $-c^+ + hS^+ = -c^- + hS^-$—and solve for the appropriate hedge ratio:

$$h = \frac{c^+ - c^-}{S^+ - S^-} \geq 0 \tag{1}$$

We determine the hedge ratio such that we are indifferent to the underlying going up or down. Thus, we are hedged against moves in the underlying. A simple rule for remembering this formula is that the hedge ratio is the value of the call if the underlying goes up minus the value of the call if the underlying goes down divided by the value of the underlying if it goes up minus the value of the underlying if it goes down. The up and down patterns are the same in the numerator and denominator, but the numerator contains the option and the denominator contains the underlying.

Because call prices are positively related to changes in the underlying price, we know that h is non-negative. As shown in Exhibit 3, we will buy h underlying units as depicted in the second trade, and we will finance the present value of the net cash flows as depicted in the third trade. If we assume r denotes the per period risk-free interest rate, then the present value calculation, denoted as PV, is equal to $1/(1 + r)$. We need to borrow or lend an amount such that the future net cash flows are equal to zero. Therefore, we finance today the present value of $-hS^- + c^-$ which also equals $-hS^+ + c^+$. At this point we do not know if the finance term is positive or negative, thus we may be either borrowing or lending, which will depend on c, h, and S.

Exhibit 3 Writing One Call Hedge with h Units of the Underlying and Finance

Strategy	Time Step 0	Time Step 1 Down Occurs	Time Step 1 Up Occurs
1) Write one call option	$+c$	$-c^-$	$-c^+$
2) Buy h underlying units	$-hS$	$+hS^-$	$+hS^+$
3) Borrow or lend	$-PV(-hS^- + c^-)$ $= -PV(-hS^+ + c^+)$	$-hS^- + c^-$	$-hS^+ + c^+$
Net Cash Flow	$+c - hS$ $-PV(-hS^- + c^-)$	0	0

The value of the net portfolio at Time Step 0 should be zero or there is an arbitrage opportunity. If the net portfolio has positive value, then arbitrageurs will engage in this strategy, which will push the call price down and the underlying price up until the net is no longer positive. We assume the size of the borrowing will not influence interest rates. If the net portfolio has negative value, then arbitrageurs will engage in the opposite strategy—buy calls, short sell the underlying, and lend—pushing the call price up and the underlying price down until the net cash flow at Time 0 is no longer positive. Therefore, within the single-period binomial model, we have

$$+c - hS - PV(-hS^- + c^-) = 0$$

or, equivalently,

$$+c - hS - PV(-hS^+ + c^+) = 0.$$

Therefore, the **no-arbitrage approach** leads to the following single-period call option valuation equation:

$$c = hS + PV(-hS^- + c^-) \tag{2}$$

or, equivalently, $c = hS + PV(-hS^+ + c^+)$. In words, long a call option is equal to owning h shares of stock partially financed, where the financed amount is $PV(-hS^- + c^-)$, or using the per period rate, $(-hS^- + c^-)/(1 + r)$.[3]

We will refer to Equation 2 as the no-arbitrage single-period binomial option valuation model. This equation is foundational to understanding the two-period binomial as well as other option valuation models. The option can be replicated with the underlying and financing, a point illustrated in the following example.

EXAMPLE 1

Long Call Option Replicated with Underlying and Financing

Identify the trading strategy that will generate the payoffs of taking a long position in a call option within a single-period binomial framework.

A Buy $h = (c^+ + c^-)/(S^+ + S^-)$ units of the underlying and financing of $-PV(-hS^- + c^-)$

B Buy $h = (c^+ - c^-)/(S^+ - S^-)$ units of the underlying and financing of $-PV(-hS^- + c^-)$

C Short sell $h = (c^+ - c^-)/(S^+ - S^-)$ units of the underlying and financing of $+PV(-hS^- + c^-)$

Solution:

B is correct. The following table shows the terminal payoffs to be identical between a call option and buying the underlying with financing.

Strategy	Time Step 0	Time Step 1 Down Occurs	Time Step 1 Up Occurs
Buy 1 call option	$-c$	$+c^-$	$+c^+$
OR A REPLICATING PORTFOLIO			
Buy h underlying units	$-hS$	$+hS^-$	$+hS^+$

3 Or, by the same logic, $PV(-hS^+ + c^+)$, which is $(-hS^+ + c^+)/(1 + r)$.

Strategy	Time Step 0	Time Step 1 Down Occurs	Time Step 1 Up Occurs
Borrow or lend	$-PV(-hS^- + c^-) = -PV(-hS^+ + c^+)$	$-hS^- + c^-$	$-hS^+ + c^+$
Net	$-hS - PV(-hS^- + c^-)$	$+c^-$	$+c^+$

Recall that by design, h is selected such that $-hS^- + c^- = -hS^+ + c^+$ or $h = (c^+ - c^-)/(S^+ - S^-)$. Therefore, a call option can be replicated with the underlying and financing. Specifically, the call option is equivalent to a leveraged position in the underlying.

Thus, the no-arbitrage approach is a replicating strategy: A call option is synthetically replicated with the underlying and financing. Following a similar strategy with puts, the no-arbitrage approach leads to the following no-arbitrage single-period put option valuation equation:

$$p = hS + PV(-hS^- + p^-) \tag{3}$$

or, equivalently, $p = hS + PV(-hS^+ + p^+)$ where

$$h = \frac{p^+ - p^-}{S^+ - S^-} \leq 0 \tag{4}$$

Because p^+ is less than p^-, the hedge ratio is negative. Hence, to replicate a long put position, the arbitrageur will short sell the underlying and lend a portion of the proceeds. Note that a long put position would be replicated by trading h units of the underlying. With h negative, this trade is a short sale, and because $-h$ is positive, the value $-hS$ results in a positive cash flow at Time Step 0.

EXAMPLE 2

Long Put Option Replicated with Underlying and Financing

Identify the trading strategy that will generate the payoffs of taking a long position in a put option within a single-period binomial framework.

A Short sell $-h = -(p^+ - p^-)/(S^+ - S^-)$ units of the underlying and financing of $-PV(-hS^- + p^-)$

B Buy $-h = (p^+ - p^-)/(S^+ - S^-)$ units of the underlying and financing of $-PV(-hS^- + p^-)$

C Short sell $h = (p^+ - p^-)/(S^+ - S^-)$ units of the underlying and financing of $+PV(-hS^- + p^-)$

Solution:

A is correct. Before illustrating the replicating portfolio, we make a few observations regarding the hedge ratio. Note that by design, h is selected such that $-hS^- + p^- = -hS^+ + p^+$ or $h = (p^+ - p^-)/(S^+ - S^-)$. Unlike calls, the put hedge ratio is not positive (note that $p^+ < p^-$ but $S^+ > S^-$). Remember that taking a position in $-h$ units of the underlying is actually short selling the underlying rather than buying it. The following table shows the terminal payoffs to be identical between a put option and a position in the underlying with financing.

Strategy	Time Step 0	Time Step 1 Down Occurs	Time Step 1 Up Occurs
Buy 1 Put Option	$-p$	$+p^-$	$+p^+$
OR A REPLICATING PORTFOLIO			
Short sell $-h$ Underlying Units	$-hS$	$+hS^-$	$+hS^+$
Borrow or Lend	$-PV(-hS^- + p^-)$ $= -PV(-hS^+ + p^+)$	$-hS^- + p^-$	$-hS^+ + p^+$
Net	$-hS - PV(-hS^- + p^-)$	$+p^-$	$+p^+$

Therefore, a put option can be replicated with the underlying and financing. Specifically, the put option is simply equivalent to a short position in the underlying with financing in the form of lending.

What we have shown to this point is the no-arbitrage approach. Before turning to the expectations approach, we mention, for the sake of completeness, that the transactions for replicating the payoffs for writing options are the reverse for those of buying them. Thus, for writing a call option, the writer will be selling stock short and investing proceeds (i.e. lending), whereas for a put, the writer will be purchasing stock on margin (i.e. borrowing). Once again, we see the powerful result that the same basic conceptual structure is used for puts and calls, whether written or purchased. Only the exercise and expiration conditions vary.

The no-arbitrage results that have been presented can be expressed as the present value of a unique expectation of the option payoffs.[4] Specifically, the **expectations approach** results in an identical value as the no-arbitrage approach, but it is usually easier to compute. The formulas are viewed as follows:

$$c = PV[\pi c^+ + (1 - \pi)c^-] \text{ and} \tag{5}$$

$$p = PV[\pi p^+ + (1 - \pi)p^-] \tag{6}$$

where the probability of an up move is

$$\pi = [FV(1) - d]/(u - d)$$

Recall the future value is simply the reciprocal of the present value or $FV(1) = 1/PV(1)$. Thus, if $PV(1) = 1/(1 + r)$, then $FV(1) = (1 + r)$. Note that the option values are simply the present value of the expected terminal option payoffs. The expected terminal option payoffs can be expressed as

$$E(c_1) = \pi c^+ + (1 - \pi)c^- \text{ and}$$

$$E(p_1) = \pi p^+ + (1 - \pi)p^-$$

where c_1 and p_1 are the values of the options at Time 1. The present value and future value calculations are based on the risk-free rate, denoted r.[5] Thus, the option values based on the expectations approach can be written and remembered concisely as

$$c = PV_r[E(c_1)] \text{ and}$$

$$p = PV_r[E(p_1)]$$

4 It takes a bit of algebra to move from the no-arbitrage expression to the present value of the expected future payoffs, but the important point is that both expressions yield exactly the same result.

5 We will suppress "r" most of the time and simply denote the calculation as PV. The "r" will be used at times to reinforce that the present value calculation is based on the risk-free interest rate.

The expectations approach to option valuation differs in two significant ways from the discounted cash flow approach to securities valuation. First, the expectation is not based on the investor's beliefs regarding the future course of the underlying. That is, the probability, π, is objectively determined and not based on the investor's personal view. This probability has taken several different names, including risk-neutral (RN) probability. Importantly, we did not make any assumption regarding the arbitrageur's risk preferences: The expectations approach is a result of this arbitrage process, not an assumption regarding risk preferences. Hence, they are called risk-neutral probabilities. Although we called them probabilities from the very start, they are not the true probabilities of up and down moves.

Second, the discount rate is *not* risk adjusted. The discount rate is simply based on the estimated risk-free interest rate. The expectations approach here is often viewed as superior to the discounted cash flow approach because both the subjective future expectation as well as the subjective risk-adjusted discount rate have been replaced with more objective measures.

EXAMPLE 3

Single-Period Binomial Call Value

A non-dividend-paying stock is currently trading at €100. A call option has one year to mature, the periodically compounded risk-free interest rate is 5.15%, and the exercise price is €100. Assume a single-period binomial option valuation model, where u = 1.35 and d = 0.74.

1 The optimal hedge ratio will be *closest* to:

A 0.57.

B 0.60.

C 0.65.

2 The call option value will be *closest* to:

A €13.

B €15.

C €17.

Solution to 1:

A is correct. Given the information provided, we know the following:

$S^+ = uS = 1.35(100) = 135$

$S^- = dS = 0.74(100) = 74$

$c^+ = \text{Max}(0,uS - X) = \text{Max}(0,135 - 100) = 35$

$c^- = \text{Max}(0,dS - X) = \text{Max}(0,74 - 100) = 0$

With this information, we can compute both the hedge ratio as well as the call option value. The hedge ratio is:

$$h = \frac{c^+ - c^-}{S^+ - S^-} = \frac{35 - 0}{135 - 74} = 0.573770$$

Solution to 2:

C is correct. The risk-neutral probability of an up move is

$\pi = [FV(1) - d]/(u - d) = (1.0515 - 0.74)/(1.35 - 0.74) = 0.510656,$

where $FV(1) = (1 + r) = 1.0515$.

Thus the call value by the expectations approach is

$$c = PV[\pi c^+ + (1 - \pi)c^-] = 0.951022[(0.510656)35 + (1 - 0.510656)0] = €16.998,$$

where PV(1) = 1/(1 + r) = 1/(1.0515) = 0.951022.

Note that the call value by the no-arbitrage approach yields the same answer:

$$c = hS + PV(-hS^- + c^-) = 0.573770(100) + 0.951022[-0.573770(74) + 0] = €16.998.$$

The value of a put option can also be found based on put–call parity. Put–call parity can be remembered as simply two versions of portfolio insurance, long stock and long put or lend and long call, where the exercise prices for the put and call are identical. Put–call parity with symbols is

$$S + p = PV(X) + c \tag{7}$$

Put–call parity holds regardless of the particular valuation model being used. Depending on the context, this equation can be rearranged. For example, a call option can be expressed as a position in a stock, financing, and a put, or

$$c = S - PV(X) + p$$

EXAMPLE 4

Single-Period Binomial Put Value

You again observe a €100 price for a non-dividend-paying stock with the same inputs as the previous box. That is, the call option has one year to mature, the periodically compounded risk-free interest rate is 5.15%, the exercise price is €100, u = 1.35, and d = 0.74. The put option value will be *closest* to:

A €12.00.

B €12.10.

C €12.20.

Solution:

B is correct. For puts, we know the following:

$$p^+ = Max(0,100 - uS) = Max(0,100 - 135) = 0$$

$$p^- = Max(0,100 - dS) = Max(0,100 - 74) = 26$$

With this information, we can compute the put option value based on risk-neutral probability from the previous example or [recall that PV(1) = 0.951022]

$$p = PV[\pi p^+ + (1 - \pi)p^-] = 0.951022[(0.510656)0 + (1 - 0.510656)26] = €12.10$$

Therefore, in summary, option values can be expressed either in terms of replicating portfolios or as the present value of the expected future cash flows. Both expressions yield the same valuations.

BINOMIAL MODEL: TWO-PERIOD (CALL OPTIONS)

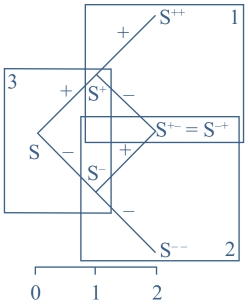
4

a describe and interpret the binomial option valuation model and its component terms;

c identify an arbitrage opportunity involving options and describe the related arbitrage;

The two-period binomial lattice can be viewed as three one-period binomial lattices, as illustrated in Exhibit 4. Clearly, if we understand the one-period model, then the process can be repeated three times. First, we analyze Box 1 and Box 2. Finally, based on the results of Box 1 and Box 2, we analyze Box 3.

Exhibit 4 Two-Period Binomial Lattice as Three One-Period Binomial Lattices



At Time 2, there are only three values of the underlying, S^{++} (an up move occurs twice), S^{--} (a down move occurs twice), and $S^{+-} = S^{-+}$ (either an up move occurs and then a down move or a down move occurs and then an up move). For computational reasons, it is extremely helpful that the lattice recombines—that is, $S^{+-} = S^{-+}$, meaning that if the underlying goes up and then down, it ends up at the same price as if it goes down and then up. A recombining binomial lattice will always have just one more ending node in the final period than the number of time steps. In contrast, a non-recombining lattice of n time steps will have 2^n ending nodes, which poses a tremendous computational challenge even for powerful computers.

For our purposes here, we assume the up and down factors are constant throughout the lattice, ensuring that the lattice recombines—that is $S^{+-} = S^{-+}$. For example, assume u = 1.25, d = 0.8, and S_0 = 100. Note that S^{+-} = 1.25(0.8)100 = 100 and S^{-+} = 0.8(1.25)100 = 100. So the middle node at Time 2 is 100 and can be reached from either of two paths.

The two-period binomial option valuation model illustrates two important concepts, self-financing and dynamic replication. Self-financing implies that the replicating portfolio will not require any additional funds from the arbitrageur during the life of this dynamically rebalanced portfolio. If additional funds are needed, then they are financed externally. Dynamic replication means that the payoffs from the option can be exactly replicated through a planned trading strategy. Option valuation relies on self-financing, dynamic replication.

Mathematically, the no-arbitrage approach for the two-period binomial model is best understood as working backward through the binomial tree. At Time 2, the payoffs are driven by the option's exercise value.

For calls:

$$c^{++} = \text{Max}(0,S^{++} - X) = \text{Max}(0,u^2S - X),$$
$$c^{+-} = \text{Max}(0,S^{+-} - X) = \text{Max}(0,udS - X), \text{ and}$$
$$c^{--} = \text{Max}(0,S^{--} - X) = \text{Max}(0,d^2S - X)$$

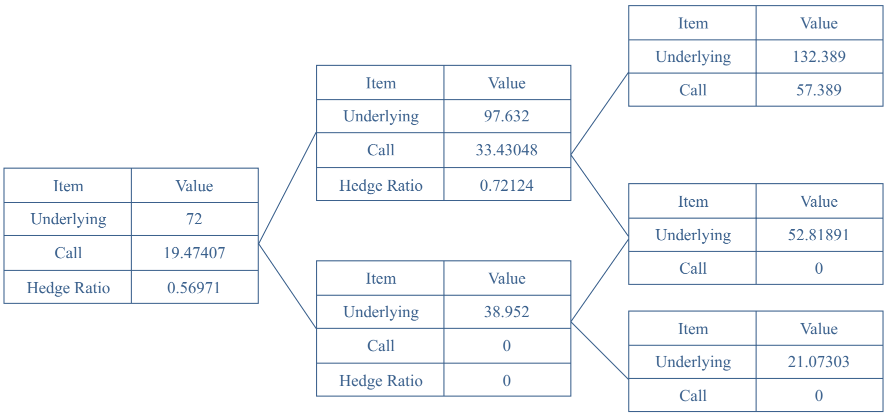
For puts:

$$p^{++} = \text{Max}(0,X - S^{++}) = \text{Max}(0,X - u^2S),$$
$$p^{+-} = \text{Max}(0,X - S^{+-}) = \text{Max}(0,X - udS), \text{ and}$$
$$p^{--} = \text{Max}(0,X - S^{--}) = \text{Max}(0,X - d^2S)$$

At Time 1, the option values are driven by the arbitrage transactions that synthetically replicate the payoffs at Time 2. We can compute the option values at Time 1 based on the option values at Time 2 using the no-arbitrage approach based on Equations 1 and 2. At Time 0, the option values are driven by the arbitrage transactions that synthetically replicate the value of the options at Time 1 (again based on Equations 1 and 2).

We illustrate the no-arbitrage approach for solving the two-period binomial call value. Suppose the annual interest rate is 3%, the underlying stock is S = 72, u = 1.356, d = 0.541, and the exercise price is X = 75. The stock does not pay dividends. Exhibit 5 illustrates the results.

Exhibit 5 Two-Period Binomial Tree with Call Values and Hedge Ratios



We now verify selected values reported in Exhibit 5. At Time Step 2 and assuming up occurs twice, the underlying stock value is $u^2S = (1.356)^2 72 = 132.389$, and hence, the call value is 57.389 [= Max(0,132.389 – 75)]. The hedge ratio at Time Step 1, assuming up occurs once, is

$$h^+ = \frac{c^{++} - c^{+-}}{S^{++} - S^{+-}} = \frac{57.389 - 0}{132.389 - 52.819} = 0.72124$$

The RN probability of an up move throughout this tree is

$$\pi = [FV(1) - d]/(u - d) = (1.03 - 0.541)/(1.356 - 0.541) = 0.6$$

With this information, we can compute the call price at Time 1 when an up move occurs as

$$c = PV[\pi c^{++} + (1 - \pi)c^{+-}] = (1/1.03)[(0.6)57.389 + (1 - 0.6)0] = 33.43048$$

and at Time Step 0,

$$h = \frac{c^+ - c^-}{S^+ - S^-} = \frac{33.43048 - 0}{97.632 - 38.952} = 0.56971$$

Thus, the call price at the start is

$$c = PV[\pi c^+ + (1 - \pi)c^-] = (1/1.03)[(0.6)33.43048 + (1 - 0.6)0] = 19.47$$

From the no-arbitrage approach, the call payoffs can be replicated by purchasing h shares of the underlying and financing $-PV(-hS^- + c^-)$. Therefore, we purchase 0.56971 shares of stock for 41.019 [= 0.56971(72)] and borrow 21.545 {or in cash flow terms, –21.545 = (1/1.03)[–0.56971(38.952) + 0]}, replicating the call values at Time 0. We then illustrate Time 1 assuming that an up move occurs. The stock position will now be worth 55.622 [= 0.56971(97.632)], and the borrowing must be repaid with interest or 22.191 [= 1.03(21.545)]. Note that the portfolio is worth 33.431 (55.622 – 22.191), the same value as the call except for a small rounding error. Therefore, the portfolio of stock and the financing dynamically replicates the value of the call option.

The final task is to demonstrate that the portfolio is self-financing. Self-financing can be shown by observing that the new portfolio at Time 1, assuming an up move occurs, is equal to the old portfolio that was formed at Time 0 and liquidated at Time 1. Notice that the hedge ratio rose from 0.56971 to 0.72124 as we moved from Time 0 to Time 1, assuming an up move occurs, requiring the purchase of additional shares. These additional shares will be financed with additional borrowing. The total borrowing is 36.98554 {$= -PV(-hS^{+-} + c^{+-}) = -(1/1.03)[-0.72124(52.81891) + 0]$}. The borrowing at Time 0 that is due at Time 1 is 22.191. The funds borrowed at Time 1 grew to 36.98554. Therefore, the strategy is self-financing.

The two-period binomial model can also be represented as the present value of an expectation of future cash flows. Based on the one-period results, it follows by repeated substitutions that

$$c = PV[\pi^2 c^{++} + 2\pi(1 - \pi)c^{+-} + (1 - \pi)^2 c^{--}] \tag{8}$$

and

$$p = PV[\pi^2 p^{++} + 2\pi(1 - \pi)p^{+-} + (1 - \pi)^2 p^{--}] \tag{9}$$

Therefore, the two-period binomial model is again simply the present value of the expected future cash flows based on the RN probability. Again, the option values are simply the present value of the expected terminal option payoffs. The expected terminal option payoffs can be expressed as

$$E(c_2) = \pi^2 c^{++} + 2\pi(1 - \pi)c^{+-} + (1 - \pi)^2 c^{--}$$

and

$$E(p_2) = \pi^2 p^{++} + 2\pi(1 - \pi)p^{+-} + (1 - \pi)^2 p^{--}$$

Thus, the two-period binomial option values based on the expectations approach can be written and remembered concisely as

$c = PV_r[E\pi(c_2)]$ and

$p = PV_r[E\pi(p_2)]$

It is vital to remember that this present value is over two periods, so the discount factor with discrete rates is $PV = [1/(1 + r)^2]$. Recall the subscript "r" just emphasizes the present value calculation and is based on the risk-free interest rate.

EXAMPLE 5

Two-Period Binomial Model Call Valuation

You observe a €50 price for a non-dividend-paying stock. The call option has two years to mature, the periodically compounded risk-free interest rate is 5%, the exercise price is €50, u = 1.356, and d = 0.744. Assume the call option is European-style.

1 The probability of an up move based on the risk-neutral probability is *closest* to:

A 30%.

B 40%.

C 50%.

2 The current call option value is *closest* to:

A €9.53.

B €9.71.

C €9.87.

3 The current put option value is *closest* to:

A €5.06.

B €5.33.

C €5.94.

Solution to 1:

C is correct. Based on the RN probability equation, we have:

$$\pi = [FV(1) - d]/(u - d) = [(1 + 0.05) - 0.744]/(1.356 - 0.744) = 0.5 \text{ or } 50\%$$

Solution to 2:

B is correct. The current call option value calculations are as follows:

$$c^{++} = Max(0,u^2S - X) = Max[0,1.356^2(50) - 50] = 41.9368$$

$$c^{-+} = c^{+-} = Max(0,udS - X) = Max[0,1.356(0.744)(50) - 50] = 0.44320$$

$$c^{--} = Max(0,d^2S - X) = Max[0,0.744^2(50) - 50] = 0.0$$

With this information, we can compute the call option value:

$$\begin{aligned} c &= PV[E(c_2)] = PV[\pi^2c^{++} + 2\pi(1 - \pi)c^{+-} + (1 - \pi)^2c^{--}] \\ &= [1/(1 + 0.05)]^2[0.5^2 41.9368 + 2(0.5)(1 - 0.5)0.44320 + (1 - 0.5)^2 0.0] \\ &= 9.71 \end{aligned}$$

It is vital to remember that the present value is over two periods, hence the single-period PV is squared. Thus, the current call price is €9.71.

Solution to 3:

A is correct. The put option value can be computed simply by applying put–call parity or $p = c + PV(X) - S = 9.71 + [1/(1 + 0.05)]^2 50 - 50 = 5.06$. Thus, the current put price is €5.06.

BINOMIAL MODEL: TWO-PERIOD (PUT OPTIONS)

5

a describe and interpret the binomial option valuation model and its component terms;

b calculate the no-arbitrage values of European and American options using a two-period binomial model;

We now turn to consider American-style options. It is well-known that non-dividend-paying call options on stock will not be exercised early because the minimum price of the option exceeds its exercise value. To illustrate by example, consider a call on a US$100 stock, with an exercise price of US$10 (that is, very deep in the money). Suppose the call is worth its exercise value of only US$90. To get stock exposure, one could fund and pay US$100 to buy the stock, or fund and pay only US$90 for the call and pay the last US$10 at expiration only if the stock is at or above US$100 at that time. Because the latter choice is preferable, the call must be worth more than the US$90 exercise value. Another way of looking at it is that it would make no sense to exercise this call because you do not believe the stock can go any higher and you would thus simply be obtaining a stock that you believe would go no higher. Moreover, the stock would require that you pay far more money than you have tied up in the call. It is always better to just sell the call in this situation because it will be trading for more than the exercise value.

The same is not true for put options. By early exercise of a put, particularly a deep in-the-money put, the sale proceeds can be invested at the risk-free rate and earn interest worth more than the time value of the put. Thus, we will examine how early exercise influences the value of an American-style put option. As we will see, when early exercise has value, the no-arbitrage approach is the only way to value American-style options.

Suppose the periodically compounded interest rate is 3%, the non-dividend-paying underlying stock is currently trading at 72, the exercise price is 75, u = 1.356, d = 0.541, and the put option expires in two years. Exhibit 6 shows the results for a European-style put option.

Exhibit 6 Two-Period Binomial Model for a European-Style Put Option

Item	Value
Underlying	72
Put	18.16876
Hedge Ratio	–0.43029

Item	Value
Underlying	97.632
Put	8.61401
Hedge Ratio	–0.27876

Item	Value
Underlying	38.952
Put	33.86353
Hedge Ratio	–1

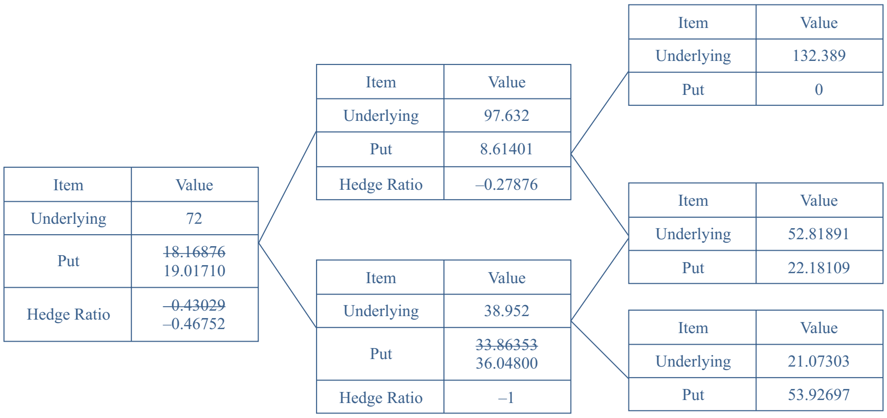
Item	Value
Underlying	132.389
Put	0

Item	Value
Underlying	52.81891
Put	22.18109

Item	Value
Underlying	21.07303
Put	53.92697

The Time 1 down move is of particular interest. The exercise value for this put option is 36.048 [= Max(0,75 – 38.952)]. Therefore, the exercise value is higher than the put value. So, if this same option were American-style, then the option would be worth more exercised than not exercised. Thus, the put option should be exercised. Exhibit 7 illustrates how the analysis changes if this put option were American-style. Clearly, the right to exercise early translates into a higher value.

Exhibit 7 Two-Period Binomial Model for an American-Style Put Option



American-style option valuation requires that one work backward through the binomial tree and address whether early exercise is optimal at each step. In Exhibit 7, the early exercise premium at Time 1 when a down move occurs is 2.18447 (36.048 – 33.86353). Also, if we replace 33.86353 with 36.048—in bold below for emphasis—in the Time 0 calculation, we obtain a put value of

$$p = PV[\pi p^+ + (1 - \pi)p^-] = (1/1.03)[(0.6)8.61401 + (1 - 0.6)\mathbf{36.048}] = 19.02$$

Thus, the early exercise premium at Time 0 is 0.85 (19.02 – 18.17). From this illustration, we see clearly that in a multiperiod setting, American-style put options cannot be valued simply as the present value of the expected future option payouts, as shown in Equation 9. American-style put options can be valued as the present value of the expected future option payout in a single-period setting. Hence, when early exercise is a consideration, we must address the possibility of early exercise as we work backward through the binomial tree.

EXAMPLE 6

Two-Period Binomial American-Style Put Option Valuation

Suppose you are given the following information: S_0 = 26, X = 25, u = 1.466, d = 0.656, n = 2 (time steps), r = 2.05% (per period), and no dividends. The tree is provided in Exhibit 8.

Exhibit 8 Two-Period Binomial American-Style Put Option

Item	Value
Underlying	26
Put	4.01174
Hedge Ratio	–0.35345

Item	Value
Underlying	38.116
Put	0
Hedge Ratio	0

Item	Value
Underlying	17.056
Put	7.44360
Hedge Ratio	–0.99970

Item	Value
Underlying	55.87806
Put	0

Item	Value
Underlying	25.00410
Put	0

Item	Value
Underlying	11.18874
Put	13.81126

The early exercise premium of the above American-style put option is *closest* to:

A 0.27.

B 0.30.

C 0.35.

Solution:

A is correct. The exercise value at Time 1 with a down move is 7.944 [= Max(0,25 – 17.056)]. Thus, we replace this value in the binomial tree and compute the hedge ratio at Time 0. The resulting put option value at Time 0 is thus 4.28143 (see Exhibit 9).

Exhibit 9 Solution

Item	Value
Underlying	26
Put	~~4.01174~~ 4.28143
Hedge Ratio	~~–0.35345~~ –0.37721

Item	Value
Underlying	38.116
Put	0
Hedge Ratio	0

Item	Value
Underlying	17.056
Put	~~7.44360~~ 7.94400
Hedge Ratio	–0.99970

Item	Value
Underlying	55.87806
Put	0

Item	Value
Underlying	25.00410
Put	0

Item	Value
Underlying	11.18874
Put	13.81126

In Exhibit 9, the early exercise premium at Time 1 when a down move occurs is 0.5004 (7.944 – 7.44360). Thus, if we replace 7.44360 with 7.944—in bold below for emphasis—in the Time 0 calculation, we have the put value of

$$p = PV[\pi p^{+} + (1 - \pi)p^{-}] = (1/1.0205)[(0.45)0 + (1 - 0.45)\mathbf{7.944}] = 4.28$$

Thus, the early exercise premium at Time 0 when a down move occurs 0.27 (= 4.28 – 4.01).

6 BINOMIAL MODEL: TWO-PERIOD (ROLE OF DIVIDENDS & COMPREHENSIVE EXAMPLE)

a describe and interpret the binomial option valuation model and its component terms;

We now briefly introduce the role of dividend payments within the binomial model. Our approach here is known as the escrow method. Because dividends lower the value of the stock, a call option holder is hurt. Although it is possible to adjust the option terms to offset this effect, most option contracts do not provide protection against dividends. Thus, dividends affect the value of an option. We assume dividends are perfectly predictable; hence, we split the underlying instrument into two components: the underlying instrument without the known dividends and the known dividends.[6] For example, the current value of the underlying instrument without dividends can be expressed as

$$\hat{S} = S - \gamma$$

where γ denotes the present value of dividend payments. We use the ^ symbol to denote the underlying instrument without dividends. In this case, we model the uncertainty of the stock based on $\hat{S}$ and not S. At expiration, the underlying instrument value is the same, $\hat{S}_T = S_T$, because we assume any dividends have already been paid. The value of an investment in the stock, however, would be $S_T + \gamma_T$, which assumes the dividend payments are reinvested at the risk-free rate.

To illustrate by example, consider a call on a US$100 stock with exercise price of US$95. The periodically compounded interest rate is 1.0%, the stock will pay a US$3 dividend at Time Step 1, u = 1.224, d = 0.796, and the call option expires in two years. Exhibit 10 shows some results for an American-style call option. The computations in Exhibit 10 involve several technical nuances that are beyond the scope of our objectives. The key objective here is to see how dividend-motivated early exercise influences American options.

The Time 1 up move is particularly interesting. At Time 0, the present value of the US$3 dividend payment is US$2.970297 (= 3/1.01). Therefore, 118.7644 = (100 – 2.970297)1.224 is the stock value without dividends at Time 1, assuming an up move occurs. The exercise value for this call option, including dividends, is 26.7644 [= Max(0,118.7644 + 3 – 95)], whereas the value of the call option per the binomial model is 24.9344. In other words, the stock price just before it goes ex-dividend is 118.7644 + 3 = 121.7644, so the option can be exercised for 121.7644 – 95 = 26.7644. If not exercised, the stock drops as it goes ex-dividend and the option becomes worth 24.9344 at the ex-dividend price. Thus, by exercising early, the call buyer acquires the stock just before it goes ex-dividend and thus is able to capture the dividend. If the call is not exercised, the call buyer will not receive this dividend. The American-style call option is worth more than the European-style call option because at Time Step 1 when an up move occurs, the call is exercised early, capturing additional value.

6 The reading focuses on regular, "known" dividends. In the case of large, special dividends, option exchanges may adjust the exercise price.

Exhibit 10 Two-Period Binomial Model for an American-Style Call Option with Dividends

Item	Value
Underlying	100
Call	~~12.3438~~ 13.2497
Hedge Ratio	~~0.6004~~ 0.6445

Item	Value
Underlying	118.7644
Call	~~24.9344~~ 26.7644
Hedge Ratio	0.9909

Item	Value
Underlying	77.2356
Call	0
Hedge Ratio	0

Item	Value
Underlying	145.3676
Call	50.3676

Item	Value
Underlying	94.5364
Call	0

Item	Value
Underlying	61.4796
Call	0

We now provide a comprehensive binomial option valuation example. In this example, we contrast European-style exercise with American-style exercise.

EXAMPLE 7

Comprehensive Two-Period Binomial Option Valuation Model Exercise

Suppose you observe a non-dividend-paying Australian equity trading for A$7.35. The call and put options have two years to mature, the periodically compounded risk-free interest rate is 4.35%, and the exercise price is A$8.0. Based on an analysis of this equity, the estimates for the up and down moves are u = 1.445 and d = 0.715, respectively.

1 Calculate the European-style call and put option values at Time Step 0 and Time Step 1. Describe and interpret your results.

2 Calculate the European-style call and put option hedge ratios at Time Step 0 and Time Step 1. Based on these hedge ratios, interpret the component terms of the binomial option valuation model.

3 Calculate the American-style call and put option values and hedge ratios at Time Step 0 and Time Step 1. Explain how your results differ from the European-style results.

Solution to 1:

The expectations approach requires the following preliminary calculations:

$$
\begin{aligned}
\text{RN probability: } \pi &= [FV(1) - d]/(u - d) \\
&= [(1 + 0.0435) - 0.715]/(1.445 - 0.715) = 0.45 \\
c^{++} &= \text{Max}(0,u^2S - X) \\
&= \text{Max}[0,1.445^2(7.35) - 8.0] = 7.347 \\
c^{+-} &= \text{Max}(0,udS - X) \\
&= \text{Max}[0,1.445(0.715)7.35 - 8.0] = 0 \\
c^{--} &= \text{Max}(0,d^2S - X) \\
&= \text{Max}[0,0.715^2(7.35) - 8.0] = 0 \\
p^{++} &= \text{Max}(0,X - u^2S) \\
&= \text{Max}[0,8.0 - 1.445^2(7.35)] = 0 \\
p^{+-} &= \text{Max}(0,X - udS) \\
&= \text{Max}[0,8.0 - 1.445(0.715)7.35] = 0.406 \\
p^{--} &= \text{Max}(0,X - d^2S) \\
&= \text{Max}[0,8.0 - 0.715^2(7.35)] = 4.24
\end{aligned}
$$

Therefore, at Time Step 1, we have (note that $c_2\big|_1^+$ is read as the call value expiring at Time Step 2 observed at Time Step 1, assuming an up move occurs)

$$E\left(c_2\big|_1^+\right) = \pi c^{++} + (1 - \pi)c^{+-} = 0.45(7.347) + (1 - 0.45)0 = 3.31$$

$$E\left(c_2\big|_1^-\right) = \pi c^{-+} + (1 - \pi)c^{--} = 0.45(0.0) + (1 - 0.45)0.0 = 0.0$$

$$E\left(p_2\big|_1^+\right) = \pi p^{++} + (1 - \pi)p^{+-} = 0.45(0.0) + (1 - 0.45)0.406 = 0.2233$$

$$E\left(p_2\big|_1^-\right) = \pi p^{-+} + (1 - \pi)p^{--} = 0.45(0.406) + (1 - 0.45)4.24 = 2.51$$

Thus, because $PV_{1,2}(1) = 1/(1 + 0.0435) = 0.958313$, we have the Time Step 1 option values of

$$c^+ = PV_{1,2}\left[E\left(c_2\big|_1^+\right)\right] = 0.958313(3.31) = 3.17$$

$$c^- = PV_{1,2}\left[E\left(c_2\big|_1^-\right)\right] = 0.958313(0.0) = 0.0$$

$$p^+ = PV_{1,2}\left[E\left(p_2\big|_1^+\right)\right] = 0.\ 958313(0.2233) = 0.214$$

$$p^- = PV_{1,2}\left[E\left(p_2\big|_1^-\right)\right] = 0.958313(2.51) = 2.41$$

At Time Step 0, we have

$$
\begin{aligned}
E\left(c_2\big|_0\right) &= \pi^2c^{++} + 2\pi(1 - \pi)c^{+-} + (1 - \pi)^2c^{--} \\
&= 0.45^2(7.347) + 2(0.45)(1 - 0.45)0 + (1 - 0.45)^2 0 = 1.488
\end{aligned}
$$

$$
\begin{aligned}
E\left(p_2\big|_0\right) &= \pi^2p^{++} + 2\pi(1 - \pi)p^{+-} + (1 - \pi)^2p^{--} \\
&= 0.45^2(0) + 2(0.45)(1 - 0.45)0.406 + (1 - 0.45)^2 4.24 = 1.484
\end{aligned}
$$

Thus,

$$c = PV_{rf,0,2}\left[E\left(c_2 \middle| 0\right)\right] = 0.91836(1.488) = 1.37 \text{ and}$$

$$p = PV_{rf,0,2}\left[E\left(p_2 \middle| 0\right)\right] = 0.91836(1.484) = 1.36$$

With the two-period binomial model, the call and put values based on the expectations approach are simply the present values of the expected payoffs. The present value of the expected payoffs is based on the risk-free interest rate and the expectations approach is based on the risk-neutral probability. The parameters in this example were selected so that the European-style put and call would have approximately the same value. Notice that the stock price is less than the exercise price by roughly the present value factor or $7.35 = 8.0/1.0435^2$. One intuitive explanation is put–call parity, which can be expressed as $c - p = S - PV(X)$. Thus, if $S = PV(X)$, then $c = p$.

Solution to 2:

The computation of the hedge ratios at Time Step 1 and Time Step 0 will require the option values at Time Step 1 and Time Step 2. The terminal values of the options are given in Solution 1.

$$\begin{aligned}
S^{++} &= u^2S = 1.445^2(7.35) = 15.347 \\
S^{+-} &= udS = 1.445(0.715)7.35 = 7.594 \\
S^{--} &= d^2S = 0.715^2(7.35) = 3.758 \\
S^{+} &= uS = 1.445(7.35) = 10.621 \\
S^{-} &= dS = 0.715(7.35) = 5.255
\end{aligned}$$

Therefore, the hedge ratios at Time 1 are

$$h_c^+ = \frac{c^{++} - c^{+-}}{S^{++} - S^{+-}} = \frac{7.347 - 0.0}{15.347 - 7.594} = 0.9476$$

$$h_c^- = \frac{c^{-+} - c^{--}}{S^{-+} - S^{--}} = \frac{0.0 - 0.0}{7.594 - 3.758} = 0.0$$

$$h_p^+ = \frac{p^{++} - p^{+-}}{S^{++} - S^{+-}} = \frac{0.0 - 0.406}{15.347 - 7.594} = -0.05237$$

$$h_p^- = \frac{p^{-+} - p^{--}}{S^{-+} - S^{--}} = \frac{0.406 - 4.24}{7.594 - 3.758} = -1.0$$

In the last hedge ratio calculation, both put options are in the money (p^{-+} and p^{--}). In this case, the hedge ratio will be –1, subject to a rounding error. We now turn to interpreting the model's component terms. Based on the no-arbitrage approach, we have for the call price, assuming an up move has occurred, at Time Step 1,

$$\begin{aligned}
c^+ &= h_c^+S^+ + PV_{1,2}\left(-h_c^+S^{+-} + c^{+-}\right) \\
&= 0.9476(10.621) + (1/1.0435)[-0.9476(7.594) + 0.0] = 3.1684
\end{aligned}$$

Thus, the call option can be interpreted as a leveraged position in the stock. Specifically, long 0.9476 shares for a cost of 10.0645 [= 0.9476(10.621)] partially financed with a 6.8961 {= (1/1.0435)[–0.9476(7.594) + 0.0]} loan. Note that the loan amount can be found simply as the cost of the position in shares less the option value [6.8961 = 0.9476(10.621) – 3.1684]. Similarly, we have

$$c^- = h_c^- S^- + PV_{1,2}\left(-h_c^- S^{--} + c^{--}\right)$$
$$= 0.0(5.255) + (1/1.0435)[-0.0(3.758) + 0.0] = 0.0$$

Specifically, long 0.0 shares for a cost of 0.0 [= 0.0(5.255)] with no financing. For put options, the interpretation is different. Specifically, we have

$$p^+ = PV_{1,2}\left(-h_p^+ S^{++} + p^{++}\right) + h_p^+ S^+$$
$$= (1/1.0435)[-(-0.05237)15.347 + 0.0] + (-0.05237)10.621 = 0.2140$$

Thus, the put option can be interpreted as lending that is partially financed with a short position in shares. Specifically, short 0.05237 shares for a cost of 0.55622 [= (–0.05237)10.621] with financing of 0.77022 {= (1/1.0435)[–(–0.05237)15.347 + 0.0]}. Note that the lending amount can be found simply as the proceeds from the short sale of shares plus the option value [0.77022 = (0.05237)10.621 + 0.2140]. Again, we have

$$p^- = PV_{1,2}\left(-h_p^- S^{-+} + p^{-+}\right) + h_p^- S^-$$
$$= (1/1.0435)[-(-1.0)7.594 + 0.406] + (-1.0)5.255 = 2.4115$$

Here, we short 1.0 shares for a cost of 5.255 [= (–1.0)5.255] with financing of 7.6665 {= (1/1.0435)[–(–1.0)7.594 + 0.406]}. Again, the lending amount can be found simply as the proceeds from the short sale of shares plus the option value [7.6665 = (1.0)5.255 + 2.4115].

Finally, we have at Time Step 0

$$h_c = \frac{c^+ - c^-}{S^+ - S^-} = \frac{3.1684 - 0}{10.621 - 5.255} = 0.5905$$

$$h_p = \frac{p^+ - p^-}{S^+ - S^-} = \frac{0.2140 - 2.4115}{10.621 - 5.255} = -0.4095$$

The interpretations remain the same at Time Step 0:

$$c = h_c S + PV_{0,1}(-h_c S^- + c^-)$$
$$= 0.5905(7.35) + (1/1.0435)[-0.5905(5.255) + 0.0] = 1.37$$

Here, we are long 0.5905 shares for a cost of 4.3402 [=0.5905(7.35)] partially financed with a 2.97 {= (1/1.0435)[–0.5905(5.255) + 0.0] or = 0.5905(7.35) – 1.37} loan.

$$p = PV_{0,1}(-h_p S^+ + p^+) + h_p S$$
$$= (1/1.0435)\{-[-0.4095(10.621)] + 0.214\} + (-0.4095)7.35 = 1.36$$

Here, we short 0.4095 shares for a cost of 3.01 [= (–0.4095)7.35] with financing of 4.37 (= (1/1.0435){–[–0.4095(10.621)] + 0.214} or = (0.4095)7.35 + 1.36).

Solution to 3:

We know that American-style call options on non-dividend-paying stock are worth the same as European-style call options because early exercise will not occur. Thus, as previously computed, $c^+ = 3.17$, $c^- = 0.0$, and $c = 1.37$. Recall that the call exercise value (denoted with EV) is simply the maximum of zero or the stock price minus the exercise price. We note that the EVs are less than or equal to the call model values; that is,

$$c_{EV}^{+} = \text{Max}(0,S^{+} - X) = \text{Max}(0,10.621 - 8.0) = 2.621\ (< 3.1684)$$

$$c_{EV}^{-} = \text{Max}(0,S^{-} - X) = \text{Max}(0,5.255 - 8.0) = 0.0\ (= 0.0)$$

$$c_{EV} = \text{Max}(0,S - X) = \text{Max}(0,7.35 - 8.0) = 0.0\ (< 1.37)$$

Therefore, the American-style feature for non-dividend-paying stocks has no effect on either the hedge ratio or the option value. The binomial model for American-style calls on non-dividend-paying stocks can be described and interpreted the same as a similar European-style call. This point is consistent with what we said earlier. If there are no dividends, an American-style call will not be exercised early.

This result is not true for puts. We know that American-style put options on non-dividend-paying stock may be worth more than the analogous European-style put options. The hedge ratios at Time Step 1 will be the same as European-style puts because there is only one period left. Therefore, as previously shown, $p^{+} = 0.214$ and $p^{-} = 2.41$.

The put exercise values are

$$p_{EV}^{+} = \text{Max}(0,X - S^{+}) = \text{Max}(0,8.0 - 10.621) = 0\ (< 0.214)$$

$$p_{EV}^{-} = \text{Max}(0,X - S^{-}) = \text{Max}(0,8.0 - 5.255) = 2.745\ (> 2.41)$$

Because the exercise value for the put at Time Step 1, assuming a down move occurred, is greater than the model value, we replace the model value with the exercise value. Hence,

$$p^{-} = 2.745$$

and the hedge ratio at Time Step 0 will be affected. Specifically, we now have

$$h_p = \frac{p^{+} - p^{-}}{S^{+} - S^{-}} = \frac{0.2140 - 2.745}{10.621 - 5.255} = -0.4717$$

and thus the put model value is

$$p = (1/1.0435)[0.45(0.214) + 0.55(2.745)] = 1.54$$

Clearly, the early exercise feature has a significant impact on both the hedge ratio and the put option value in this case. The hedge ratio goes from –0.4095 to –0.4717. The put value is raised from 1.36 to 1.54.

We see through the simple two-period binomial model that an option can be viewed as a position in the underlying with financing. Furthermore, this valuation model can be expressed as the present value of the expected future cash flows, where the expectation is taken under the RN probability and the discounting is at the risk-free rate.

Up to this point, we have focused on equity options. The binomial model can be applied to any underlying instrument though often requiring some modifications. For example, currency options would require incorporating the foreign interest rate. Futures options would require a binomial lattice of the futures prices. Interest rate options, however, require somewhat different tools that we now examine.

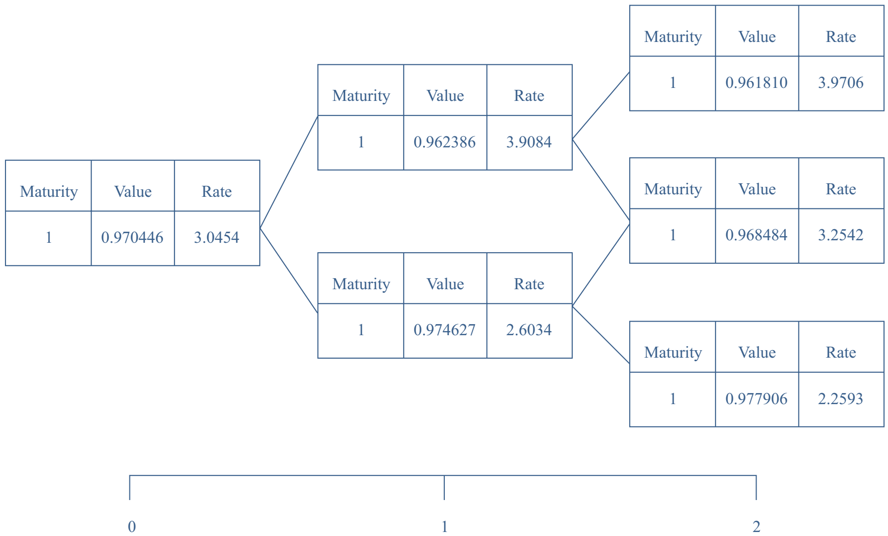
7 INTEREST RATE OPTIONS & MULTIPERIOD MODEL

d calculate and interpret the value of an interest rate option using a two-period binomial model;

In this section, we will briefly illustrate how to value interest rate options. There are a wide variety of approaches to valuing interest rate options. We do not delve into how arbitrage-free interest rate trees are generated. The particular approach used here assumes the RN probability of an up move at each node is 50%.

Exhibit 11 presents a binomial lattice of interest rates covering two years along with the corresponding zero-coupon bond values. The rates are expressed in annual compounding. Therefore, at Time 0, the spot rate is (1.0/0.970446) – 1 or 3.04540%.[7] Note that at Time 1, the value in the column labeled "Maturity" reflects time to maturity not calendar time. The lattice shows the rates on one-period bonds, so all bonds have a maturity of 1. The column labeled "Value" is the value of a zero-coupon bond with the stated maturity based on the rates provided.

Exhibit 11 Two-Year Binomial Interest Rate Lattice by Year



The underlying instrument for interest rate options here is the spot rate. A call option on interest rates will be in the money when the current spot rate is above the exercise rate. A put option on interest rates will be in the money when the current spot rate is below the exercise rate. Thus, based on the notation in the previous section, the current spot rate is denoted S. Option valuation follows the expectations approach discussed in the previous section but taken only one period at a time. The procedure is illustrated with an example.

7 The values in the first box from the left are observed at t = 0. The values in the remainder of the lattice are derived by using a technique that is outside the scope of this reading.

EXAMPLE 8

Option on Interest Rates

This example is based on Exhibit 11. Suppose we seek to value two-year European-style call and put options on the periodically compounded one-year spot interest rate (the underlying). Assume the notional amount of the options is US$1,000,000 and the call and put exercise rate is 3.25% of par. Assume the RN probability is 50% and these option cash settle at Time 2 based on the observed rates.[8]

Solution:

Using the expectations approach introduced in the last section, we have (per US$1) at Time Step 2

$$c^{++} = \text{Max}(0,S^{++} - X) = \text{Max}[0,0.039706 - 0.0325] = 0.007206$$
$$c^{+-} = \text{Max}(0,S^{+-} - X) = \text{Max}[0,0.032542 - 0.0325] = 0.000042$$
$$c^{--} = \text{Max}(0,S^{--} - X) = \text{Max}[0,0.022593 - 0.0325] = 0.0$$
$$p^{++} = \text{Max}(0,X - S^{++}) = \text{Max}[0,0.0325 - 0.039706] = 0.0$$
$$p^{+-} = \text{Max}(0,X - S^{+-}) = \text{Max}[0,0.0325 - 0.032542] = 0.0$$
$$p^{--} = \text{Max}(0,X - S^{--}) = \text{Max}[0,0.0325 - 0.022593] = 0.009907$$

At Time Step 1, we have

$$\begin{aligned} c^{+} &= PV_{1,2}[\pi c^{++} + (1 - \pi)c^{+-}] \\ &= 0.962386[0.5(0.007206) + (1 - 0.5)0.000042] \\ &= 0.003488 \\ c^{-} &= PV_{1,2}[\pi c^{+-} + (1 - \pi)c^{--}] \\ &= 0.974627[0.5(0.000042) + (1 - 0.5)0.0] \\ &= 0.00002 \\ p^{+} &= PV_{1,2}[\pi p^{++} + (1 - \pi)p^{+-}] \\ &= 0.962386[0.5(0.0) + (1 - 0.5)0.0] \\ &= 0.0 \\ p^{-} &= PV_{1,2}[\pi p^{+-} + (1 - \pi)p^{--}] \\ &= 0.974627[0.5(0.0) + (1 - 0.5)0.009907] \\ &= 0.004828 \end{aligned}$$

Notice how the present value factors are different for the up and down moves. At Time Step 1 in the + outcome, we discount by a factor of 0.962386, and in the – outcome, we discount by the factor 0.974627. Because this is an option on interest rates, it should not be surprising that we have to allow the interest rate to vary.

8 In practice, interest rate options usually have a settlement procedure that results in a deferred payoff. The deferred payoff arises from the fact that the underlying interest rate is based on an instrument that pays interest at the end of its life. For the instrument underlying the interest rate, the interest payment occurs after the interest has accrued. To accommodate this reality in this problem, we would have to introduce an instrument that matures at time three. The purpose of this example is merely to illustrate the procedure for rolling backward through an interest rate tree when the underlying is the interest rate. We simplify this example by omitting this deferred settlement. In Section 12, we discuss in detail the deferred settlement procedure and incorporate it into the pricing model.

Therefore, at Time Step 0, we have

$$c = PV_{rf,0,1}[\pi c^{+} + (1 - \pi)c^{-}]$$
$$= 0.970446[0.5(0.003488) + (1 - 0.5)0.00002]$$
$$= 0.00170216$$
$$p = PV_{rf,0,1}[\pi p^{+} + (1 - \pi)p^{-}]$$
$$= 0.970446[0.5(0.0) + (1 - 0.5)0.004828]$$
$$= 0.00234266$$

Because the notional amount is US$1,000,000, the call value is US$1,702.16 [= US$1,000,000(0.00170216)] and the put value is US$2,342.66 [= US$1,000,000(0.00234266)]. The key insight is to just work a two-period binomial model as three one-period binomial models.

We turn now to briefly generalize the binomial model as it leads naturally to the Black–Scholes–Merton option valuation model.

7.1 Multiperiod Model

The multiperiod binomial model provides a natural bridge to the Black–Scholes–Merton option valuation model presented in the next section. The idea is to take the option's expiration and slice it up into smaller and smaller periods. The two-period model divides the expiration into two periods. The three-period model divides expiration into three periods and so forth. The process continues until you have a large number of time steps. The key feature is that each time step is of equal length. Thus, with a maturity of T, if there are n time steps, then each time step is T/n in length.

For American-style options, we must also test at each node whether the option is worth more exercised or not exercised. As in the two-period case, we work backward through the binomial tree testing the model value against the exercise value and always choosing the higher one.

The binomial model is an important and useful methodology for valuing options. The expectations approach can be applied to European-style options and will lead naturally to the BSM model in the next section. This approach simply values the option as the present value of the expected future payoffs, where the expectation is taken under the risk-neutral probability and the discounting is based on the risk-free rate. The no-arbitrage approach can be applied to either European-style or American-style options because it provides the intuition for the fair value of options.

8 BLACK-SCHOLES-MERTON (BSM) OPTION VALUATION MODEL, INTRODUCTION AND ASSUMPTIONS OF THE BSM MODEL

f identify assumptions of the Black–Scholes–Merton option valuation model;

The BSM model, although very complex in its derivation, is rather simple to use and interpret. The objective here is to illustrate several facets of the BSM model with the objective of highlighting its practical usefulness. After a brief introduction, we examine the assumptions of the BSM model and then delve into the model itself.

8.1 Introductory Material

Louis Bachelier published the first known mathematically rigorous option valuation model in 1900. By the late 1960s, there were several published quantitative option models. Fischer Black, Myron Scholes, and Robert Merton introduced the BSM model in 1973 in two published papers, one by Black and Scholes and the other by Merton. The innovation of the BSM model is essentially the no-arbitrage approach introduced in the previous section but applied with a continuous time process, which is equivalent to a binomial model in which the length of the time step essentially approaches zero. It is also consistent with the basic statistical fact that the binomial process with a "large" number of steps converges to the standard normal distribution. Myron Scholes and Robert Merton won the 1997 Nobel Prize in Economics based, in part, on their work related to the BSM model.[9] Let us now examine the BSM model assumptions.

8.2 Assumptions of the BSM Model

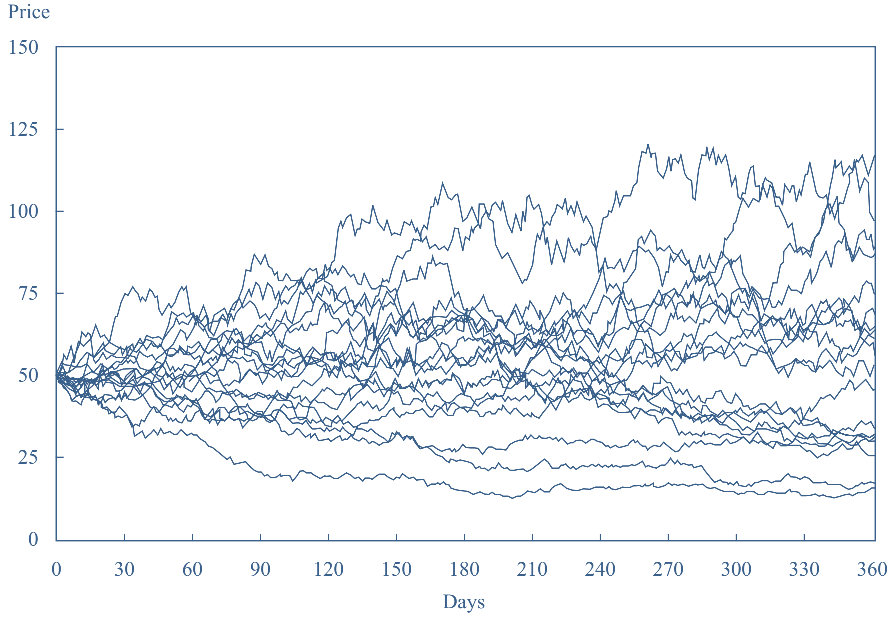
The key assumption for option valuation models is how to model the random nature of the underlying instrument. This characteristic of how an asset evolves randomly is called a stochastic process. Many financial instruments enjoy limited liability; hence, the values of instruments cannot be negative, but they certainly can be zero. In 1900, Bachelier proposed the normal distribution. The key advantages of the normal distribution are that zero is possible, meaning that bankruptcy is allowable, it is symmetric, it is relatively easy to manipulate, and it is additive (which means that sums of normal distributions are normally distributed). The key disadvantage is that negative stock values are theoretically possible, which violates the limited liability principal of stock ownership. Based on research on stock prices in the 1950s and 1960s, a preference emerged for the lognormal distribution, which means that log returns are distributed normally. Black, Scholes, and Merton chose to use the lognormal distribution.

Recall that the no-arbitrage approach requires self-financing and dynamic replication; we need more than just an assumption regarding the terminal distribution of the underlying instrument. We need to model the value of the instrument as it evolves over time, which is what we mean by a stochastic process. The stochastic process chosen by Black, Scholes, and Merton is called geometric Brownian motion (GBM).

Exhibit 12 illustrates GBM, assuming the initial stock price is $S = 50$. We assume the stock will grow at 3% ($\mu = 3\%$ annually, geometrically compounded rate). This GBM process also reflects a random component that is determined by a volatility (σ) of 45%. This volatility is the annualized standard deviation of continuously compounded percentage change in the underlying, or in other words, the log return. Note that as a particular sample path drifts upward, we observe more variability on an absolute basis, whereas when the particular sample path drifts downward, we observe less variability on an absolute basis. For example, examine the highest and lowest lines shown in Exhibit 12. The highest line is much more erratic than the lowest line. Recall that a 10% move in a stock with a price of 100 is 10 whereas a 10% move in a stock with a price of 10 is only 1. Thus, GBM can never hit zero nor go below it. This property is appealing because many financial instruments enjoy limited liability and cannot be negative. Finally, note that although the stock movements are rather erratic, there are no large jumps—a common feature with marketable financial instruments.

9 Fischer Black passed away in 1995 and the Nobel Prize is not awarded posthumously.

Exhibit 12 Geometric Brownian Motion Simulation (S = 50, μ = 3%, σ = 45%)



Within the BSM model framework, it is assumed that all investors agree on the distributional characteristics of GBM except the assumed growth rate of the underlying. This growth rate depends on a number of factors, including other instruments and time. The standard BSM model assumes a constant growth rate and constant volatility.

The specific assumptions of the BSM model are as follows:

- The underlying follows a statistical process called geometric Brownian motion, which implies that the continuously compounded return is normally distributed.
- Geometric Brownian motion implies continuous prices, meaning that the price of underlying instrument does not jump from one value to another; rather, it moves smoothly from value to value.
- The underlying instrument is liquid, meaning that it can be easily bought and sold.
- Continuous trading is available, meaning that in the strictest sense one must be able to trade at every instant.
- Short selling of the underlying instrument with full use of the proceeds is permitted.
- There are no market frictions, such as transaction costs, regulatory constraints, or taxes.
- No arbitrage opportunities are available in the marketplace.
- The options are European-style, meaning that early exercise is not allowed.
- The continuously compounded risk-free interest rate is known and constant; borrowing and lending is allowed at the risk-free rate.
- The volatility of the return on the underlying is known and constant.
- If the underlying instrument pays a yield, it is expressed as a continuous known and constant yield at an annualized rate.

Naturally, the foregoing assumptions are not absolutely consistent with real financial markets, but, as in all financial models, the question is whether they produce models that are tractable and useful in practice, which they do.

EXAMPLE 9

BSM Model Assumptions

Which is the *correct* pair of statements? The BSM model assumes:

A the return on the underlying has a normal distribution. The price of the underlying can jump abruptly to another price.

B brokerage costs are factored into the BSM model. It is impossible to trade continuously.

C volatility can be predicted with certainty. Arbitrage is non-existent in the marketplace.

Solution:

C is correct. All four of the statements in A and B are incorrect within the BSM model paradigm.

BSM MODEL: COMPONENTS

9

g interpret the components of the Black–Scholes–Merton model as applied to call options in terms of a leveraged position in the underlying;

We turn now to a careful examination of the BSM model.

The BSM model is a continuous time version of the discrete time binomial model. Given that the BSM model is based on continuous time, it is customary to use a continuously compounded interest rate rather than some discretely compounded alternative. Thus, when an interest rate is used here, denoted simply as r, we mean solely the annualized continuously compounded rate.[10] The volatility, denoted as σ, is also expressed in annualized percentage terms. Initially, we focus on a non-dividend-paying stock. The BSM model, with some adjustments, applies to other underlying instruments, which will be examined later.

The BSM model for stocks can be expressed as

$$c = SN(d_1) - e^{-rT}XN(d_2) \quad (10)$$

and

$$p = e^{-rT}XN(-d_2) - SN(-d_1) \quad (11)$$

where

$$d_1 = \frac{\ln(S/X) + \left(r + \sigma^2/2\right)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

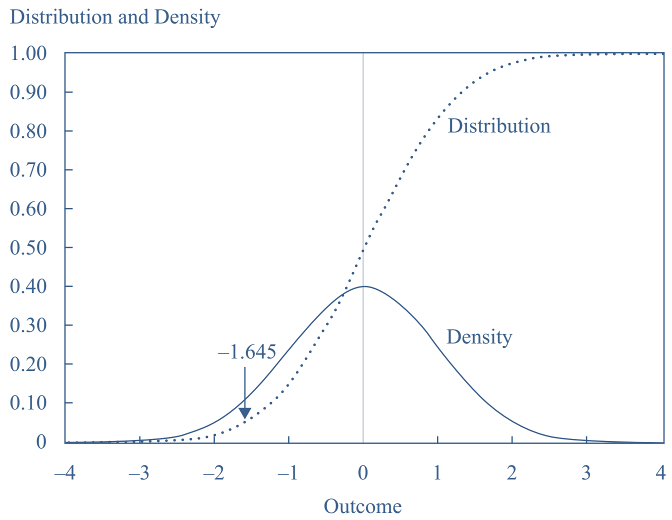
10 Note $e^r = 1 + r_d$, where r_d is the annually compounded rate.

N(x) denotes the standard normal cumulative distribution function, which is the probability of obtaining a value of less than x based on a standard normal distribution. In our context, x will have the value of d_1 or d_2. N(x) reflects the likelihood of observing values less than x from a random sample of observations taken from the standard normal distribution.

Although the BSM model appears very complicated, it has straightforward interpretations that will be explained. N(x) can be estimated by a computer program or a spreadsheet or approximated from a lookup table. The normal distribution is a symmetric distribution with two parameters, the mean and standard deviation. The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1.

Exhibit 13 illustrates the standard normal probability density function (the standard bell curve) and the cumulative distribution function (the accumulated probability and range of 0 to 1). Note that even though GBM is lognormally distributed, the N(x) functions in the BSM model are based on the standard normal distribution. In Exhibit 13, we see that if x = –1.645, then N(x) = N(–1.645) = 0.05. Thus, if the model value of d is –1.645, the corresponding probability is 5%. Clearly, values of d that are less than 0 imply values of N(x) that are less than 0.5. As a result of the symmetry of the normal distribution, we note that N(–x) = 1 – N(x).

Exhibit 13 Standard Normal Distribution



The BSM model can be described as the present value of the expected option payoff at expiration. Specifically, we can express the BSM model for calls as $c = PV_r[E(c_T)]$ and for puts as $p = PV_r[E(p_T)]$, where $E(c_T) = Se^{rT}N(d_1) - XN(d_2)$ and $E(p_T) = XN(-d_2) - Se^{rT}N(-d_1)$. The present value term in this context is simply e^{-rT}. As with most valuation tasks in finance, the value today is simply the present value of the expected future cash flows. It is important to note that the expectation is based on the risk-neutral probability measure defined in Section 3. The expectation is not based on the investor's subjective beliefs, which reflect an aversion to risk. Also, the present value function is based on the risk-free interest rate not on the investor's required return on invested capital, which of course is a function of risk.

Alternatively, the BSM model can be described as having two components: a stock component and a bond component. For call options, the stock component is $SN(d_1)$ and the bond component is $e^{-rT}XN(d_2)$. The BSM model call value is the stock component minus the bond component. For put options, the stock component is $SN(-d_1)$ and the bond component is $e^{-rT}XN(-d_2)$. The BSM model put value is the bond component minus the stock component.

The BSM model can be interpreted as a dynamically managed portfolio of the stock and zero-coupon bonds.[11] The goal is to replicate the option payoffs with stocks and bonds. For both call and put options, we can represent the initial cost of this replicating strategy as

Replicating strategy cost = $n_S S + n_B B$

where the equivalent number of underlying shares is $n_S = N(d_1) > 0$ for calls and $n_S = -N(-d_1) < 0$ for puts. The equivalent number of bonds is $n_B = -N(d_2) < 0$ for calls and $n_B = N(-d_2) > 0$ for puts. The price of the zero-coupon bond is $B = e^{-rT}X$. Note, if n is positive, we are buying the underlying and if n is negative we are selling (short selling) the underlying. The cost of the portfolio will exactly equal either the BSM model call value or the BSM model put value.

For calls, we are simply buying stock with borrowed money because $n_S > 0$ and $n_B < 0$. Again the cost of this portfolio will equal the BSM model call value, and if appropriately rebalanced, then this portfolio will replicate the payoff of the call option. Therefore, a call option can be viewed as a leveraged position in the stock.

Similarly, for put options, we are simply buying bonds with the proceeds from short selling the underlying because $n_S < 0$ and $n_B > 0$. The cost of this portfolio will equal the BSM model put value, and if appropriately rebalanced, then this portfolio will replicate the payoff of the put option. Note that a short position in a put will result in receiving money today and $n_S > 0$ and $n_B < 0$. Therefore, a short put can be viewed as an over-leveraged or over-geared position in the stock because the borrowing exceeds 100% of the cost of the underlying.

Exhibit 14 illustrates the direct comparison between the no-arbitrage approach to the single-period binomial option valuation model and the BSM option valuation model. The parallel between the h term in the binomial model and $N(d_1)$ is easy to see. Recall that the term hedge ratio was used with the binomial model because we were creating a no-arbitrage portfolio. Note for call options, $-N(d_2)$ implies borrowing money or short selling $N(d_2)$ shares of a zero-coupon bond trading at $e^{-rT}X$. For put options, $N(-d_2)$ implies lending money or buying $N(-d_2)$ shares of a zero-coupon bond trading at $e^{-rT}X$.

Exhibit 14 BSM and Binomial Option Valuation Model Comparison

	Call Option		Put Option	
Option Valuation Model Terms	**Underlying**	**Financing**	**Underlying**	**Financing**
Binomial Model	hS	$PV(-hS^- + c^-)$	hS	$PV(-hS^- + p^-)$
BSM Model	$N(d_1)S$	$-N(d_2)e^{-rT}X$	$-N(-d_1)S$	$N(-d_2)e^{-rT}X$

If the value of the underlying, S, increases, then the value of $N(d_1)$ also increases because S has a positive effect on d_1. Thus, the replicating strategy for calls requires continually buying shares in a rising market and selling shares in a falling market.

11 When covering the binomial model, the bond component was generically termed financing. This component is typically handled with bank borrowing or lending. With the BSM model, it is easier to understand as either buying or short selling a risk-free zero-coupon bond.

Within the BSM model theory, the aggregate losses from this "buy high/sell low" strategy, over the life of the option, adds up exactly to the BSM model option premium received for the option at inception.[12] This result must be the case; otherwise there would be arbitrage profits available. Because transaction costs are not, in fact, zero, the frequent rebalancing by buying and selling the underlying adds significant costs for the hedger. Also, markets can often move discontinuously, contrary to the BSM model's assumption that prices move continuously, thus allowing for continuous hedging adjustments. Hence, in reality, hedges are imperfect. For example, if a company announces a merger, then the company's stock price may jump substantially higher, contrary to the BSM model's assumption.

In addition, volatility cannot be known in advance. For these reasons, options are typically more expensive than they would be as predicted by the BSM model theory. In order to continue using the BSM model, the volatility parameter used in the formula is usually higher (by, say, 1% or 2%, but this can vary a lot) than the volatility of the stock actually expected by market participants. We will ignore this point for now, however, as we focus on the mechanics of the model.

EXAMPLE 10

Illustration of BSM Model Component Interpretation

Suppose we are given the following information on call and put options on a stock: S = 100, X = 100, r = 5%, T = 1.0, and σ = 30%. Thus, based on the BSM model, it can be demonstrated that PV(X) = 95.123, d_1 = 0.317, d_2 = 0.017, $N(d_1)$ = 0.624, $N(d_2)$ = 0.507, $N(-d_1)$ = 0.376, $N(-d_2)$ = 0.493, c = 14.23, and p = 9.35.

1. The initial trading strategy required by the no-arbitrage approach to replicate the call option payoffs for a buyer of the option is:
   - **A** buy 0.317 shares of stock and short sell –0.017 shares of zero-coupon bonds.
   - **B** buy 0.624 shares of stock and short sell 0.507 shares of zero-coupon bonds.
   - **C** short sell 0.317 shares of stock and buy 0.017 shares of zero-coupon bonds.
2. Identify the initial trading strategy required by the no-arbitrage approach to replicate the put option payoffs for a buyer of the put.
   - **A** Buy 0.317 shares of stock and short sell –0.017 shares of zero-coupon bonds.
   - **B** Buy 0.624 shares of stock and short sell 0.507 shares of zero-coupon bonds.
   - **C** Short sell 0.376 shares of stock and buy 0.493 shares of zero-coupon bonds.

Solution to 1:

B is correct. The no-arbitrage approach to replicating the call option involves purchasing $n_S = N(d_1)$ = 0.624 shares of stock partially financed with $n_B = -N(d_2)$ = –0.507 shares of zero-coupon bonds priced at $B = Xe^{-rT}$ = 95.123 per bond.

12 The validity of this claim does not rest on the validity of the BSM model assumptions; rather the validity depends only on whether the BSM model accurately predicts the replication cost.

Note that by definition the cost of this replicating strategy is the BSM call model value or $n_S S + n_B B = 0.624(100) + (-0.507)95.123 = 14.17$. Without rounding errors, the option value is 14.23.

Solution to 2:

C is correct. The no-arbitrage approach to replicating the put option is similar. In this case, we trade $n_S = -N(-d_1) = -0.376$ shares of stock—specifically, short sell 0.376 shares—and buy $n_B = N(-d_2) = 0.493$ shares of zero-coupon bonds. Again, the cost of the replicating strategy is $n_S S + n_B B = -0.376(100) + (0.493)95.123 = 9.30$. Without rounding errors, the option value is 9.35. Thus, to replicate a call option based on the BSM model, we buy stock on margin. To replicate a put option, we short the stock and buy zero-coupon bonds.

Note that the $N(d_2)$ term has an additional important interpretation. It is a unique measure of the probability that the call option expires in the money, and correspondingly, $1 - N(d_2) = N(-d_2)$ is the probability that the put option expires in the money. Specifically, the probability based on the RN probability of being in the money, not one's own estimate of the probability of being in the money nor the market's estimate. That is, $N(d_2) = \text{Prob}(S_T > X)$ based on the unique RN probability.

BSM MODEL: CARRY BENEFITS AND APPLICATIONS

10

h describe how the Black–Scholes–Merton model is used to value European options on equities and currencies;

We now turn to incorporating various carry benefits into the BSM model. Carry benefits include dividends for stock options, foreign interest rates for currency options, and coupon payments for bond options. For other underlying instruments, there are carry costs that can easily be treated as negative carry benefits, such as storage and insurance costs for agricultural products. Because the BSM model is established in continuous time, it is common to model these carry benefits as a continuous yield, denoted generically here as γ^c or simply γ.

The BSM model requires a few adjustments to accommodate carry benefits. The carry benefit-adjusted BSM model is

$$c = Se^{-\gamma T}N(d_1) - e^{-rT}XN(d_2) \tag{12}$$

and

$$p = e^{-rT}XN(-d_2) - Se^{-\gamma T}N(-d_1) \tag{13}$$

where

$$d_1 = \frac{\ln(S/X) + \left(r - \gamma + \sigma^2/2\right)T}{\sigma\sqrt{T}}$$

Note that d_2 can be expressed again simply as $d_2 = d_1 - \sigma\sqrt{T}$. The value of a put option can also be found based on the carry benefit-adjusted put–call parity:

$$p + Se^{-\gamma T} = c + e^{-rT}X \tag{14}$$

The carry benefit-adjusted BSM model can again be described as the present value of the expected option payoff at expiration. Now, however, $E(c_T) = Se^{(r-\gamma)T}N(d_1) - XN(d_2)$ and $E(p_T) = XN(-d_2) - Se^{(r-\gamma)T}N(-d_1)$. The present value term remains simply e^{-rT}. Carry benefits will have the effect of lowering the expected future value of the underlying

Again, the carry benefit adjusted BSM model can be described as having two components, a stock component and a bond component. For call options, the stock component is $Se^{-\gamma T}N(d_1)$ and the bond component is again $e^{-rT}XN(d_2)$. For put options, the stock component is $Se^{-\gamma T}N(-d_1)$ and the bond component is again $e^{-rT}XN(-d_2)$. Although both d_1 and d_2 are reduced by carry benefits, the general approach to valuation remains the same. An increase in carry benefits will lower the value of the call option and raise the value of the put option.

Note that $N(d_2)$ term continues to be interpreted as the RN probability of a call option being in the money. The existence of carry benefits has the effect of lowering d_1 and d_2, hence the probability of being in the money with call options declines as the carry benefit rises. This RN probability is an important element to describing how the BSM model is used in various valuation tasks.

For stock options, $\gamma = \delta$, which is the continuously compounded dividend yield. The dividend-yield BSM model can again be interpreted as a dynamically managed portfolio of the stock and zero coupon bonds. Based on the call model above applied to a dividend yielding stock, the equivalent number of units of stock is now $n_S = e^{-\delta T}N(d_1) > 0$ and the equivalent number of units of bonds remains $n_B = -N(d_2) < 0$. Similarly with puts, the equivalent number of units of stock is now $n_S = -e^{-\delta T}N(-d_1) < 0$ and the equivalent number of units of bonds again remains $n_B = N(-d_2) > 0$.

With dividend paying stocks, the arbitrageur is able to receive the benefits of dividend payments when long the stock and has to pay dividends when short the stock. Thus, the burden of carrying the stock is diminished for a long position. The key insight is that dividends influence the dynamically managed portfolio by lowering the number of shares to buy for calls and raising the number of shares to short sell for puts. Higher dividends will lower the value of d_1, thus lowering $N(d_1)$. Also, higher dividends will lower the number of bonds to short sell for calls and raise the number of bonds to buy for puts.

EXAMPLE 11

BSM Model Applied to Equities

Suppose we are given the following information on an underlying stock and options: S = 60, X = 60, r = 2%, T = 0.5, δ = 2%, and σ = 45%. Assume we are examining European-style options.

1 Which answer *best* describes how the BSM model is used to value a call option with the parameters given?

A The BSM model call value is the exercise price times $N(d_1)$ less the present value of the stock price times $N(d_2)$.

B The BSM model call value is the stock price times $e^{-\delta T}N(d_1)$ less the exercise price times $e^{-rT}N(d_2)$.

C The BSM model call value is the stock price times $e^{-\delta T}N(-d_1)$ less the present value of the exercise price times $e^{-rT}N(-d_2)$.

2 Which answer *best* describes how the BSM model is used to value a put option with the parameters given?

A The BSM model put value is the exercise price times $N(d_1)$ less the present value of the stock price times $N(d_2)$.

B The BSM model put value is the exercise price times $e^{-\delta T}N(-d_2)$ less the stock price times $e^{-rT}N(-d_2)$.

C The BSM model put value is the exercise price times $e^{-rT}N(-d_2)$ less the stock price times $e^{-\delta T}N(-d_1)$.

3 Suppose now that the stock does not pay a dividend—that is, δ = 0%. Identify the correct statement.

A The BSM model option value is the same as the previous problems because options are not dividend adjusted.

B The BSM model option values will be different because there is an adjustment term applied to the exercise price, that is $e^{-\delta T}$, which will influence the option values.

C The BSM model option value will be different because d_1, d_2, and the stock component are all adjusted for dividends.

Solution to 1:

B is correct. The BSM call model for a dividend-paying stock can be expressed as $Se^{-\delta T}N(d_1) - Xe^{-rT}N(d_2)$.

Solution to 2:

C is correct. The BSM put model for a dividend-paying stock can be expressed as $Xe^{-rT}N(-d_2) - Se^{-\delta T}N(-d_1)$.

Solution to 3:

C is correct. The BSM model option value will be different because d_1, d_2, and the stock component are all adjusted for dividends.

EXAMPLE 12

How the BSM Model Is Used to Value Stock Options

Suppose that we have some Bank of China shares that are currently trading on the Hong Kong Stock Exchange at HKD4.41. Our view is that the Bank of China's stock price will be steady for the next three months, so we decide to sell some three-month out-of-the-money calls with exercise price at 4.60 in order to enhance our returns by receiving the option premium. Risk-free government securities are paying 1.60% and the stock is yielding HKD 0.24%. The stock volatility is 28%. We use the BSM model to value the calls.

Which statement is correct? The BSM model inputs (underlying, exercise, expiration, risk-free rate, dividend yield, and volatility) are:

A 4.60, 4.41, 3, 0.0160, 0.0024, and 0.28.

B 4.41, 4.60, 0.25, 0.0160, 0.0024, and 0.28.

C 4.41, 4.41, 0.3, 0.0160, 0.0024, and 0.28.

Solution:

B is correct. The spot price of the underlying is HKD4.41. The exercise price is HKD4.60. The expiration is 0.25 years (three months). The risk-free rate is 0.016. The dividend yield is 0.0024. The volatility is 0.28.

For foreign exchange options, $\gamma = r^f$, which is the continuously compounded foreign risk-free interest rate. When quoting an exchange rate, we will give the value of the domestic currency per unit of the foreign currency. For example, Japanese yen (¥) per unit of the euro (€) will be expressed as the euro trading for ¥135 or succinctly 135¥/€. This is called the foreign exchange spot rate. Thus, the foreign currency, the euro, is expressed in terms of the Japanese yen, which is in this case the domestic currency. This is logical, for example, when a Japanese firm would want to express its foreign euro holdings in terms of its domestic currency, Japanese yen.

With currency options, the underlying instrument is the foreign exchange spot rate. Again, the carry benefit is the interest rate in the foreign country because the foreign currency could be invested in the foreign country's risk-free instrument. Also, with currency options, the underlying and the exercise price must be quoted in the same currency unit. Lastly, the volatility in the model is the volatility of the log return of the spot exchange rate. Each currency option is for a certain quantity of foreign currency, termed the notional amount, a concept analogous to the number of shares of stock covered in an option contract. The total cost of the option would be obtained by multiplying the formula value by the notional amount in the same way that one would multiply the formula value of an option on a stock by the number of shares the option contract covers.

The BSM model applied to currencies can be described as having two components, a foreign exchange component and a bond component. For call options, the foreign exchange component is $Se^{-r^f T}N(d_1)$ and the bond component is $e^{-rT}XN(d_2)$, where r is the domestic risk-free rate. The BSM call model applied to currencies is simply the foreign exchange component minus the bond component. For put options, the foreign exchange component is $Se^{-r^f T}N(-d_1)$ and the bond component is $e^{-rT}XN(-d_2)$. The BSM put model applied to currencies is simply the bond component minus the foreign exchange component. Remember that the underlying is expressed in terms of the domestic currency.

EXAMPLE 13

BSM Model Applied to Value Options on Currency

A Japanese camera exporter to Europe has contracted to receive fixed euro (€) amounts each quarter for his goods. The spot price of the currency pair is 135¥/€. If the exchange rate falls to, say, 130¥/€, then the yen will have strengthened because it will take fewer yen to buy one euro. The exporter is concerned that the yen will strengthen because in this case, his forthcoming fixed euro will buy fewer yen. Hence, the exporter is considering buying an at-the-money spot euro put option to protect against this fall; this in essence is a call on yen. The Japanese risk-free rate is 0.25% and the European risk-free rate is 1.00%.

1 What are the underlying and exercise prices to use in the BSM model to get the euro put option value?

A 1/135; 1/135

B 135; 135

C 135; 130

2 What are the risk-free rate and the carry rate to use in the BSM model to get the euro put option value?

A 0.25%; 1.00%

B 0.25%; 0.00%

C 1.00%; 0.25%

Solution to 1:

B is correct. The underlying is the spot FX price of 135 ¥/€. Because the put is at-the-money spot, the exercise price equals the spot price.

Solution to 2:

A is correct. The risk-free rate to use is the Japanese rate because the Japanese yen is the domestic currency unit per the exchange rate quoting convention. The carry rate is the foreign currency's risk-free rate, which is the European rate.

BLACK OPTION VALUATION MODEL AND EUROPEAN OPTIONS ON FUTURES 11

i. describe how the Black model is used to value European options on futures;

We turn now to examine a modification of the BSM model when the underlying is a forward or futures contract.

In 1976, Fischer Black introduced a modified version of the BSM model approach that is applicable to options on underlying instruments that are costless to carry, such as options on futures contracts—for example, equity index futures—and options on forward contracts. The latter include interest rate-based options, such as caps, floors, and swaptions.

11.1 European Options on Futures

We assume that the futures price also follows geometric Brownian motion. We ignore issues like margin requirements and marking to market. Black proposed the following model for European-style futures options:

$$c = e^{-rT}[F_0(T)N(d_1) - XN(d_2)] \tag{15}$$

and

$$p = e^{-rT}[XN(-d_2) - F_0(T)N(-d_1)] \tag{16}$$

where

$$d_1 = \frac{\ln\left[F_0(T)/X\right] + \left(\sigma^2/2\right)T}{\sigma\sqrt{T}} \text{ and}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

Note that $F_0(T)$ denotes the futures price at Time 0 that expires at Time T, and σ denotes the volatility related to the futures price. The other terms are as previously defined. Black's model is simply the BSM model in which the futures contract is assumed to reflect the carry arbitrage model. Futures option put–call parity can be expressed as

$$c = e^{-rT}[F_0(T) - X] + p \tag{17}$$

As we have seen before, put–call parity is a useful tool for describing the valuation relationship between call and put values within various option valuation models.

The Black model can be described in a similar way to the BSM model. The Black model has two components, a futures component and a bond component. For call options, the futures component is $F_0(T)e^{-rT}N(d_1)$ and the bond component is again $e^{-rT}XN(d_2)$. The Black call model is simply the futures component minus the bond component. For put options, the futures component is $F_0(T)e^{-rT}N(-d_1)$ and the bond component is again $e^{-rT}XN(-d_2)$. The Black put model is simply the bond component minus the futures component.

Alternatively, futures option valuation, based on the Black model, is simply computing the present value of the difference between the futures price and the exercise price. The futures price and exercise price are appropriately adjusted by the N(d) functions. For call options, the futures price is adjusted by $N(d_1)$ and the exercise price is adjusted by $-N(d_2)$ to arrive at difference. For put options, the futures price is adjusted by $-N(-d_1)$ and the exercise price is adjusted by $+N(-d_2)$.

EXAMPLE 14

European Options on Futures Index

The S&P 500 Index (a spot index) is presently at 1,860 and the 0.25 expiration futures contract is trading at 1,851.65. Suppose further that the exercise price is 1,860, the continuously compounded risk-free rate is 0.2%, time to expiration is 0.25, volatility is 15%, and the dividend yield is 2.0%. Based on this information, the following results are obtained for options on the futures contract.[13]

Options on Futures	
Calls	**Puts**
$N(d_1)$ =0.491	$N(-d_1)$ = 0.509
$N(d_2)$ = 0.461	$N(-d_2)$ = 0.539
c = US$51.41	p = US$59.76

1 Identify the statement that *best* describes how the Black model is used to value a European call option on the futures contract just described.

A The call value is the present value of the difference between the exercise price times 0.461 and the current futures price times 0.539.

B The call value is the present value of the difference between the current futures price times 0.491 and the exercise price times 0.461.

C The call value is the present value of the difference between the current spot price times 0.491 and the exercise price times 0.461.

2 Which statement *best* describes how the Black model is used to value a European put options on the futures contract just described?

A The put value is the present value of the difference between the exercise price times 0.539 and the current futures price times 0.509.

B The put value is the present value of the difference between the current futures price times 0.491 and the exercise price times 0.461.

C The put value is the present value of the difference between the current spot price times 0.491 and the exercise price times 0.461.

13 We ignore the effect of the multiplier. As of this writing, the S&P 500 futures option contract has a multiplier of 250. The prices reported here have not been scaled up by this amount. In practice, the option cost would by 250 times the option value.

3 What are the underlying and exercise prices to use in the Black futures option model?

A 1,851.65; 1,860

B 1,860; 1,860

C 1,860; 1,851.65

Solution to 1:

B is correct. Recall Black's model for call options can be expressed as $c = e^{-rT}[F_0(T)N(d_1) - XN(d_2)]$.

Solution to 2:

A is correct. Recall Black's model for put options can be expressed as $p = e^{-rT}[XN(-d_2) - F_0(T)N(-d_1)]$.

Solution to 3:

A is correct. The underlying is the futures price of 1,851.65 and the exercise price was given as 1,860.

INTEREST RATE OPTIONS

12

j describe how the Black model is used to value European interest rate options and European swaptions;

With interest rate options, the underlying instrument is a reference interest rate, such as three-month Libor. An interest rate call option gains when the reference interest rate rises and an interest rate put option gains when the reference interest rate falls. Interest rate options are the building blocks of many other instruments.

For an interest rate call option on three-month Libor with one year to expiration, the underlying interest rate is a forward rate agreement (FRA) rate that expires in one year. This FRA is observed today and is the underlying rate used in the Black model. The underlying rate of the FRA is a 3-month Libor deposit that is investable in 12 months and matures in 15 months. Thus, in one year, the FRA rate typically converges to the three-month spot Libor.

Interest rates are typically set in advance, but interest payments are made in arrears, which is referred to as advanced set, settled in arrears. For example, with a bank deposit, the interest rate is usually set when the deposit is made, say t_{j-1}, but the interest payment is made when the deposit is withdrawn, say t_j. The deposit, therefore, has $t_m = t_j - t_{j-1}$ time until maturity. Thus, the rate is advanced set, but the payment is settled in arrears. Likewise with a floating rate loan, the rate is usually set and the interest accrues at this known rate, but the payment is made later. Similarly, with some interest rate options, the time to option expiration (t_{j-1}) when the interest rate is set does not correspond to the option settlement (t_j) when the cash payment is made, if any. For example, if an interest rate option payment based on three-month Libor is US$5,000 determined on January 15th, the actual payment of the US$5,000 would occur on April 15.

Interest rates are quoted on an annual basis, but the underlying implied deposit is often less than a year. Thus, the annual rates must be adjusted for the accrual period. Recall that the accrual period for a quarterly reset 30/360 day count FRA is 0.25 (= 90/360). If the day count is on an actual (ACT) number of days divided by 360 (ACT/360), then the accrual period may be something like 0.252778 (= 91/360), assuming 91 days in the period. Typically, the accrual period in FRAs is based on

30/360 whereas the accrual period based on the option is actual number of days in the contract divided by the actual number of days in the year (identified as ACT/ACT or ACT/365).

The model presented here is known as the standard market model and is a variation of Black's futures option valuation model. Again, let t_{j-1} denote the time to option expiration (ACT/365), whereas let t_j denote the time to the maturity date of the underlying FRA. Note that the interest accrual on the underlying begins at the option expiration (Time t_{j-1}). Let $FRA(0,t_{j-1},t_m)$ denote the fixed rate on a FRA at Time 0 that expires at Time t_{j-1}, where the underlying matures at Time t_j (= $t_{j-1} + t_m$), with all times expressed on an annual basis. We assume the FRA is 30/360 day count. For example, FRA(0,0.25,0.5) = 2% denotes the 2% fixed rate on a forward rate agreement that expires in 0.25 years with settlement amount being paid in 0.75 (= 0.25 + 0.5) years.[14] Let R_X denote the exercise rate expressed on an annual basis. Finally, let σ denote the interest rate volatility. Specifically, σ is the annualized standard deviation of the continuously compounded percentage change in the underlying FRA rate.

Interest rate options give option buyers the right to certain cash payments based on observed interest rates. For example, an interest rate call option gives the call buyer the right to a certain cash payment when the underlying interest rate exceeds the exercise rate. An interest rate put option gives the put buyer the right to a certain cash payment when the underlying interest rate is below the exercise rate.

With the standard market model, the prices of interest rate call and put options can be expressed as

$$c = (AP)e^{-r(t_{j-1}+t_m)}\left[FRA(0,t_{j-1},t_m)N(d_1) - R_X N(d_2)\right] \quad (18)$$

and

$$p = (AP)e^{-r(t_{j-1}+t_m)}\left[R_X N(-d_2) - FRA(0,t_{j-1},t_m)N(-d_1)\right] \quad (19)$$

where

AP denotes the accrual period in years

$$d_1 = \frac{\ln\left[FRA(0,t_{j-1},t_m)/R_X\right] + (\sigma^2/2)t_{j-1}}{\sigma\sqrt{t_{j-1}}}$$

$$d_2 = d_1 - \sigma\sqrt{t_{j-1}}$$

The formulas here give the value of the option for a notional amount of 1. In practice, the notional would be more than one, so the full cost of the option is obtained by multiplying these formula amounts by the notional amount. Of course, this point is just the same as finding the value of an option on a single share of stock and then multiplying that value by the number of shares covered by the option contract.

Immediately, we note that the standard market model requires an adjustment when compared with the Black model for the accrual period. In other words, a value such as $FRA(0,t_{j-1},t_m)$ or the strike rate, R_X, as appearing in the formula given earlier, is stated on an annual basis, as are interest rates in general. The actual option premium would have to be adjusted for the accrual period. After accounting for this adjustment, this model looks very similar to the Black model, but there are important but subtle differences. First, the discount factor, $e^{-r(t_{j-1}+t_m)}$, does not apply to the option expiration, t_{j-1}. Rather, the discount factor is applied to the maturity date of the FRA or

14 Note that in other contexts the time periods are expressed in months. For example with months, this FRA would be expressed as FRA(0,3,6). Note that the third term in parentheses denotes the maturity of the underlying deposit from the expiration of the FRA.

t_j (= $t_{j-1} + t_m$). We express this maturity as ($t_{j-1} + t_m$) rather than t_j to emphasize the settlement in arrears nature of this option. Second, rather than the underlying being a futures price, the underlying is an interest rate, specifically a forward rate based on a forward rate agreement or FRA($0,t_{j-1},t_m$). Third, the exercise price is really a rate and reflects an interest rate, not a price. Fourth, the time to the option expiration, t_{j-1}, is used in the calculation of d_1 and d_2. Finally, both the forward rate and the exercise rate should be expressed in decimal form and not as percent (for example, 0.02 and not 2.0). Alternatively, if expressed as a percent, then the notional amount adjustment could be divided by 100.

As with other option models, the standard market model can be described as simply the present value of the expected option payoff at expiration. Specifically, we can express the standard market model for calls as $c = PV[E(c_{tj})]$ and for puts as $p = PV[E(p_{tj})]$, where $E(c_{tj}) = (AP)[FRA(0,t_{j-1},t_m)N(d_1) - R_XN(d_2)]$ and $E(p_{tj}) = (AP)[R_XN(-d_2) - FRA(0,t_{j-1},t_m)N(-d_1)]$. The present value term in this context is simply $e^{-rt_j} = e^{-r\left(t_{j-1}+t_m\right)}$. Again, note we discount from Time t_j, the time when the cash flows are settled on the FRA.

There are several interesting and useful combinations that can be created with interest rate options. We focus on a few that will prove useful for understanding swaptions in the next section. First, if the exercise rate is selected so as to equal the current FRA rate, then long an interest rate call option and short an interest rate put option is equivalent to a receive-floating, pay-fixed FRA.

Second, if the exercise rate is again selected so it is equal to the current FRA rate, then long an interest rate put option and short an interest rate call option is equivalent to a receive-fixed, pay-floating FRA. Note that FRAs are the building blocks of interest rate swaps.

Third, an interest rate cap is a portfolio or strip of interest rate call options in which the expiration of the first underlying corresponds to the expiration of the second option and so forth. The underlying interest rate call options are termed caplets. Thus, a set of floating-rate loan payments can be hedged with a long position in an interest rate cap encompassing a series of interest rate call options.

Fourth, an interest rate floor is a portfolio or strip of interest rate put options in which the expiration of the first underlying corresponds with the expiration of the second option and so forth. The underlying interest rate put options are termed floorlets. Thus, a floating-rate bond investment or any other floating-rate lending situation can be hedged with an interest rate floor encompassing a series of interest rate put options.

Fifth, applying put–call parity as discussed earlier, long an interest rate cap and short an interest rate floor with the exercise prices set at the swap rate is equivalent to a receive-floating, pay-fixed swap. On a settlement date, when the underlying rate is above the strike, both the cap and the swap pay off to the party. When the underlying rate is below the strike on a settlement date, the party must make a payment on the short floor, just as the case with a swap. For the opposite position, long an interest rate floor and short an interest rate cap result in the party making a payment when the underlying rate is above the strike and receiving one when the underlying rate is below the strike, just as is the case for a pay-floating, receive-fixed swap.

Finally, if the exercise rate is set equal to the swap rate, then the value of the cap must be equal to the value of the floor at the start. When an interest rate swap is initiated, its current value is zero and is known as an at-market swap. When an exercise rate is selected such that the cap value equals the floor value, then the initial cost of being long a cap and short the floor is also zero. This occurs when the cap and floor strike are equal to the swap rate.

EXAMPLE 15

European Interest Rate Options

Suppose you are a speculative investor in Singapore. On 15 May, you anticipate that some regulatory changes will be enacted, and you want to profit from this forecast. On 15 June, you intend to borrow 10,000,000 Singapore dollars to fund the purchase of an asset, which you expect to resell at a profit three months after purchase, say on 15 September. The current three-month Sibor (that is, Singapore Libor) is 0.55%. The appropriate FRA rate over the period of 15 June to 15 September is currently 0.68%. You are concerned that rates will rise, so you want to hedge your borrowing risk by purchasing an interest rate call option with an exercise rate of 0.60%.

1 In using the Black model to value this interest rate call option, what would the underlying rate be?

   A 0.55%

   B 0.68%

   C 0.60%

2 The discount factor used in pricing this option would be over what period of time?

   A 15 May–15 June

   B 15 June–15 September

   C 15 May–15 September

Solution to 1:

B is correct. In using the Black model, a forward or futures price is used as the underlying. This approach is unlike the BSM model in which a spot price is used as the underlying.

Solution to 2:

C is correct. You are pricing the option on 15 May. An option expiring 15 June when the underlying is three-month Sibor will have its payoff determined on 15 June, but the payment will be made on 15 September. Thus, the expected payment must be discounted back from 15 September to 15 May.

Interest rate option values are linked in an important way with interest rate swap values through caps and floors. As we will see in the next section, an interest rate swap serves as the underlying for swaptions. Thus, once again, we see that important links exist between interest rate options, swaps, and swaptions.

13 SWAPTIONS

j describe how the Black model is used to value European interest rate options and European swaptions;

A swap option or swaption is simply an option on a swap. It gives the holder the right, but not the obligation, to enter a swap at the pre-agreed swap rate—the exercise rate. Interest rate swaps can be either receive fixed, pay floating or receive floating, pay fixed. A payer swaption is an option on a swap to pay fixed, receive floating. A receiver swaption is an option on a swap to receive fixed, pay floating. Note that the terms "call" and "put" are often avoided because of potential confusion over the nature of the underlying. Notice also that the terminology focuses on the fixed swap rate.

A payer swaption buyer hopes the fixed rate goes up before the swaption expires. When exercised, the payer swaption buyer is able to enter into a pay-fixed, receive-floating swap at the predetermined exercise rate, R_X. The buyer can then immediately enter an offsetting at-market receive-fixed, pay-floating swap at the current fixed swap rate. The floating legs of both swaps will offset, leaving the payer swaption buyer with an annuity of the difference between the current fixed swap rate and the swaption exercise rate. Thus, swaption valuation will reflect an annuity.

Swap payments are advanced set, settled in arrears. Let the swap reset dates be expressed as t_0, t_1, t_2, ..., t_n. Let R_{FIX} denote the fixed swap rate starting when the swaption expires, denoted as before with T, quoted on an annual basis, and R_X denote the exercise rate starting at Time T, again quoted on an annual basis. As before, we will assume a notional amount of 1.

Because swap rates are quoted on an annual basis, let AP denote the accrual period. Finally, we need some measure of uncertainty. Let σ denote the volatility of the forward swap rate. More precisely, σ denotes annualized, standard deviation of the continuously compounded percentage changes in the forward swap rate.

The swaption model presented here is a modification of the Black model. Let the present value of an annuity matching the forward swap payment be expressed as

$$PVA = \sum_{j=1}^{n} PV_{0,t_j}(1)$$

This term is equivalent to what is sometimes referred to as an annuity discount factor. It applies here because a swaption creates a series of equal payments of the difference in the market swap rate at expiration and the chosen exercise rate. Therefore, the payer swaption valuation model is

$$PAY_{SWN} = (AP)PVA[R_{FIX}N(d_1) - R_XN(d_2)] \tag{20}$$

and the receiver swaption valuation model

$$REC_{SWN} = (AP)PVA[R_XN(-d_2) - R_{FIX}N(-d_1)] \tag{21}$$

where

$$d_1 = \frac{\ln(R_{FIX}/R_X) + (\sigma^2/2)T}{\sigma\sqrt{T}}, \text{ and as always,}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

As noted with interest rate options, the actual premium would need to be scaled by the notional amount. Once again, we can see the similarities to the Black model. We note that the swaption model requires two adjustments, one for the accrual period and one for the present value of an annuity. After accounting for these adjustments, this model looks very similar to the Black model but there are important subtle differences. First, the discount factor is absent. The payoff is not a single payment but a series of payments. Thus, the present value of an annuity used here embeds the option-related discount factor. Second, rather than the underlying being a futures price, the underlying is the fixed rate on a forward interest rate swap. Third, the exercise price is really expressed as an interest rate. Finally, both the forward swap rate and the exercise rate should be expressed in decimal form and not as percent (for example, 0.02 and not 2.0).

As with other option models, the swaption model can be described as simply the present value of the expected option payoff at expiration. Specifically, we can express the payer swaption model value as

$$PAY_{SWN} = PV[E(PAY_{SWN,T})]$$

and the receiver swaption model value as

$$REC_{SWN} = PV[E(REC_{SWN,T})],$$

where

$$E(PAY_{SWN,T}) = e^{rT}PAY_{SWN} \text{ and}$$
$$E(REC_{SWN,T}) = e^{rT}REC_{SWN}.$$

The present value term in this context is simply e^{-rT}. Because the annuity term embedded the discounting over the swaption life, the expected swaption values are the current swaption values grossed up by the current risk-free interest rate.

Alternatively, the swaption model can be described as having two components, a swap component and a bond component. For payer swaptions, the swap component is $(AP)PVA(R_{FIX})N(d_1)$ and the bond component is $(AP)PVA(R_X)N(d_2)$. The payer swaption model value is simply the swap component minus the bond component. For receiver swaptions, the swap component is $(AP)PVA(R_{FIX})N(-d_1)$ and the bond component is $(AP)PVA(R_X)N(-d_2)$. The receiver swaption model value is simply the bond component minus the swap component.

As with nearly all derivative instruments, there are many useful equivalence relationships. Recall that long an interest rate cap and short an interest rate floor with the same exercise rate is equal to a receive-floating, pay-fixed interest rate swap. Also, short an interest rate cap and long an interest rate floor with the same exercise rate is equal to a pay-floating, receive-fixed interest rate swap. There are also equivalence relationships with swaptions. In a similar way, long a receiver swaption and short a payer swaption with the same exercise rate is equivalent to entering a receive-fixed, pay-floating forward swap. Long a payer swaption and short a receiver swaption with the same exercise rate is equivalent to entering a receive-floating, pay-fixed forward swap. Note that if the exercise rate is selected such that the receiver and payer swaptions have the same value, then the exercise rate is equal to the at-market forward swap rate. Thus, there is again a put–call parity relationship important for valuation.

In addition, being long a callable fixed-rate bond can be viewed as being long a straight fixed-rate bond and short a receiver swaption. A receiver swaption gives the buyer the right to receive a fixed rate. Hence, the seller will have to pay the fixed rate when this right is exercised in a lower rate environment. Recall that the bond issuer has the right to call the bonds. If the bond issuer sells a receiver swaption with similar terms, then the bond issuer has essentially converted the callable bond into a straight bond. The bond issuer will now pay the fixed rate on the underlying swap and the floating rate received will be offset by the floating-rate loan created when the bond was refinanced. Specifically, the receiver swaption buyer will benefit when rates fall and the swaption is exercised. Thus, the embedded call feature is similar to a receiver swaption.

EXAMPLE 16

European Swaptions

Suppose you are an Australian company and have ongoing floating-rate debt. You have profited for some time by paying at a floating rate because rates have been falling steadily for the last few years. Now, however, you are concerned that within three months the Australian central bank may tighten its monetary policy and your debt costs will thus increase. Rather than lock in your borrowing via a swap, you prefer to hedge by buying a swaption expiring in three months, whereby you will have the choice, but not the obligation, to enter a five-year swap locking in your borrowing costs. The current three-month forward, five-year swap rate is 2.65%. The current five-year swap rate is 2.55%. The current three-month risk-free rate is 2.25%.

With reference to the Black model to value the swaption, which statement is correct?

A The underlying is the three-month forward, five-year swap rate.

B The discount rate to use is 2.55%.

C The swaption time to expiration, T, is five years.

Solution:

A is correct. The current five-year swap rate is not used as a discount rate with swaptions. The swaption time to expiration is 0.25, not the life of the swap.

OPTION GREEKS AND IMPLIED VOLATILITY: DELTA

14

k interpret each of the option Greeks;

l describe how a delta hedge is executed;

With option valuation models, such as the binomial model, BSM model, and Black's model, we are able to estimate a wide array of comparative information, such as how much the option value will change for a small change in a particular parameter.[15] We will explore this derived information as well as implied volatility in this section. These topics are essential for those managing option positions and in general in obtaining a solid understanding of how option prices change. Our discussion will be based on stock options, though the material covered in this section applies to all types of options.

The measures examined here are known as the Greeks and include, delta, gamma, theta, vega, and rho. With these calculations, we seek to address how much a particular portfolio will change for a given small change in the appropriate parameter. These measures are sometimes referred to as static risk measures in that they capture movements in the option value for a movement in one of the factors that affect the option value, while holding all other factors constant.

Our focus here is on European stock options in which the underlying stock is assumed to pay a dividend yield (denoted δ). Note that for non-dividend-paying stocks, δ = 0.

14.1 Delta

Delta is defined as the change in a given instrument for a given small change in the value of the stock, holding everything else constant. Thus, the delta of long one share of stock is by definition +1.0, and the delta of short one share of stock is by definition

15 Parameters in the BSM model, for example, include the stock price, exercise price, volatility, time to expiration, and the risk-free interest rate.

−1.0. The concept of the option delta is similarly the change in an option value for a given small change in the value of the underlying stock, holding everything else constant. The option deltas for calls and puts are, respectively,

$$\text{Delta}_c = e^{-\delta T}N(d_1) \tag{22}$$

and

$$\text{Delta}_p = -e^{-\delta T}N(-d_1) \tag{23}$$

Note that the deltas are a simple function of $N(d_1)$. The delta of an option answers the question of how much the option will change for a given change in the stock, holding everything else constant. Therefore, delta is a static risk measure. It does not address how likely this particular change would be. Recall that $N(d_1)$ is a value taken from the cumulative distribution function of a standard normal distribution. As such, the range of values is between 0 and 1. Thus, the range of call delta is 0 and $e^{-\delta T}$ and the range of put delta is $-e^{-\delta T}$ and 0. As the stock price increases, the call option goes deeper in the money and the value of $N(d_1)$ is moving toward 1. As the stock price decreases, the call option goes deeper out of the money and the value of $N(d_1)$ is moving toward zero. When the option gets closer to maturity, the delta will drift either toward 0 if it is out of the money or drift toward 1 if it is in the money. Clearly, as the stock price changes and as time to maturity changes, the deltas are also changing.

Delta hedging an option is the process of establishing a position in the underlying stock of a quantity that is prescribed by the option delta so as to have no exposure to very small moves up or down in the stock price. Hence, to execute a single option delta hedge, we first calculate the option delta and then buy or sell delta units of stock. In practice, rarely does one have only one option position to manage. Thus, in general, delta hedging refers to manipulating the underlying portfolio delta by appropriately changing the positions in the portfolio. A delta neutral portfolio refers to setting the portfolio delta all the way to zero. In theory, the delta neutral portfolio will not change in value for small changes in the stock instrument. Let N_H denote the number of units of the hedging instrument and Delta_H denote the delta of the hedging instrument, which could be the underlying stock, call options, or put options. Delta neutral implies the portfolio delta plus $N_H\text{Delta}_H$ is equal to zero. The optimal number of hedging units, N_H, is

$$N_H = -\frac{\text{Portfolio delta}}{\text{Delta}_H}$$

Note that if N_H is negative, then one must short the hedging instrument, and if N_H is positive, then one must go long the hedging instrument. Clearly, if the portfolio is options and the hedging instrument is stock, then we will buy or sell shares to offset the portfolio position. For example, if the portfolio consists of 100,000 shares of stock at US$10 per share, then the portfolio delta is 100,000. The delta of the hedging instrument, stock, is +1. Thus, the optimal number of hedging units, N_H, is −100,000 (= −100,000/1) or short 100,000 shares. Alternatively, if the portfolio delta is 5,000 and a particular call option with delta of 0.5 is used as the hedging instrument, then to arrive at a delta neutral portfolio, one must sell 10,000 call options (= −5,000/0.5). Alternatively, if a portfolio of options has a delta of −1,500, then one must buy 1,500 shares of stock to be delta neutral [= −(−1,500)/1]. If the hedging instrument is stock, then the delta is +1 per share.

EXAMPLE 17

Delta Hedging

Apple stock is trading at US$125. We write calls (that is, we sell calls) on 1,000 Apple shares and now are exposed to an increase in the price of the Apple stock. That is, if Apple rises, we will lose money because the calls we sold will go up in value, so our liability will increase. Correspondingly, if Apple falls, we will make money. We want to neutralize our exposure to Apple. Say the call delta is 0.50, which means that if Apple goes up by US$0.10, a call on one Apple share will go up US$0.05. We need to trade in such a way as to make money if Apple goes up, to offset our exposure. Hence, we buy 500 Apple shares to hedge. Now, if Apple goes up US$0.10, the sold calls will go up US$50 (our liability goes up), but our long 500 Apple hedge will profit by US$50. Hence, we are delta hedged.

Identify the *incorrect* statement:

A If we sell Apple puts, we need to buy Apple stock to delta hedge.

B Call delta is non-negative (≥ 0); put delta is non-positive (≤ 0).

C Delta hedging is the process of neutralizing exposure to the underlying.

Solution:

A is the correct answer because statement A is incorrect. If we sell puts, we need to short sell stock to delta hedge.

One final interpretation of option delta is related to forecasting changes in option prices. Let $\hat{c}$, $\hat{p}$, and $\hat{S}$ denote some new value for the call, put, and stock. Based on an approximation method, the change in the option price can be estimated with a concept known as a delta approximation or

$$\hat{c} - c \cong \text{Delta}_c\left(\hat{S} - S\right) \text{ for calls and}$$

$$\hat{p} - p \cong \text{Delta}_p\left(\hat{S} - S\right) \text{ for puts.}[16]$$

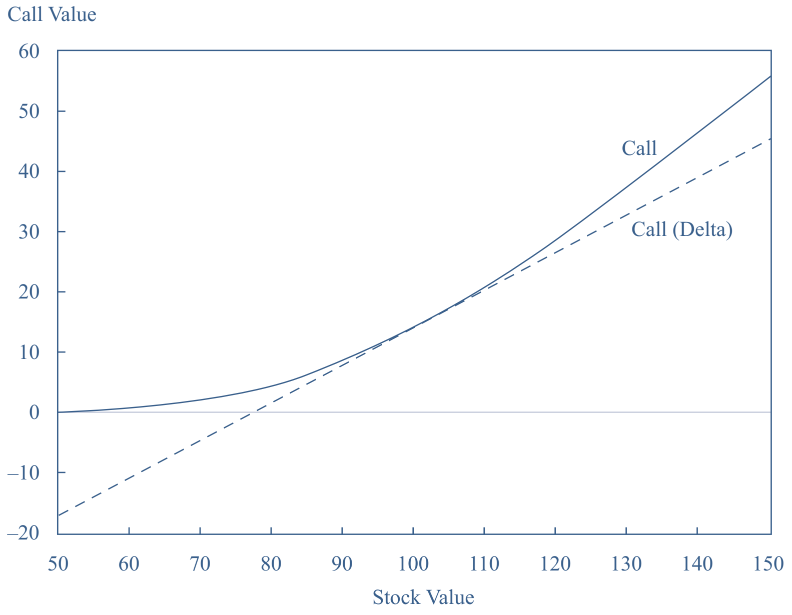
We can now illustrate the actual call values as well as the estimated call values based on delta. Exhibit 15 illustrates the call value based on the BSM model and the call value based on the delta approximation,

$$\hat{c} = c + \text{Delta}_c\left(\hat{S} - S\right)$$

Notice for very small changes in the stock, the delta approximation is fairly accurate. For example, if the stock value rises from 100 to 101, notice that both the call line and the call (delta) estimated line are almost the same value. If, however, the stock value rises from 100 to 150, the call line is now significantly above the call (delta) estimated line. Thus, we see that as the change in the stock increases, the estimation error also increases. The delta approximation is biased low for both a down move and an up move.

16 The symbol ≅ denotes approximately. The approximation method is known as a Taylor series. Also note that the put delta is non-positive (≤ 0).

Exhibit 15 Call Values and Delta Estimated Call Values (S = 100 = X, r = 5%, σ = 30%, δ = 0)



We see that delta hedging is imperfect and gets worse as the underlying moves further away from its original value of 100. Based on the graph, the BSM model assumption of continuous trading is essential to avoid hedging risk. This hedging risk is related to the difference between these two lines and the degree to which the underlying price experiences large changes.

EXAMPLE 18

Delta Hedging

Suppose we know S = 100, X = 100, r = 5%, T = 1.0, σ = 30%, and δ = 5%. We have a short position in put options on 10,000 shares of stock. Based on this information, we note $\text{Delta}_c = 0.532$, and $\text{Delta}_p = -0.419$. Assume each stock option contract is for one share of stock.

1 The appropriate delta hedge, assuming the hedging instrument is stock, is executed by which of the following transactions? Select the *closest* answer.
   - A Buy 5,320 shares of stock.
   - B Short sell 4,190 shares of stock.
   - C Buy 4,190 shares of stock.

2 The appropriate delta hedge, assuming the hedging instrument is calls, is executed by which of the following transactions? Select the *closest* answer.
   - A Sell 7,876 call options.
   - B Sell 4,190 call options.
   - C Buy 4,190 call options.

3 Identify the correct interpretation of an option delta.
   - A Option delta measures the curvature in the option price with respect to the stock price.

B Option delta is the change in an option value for a given small change in the stock's value, holding everything else constant.

C Option delta is the probability of the option expiring in the money.

Solution to 1:

B is correct. Recall that $N_H = -\frac{\text{Portfolio delta}}{\text{Delta}_H}$. The put delta is given as –0.419, thus the short put delta is 0.419. In this case, Portfolio delta = 10,000(0.419) = 4,190 and Delta_H = 1.0. Thus, the number of number of hedging units is –4,190 [= –(4,190/1)] or short sell 4,190 shares of stock.

Solution to 2:

A is correct. Again the Portfolio delta = 4,190 but now Delta_H = 0.532. Thus, the number of hedging units is –7,875.9 [= –(4,190/0.532)] or sell 7,876 call options.

Solution to 3:

B is correct. Delta is defined as the change in a given portfolio for a given small change in the stock's value, holding everything else constant. Option delta is defined as the change in an option value for a given small change in the stock's value, holding everything else constant.

15 GAMMA

k interpret each of the option Greeks;

m describe the role of gamma risk in options trading;

Recall that delta is a good approximation of how an option price will change for a small change in the stock. For larger changes in the stock, we need better accuracy. **Gamma** is defined as the change in a given instrument's delta for a given small change in the stock's value, holding everything else constant. Option gamma is similarly defined as the change in a given option delta for a given small change in the stock's value, holding everything else constant. Option gamma is a measure of the curvature in the option price in relationship to the stock price. Thus, the gamma of a long or short position in one share of stock is zero because the delta of a share of stock never changes. A stock always moves one-for-one with itself. Thus, its delta is always +1 and, of course, –1 for a short position in the stock. The gamma for a call and put option are the same and can be expressed as

$$\text{Gamma}_c = \text{Gamma}_p = \frac{e^{-\delta T}}{S\sigma\sqrt{T}}n(d_1) \qquad (24)$$

where $n(d_1)$ is the standard normal probability density function. The lowercase "n" is distinguished from the cumulative normal distribution—which the density function generates—and that we have used elsewhere in this reading denoted by uppercase "N". The gamma of a call equals the gamma of a similar put based on put–call parity or $c - p = S_0 - e^{-rT}X$. Note that neither S_0 nor $e^{-rT}X$ is a direct function of delta. Hence, the right-hand side of put–call parity has a delta of 1. Thus, the right-hand side delta is not sensitive to changes in the underlying. Therefore, the gamma of a call must equal the gamma of a put.

Gamma is always non-negative. Gamma takes on its largest value near at the money. Options deltas do not change much for small changes in the stock price if the option is either deep in or deep out of the money. Also, as the stock price changes and as time to expiration changes, the gamma is also changing.

Gamma measures the rate of change of delta as the stock changes. Gamma approximates the estimation error in delta for options because the option price with respect to the stock is non-linear and delta is a linear approximation. Thus, gamma is a risk measure; specifically, gamma measures the non-linearity risk or the risk that remains once the portfolio is delta neutral. A gamma neutral portfolio implies the gamma is zero. For example, gamma can be managed to an acceptable level first and then delta is neutralized as a second step. This hedging approach is feasible because options have gamma but a stock does not. Thus, in order to modify gamma, one has to include additional option trades in the portfolio. Once the revised portfolio, including any new option trades, has the desired level of gamma, then the trader can get the portfolio delta to its desired level as step two. To alter the portfolio delta, the trader simply buys or sells stock. Because stock has a positive delta, but zero gamma, the portfolio delta can be brought to its desired level with no impact on the portfolio gamma.

One final interpretation of gamma is related to improving the forecasted changes in option prices. Again, let $\hat{c}$, $\hat{p}$, and $\hat{S}$ denote new values for the call, put, and stock. Again based on an approximation method, the change in the option price can be estimated by a delta-plus-gamma approximation or

$$\hat{c} - c \approx \text{Delta}_c\left(\hat{S} - S\right) + \frac{\text{Gamma}_c}{2}\left(\hat{S} - S\right)^2 \text{ for calls and}$$

$$\hat{p} - p \approx \text{Delta}_p\left(\hat{S} - S\right) + \frac{\text{Gamma}_p}{2}\left(\hat{S} - S\right)^2 \text{ for puts.}$$

Exhibit 16 illustrates the call value based on the BSM model; the call value based on the delta approximation,

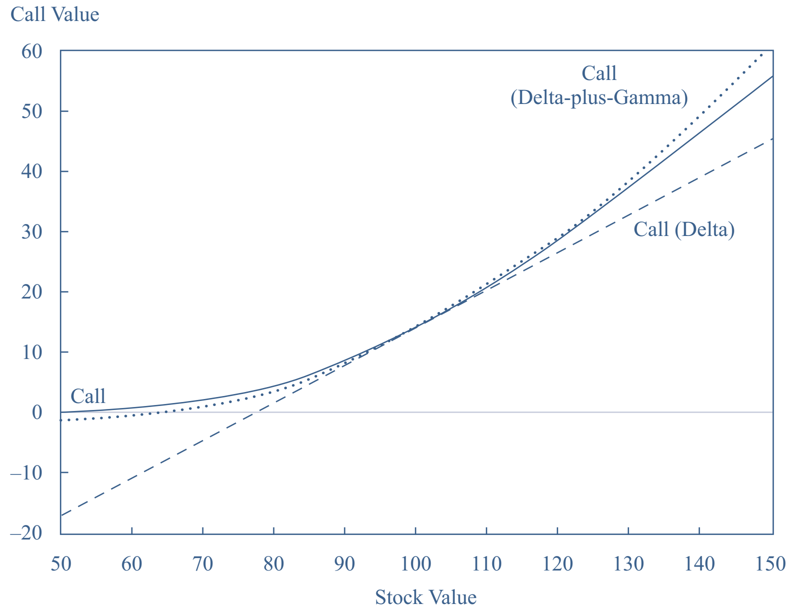
$$\hat{c} = c + \text{Delta}_c\left(\hat{S} - S\right)$$

and the call value based on the delta-plus-gamma approximation,

$$\hat{c} = c + \text{Delta}_c\left(\hat{S} - S\right) + \frac{\text{Gamma}_c}{2}\left(\hat{S} - S\right)^2$$

Notice again that for very small changes in the stock, the delta approximation and the delta-plus-gamma approximations are fairly accurate. If the stock value rises from 100 to 150, the call line is again significantly above the delta estimated line but is below the delta-plus-gamma estimated line. Importantly, the call delta-plus-gamma estimated line is significantly closer to the BSM model call values. Thus, we see that even for fairly large changes in the stock, the delta-plus-gamma approximation is accurate. As the change in the stock increases, the estimation error also increases. From Exhibit 16, we see the delta-plus-gamma approximation is biased low for a down move but biased high for an up move. Thus, when estimating how the call price changes when the underlying changes, we see how the delta-plus-gamma approximation is an improvement when compared with using the delta approximation on its own.

Exhibit 16 Call Values, Delta Estimated Call Values, and Delta-Plus-Gamma Estimated Call Values (S = 100 = X, r = 5%, σ = 30%, δ = 0)



If the BSM model assumptions hold, then we would have no risk in managing option positions. In reality, however, stock prices often jump rather than move continuously and smoothly, which creates "gamma risk." Gamma risk is so-called because gamma measures the risk of stock prices jumping when hedging an option position, and thus leaving us suddenly unhedged.

EXAMPLE 19

Gamma Risk in Option Trading

Suppose we are options traders and have only one option position—a short call option. We also hold some stock such that we are delta hedged. Which one of the following statements is true?

A We are gamma neutral.

B Buying a call will increase our overall gamma.

C Our overall position is a positive gamma, which will make large moves profitable for us, whether up or down.

Solution:

B is correct. Buying options (calls or puts) will always increase net gamma. A is incorrect because we are short gamma, not gamma neutral. C is also incorrect because we are short gamma. We can only become gamma neutral from a short gamma position by purchasing options.

16 THETA

k interpret each of the option Greeks;

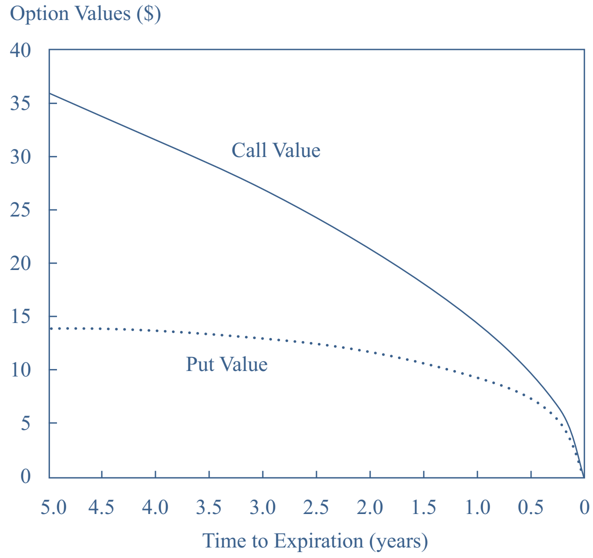
Theta is defined as the change in a portfolio for a given small change in calendar time, holding everything else constant. Option theta is similarly defined as the change in an option value for a given small change in calendar time, holding everything else constant. Option theta is the rate at which the option time value declines as the option approaches expiration. To understand theta, it is important to remember the "holding everything else constant" assumption. Specifically, the theta calculation assumes nothing changes except calendar time. Clearly, if calendar time passes, then time to expiration declines. Because stocks do not have an expiration date, the stock theta is zero. Like gamma, theta cannot be adjusted with stock trades.

The gain or loss of an option portfolio in response to the mere passage of calendar time is known as time decay. Particularly with long options positions, often the mere passage of time without any change in other variables, such as the stock, will result is significant losses in value. Therefore, investment managers with significant option positions carefully monitor theta and their exposure to time decay. Time decay is essentially the measure of profit and loss of an option position as time passes, holding everything else constant.

Note that theta is fundamentally different from delta and gamma in the sense that the passage of time does not involve any uncertainty. There is no chance that time will go backward. Time marches on, but it is important to understand how your investment position will change with the mere passage of time.

Typically, theta is negative for options. That is, as calendar time passes, expiration time declines and the option value also declines. Exhibit 17 illustrates the option value with respect to time to expiration. Remember, as calendar time passes, the time to expiration declines. Both the call and the put option are at the money and eventually are worthless if the stock does not change. Notice, however, how the speed of the option value decline increases as time to expiration decreases.

Exhibit 17 Option Values and Time to Expiration ($S = 100 = X$, $r = 5\%$, $\sigma = 30\%$, $\delta = 0$)



VEGA

17

k interpret each of the option Greeks;

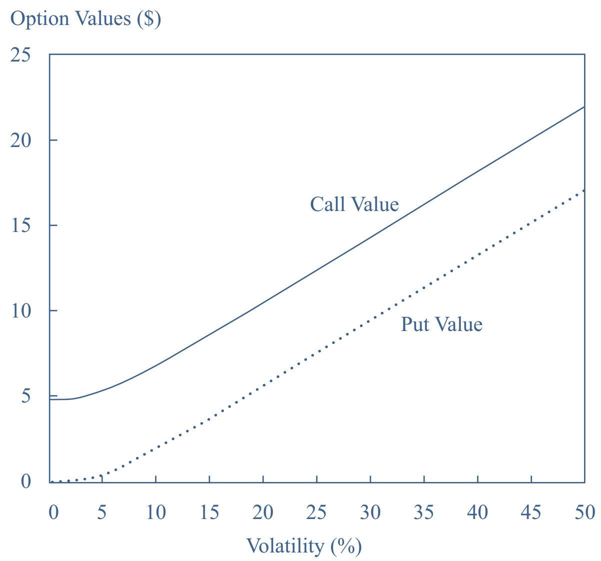
Vega is defined as the change in a given portfolio for a given small change in volatility, holding everything else constant. Vega measures the sensitivity of a given portfolio to volatility. The vega of an option is positive. An increase in volatility results in an increase in the option value for both calls and puts.

The vega of a call equals the vega of a similar put based on put–call parity or $c - p = S_0 - e^{-rT}X$. Note that neither S_0 nor $e^{-rT}X$ is a direct function of volatility. Therefore, the vega of a call must offset the vega of a put so that the vega of the right-hand side is zero.

Unlike the Greeks we have already discussed, vega is based on an unobservable parameter, future volatility. Although historical volatility can be calculated, there is no objective measure of future volatility. Similar to the concept of expected value, future volatility is subjective. Thus, vega measures the sensitivity of a portfolio to changes in the volatility used in the option valuation model. Option values are generally quite sensitive to volatility. In fact, of the five variables in the BSM, an option's value is most sensitive to volatility changes.

At extremely low volatility, the option values tend toward their lower bounds. The lower bound of a European-style call option is zero or the stock less the present value of the exercise price, whichever is greater. The lower bound of a European-style put option is zero or the present value of the exercise price less the stock, whichever is greater. Exhibit 18 illustrates the option values with respect to volatility. In this case, the call lower bound is 4.88 and the put lower bound is 0. The difference between the call and put can be explained by put–call parity.

Exhibit 18 Option Values and Volatility ($S = 100 = X$, $r = 5\%$, $T = 1$, $\delta = 0$)



Vega is very important in managing an options portfolio because option values can be very sensitive to volatility changes. Vega is high when options are at or near the money. Volatility is usually only hedged with other options and volatility itself can be quite volatile. Volatility is sometimes considered a separate asset class or a separate risk factor. Because it is rather exotic and potentially dangerous, exposure to volatility needs to be managed, bearing in mind that risk managers, board members, and clients may not understand or appreciate losses if volatility is the source.

18 RHO

k interpret each of the option Greeks;

Rho is defined as the change in a given portfolio for a given small change in the risk-free interest rate, holding everything else constant. Thus, rho measures the sensitivity of the portfolio to the risk-free interest rate.

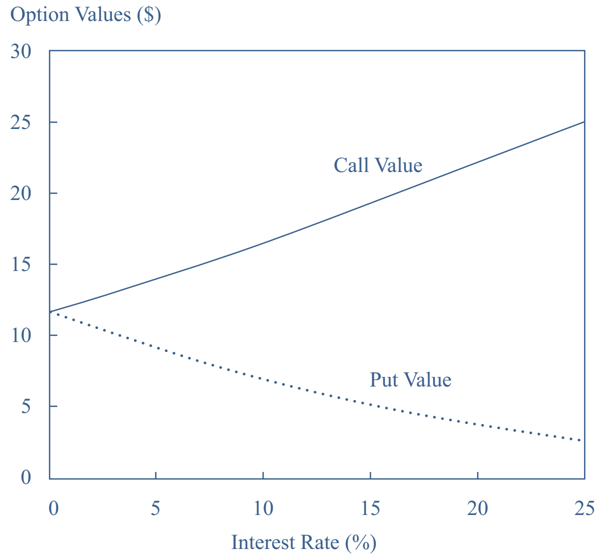
The rho of a call is positive. Intuitively, buying an option avoids the financing costs involved with purchasing the stock. In other words, purchasing a call option allows an investor to earn interest on the money that otherwise would have gone to purchasing the stock. The higher the interest rate, the higher the call value.

The rho of a put is negative. Intuitively, the option to sell the stock delays the opportunity to earn interest on the proceeds from the sale. For example, purchasing a put option rather than selling the stock deprives an investor of the potential interest that would have been earned from the proceeds of selling the stock. The higher the interest rate, the lower the put value.

When interest rates are zero, the call and put option values are the same for at-the-money options. Recall that with put–call parity, we have $c - p = S_0 - e^{-rT}X$, and when interest rates are zero, then the present value function has no effect. As interest rates rise, the difference between call and put options increases as illustrated in Exhibit 19. The impact on option prices when interest rates change is relatively small when compared with that for volatility changes and that for changes in the stock. Hence, the influence of interest rates is generally not a major concern.[17]

17 An exception to this rule is that with interest rate options, the interest rate is not constant and serves as the underlying. The relationship between the option value and the underlying interest rate is, therefore, captured by the delta, not the rho. Rho is really more generally the relationship between the option value and the rate used to discount cash flows.

Exhibit 19 Option Values and Interest Rates (S = 100 = X, r = 5%, T = 1, δ = 0)



IMPLIED VOLATILITY

19

- define implied volatility and explain how it is used in options trading.

As we have already touched on in Section 17, for most options, the value is particularly sensitive to volatility. Unlike the price of the underlying, however, volatility, is not an observable value in the marketplace. Volatility can be, and often is estimated, based on a sample of historical data. For example, for a three-month option, we might look back over the last three months and calculate the actual historical stock volatility. We can then use this figure as an estimate of volatility over the next three months. The volatility parameter in the BSM model, however, is the *future* volatility. As we know, history is a very frail guide of the future, so the option may appear to be "mispriced" with respect to the actual future volatility experienced. Different investors will have different views of the future volatility. The one with the most accurate forecast will have the most accurate assessment of the option value.

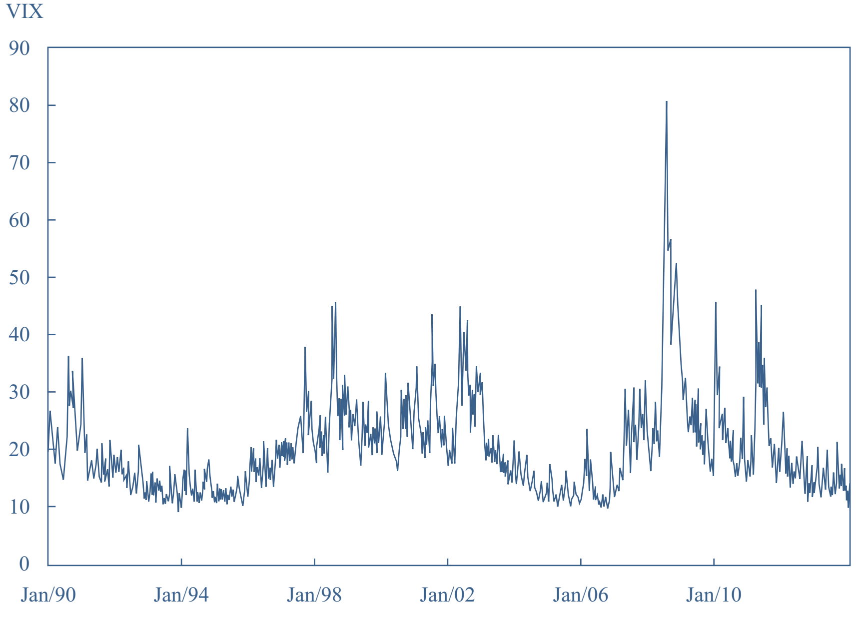
Much like yield to maturity with bonds, volatility can be inferred from option prices. This inferred volatility is called the **implied volatility**. Thus, one important use of the BSM model is to invert the model and estimate implied volatility. The key advantage is that implied volatility provides information regarding the perceived uncertainty going forward and thereby allows us to gain an understanding of the collective opinions of investors on the volatility of the underlying and the demand for options. If the demand for options increases and the no-arbitrage approach is not perfectly reflected in market prices—for example, because of transaction costs—then the preference for buying options will drive option prices up, and hence, the observed implied volatility. This kind of information is of great value to traders in options.

Recall that one assumption of the BSM model is that all investors agree on the value of volatility and that this volatility is non-stochastic. Note that the original BSM model assumes the underlying instrument volatility is constant in our context. That is, when we calculate option values, we have assumed a single volatility number, like 30%. In practice, it is very common to observe different implied volatilities for different exercise prices and observe different implied volatilities for calls and puts with the same terms. Implied volatility also varies across time to expiration as well as across exercise prices. The implied volatility with respect to time to expiration is known as the term structure of volatility, whereas the implied volatility with respect to the exercise price is known as the volatility smile or sometimes skew depending on the particular shape. It is common to construct a three dimensional plot of the implied volatility with respect to both expiration time and exercise prices, a visualization known as the volatility surface. If the BSM model assumptions were true, then one would expect to find the volatility surface flat.

Implied volatility is also not constant through calendar time. As implied volatility increases, market participants are communicating an increased market price of risk. For example, if the implied volatility of a put increases, it is more expensive to buy downside protection with a put. Hence, the market price of hedging is rising. With index options, various volatility indexes have been created, and these indexes measure the collective opinions of investors on the volatility in the market. Investors can now trade futures and options on various volatility indexes in an effort to manage their vega exposure in other options.

Exhibit 20 provides a look at a couple of decades of one such volatility index, the Chicago Board Options Exchange S&P 500 Volatility Index, known as the VIX. The VIX is quoted as a percent and is intended to approximate the implied volatility of the S&P 500 over the next 30 days. VIX is often termed the fear index because it is viewed as a measure of market uncertainty. Thus, an increase in the VIX index is regarded as greater investor uncertainty. From this figure, we see that the implied volatility of the S&P 500 is not constant and goes through periods when the VIX is low and periods when the VIX is high. In the 2008 global financial crisis, the VIX was extremely high, indicating great fear and uncertainty in the equity market. Remember that implied volatility reflects both beliefs regarding future volatility as well as a preference for risk mitigating products like options. Thus, during the crisis, the higher implied volatility reflected both higher expected future volatility as well as increased preference for buying rather than selling options.

Exhibit 20 VIX Daily Values, 2 January 1990–18 July 2014



Implied volatility has several uses in option trading. An understanding of implied volatility is essential in managing an options portfolio. This reading explains the valuation of options as a function of the value of the underlying, the exercise price, the expiration date, the risk-free rate, dividends or other benefits paid by the underlying, and the volatility of the underlying. Note that each of these parameters is observable except the volatility of the underlying over the option term looking ahead. This volatility has to be estimated in some manner, such as by calculating historical volatility. But as noted, historical volatility involves looking back in time. There are, however, a vast number of liquid options traded on exchanges around the world so that a wide variety of option prices are observable. Because we know the price and all the parameters except the volatility, we can back out the volatility needed by the option valuation model to get the known price. This volatility is the implied volatility.

Hence, implied volatility can be interpreted as the market's view of how to value options. In the option markets, participants use volatility as the medium in which to quote options. The price is simply calculated by the use of an agreed model with the quoted volatility. For example, rather than quote a particular call option as trading for €14.23, it may be quoted as 30.00, where 30.00 denotes in percentage points the implied volatility based on a €14.23 option price. Note that there is a one-to-one relationship between the implied volatility and the option price, ignoring rounding errors.

The benefit of quoting via implied volatility (or simply volatility), rather than price, is that it allows volatility to be traded in its own right. Volatility is the "guess factor" in option pricing. All other inputs—value of the underlying, exercise price, expiration, risk-free rate, and dividend yield—are agreed.[18] Volatility is often the same order of magnitude across exercise prices and expiration dates. This means that traders can compare the values of two options, which may have markedly different exercise prices and expiration dates, and therefore, markedly different prices in a common unit of measure, specifically implied volatility.

EXAMPLE 20

Implied Volatility in Option Trading within One Market

Suppose we hold portfolio of options all tied to FTSE 100 futures contracts. Let the current futures price be 6,850. A client calls to request our offer prices on out-of-the-money puts and at-the-money puts, both with the same agreed expiration date. We calculate the prices to be respectively, 190 and 280 futures points. The client wants these prices quoted in implied volatility as well as in futures points because she wants to compare prices by comparing the quoted implied volatilities. The implied volatilities are 16% for the out-of-the-money puts and 15.2% for the at-the-money puts. Why does the client want the quotes in implied volatility?

A Because she can better compare the two options for value—that is, she can better decide which is cheap and which is expensive.

B Because she can assess where implied volatility is trading at that time, and thus consider revaluing her options portfolio at the current market implied volatilities for the FTSE 100.

C Both A and B are valid reasons for quoting options in volatility units.

18 The risk-free rate and dividend yield may not be entirely agreed, but the impact of variations to these parameters is generally very small compared with the other inputs.

Solution:

C is correct. Implied volatility can be used to assess the relative value of different options, neutralizing the moneyness and time to expiration effects. Also, implied volatility is useful for revaluing existing positions over time.

EXAMPLE 21

Implied Volatility in Option Trading Across Markets

Suppose an options dealer offers to sell a three-month at-the-money call on the FTSE index option at 19% implied volatility and a one-month in-the-money put on Vodaphone (VOD) at 24%. An option trader believes that based on the current outlook, FTSE volatility should be closer to 25% and VOD volatility should be closer to 20%. What actions might the trader take to benefit from her views?

A Buy the FTSE call and the VOD put.

B Buy the FTSE call and sell the VOD put.

C Sell the FTSE call and sell the VOD puts.

Solution:

B is correct. The trader believes that the FTSE call volatility is understated by the dealer and that the VOD put volatility is overstated. Thus, the trader would expect FTSE volatility to rise and VOD volatility to fall. As a result, the FTSE call would be expected to increase in value and the VOD put would be expected to decrease in value. The trader would take the positions as indicated in B.

Regulators, banks, compliance officers, and most option traders use implied volatilities to communicate information related to options portfolios. This is because implied volatilities, together with standard pricing models, give the "market consensus" valuation, in the same way that other assets are valued using market prices.

In summary, as long as all market participants agree on the underlying option model and how other parameters are calculated, then implied volatility can be used as a quoting mechanism. Recall that there are calls and puts, various exercise prices, various maturities, American and European, and exchange-traded and OTC options. Thus, it is difficult to conceptualize all these different prices. For example, if two call options on the same stock had different prices, but one had a longer expiration and lower exercise price and the other had a shorter expiration and higher exercise, which should be the higher priced option? It is impossible to tell on the surface. But if one option implied a higher volatility than the other, we know that after taking into account the effects of time and exercise, one option is more expensive than the other. Thus, by converting the quoted price to implied volatility, it is easier to understand the current market price of various risk exposures.

SUMMARY

This reading on the valuation of contingent claims provides a foundation for understanding how a variety of different options are valued. Key points include the following:

- The arbitrageur would rather have more money than less and abides by two fundamental rules: Do not use your own money and do not take any price risk.
- The no-arbitrage approach is used for option valuation and is built on the key concept of the law of one price, which says that if two investments have the same future cash flows regardless of what happens in the future, then these two investments should have the same current price.
- Throughout this reading, the following key assumptions are made:
  - Replicating instruments are identifiable and investable.
  - Market frictions are nil.
  - Short selling is allowed with full use of proceeds.
  - The underlying instrument price follows a known distribution.
  - Borrowing and lending is available at a known risk-free rate.
- The two-period binomial model can be viewed as three one-period binomial models, one positioned at Time 0 and two positioned at Time 1.
- In general, European-style options can be valued based on the expectations approach in which the option value is determined as the present value of the expected future option payouts, where the discount rate is the risk-free rate and the expectation is taken based on the risk-neutral probability measure.
- Both American-style options and European-style options can be valued based on the no-arbitrage approach, which provides clear interpretations of the component terms; the option value is determined by working backward through the binomial tree to arrive at the correct current value.
- For American-style options, early exercise influences the option values and hedge ratios as one works backward through the binomial tree.
- Interest rate option valuation requires the specification of an entire term structure of interest rates, so valuation is often estimated via a binomial tree.
- A key assumption of the Black–Scholes–Merton option valuation model is that the return of the underlying instrument follows geometric Brownian motion, implying a lognormal distribution of the price.
- The BSM model can be interpreted as a dynamically managed portfolio of the underlying instrument and zero-coupon bonds.
- BSM model interpretations related to $N(d_1)$ are that it is the basis for the number of units of underlying instrument to replicate an option, that it is the primary determinant of delta, and that it answers the question of how much the option value will change for a small change in the underlying.
- BSM model interpretations related to $N(d_2)$ are that it is the basis for the number of zero-coupon bonds to acquire to replicate an option and that it is the basis for estimating the risk-neutral probability of an option expiring in the money.
- The Black futures option model assumes the underlying is a futures or a forward contract.

- Interest rate options can be valued based on a modified Black futures option model in which the underlying is a forward rate agreement (FRA), there is an accrual period adjustment as well as an underlying notional amount, and that care must be given to day-count conventions.
- An interest rate cap is a portfolio of interest rate call options termed caplets, each with the same exercise rate and with sequential maturities.
- An interest rate floor is a portfolio of interest rate put options termed floorlets, each with the same exercise rate and with sequential maturities.
- A swaption is an option on a swap.
- A payer swaption is an option on a swap to pay fixed and receive floating.
- A receiver swaption is an option on a swap to receive fixed and pay floating.
- Long a callable fixed-rate bond can be viewed as long a straight fixed-rate bond and short a receiver swaption.
- Delta is a static risk measure defined as the change in a given portfolio for a given small change in the value of the underlying instrument, holding everything else constant.
- Delta hedging refers to managing the portfolio delta by entering additional positions into the portfolio.
- A delta neutral portfolio is one in which the portfolio delta is set and maintained at zero.
- A change in the option price can be estimated with a delta approximation.
- Because delta is used to make a linear approximation of the non-linear relationship that exists between the option price and the underlying price, there is an error that can be estimated by gamma.
- Gamma is a static risk measure defined as the change in a given portfolio delta for a given small change in the value of the underlying instrument, holding everything else constant.
- Gamma captures the non-linearity risk or the risk—via exposure to the underlying—that remains once the portfolio is delta neutral.
- A gamma neutral portfolio is one in which the portfolio gamma is maintained at zero.
- The change in the option price can be better estimated by a delta-plus-gamma approximation compared with just a delta approximation.
- Theta is a static risk measure defined as the change in the value of an option given a small change in calendar time, holding everything else constant.
- Vega is a static risk measure defined as the change in a given portfolio for a given small change in volatility, holding everything else constant.
- Rho is a static risk measure defined as the change in a given portfolio for a given small change in the risk-free interest rate, holding everything else constant.
- Although historical volatility can be estimated, there is no objective measure of future volatility.
- Implied volatility is the BSM model volatility that yields the market option price.
- Implied volatility is a measure of future volatility, whereas historical volatility is a measure of past volatility.
- Option prices reflect the beliefs of option market participant about the future volatility of the underlying.

- The volatility smile is a two dimensional plot of the implied volatility with respect to the exercise price.
- The volatility surface is a three dimensional plot of the implied volatility with respect to both expiration time and exercise prices.
- If the BSM model assumptions were true, then one would expect to find the volatility surface flat, but in practice, the volatility surface is not flat.

PRACTICE PROBLEMS

The following information relates to Questions 1–9

Bruno Sousa has been hired recently to work with senior analyst Camila Rocha. Rocha gives him three option valuation tasks.

Alpha Company

Sousa's first task is to illustrate how to value a call option on Alpha Company with a one-period binomial option pricing model. It is a non-dividend-paying stock, and the inputs are as follows.

- The current stock price is 50, and the call option exercise price is 50.
- In one period, the stock price will either rise to 56 or decline to 46.
- The risk-free rate of return is 5% per period.

Based on the model, Rocha asks Sousa to estimate the hedge ratio, the risk-neutral probability of an up move, and the price of the call option. In the illustration, Sousa is also asked to describe related arbitrage positions to use if the call option is overpriced relative to the model.

Beta Company

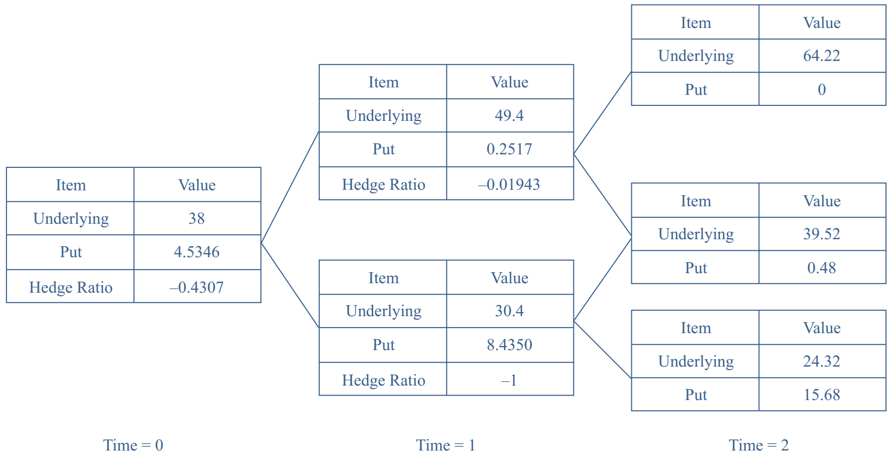
Next, Sousa uses the two-period binomial model to estimate the value of a European-style call option on Beta Company's common shares. The inputs are as follows.

- The current stock price is 38, and the call option exercise price is 40.
- The up factor (u) is 1.300, and the down factor (d) is 0.800.
- The risk-free rate of return is 3% per period.

Sousa then analyzes a put option on the same stock. All of the inputs, including the exercise price, are the same as for the call option. He estimates that the value of a European-style put option is 4.53. Exhibit 1 summarizes his analysis. Sousa next must determine whether an American-style put option would have the same value.

© 2016 CFA Institute. All rights reserved.

Exhibit 1 Two-Period Binomial European-Style Put Option on Beta Company



Sousa makes two statements with regard to the valuation of a European-style option under the expectations approach.

Statement 1 The calculation involves discounting at the risk-free rate.

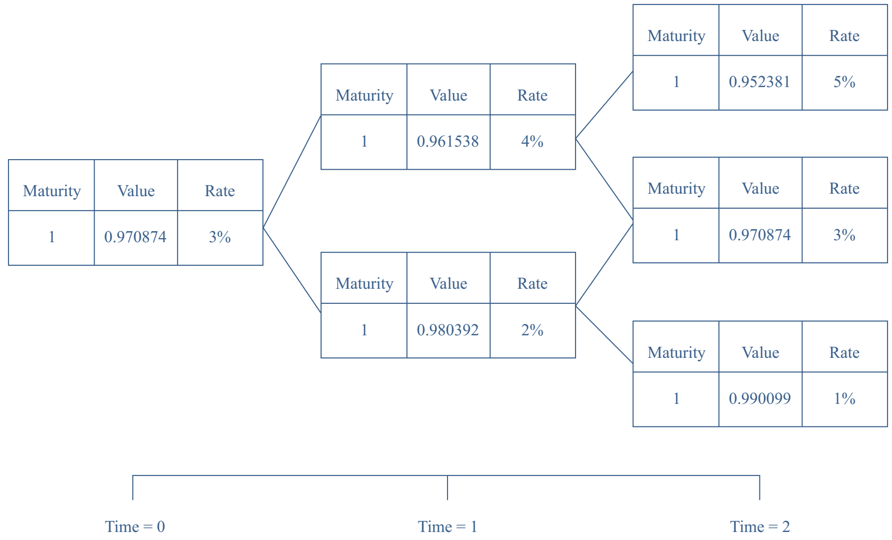
Statement 2 The calculation uses risk-neutral probabilities instead of true probabilities.

Rocha asks Sousa whether it is ever profitable to exercise American options prior to maturity. Sousa answers, "I can think of two possible cases. The first case is the early exercise of an American call option on a dividend-paying stock. The second case is the early exercise of an American put option."

Interest Rate Option

The final option valuation task involves an interest rate option. Sousa must value a two-year, European-style call option on a one-year spot rate. The notional value of the option is 1 million, and the exercise rate is 2.75%. The risk-neutral probability of an up move is 0.50. The current and expected one-year interest rates are shown in Exhibit 2, along with the values of a one-year zero-coupon bond of 1 notional value for each interest rate.

Exhibit 2 Two-Year Interest Rate Lattice for an Interest Rate Option



Rocha asks Sousa why the value of a similar in-the-money interest rate call option decreases if the exercise price is higher. Sousa provides two reasons.

Reason 1 The exercise value of the call option is lower.

Reason 2 The risk-neutral probabilities are changed.

1 The optimal hedge ratio for the Alpha Company call option using the one-period binomial model is *closest* to:

A 0.60.

B 0.67.

C 1.67.

2 The risk-neutral probability of the up move for the Alpha Company stock is *closest* to:

A 0.06.

B 0.40.

C 0.65.

3 The value of the Alpha Company call option is *closest* to:

A 3.71.

B 5.71.

C 6.19.

4 For the Alpha Company option, the positions to take advantage of the arbitrage opportunity are to write the call and:

A short shares of Alpha stock and lend.

B buy shares of Alpha stock and borrow.

C short shares of Alpha stock and borrow.

5 The value of the European-style call option on Beta Company shares is *closest* to:

A 4.83.

B 5.12.

C 7.61.

6 The value of the American-style put option on Beta Company shares is *closest* to:

A 4.53.

B 5.15.

C 9.32.

7 Which of Sousa's statements about binomial models is correct?

A Statement 1 only

B Statement 2 only

C Both Statement 1 and Statement 2

8 Based on Exhibit 2 and the parameters used by Sousa, the value of the interest rate option is *closest* to:

A 5,251.

B 6,236.

C 6,429.

9 Which of Sousa's reasons for the decrease in the value of the interest rate option is correct?

A Reason 1 only

B Reason 2 only

C Both Reason 1 and Reason 2

The following information relates to Questions 10–17

Trident Advisory Group manages assets for high-net-worth individuals and family trusts.

Alice Lee, chief investment officer, is meeting with a client, Noah Solomon, to discuss risk management strategies for his portfolio. Solomon is concerned about recent volatility and has asked Lee to explain options valuation and the use of options in risk management.

Options on Stock

Lee uses the BSM model to price TCB, which is one of Solomon's holdings. Exhibit 1 provides the current stock price (S), exercise price (X), risk-free interest rate (r), volatility (σ), and time to expiration (T) in years as well as selected outputs from the BSM model. TCB does not pay a dividend.

Exhibit 1 BSM Model for European Options on TCB

BSM Inputs

S	*X*	*r*	*Σ*	*T*
$57.03	55	0.22%	32%	0.25

(continued)

Exhibit 1 (Continued)

BSM Outputs

d_1	$N(d_1)$	d_2	$N(d_2)$	BSM Call Price	BSM Put Price
0.3100	0.6217	0.1500	0.5596	$4.695	$2.634

Options on Futures

The Black model valuation and selected outputs for options on another of Solomon's holdings, the GPX 500 Index (GPX), are shown in Exhibit 2. The spot index level for the GPX is 187.95, and the index is assumed to pay a continuous dividend at a rate of 2.2% (δ) over the life of the options being valued, which expire in 0.36 years. A futures contract on the GPX also expiring in 0.36 years is currently priced at 186.73.

Exhibit 2 Black Model for European Options on the GPX Index

Black Model Inputs

GPX Index	X	r	σ	T	δ Yield
187.95	180	0.39%	24%	0.36	2.2%

Black Model Call Value	Black Model Put Value	Market Call Price	Market Put Price
$14.2089	$7.4890	$14.26	$7.20

Option Greeks

Delta (call)	Delta (put)	Gamma (call or put)	Theta (call) daily	Rho (call) per %	Vega per % (call or put)
0.6232	–0.3689	0.0139	–0.0327	0.3705	0.4231

After reviewing Exhibit 2, Solomon asks Lee which option Greek letter best describes the changes in an option's value as time to expiration declines.

Solomon observes that the market price of the put option in Exhibit 2 is $7.20. Lee responds that she used the historical volatility of the GPX of 24% as an input to the BSM model, and she explains the implications for the implied volatility for the GPX.

Options on Interest Rates

Solomon forecasts the three-month Libor will exceed 0.85% in six months and is considering using options to reduce the risk of rising rates. He asks Lee to value an interest rate call with a strike price of 0.85%. The current three-month Libor is 0.60%, and an FRA for a three-month Libor loan beginning in six months is currently 0.75%.

Hedging Strategy for the Equity Index

Solomon's portfolio currently holds 10,000 shares of an exchange-traded fund (ETF) that tracks the GPX. He is worried the index will decline. He remarks to Lee, "You have told me how the BSM model can provide useful information for reducing the risk of my GPX position." Lee suggests a delta hedge as a strategy to protect against small moves in the GPX Index.

Lee also indicates that a long position in puts could be used to hedge larger moves in the GPX. She notes that although hedging with either puts or calls can result in a delta-neutral position, they would need to consider the resulting gamma.

10 Based on Exhibit 1 and the BSM valuation approach, the initial portfolio required to replicate the long call option payoff is:

- **A** long 0.3100 shares of TCB stock and short 0.5596 shares of a zero-coupon bond.
- **B** long 0.6217 shares of TCB stock and short 0.1500 shares of a zero-coupon bond.
- **C** long 0.6217 shares of TCB stock and short 0.5596 shares of a zero-coupon bond.

11 To determine the long put option value on TCB stock in Exhibit 1, the correct BSM valuation approach is to compute:

- **A** 0.4404 times the present value of the exercise price minus 0.6217 times the price of TCB stock.
- **B** 0.4404 times the present value of the exercise price minus 0.3783 times the price of TCB stock.
- **C** 0.5596 times the present value of the exercise price minus 0.6217 times the price of TCB stock.

12 What are the correct spot value (S) and the risk-free rate (r) that Lee should use as inputs for the Black model?

- **A** 186.73 and 0.39%, respectively
- **B** 186.73 and 2.20%, respectively
- **C** 187.95 and 2.20%, respectively

13 Which of the following is the correct answer to Solomon's question regarding the option Greek letter?

- **A** Vega
- **B** Theta
- **C** Gamma

14 Based on Solomon's observation about the model price and market price for the put option in Exhibit 2, the implied volatility for the GPX is *most likely*:

- **A** less than the historical volatility.
- **B** equal to the historical volatility.
- **C** greater than the historical volatility.

15 The valuation inputs used by Lee to price a call reflecting Solomon's interest rate views should include an underlying FRA rate of:

- **A** 0.60% with six months to expiration.
- **B** 0.75% with nine months to expiration.
- **C** 0.75% with six months to expiration.

16 The strategy suggested by Lee for hedging small moves in Solomon's ETF position would *most likely* involve:

A selling put options.

B selling call options.

C buying call options.

17 Lee's put-based hedge strategy for Solomon's ETF position would *most likely* result in a portfolio gamma that is:

A negative.

B neutral.

C positive.

SOLUTIONS

1 A is correct. The hedge ratio requires the underlying stock and call option values for the up move and down move. $S^+ = 56$, and $S^- = 46$. $c^+ = \text{Max}(0,S^+ - X) = \text{Max}(0,56 - 50) = 6$, and $c^- = \text{Max}(0,S^- - X) = \text{Max}(0,46 - 50) = 0$. The hedge ratio is

$$h = \frac{c^+ - c^-}{S^+ - S^-} = \frac{6 - 0}{56 - 46} = \frac{6}{10} = 0.60$$

2 C is correct. For this approach, the risk-free rate is $r = 0.05$, the up factor is $u = S^+/S = 56/50 = 1.12$, and the down factor is $d = S^-/S = 46/50 = 0.92$. The risk-neutral probability of an up move is

$$\pi = [\text{FV}(1) - d]/(u - d) = (1 + r - d]/(u - d)$$
$$\pi = (1 + 0.05 - 0.92)/(1.12 - 0.92) = 0.13/0.20 = 0.65$$

3 A is correct. The call option can be estimated using the no-arbitrage approach or the expectations approach. With the no-arbitrage approach, the value of the call option is

$$c = hS + \text{PV}(-hS^- + c^-).$$
$$h = (c^+ - c^-)/(S^+ - S^-) = (6 - 0)/(56 - 46) = 0.60.$$
$$c = (0.60 \times 50) + (1/1.05) \times [(-0.60 \times 46) + 0].$$
$$c = 30 - [(1/1.05) \times 27.6] = 30 - 26.286 = 3.714.$$

Using the expectations approach, the risk-free rate is $r = 0.05$, the up factor is $u = S^+/S = 56/50 = 1.12$, and the down factor is $d = S^-/S = 46/50 = 0.92$. The value of the call option is

$$c = \text{PV} \times [\pi c^+ + (1 - \pi)c^-].$$
$$\pi = [\text{FV}(1) - d]/(u - d) = (1.05 - 0.92)/(1.12 - 0.92) = 0.65.$$
$$c = (1/1.05) \times [0.65(6) + (1 - 0.65)(0)] = (1/1.05)(3.9) = 3.714.$$

Both approaches are logically consistent and yield identical values.

4 B is correct. You should sell (write) the overpriced call option and then go long (buy) the replicating portfolio for a call option. The replicating portfolio for a call option is to buy h shares of the stock and borrow the present value of $(hS^- - c^-)$.

$$c = hS + \text{PV}(-hS^- + c^-).$$
$$h = (c^+ - c^-)/(S^+ - S^-) = (6 - 0)/(56 - 46) = 0.60.$$

For the example in this case, the value of the call option is 3.714. If the option is overpriced at, say, 4.50, you short the option and have a cash flow at Time 0 of +4.50. You buy the replicating portfolio of 0.60 shares at 50 per share (giving you a cash flow of –30) and borrow (1/1.05) × [(0.60 × 46) – 0] = (1/1.05) × 27.6 = 26.287. Your cash flow for buying the replicating portfolio is –30 + 26.287 = –3.713. Your net cash flow at Time 0 is + 4.50 – 3.713 = 0.787. Your net cash flow at Time 1 for either the up move or down move is zero. You have made an arbitrage profit of 0.787.

In tabular form, the cash flows are as follows:

Transaction	Time Step 0	Time Step 1 Down Occurs	Time Step 1 Up Occurs
Sell the call option	4.50	0	–6.00
Buy h shares	–0.6 × 50 = –30	0.6 × 46 = 27.6	0.6 × 56 = 33.6
Borrow $-\text{PV}(-hS^- + c^-)$	–(1/1.05) × [(–0.6 × 46) + 0] = 26.287	–0.6 × 46 = –27.6	–0.6 × 46 = –27.6
Net cash flow	0.787	0	0

5 A is correct. Using the expectations approach, the risk-neutral probability of an up move is

$$\pi = [\text{FV}(1) - d]/(u - d) = (1.03 - 0.800)/(1.300 - 0.800) = 0.46.$$

The terminal value calculations for the exercise values at Time Step 2 are

$$c^{++} = \text{Max}(0,u^2S - X) = \text{Max}[0,1.30^2(38) - 40] = \text{Max}(0,24.22) = 24.22.$$
$$c^{-+} = \text{Max}(0,udS - X) = \text{Max}[0,1.30(0.80)(38) - 40] = \text{Max}(0,-0.48) = 0.$$
$$c^{--} = \text{Max}(0,d^2S - X) = \text{Max}[0,0.80^2(38) - 40] = \text{Max}(0,-15.68) = 0.$$

Discounting back for two years, the value of the call option at Time Step 0 is

$$c = \text{PV}[\pi^2c^{++} + 2\pi(1 - \pi)c^{-+} + (1 - \pi)^2c^{--}].$$
$$c = [1/(1.03)]^2[0.46^2(24.22) + 2(0.46)(0.54)(0) + 0.54^2(0)].$$
$$c = [1/(1.03)]^2[5.1250] = 4.8308.$$

6 A is correct. Using the expectations approach, the risk-neutral probability of an up move is

$$\pi = [\text{FV}(1) - d]/(u - d) = (1.03 - 0.800)/(1.300 - 0.800) = 0.46.$$

An American-style put can be exercised early. At Time Step 1, for the up move, p^+ is 0.2517 and the put is out of the money and should not be exercised early ($X < S$, 40 < 49.4). However, at Time Step 1, p^- is 8.4350 and the put is in the money by 9.60 ($X - S$ = 40 – 30.40). So, the put is exercised early, and the value of early exercise (9.60) replaces the value of not exercising early (8.4350) in the binomial tree. The value of the put at Time Step 0 is now

$$p = \text{PV}[\pi p^+ + (1 - \pi)p^-] = [1/(1.03)][0.46(0.2517) + 0.54(9.60)] = 5.1454.$$

Following is a supplementary note regarding Exhibit 1.

The values in Exhibit 1 are calculated as follows.

At Time Step 2:

$$p^{++} = \text{Max}(0,X - u^2S) = \text{Max}[0,40 - 1.300^2(38)] = \text{Max}(0,40 - 64.22) = 0.$$
$$p^{-+} = \text{Max}(0,X - udS) = \text{Max}[0,40 - 1.300(0.800)(38)] = \text{Max}(0,40 - 39.52) = 0.48.$$
$$p^{--} = \text{Max}(0,X - d^2S) = \text{Max}[0,40 - 0.800^2(38)] = \text{Max}(0,40 - 24.32) = 15.68.$$

At Time Step 1:

$$p^+ = \text{PV}[\pi p^{++} + (1 - \pi)p^{-+}] = [1/(1.03)][0.46(0) + 0.54(0.48)] = 0.2517.$$
$$p^- = \text{PV}[\pi p^{-+} + (1 - \pi)p^{--}] = [1/(1.03)][0.46(0.48) + 0.54(15.68)] = 8.4350.$$

At Time Step 0:

$$p = \text{PV}[\pi p^+ + (1 - \pi)p^-] = [1/(1.03)][0.46(0.2517) + 0.54(8.4350)] = 4.5346.$$

7 C is correct. Both statements are correct. The expected future payoff is calculated using risk-neutral probabilities, and the expected payoff is discounted at the risk-free rate.

8 C is correct. Using the expectations approach, per 1 of notional value, the values of the call option at Time Step 2 are

$$c^{++} = \text{Max}(0,S^{++} - X) = \text{Max}(0,0.050 - 0.0275) = 0.0225.$$
$$c^{+-} = \text{Max}(0,S^{+-} - X) = \text{Max}(0,0.030 - 0.0275) = 0.0025.$$
$$c^{--} = \text{Max}(0,S^{--} - X) = \text{Max}(0,0.010 - 0.0275) = 0.$$

At Time Step 1, the call values are

$$c^{+} = \text{PV}[\pi c^{++} + (1 - \pi)c^{+-}] .$$
$$c^{+} = 0.961538[0.50(0.0225) + (1 - 0.50)(0.0025)] = 0.012019.$$
$$c^{-} = \text{PV}[\pi c^{+-} + (1 - \pi)c^{--}].$$
$$c^{-} = 0.980392[0.50(0.0025) + (1 - 0.50)(0)] = 0.001225.$$

At Time Step 0, the call option value is

$$c = \text{PV}[\pi c^{+} + (1 - \pi)c^{-}].$$
$$c = 0.970874[0.50(0.012019) + (1 - 0.50)(0.001225)] = 0.006429.$$

The value of the call option is this amount multiplied by the notional value, or $0.006429 \times 1{,}000{,}000 = 6{,}429$.

9 A is correct. Reason 1 is correct: A higher exercise price does lower the exercise value (payoff) at Time 2. Reason 2 is not correct because the risk-neutral probabilities are based on the paths that interest rates take, which are determined by the market and not the details of a particular option contract.

10 C is correct. The no-arbitrage approach to creating a call option involves buying Delta = $N(d_1) = 0.6217$ shares of the underlying stock and financing with $-N(d_2) = -0.5596$ shares of a risk-free bond priced at $\exp(-rt)(X) = \exp(-0.0022 \times 0.25)(55) = \$54.97$ per bond. Note that the value of this replicating portfolio is $n_S S + n_B B = 0.6217(57.03) - 0.5596(54.97) = \$4.6943$ (the value of the call option with slight rounding error).

11 B is correct. The formula for the BSM price of a put option is $p = e^{-rt}XN(-d_2) - SN(-d_1)$. $N(-d_1) = 1 - N(d_1) = 1 - 0.6217 = 0.3783$, and $N(-d_2) = 1 - N(d_2) = 1 - 0.5596 = 0.4404$.

Note that the BSM model can be represented as a portfolio of the stock ($n_S S$) and zero-coupon bonds ($n_B B$). For a put, the number of shares is $n_S = -\text{N}(-d_1) < 0$ and the number of bonds is $n_B = -N(d_2) > 0$. The value of the replicating portfolio is $n_S S + n_B B = -0.3783(57.03) + 0.4404(54.97) = \$2.6343$ (the value of the put option with slight rounding error). B is a risk-free bond priced at $\exp(-rt)(X) = \exp(-0.0022 \times 0.25)(55) = \$54.97$.

12 A is correct. Black's model to value a call option on a futures contract is $c = e^{-rT}[F_0(T)N(d_1) - XN(d_2)]$. The underlying F_0 is the futures price (186.73). The correct discount rate is the risk-free rate, $r = 0.39\%$.

13 B is correct. Lee is pointing out the option price's sensitivity to small changes in time. In the BSM approach, option price sensitivity to changes in time is given by the option Greek theta.

14 A is correct. The put is priced at \$7.4890 by the BSM model when using the historical volatility input of 24%. The market price is \$7.20. The BSM model overpricing suggests the implied volatility of the put must be lower than 24%.

15 C is correct. Solomon's forecast is for the three-month Libor to exceed 0.85% in six months. The correct option valuation inputs use the six-month FRA rate as the underlying, which currently has a rate of 0.75%.

16 B is correct because selling call options creates a short position in the ETF that would hedge his current long position in the ETF.

Exhibit 2 could also be used to answer the question. Solomon owns 10,000 shares of the GPX, each with a delta of +1; by definition, his portfolio delta is +10,000. A delta hedge could be implemented by selling enough calls to make the portfolio delta neutral:

$$N_H = -\frac{\text{Portfolio delta}}{\text{Delta}_H} = -\frac{+10{,}000}{+0.6232} = -16{,}046 \text{ calls.}$$

17 C is correct. Because the gamma of the stock position is 0 and the put gamma is always non-negative, adding a long position in put options would most likely result in a positive portfolio gamma.

Gamma is the change in delta from a small change in the stock's value. A stock position always has a delta of +1. Because the delta does not change, gamma equals 0.

The gamma of a call equals the gamma of a similar put, which can be proven using put–call parity.

Alternative Investments

STUDY SESSION

Study Session 14	Alternative Investments

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to analyze and evaluate real estate, private equity, and commodities including commodity derivatives using appropriate valuation concepts and techniques.

Allocations to alternative investments such as real estate, private equity, and commodities have been growing in institutional and individual investor portfolios. Although attractive for their potential return, inflation protection, and diversification benefits, alternative investments are typically less transparent, more expensive, and inherently more complex than traditional stocks and bonds. Careful evaluation and due diligence are therefore important when considering alternative investments.

© 2021 CFA Institute. All rights reserved.

ALTERNATIVE INVESTMENTS

STUDY SESSION

14

Alternative Investments

This study session focuses on the following categories of alternative investments: real estate, private equity, and commodities. Real estate investments, both private and public, are described, and methods for analysis and evaluation are presented. Private equity, including venture capital and leveraged buyouts, is examined from the perspectives of a private equity firm evaluating equity portfolio investments and an investor considering participation in a private equity fund. The study session concludes with a discussion of commodities and commodity futures, including scenarios of contango and backwardation for futures prices.

READING ASSIGNMENTS

Reading 35	Real Estate Investments by Steven G. Bloom, CFA, Jeffrey D. Fisher, PhD, David Kruth, CFA, Bryan D. MacGregor, PhD, MRICS, MRTPI, Ian Rossa O'Reilly, CFA, and Anthony Paolone, CFA
Reading 36	Private Equity Investments by Yves Courtois, CMT, MRICS, CFA, and Tim Jenkinson, PhD
Reading 37	Introduction to Commodities and Commodity Derivatives by David Burkart, CFA, and James Alan Finnegan, CAIA, RMA, CFA

© 2021 CFA Institute. All rights reserved.

READING

35

Real Estate Investments

by Steven G. Bloom, CFA, Jeffrey D Fisher, PhD, David Kruth, CFA, Bryan D. MacGregor, PhD, MRICS, MRTPI, Ian Rossa O'Reilly, CFA, and Anthony Paolone, CFA

Steven G. Bloom, CFA, is at ARC Fiduciary (USA). Jeffrey D. Fisher, PhD, is Professor Emeritus at Indiana University and at Homer Hoyt Institute (USA). David Kruth, CFA, is at Columbia University (USA). Bryan D. MacGregor, PhD, MRICS, MRTPI, is at the University of Aberdeen, Scotland (United Kingdom). Ian Rossa O'Reilly, CFA (Canada). Anthony Paolone, CFA (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** compare the characteristics, classifications, principal risks, and basic forms of public and private real estate investments;
☐	**b.** explain portfolio roles and economic value determinants of real estate investments;
☐	**c.** discuss commercial property types, including their distinctive investment characteristics;
☐	**d.** explain the due diligence process for both private and public equity real estate investments;
☐	**e.** discuss real estate investment indexes, including their construction and potential biases;
☐	**f.** discuss the income, cost, and sales comparison approaches to valuing real estate properties;
☐	**g.** compare the direct capitalization and discounted cash flow valuation methods;
☐	**h.** estimate and interpret the inputs (for example, net operating income, capitalization rate, and discount rate) to the direct capitalization and discounted cash flow valuation methods;
☐	**i.** calculate the value of a property using the direct capitalization and discounted cash flow valuation methods;
☐	**j.** calculate and interpret financial ratios used to analyze and evaluate private real estate investments;
☐	**k.** discuss types of REITs;

(continued)

© 2021 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**l.** justify the use of net asset value per share (NAVPS) in REIT valuation and estimate NAVPS based on forecasted cash net operating income;
☐	**m.** describe the use of funds from operations (FFO) and adjusted funds from operations (AFFO) in REIT valuation;
☐	**n.** calculate and interpret the value of a REIT share using the net asset value, relative value (price-to-FFO and price-to-AFFO), and discounted cash flow approaches; and
☐	**o.** explain advantages and disadvantages of investing in real estate through publicly traded securities compared to private vehicles.

SECTION A. OVERVIEW OF TYPES OF REAL ESTATE INVESTMENT

Real estate is one of the largest and oldest investment asset classes, yet it is widely considered an alternative asset class. Nearly everyone has had experience with real estate, be it as a renter, homeowner and borrower, office space occupant, or retail shopper. Other than for homeowners, direct real estate investment has largely been beyond the reach of individual investors. The large amount of capital and expertise needed to invest in real estate, combined with its low liquidity, can be a deterrent that prevents most individuals and even many institutional investors from owning investment properties outright. During the last 20 years, however, investor acceptance of private fund vehicles and listed real estate securities, combined with the search for income during a period of declining and historically low interest rates, has contributed to steadily rising indirect real estate allocations and strong asset class performance.

1 INTRODUCTION AND BASIC FORMS OF REAL ESTATE INVESTMENT

Real estate offers investors long-term stable income, some protection from inflation, and generally low correlations with stocks and bonds. High-quality, well-managed properties with low leverage are generally expected to provide higher returns than high-grade corporate debt (albeit with higher risk) and lower returns and risk than equity. Real estate investment can be an effective means of diversification in many balanced investment portfolios. Investors can choose to have the equity, or ownership, position in properties, or they may prefer to have exposure to real estate debt as a lender or owner of mortgage-backed securities. Residential real estate constitutes by far the largest portion of the real estate market, most of which is owner occupied. Nonetheless, we will focus almost exclusively on rental, or commercial, properties. These include office buildings, shopping centers, distribution facilities, and for-rent residential properties.

Private real estate investments often hold the greatest appeal for investors with long-term investment horizons and the ability to accept relatively lower liquidity. Pension funds, sovereign wealth funds, insurance companies, and high-net-worth individuals have been among the largest investors in private real estate. Securitized real estate ownership—shares of publicly traded, pooled real estate investments, such as real estate operating companies (REOCs), real estate investment trusts (REITs), and mortgage-backed securities (MBS)—has historically provided smaller investors with ready access to the asset class because of low share prices and the benefits of higher liquidity and professional management. Institutional investors also pursue securitized real estate when the market capitalization of the vehicles can accommodate large investor demand. In fact, institutional ownership of US REITs has increased from 6.6% in 1990 to 64.5% in 2015, according to a 2019 research paper (Huerta, Ngo, and Pyles 2019).

Regardless of vehicle type, the risk profile of the underlying investment can vary significantly. High-quality properties in leading markets with long-term leases and low leverage have a conservative risk profile, as do those **mortgages** that represent only slightly more than half of the asset's value. Older properties with short-term leases in suburban markets with ample room for new development and higher leverage constitute higher-risk properties. Below-investment-grade, non-rated, and mezzanine debt similarly carries higher risk. Development property is often considered the riskiest because of long lead times and the dependence on contractors, suppliers, regulators, and future tenants for success.

Section A presents real estate as an asset class, delves into its role in portfolios, and contrasts the different characteristics of the major property types. Sections B and C explore private and public investing, respectively, with particular attention to valuation. Investment valuation and performance can be analyzed at the property and vehicle level. The end of the reading returns to the role of real estate in the portfolio and discusses whether investors' goals are best served by choosing private or public real estate vehicles.

1.1 Real Estate Market Size

a compare the characteristics, classifications, principal risks, and basic forms of public and private real estate investments

The total value of the global real estate market dwarfs other asset classes. The value of real estate reached \$281 trillion at the end of 2017, according to Savills World Research. Residential real estate accounted for nearly 80% of the total, at \$220 trillion, making it by far the largest segment of the real estate market. Commercial buildings and agricultural/forestry totaled \$33 trillion and \$20 trillion, respectively. In comparison, traditional debt securities (bonds) stood at \$105 trillion and equity securities (stocks) reached \$83 trillion at the end of 2017. For further comparison, global GDP was \$78 trillion at that time.

1.2 Real Estate Investment: Basic Forms

There are many different types of real estate property, capital position, and investment vehicle classifications. One simple way to classify property type distinguishes between residential and non-residential—typically, commercial—properties. Real estate can be categorized as either owner occupied or for rent. Real property type can further be classified as single-family residential, commercial, farmland, and timberland. Commercial real estate generally refers to the four largest rental property types: office buildings, shopping centers, industrial warehouses/distribution facilities, and for-rent residential, which can include multi-family rentals (i.e., buildings with multiple dwelling units/apartments/condominiums) and single-family detached home rentals. These four sectors make up the core property types because of market size, income stability, and low risk/return profile. In addition to the core property types, a **core real estate**

investment style or strategy is further defined as investing in high-quality, well-leased, core property types with low leverage (no more than 30% of asset value) in the largest markets with strong, diversified economies. It is a conservative strategy designed to avoid real estate–specific risks, including leasing, development, and speculation in favor of steady returns. Hotel properties are excluded from the core categories because of the higher cash flow volatility resulting from single-night leases and the greater importance of property operations, brand, and marketing. Senior housing/assisted living properties represent a specialized real estate use. Other specialized and niche property types include medical offices and facilities, self-storage, data centers, manufactured housing communities, casinos, cell towers, movie theaters, billboards, and just about any other type of for-rent real property. These other property types would fit a non-core investment strategy.

Capital position describes whether the investment is structured as equity or debt. An equity investor has an ownership interest: Such an investor may be the sole or a joint owner of the real estate property or may invest in securities of a company that owns the real estate property. The owner of the real estate property controls such decisions as whether to obtain a mortgage loan on the real estate, who should handle property management, and when to sell the real estate. In the case of a REIT, that control is delegated to the managers of the REIT by the shareholders. A debt investor is in a position of lender: Such an investor may loan funds to the entity acquiring the real estate property or may invest in securities based on real estate lending. Typically, the real estate property is used as collateral for a mortgage loan. In such cases, the mortgage lender has a priority claim on the real estate. The value of the equity investor's interest in the real estate is equal to the value of the real estate less the amount owed to the mortgage lender.

Exhibit 1 presents the various property types along these two lines of classification

Exhibit 1 Commercial Real Estate Includes Any Type of Real Property That Can Be Rented

	Owner Occupied	For Rent (Commercial)
Residential*	Single-family homes, apartments/condominiums, manufactured housing	Single-family detached homes and multi-family buildings
Non-Residential	Office, shopping centers, manufacturing facilities, warehouses, agricultural, other specialty real estate	Office, shopping centers, industrial warehouse/distribution, hotels, agricultural, other specialty real estate

*Section 2.4, "Classifications," goes into further detail about various types of residential real estate. Multi-family properties contain individual for-rent apartments or flats. Condominiums refer to owner-occupied units in multi-unit buildings.

The common investment distinction considers whether the real estate investment is made through a private or public vehicle. Private investment can be as simple as buying a property outright. Resident homeowners usually make their purchase without further structuring. Commercial property owners, whether as the sole owner or joint owner, are more likely to use a special vehicle to limit their liability. Property owners who accept capital from passive investors will form partnerships with the real estate professionals acting as the general partners (GPs) and the passive investors being admitted to the partnership as limited partners (LPs). The model commonly adopted by private equity investors has the entrepreneur/real estate professional taking the GP role and managing the partnership for the LPs. The LPs typically consist of insurance companies, pension funds, sovereign wealth funds, foundations, endowments, and high-net-worth individuals. Private investors may also invest through private companies.

Public investors can purchase common stock, partnership units, or trust units in entities that are listed on public exchanges and freely traded. By definition, investments in corporations, REITs, and other vehicles that, in turn, own properties are indirect investments. The key benefits to investing in publicly traded securities include access to professional management and a portfolio of properties combined with low minimum-purchase requirements.

Real estate operating companies are taxable corporations that operate and manage commercial real estate with few corporate-structure restrictions. They commonly own and often develop real estate. In contrast, REITs are restricted to primarily owning and operating rental properties and mortgages and are required to distribute nearly all or all of their earnings to investors to avoid paying corporate income. Mortgage-backed securities are often classified as public investments because there are often active secondary trading markets. There are some restrictions as to who is eligible to purchase the MBS and minimum trade sizes. MBS are indirect investments. The trust certificates typically own the right to receive cash flow from an underlying pool of mortgages, which, in turn, are secured by real property, rather than owning the mortgage outright.

Investment in real estate is often defined from a capital market perspective in the context of four quadrants, or areas, through which capital can be invested. The quadrants are a result of two dimensions of investment. The *first dimension* is whether the investment is made in the public or private market. The public market does not involve direct investment; rather, it involves investing in a security with claims on the underlying position(s)—for example, through investments in a REOC, a REIT, or a mortgage-backed security. The private market often involves investing directly in an asset—for example, purchasing a property or making or buying mortgages, which, in turn, have a claim on the asset. The private investment can also be made indirectly through a number of different investment vehicles that limit investor liability and tax leakage, such as a limited partnership or commingled real estate private equity fund. Regardless of the investment vehicle ownership structure, the transactions occur in the private market. The *second dimension* describes the investor's capital position in the underlying real estate. Property owners take the equity position and have rights to property profits, whereas debt investors lend capital to the owners subject to contractual interest payment and principal repayment terms or purchase mortgages or mortgage cash flow rights.

Combining the two dimensions, we have four quadrants: public equity, private equity, public debt, and private debt, as illustrated in Exhibit 2.

Exhibit 2 Examples of the Basic Forms of Real Estate Investment

	Public	Private
Equity	• Shares of REOCs • Shares of REITs, other listed trusts, exchange-traded funds (ETFs), and index funds	• Direct investments in real estate, including sole ownership and joint ventures • Indirect real estate ownership through limited partnerships, other forms of commingled funds, or private REITS and REOCs
Debt	• Mortgage REITs • MBS (residential and commercial) • Unsecured REIT debt	• Mortgages • Private debt • Bank debt

Equity investors generally expect a higher rate of return than lenders (debt investors) because they take on more risk. The lenders' claims on the cash flows and proceeds from sale must be satisfied before the equity investors can receive anything. As the amount of debt on a property, or financial leverage, increases, risk increases for both debt and equity; thus, an investor's debt or equity return expectations will increase. Of course, the risk is that the higher return will not materialize. The risk is even higher for an equity investor.

Debt investors in real estate, whether through private or public markets, expect to receive their return from promised cash flows and typically do not participate in any appreciation in value of the underlying real estate. Thus, debt investments in real estate are similar to other fixed-income investments, such as bonds. The returns to equity real estate investors have two components: an income stream resulting from such activities as renting the property and a capital appreciation component resulting from changes in the value of the underlying real estate. If the returns to equity real estate investors are less than perfectly positively correlated with the returns to stocks and bonds, then adding equity real estate investments to a traditional portfolio will potentially have diversification benefits.

Real estate markets in each of the four quadrants in Exhibit 2 have evolved and matured to create relatively efficient market structures for accessing all types of capital for real estate (i.e., public and private debt and equity). Such structures are critical for the success of the asset class for both lenders and equity investors. The categorization of real estate investment into the four quadrants helps investors identify the forms that best fit their objectives. For example, some investors may prefer to own and manage real estate. Other investors may prefer the greater liquidity and professional management associated with purchasing publicly traded REITs. Other investors may prefer mortgage lending because it involves less risk than equity investment or unsecured lending; the mortgage lender has a priority claim on the real estate used as collateral for the mortgage. Still other investors may want to invest in each quadrant or allocate more capital to one quadrant or another over time as they perceive shifts in the relative value of each. Each quadrant offers differences in risk and expected return, including the impact of taxes on the return. So, investors should explore the risk and return characteristics of each quadrant as part of their investment decisions.

EXAMPLE 1

Form of Investment

An investor is interested in adding real estate to her portfolio of equity and fixed-income securities for the first time. She has no previous real estate experience but believes adding real estate will provide some diversification benefits. She is concerned about liquidity because she may need the money in a year or so. Which form of investment is *most likely* appropriate for her?

A Shares of REITs

B Mortgage-backed securities

C Direct ownership of commercial real estate property

Solution:

A is correct. She is probably better off investing in shares of publicly traded REITs, which provide liquidity, have professional management, and require a smaller investment than direct ownership of real estate. Using REITs, she may be able to put together a diversified real estate investment portfolio. Although REITs are more correlated with stocks than direct ownership of real estate, direct ownership is much less liquid and a lot of properties are needed to have a diversified real estate portfolio. Also, adding shares of REITs to her current portfolio should provide more diversification benefits than adding debt in the form of mortgage-backed securities and will allow her to benefit from any appreciation of the real estate. Debt investments in real estate, such as MBS, are similar to other fixed-income investments, such as bonds, and can be highly sensitive to changes in interest rates. The difference is that their income streams are secured on real estate assets, which means that the risks are default risks linked to the performance of the real estate assets and the ability of mortgagees to pay interest. In contrast, adding equity real estate investments to a traditional portfolio (of equity and fixed-income investments) will potentially have diversification benefits.

1.3 Characteristics

Some of the main characteristics of real estate investment that distinguish it from the other main investment asset classes and that complicate the measurement and assessment of performance include the following:

- *Unique asset and fixed location*: Whereas all bonds of a particular issue are identical, as are stocks of a particular type in a specific company, no two properties are the same. (In real estate economics, the terms heterogeneity or non-homogeneity are used to characterize real estate land, building, and location.) Buildings differ in use, size, location, age, type of construction, quality, and tenant and leasing arrangements. Even identically constructed buildings with the same tenants and leases will be at different locations. These factors are important in trying to establish value and in the amount of specific risk in a real estate investment.
- *High unit value*: The unit value of private real estate property is large. The amount required to make an investment in private real estate limits the number of potential investors and the ability to construct a diversified real estate portfolio. Even when private equity investors pool their capital, capital requirements are generally too high for individual investors. This factor contributed to the development of publicly traded securities, such as REITs, which allow partial

ownership of an indivisible asset. Only when real estate is securitized, as with corporate debt through public bond offerings or company ownership via listed stocks, is ownership in the reach of most investors.

- *Management intensive*: An investor in bonds or stocks is not expected to be actively involved in managing the company, but a private real estate equity investor or direct owner of real estate has responsibility for management of the real estate, including maintaining the properties, negotiating leases, and collecting rents. This active management, whether carried out by the owner or by hired property managers, creates additional costs that must be taken into account when projecting returns.
- *High transaction costs*: Buying and selling real estate is also costly and time consuming because others—such as brokers, appraisers, lawyers, lenders, and construction professionals—are likely to be involved in the process until a transaction is completed.
- *Depreciation*: Buildings depreciate as a result of use and the passage of time. A building's value may also change as the desirability of its location and its design changes from the perspective of end users.
- *Need for debt capital*: Because of the large amounts required to purchase and develop real estate properties, the ability to access funds and the cost of funds in the credit markets are important. As a result, real estate values are sensitive to the cost and availability of debt capital. When debt capital is scarce or interest rates are high, the value of real estate tends to be lower than when debt capital is readily available or interest rates are low.
- *Illiquidity*: As a result of several of the listed factors, and because properties trade infrequently, real estate properties are relatively illiquid. They may take a significant amount of time to market and to sell at a price that is close to the owner's perceived fair market value. The initial spread between bid and asked prices is generally wide.
- *Price determination*: As a result of the wide differences in the characteristics of real estate properties and the low volume of transactions, estimates of value or appraisals rather than transaction prices are usually necessary to assess changes in value or expected selling price over time. However, the transaction prices of similar properties are often considered in estimating the value of or appraising a property. The limited number of participants in the market for a property, combined with the importance of local knowledge, makes it harder to know the market value of a property. In a less efficient market, those who have superior information and skill at evaluating properties may have an advantage. This situation is quite different from stocks in publicly traded companies, where many buyers and sellers with access to the same information value and transact in the shares in an active market.

These characteristics slowed widespread investor allocations to real estate. Securitization helped overcome some of these problems, especially investment size and illiquidity. In the United States, REITs were originally conceived of as a type of mutual fund to provide small investors with access to the asset class. The REIT provides or hires professional company and property managers. Similar to mutual funds, this vehicle does not pay income taxes and instead distributes dividends to investors. REITs typically allow exposure to a diversified portfolio of real estate. In regions without REIT structures or if property companies want greater flexibility, REOCs could also raise public capital. REIT and REOC shares are typically liquid, and active trading results in prices that are more likely to reflect market value. It is much easier to sell the shares of a listed company that owns real estate than to sell the underlying real estate.

EXAMPLE 2

Investment Characteristics

Question 1:

An investor states that he likes investing in private real estate because he believes the market is less efficient than other liquid asset classes and, therefore, expects to earn a return premium. What are some of the sources of real estate market inefficiency?

Solution 1:

It can be difficult to readily establish fair market value in real estate. Infrequent transactions, high transaction costs, and low transparency reduce market efficiency. There is evidence that real estate values are serially correlated, or autocorrelated, meaning prior-period values have a large influence on current-period values rather than price changes displaying the random-walk movement associated with efficient markets. When serial correlation is high, property values do not quickly incorporate new information. Market players who recognize the impact of new information on underlying property value cannot readily buy real estate when it is priced below intrinsic value and sell real estate when prices move above intrinsic value. In a less efficient market, an investor with superior knowledge and information or a better understanding of the appropriate price to pay for properties (superior valuation skills) may earn a higher return, provided that market prices adjust to intrinsic values, by making more informed investment decisions. However, there is also mounting evidence that real estate is efficient, or is at least becoming more efficient. Online data services provide real-time pricing transparency based on property location, type, size, and age. There is also information about commercial tenants, rents, lease terms, and lease expiration schedules. The large number and large size of real estate private equity funds with ample capital to deploy suggest numerous professional investors are scouring markets for the best opportunities. An investor buying relatively few properties may be able to take advantage of market inefficiencies. However, larger investors with broad real estate exposure are more likely to see diversification reduce idiosyncratic opportunities for above-market returns. Private real estate investors should expect to earn a return premium for illiquidity. Earning excess returns from market inefficiency becomes increasingly difficult as the number of knowledgeable, well-capitalized participants competing for acquisitions and spurring transaction activity increases.

Question 2:

A portfolio manager believes the entire real estate sector is trading at cyclically depressed levels because of prior overbuilding, a jump in interest rates, and a recession. The manager wants tactical exposure to real estate for what the manager expects to be a three-year recovery cycle. What would be a good real estate investment strategy for the manager?

Solution 2:

The portfolio manager could purchase the shares of a large, diversified REIT or REOC. REIT shares would provide exposure to underlying real estate, and REOCs could offer exposure to a combination of rental income, property management and brokerage income, and development profits. By investing in the shares of a larger, presumably liquid company, the portfolio manager should be able to exit the position if the sector recovers as expected or if the portfolio manager decides to raise cash. Geography- and sector-focused real estate companies (e.g., companies that own shopping centers in Australia) should be considered if the portfolio manager's views extend to specific markets.

Investing in private funds or companies may not offer as much liquidity, and entry/exit costs could be higher.

1.4 Risk Factors

Investors want an expected return that compensates them for incurring risk: The higher the risk, the higher the expected return. In this section, we consider risk factors associated with investing in commercial real estate. Most of the risk factors that follow affect the income and/or value of the real estate property and if investing indirectly, the income and value of the equity or debt investment.

The following are characteristic sources of risk or risk factors of real estate investment.

Property Demand and Supply

- *Business conditions*: Fundamentally, the real estate business involves renting space to users. The demand for space depends on a myriad of international, national, regional, and local economic conditions. GDP, employment, household income, interest rates, and inflation are particularly relevant to real estate. Changes in macroeconomic conditions will affect real estate investments because both current and expected income and real estate values may be affected.
- *Demographics*: Expanding on the already described macro factors are a variety of demographic factors, such as the size and age distribution of the population in the local market, the distribution of socioeconomic groups, and rates of new household formation. These demographic factors affect the demand for real estate.
- *Excess supply*: The real estate cycle is generally long, lasting approximately 17–18 years *on average*, albeit with a great deal of variance. Increases in the demand for space, which usually accompany the business cycle, will lead to higher occupancy, which, in turn, can support higher rents. New development usually begins once rents and property income increase to levels high enough to meet investor return requirements. Construction costs and property operating expenses generally increase later in the real estate cycle as increased competition for labor, materials, and land contribute to rising development costs, thereby increasing the minimum rent threshold required to justify new construction.

  New development requires long lead times to secure capital, land, designs, permits, and zoning approval; to start and complete construction; to lease space; and to have tenants move in. If additions to real estate supply do not keep up with demand, rents will continue to rise, which encourages even more development. As the business cycle ages, recession risks increase. When the inevitable contraction in business activity occurs and demand for space moderates or declines, new supply continuing to come to market will contribute to a decline in market occupancy, which is accompanied by falling rents and declining returns to real estate investment. Rent price swings between the lows and highs can be dramatic. When rents and returns drop, new supply will contract and remain low until space demand rises high enough to absorb the excess space and contribute to higher rents.

Valuation

- *Cost and availability of capital*: Real estate must compete with other assets for debt and equity capital. The willingness of investors to invest in real estate depends on the availability of debt capital and the cost of that capital, as well as the expected return on other investments, such as stocks and bonds, which affects the availability of equity capital. A shortage of debt capital and high interest rates can significantly reduce the demand for real estate and lower prices. Alternatively, an environment of low interest rates and easy access to debt capital can lower investors' weighted average cost of capital and increase the amount investors are willing to pay for real estate investments. These capital market forces can cause prices to increase or decrease regardless of any changes in the underlying demand for real estate from tenants.
- *Availability of information*: Of increasing importance to investors, especially when investing globally, is having adequate information to make informed investment decisions. A lack of information to conduct the property analysis adds to the risk of the investment. The amount of data available on real estate space and capital markets has improved considerably. Although some countries have much more information available to investors than others, in general, the availability of information has been increasing on a global basis because real estate investment has become more global and investors want to evaluate investment alternatives on a comparable basis. Real estate indexes have become available in many countries around the world. These indexes allow investors to benchmark their properties' performance against that of peers and also provide a better understanding of the risk and return for real estate compared with other asset classes. Indexes are discussed in more detail in Section 4.
- *Lack of liquidity*: Liquidity is the ability to convert an asset to cash quickly without a significant price discount or loss of principal. Real estate is considered to have low liquidity (high liquidity risk) because of the large value of an individual investment and the time and cost it takes to sell a property at its current value. Buyers are unlikely to make large investments without conducting adequate due diligence, which takes both time and money. Therefore, buyers are not likely to agree to a quick purchase without a significant discount to the price. Illiquidity means both a longer time to realize cash and a risk that the market may move against the investor.
- *Rising interest rates.* Fixed-income securities are usually negatively affected by higher interest rates because higher discount rates reduce the present value of the instrument. Real estate values may also decline initially when interest rates rise. Unlike a fixed-rate bond with a fixed maturity price, however, property income, land prices, and real estate values may increase over time or at least through the latter part of the real estate cycle. Increasing land and property construction costs raise the rental threshold at which new development can generate target returns. Therefore, market rent can continue to rise up to the threshold before significant development begins.

Property Operations

- *Management*: Management involves the cost of monitoring an investment. Investment management can be categorized into two levels: asset management and property management. Asset management involves monitoring the investment's financial performance and making changes as needed. Property management is exclusive to real estate investments. It involves the overall day-to-day operation of the property and the physical maintenance of the property, including the buildings. Management risk reflects the ability of the property and asset managers to make the right decisions regarding the operation of the property, such as negotiating leases, maintaining the property, marketing the property, and making renovations when necessary.

- *Lease provisions:* Lease provisions may allow landlords to recover a portion or all of the loss in purchasing power from generally rising prices—that is, inflation—to preserve real returns through a combination of contractual rent bumps and expense passthroughs. Predetermined contractual rent step-ups may not be large enough to capture unexpected inflation unless they are tied to a consumer price or other inflation-linked index. Even then, regional increases in operating expenses, especially real estate taxes and insurance costs, can rise faster than general inflation. Real rental income after expenses would be penalized in such a scenario unless leases also require lessees to reimburse landlords for property operating expenses. Expense caps, which limit how much of the annual increase is passed along to the tenant, would not perfectly protect the lessor against unforeseen increases in expenses. Short-term leases (typically six months to two years) and leases in markets that allow the property owner to require periodic rent reviews present the landlord with the opportunity to frequently reset rents in response to changing market conditions. The longer the lease or the longer the period between rent reviews, the more difficult it is to anticipate rising costs and, therefore, the more important it is for lessors to require expense reimbursements from tenants. Following a real estate market downturn, however, high vacancy rates and low rents may prevent landlords from raising rents on new leases in line with inflation.
- *Leverage*: Leverage affects returns on investments in real estate but not the value of the underlying real estate property at any given point. Leverage is the use of borrowed funds to finance some of the purchase price of an investment. The ratio of borrowed funds to total purchase price is known as the loan-to-value (LTV) ratio. Higher LTV ratios mean greater amounts of leverage. Real estate transactions can be more highly leveraged than most other types of investments. But increasing leverage also increases risk because the lender has the first claim on the cash flow and on the value of the property if there is default on the loan. A small change in property income can result in a relatively large change in the amount of cash flow available to the equity investor after making the mortgage payment.
- *Environmental*: Real estate values can be affected by environmental conditions, including soil and groundwater contaminants related to a prior owner, prior tenants, or an adjacent property owner. Such problems can significantly reduce the value because of the costs incurred to correct them. Leaking fuel tanks that are discovered during buyer due diligence require removal and remediation.
- *Obsolescence:* Changes in tenant preferences, regulations, and technology affect space demand. Ceiling heights in older buildings may not be high enough to accommodate warehouse stacking requirements or office communication networking cables and equipment. Distribution facility docks may not work with larger trucks, and paved lots may not be deep enough to allow room for large trucks to turn. Internet shopping, department store closures and consolidation, and other retail shopping trends, especially in the United States, which has the largest amount of per capita retail space, have constrained demand for large shopping center space. At the same time, many delivery companies are looking at warehouse and inventory storage space much closer to retail clients. The forecasted shift to autonomous cars and trucks will further affect real estate in ways not imagined. It may not be economically viable to upgrade, reconfigure, or repurpose older buildings to comply with energy efficiency and other modernization requirements or changing business and consumer preferences.

- *Recent and ongoing market disruption*: Rising use of the internet, cloud computing services, and offsite IT backup systems have spurred the growth of data centers while reducing the space businesses need for onsite computer and server systems. Internet sales and delivery combined with increased attention to companies' carbon footprint are contributing to shifting trade and distribution patterns. Companies may prefer to locate warehouse distribution facilities closer to customers for faster and even same-day delivery. Large shopping center owners have been partially successful at replacing former department store anchor tenants with restaurants and other forms of entertainment to attract consumer traffic and converting retail space to local distribution space.

  The COVID-19 pandemic caused tremendous shocks to the global economy. The quarantine and stay-at-home policies, which were still in effect in most Western countries at the time of this writing, inflicted great pain on the lodging and brick-and-mortar retail sectors in particular and accelerated such trends as retail consolidation and the rise of internet retailing. The prevalence of working from home during the pandemic may have helped the data center sector by increasing internet communication traffic and forcing companies to rely on business continuity services. At the same time, many employers realized they can get by with less office space by permitting some employees to work from home regularly, and many employees have grown accustomed to working from home and do not look to return to the office full time, at least until the pandemic ends. There is also evidence that urban residents are leaving large, expensive cities for suburban living. Rents across the residential, office, and retail sectors have been declining more than 10% in many gateway cities. Such events as the COVID-19 pandemic make it difficult to predict by how much and how fast demand for space will change during the next 2 to 10 years.
- *Other risk factors*: Many other risk factors exist, such as unobserved physical defects in the property, natural disasters (e.g., earthquakes and hurricanes, for which insurance and repair costs can be expensive), pandemics, acts of terrorism, and climate change. Unfortunately, the biggest risk may be one that was unidentified as a risk at the time of purchasing the property. Unidentified, difficult-to-forecast, and catastrophic risks can cause major disruptions and be devastating to investors.

Risks that are identified can be planned for to some extent and incorporated in investors' expectations. In some cases, a risk can be converted to a known dollar amount through insurance. In other cases, risk can be reduced through diversification or shifted to another party through contractual arrangements. For example, the risk of expenses increasing can be shifted to tenants by including expense reimbursement clauses in their leases. The risk that remains must be evaluated and reflected in contractual terms (e.g., rental prices) such that the expected return is equal to or greater than the required return necessary to make the investment.

EXAMPLE 3

Commercial Real Estate Risk

An investor wants to add real estate to her portfolio to benefit from its diversifying characteristics. She decides to buy a commercial property, financing at most 30% of the asset with debt in order to avoid incurring financial risk due to interest rate changes. This strategy is *most likely* to:

A limit the risk due to leverage.

B mitigate the risk due to inflation.

C eliminate the risk due to interest rate changes.

Solution:

A is correct. If less money is borrowed, there is less risk of cash flow and equity value volatility due to the use of financial leverage. C is not correct because the risk related to changes in interest rates remains. The investor may be able to accept slightly more leverage and mitigate the interest rate risk associated with debt by locking in the current interest rate with a long-term, fixed-rate amortizing loan. However, if interest rates rise, the value of real estate will likely be affected even if the investor did not borrow any money. Higher interest rates mean investors require a higher rate of return on all assets. In addition, the resale price of the property will likely depend on the cost of debt to the next buyer, who may be more likely to rely on higher leverage to finance the purchase. B is not correct because there is still risk of inflation, although real estate tends to have a low amount of inflation risk. But borrowing less money doesn't necessarily mean the property is less affected by inflation. Furthermore, inflation benefits fixed-rate borrowers who are able to repay debt in the future with cash that is worth less than cash borrowed today.

2 ECONOMIC VALUE DRIVERS, ROLE IN PORTFOLIO, AND RISK/RETURN OF REAL ESTATE INVESTMENTS RELATIVE TO STOCKS AND BONDS

b explain portfolio roles and economic value determinants of real estate investments

2.1 Economic Drivers

Real estate return drivers are straightforward. Cash flow is a function of rental income, operating expenses, leverage, and capital spending. The contributors to cash flow are, in turn, driven by the supply of space, demand for space, and other economic factors. Investment vehicle valuation depends on the risk premium investors expect from real estate.

Exhibit 3 shows major economic factors that affect demand for the major property types. The list is by no means exhaustive. The relative importance of each measure can vary by market, property type, and timing of the business and real estate cycle, especially for rapid, extreme changes in the economic factors. Over the course of a full business cycle, however, each factor's relative importance for a market or property type tends to remain stable.

Risks tend to be greatest for those property-type sectors in which tenant/occupant demand for space can fluctuate most widely in the short term (notably, hotels), leases are shorter, and dislocations between supply and demand are most likely to occur (notably, office and hotel). However, the quality and locations of properties, leasing success, and financing status/access to capital are also extremely important factors in determining the investment risk profile.

Exhibit 3 Major Factors Affecting Real Estate Demand by Sector

	Retail	Office	Industrial	Multi-Family	Hotel
GDP growth	x	x	x	x	x
Population growth	x	x	x	x	x
Job creation	x	x	x	x	x
Household formations	x			x	
Wage growth	x	x	x	x	x
Personal income growth	x			x	x
Consumer spending	x		x		x
Retail sales growth	x		x		
Demographic trends	x			x	x
Consumer confidence	x			x	x
Consumer credit	x			x	x
Industrial production			x		
Trade and transportation			x		
Advances in logistics			x		
Changing supply routes			x		
Business formations		x	x	x	x
Business investment		x	x		x
Business confidence		x	x		x
Regulatory	x	x	x	x	x
Taxes	x	x	x	x	x

Growth in the economy or national GDP is generally the most important single economic factor affecting the outlook for all property types. Similarly, population growth, job creation, regulations, and taxes affect all the major sectors. Job creation tends to be reflected in increased demand for office space and in requirements for more retail space to cater to related increases in spending. Job creation also tends to be reflected in (1) increased demand for multi-family accommodation as newly employed people gain the financial means to rent their own accommodations and (2) greater hotel room demand as leisure and business travel increase in response to an expanded workforce. Job creation is also a driver for many of the specialty sectors, including self-storage and data centers.

Household formation is one of the largest drivers of apartment demand. Income, wage growth, and consumer confidence all determine whether residents can afford to move to larger, higher-quality, better-located units or buy a home.

Wage growth, increases in income and disposable income, and improvements in consumer spending generally will support retail sales growth. Even as retail sales increase, online retailers have continued to pick up market share. During the 2019 holiday season, for example, US retail sales increased 4.1% annually whereas online and non-store sales rose 14.6% (excluding automobile dealers, gas stations, and restaurants, according to the National Retail Federation). Large regional shopping centers and department stores experienced the greatest pressure. Other retail categories, such as groceries, home furnishing stores, and drug/personal care stores saw modestly higher sales. Retailers' success or failure and landlords' ability to replace weak tenants with stronger tenants or reposition properties influence rental income and occupancy directly (through rental rates based on a percentage of sales) and indirectly (through tenants' ability to pay rent).

The share of sales claimed by online retailers jumped tremendously when shelter-at-home policies related to COVID-19 forced shopping centers, restaurants, and theaters to close in 2020. Only stores selling necessities (e.g., food and medicine) were allowed to remain open. For the first time, the only way many consumers could purchase non-essential goods was online, and even online sales of groceries and personal care products rose.

Industrial manufacturing and warehouse distribution centers, which often include small offices, have seen increased demand from global trade in and near port cities. Online sales are shifting traditional transportation patterns as retailers look to store inventory closer to customers to speed delivery times. In addition, some brick-and-mortar retailers are allocating more retail space to holding inventory for delivery of online sales.

In contrast to the near- to medium-term trend in income and spending, long-term demographic trends, along with population growth, are major drivers of real estate demand. College graduates and non-child households moving to urban centers, new families moving to suburban markets, and elderly people moving to assisted living facilities are just a few of the demand drivers that have been widely reported on.

As discussed in Section 1.4, "Risk Factors," the supply side of the real estate economic cycle is driven by periods of oversupply, characterized by low occupancy and rental rates, and undersupply, when occupancy and rents are high. Property types with long development and construction periods are more prone to supply–demand dislocations because (1) new construction typically commences in a booming economy when demand for space cannot be accommodated by existing supply and rents rise high enough to provide developers with an attractive return, (2) properties already under development continue to be completed for two or three years after a recession eventually arrives and depresses demand, and (3) the large size of many facilities (especially, trophy office properties), complicated mixed-use properties, or convention center hotels further exacerbates excess supply on completion.

EXAMPLE 4

Economic Value Drivers

1 Which of the following statistics is likely to be most relevant for all of the following: office, industrial, and hotel properties?

 A Business confidence

 B Household formation

 C Industrial production

2 In addition to the market and property-specific analysis of occupancy, rental rate, lease expiry, and financing statistics, analysts of office properties are *most likely* to pay particular attention to trends in:

 A retail sales growth.

 B household formation.

 C job creation.

3 Which of the following property sectors would be expected to experience the *greatest* cash flow volatility?

A Industrial

B Hotel

C Shopping center

Solution to 1:

A is correct. Companies are more likely to expand their business and engage in business travel when business confidence is rising. Household formations have an indirect effect on these sectors and are most relevant to the multi-family sector. Changes in industrial production are less directly tied to the office and hotel markets.

Solution to 2:

C is correct. Job creation is most significant for office REITs. Household formations are more significant for multi-family and retail REITs than for office REITs, whereas retail sales growth is more significant for shopping center/retail and industrial REITs than for office REITs.

Solution to 3:

B is correct. Hotel room demand fluctuates with economic activity and business and consumer confidence; there are no long-term leases on hotel rooms to protect hotel REITs' revenue streams from changes in demand. Industrial and shopping center REITs benefit from long-term leases on their properties and from the relatively mild dislocations between supply and demand caused by the construction of new space in these subsectors.

2.2 Role of Real Estate in an Investment Portfolio

There are many different types of equity real estate investors, ranging from individual investors to large pension funds, sovereign wealth funds, and publicly traded real estate companies. Hereafter, for simplicity, the term *investor* will refer to an equity investor in real estate. Although there may be some differences in the motivations for each type of investor, they all hope to achieve one or more of the following benefits of equity real estate investment:

- *Current income*: Investors may expect to earn current income on the property through letting, leasing, or renting the property. Investors expect that market demand for space in the property will be sufficient to produce net income after collecting rents and paying operating expenses. This income constitutes part of an investor's return. The amount available to the investor will be affected by taxes and financing costs. Historically, income has been the largest component of investor return, with the main exception occurring during the steep decline in real interest rates after the 2008 global financial crisis.
- *Price appreciation (capital appreciation)*: Investors often expect prices to rise over time. Any price increase also contributes to an investor's total return. Investors may anticipate selling properties after holding them for a period of time and realizing the capital appreciation.
- *Inflation hedge*: Investors may expect both rents and real estate prices to rise in an inflationary environment. If rents and prices do, in fact, increase with inflation, then equity real estate investments provide investors with an inflationary hedge. This means that the real rate of return, as opposed to the nominal rate of return, may be less volatile for equity real estate investments.

- *Diversification*: Investors may anticipate diversification benefits. Real estate performance has not typically been highly correlated with the performance of other asset classes—such as stocks, bonds, or money market funds—so adding real estate to a portfolio may lower the risk of the portfolio (that is, the volatility of returns) relative to the expected return.

  Correlation data suggest that real estate property exposure, both private and listed, provides significant diversification benefits. Exhibit 4 shows correlations of annual returns for the 10-year period from 2010 through 2019 that are based on various US, European, and global indexes. Correlations for global private property, global private property funds, global listed real estate equities, and global listed equities are shown in addition to global consumer price index (CPI) data. Exhibit 4 also includes data for global investment-grade Treasuries and global investment-grade fixed-income securities.

Exhibit 4 10-Year Asset Class Correlations, 2010–2019

Category		CPI	Fixed Income		Property-Level Returns		Private Real Estate Funds			Listed Real Estate		Global Equities
	Indices	Global CPI	Bloomberg Barclays Global Treasury Index	Bloomberg Barclays Global Aggregate Index	MSCI Global Property Index	US NPI	GREFI	INREV Annual Fund Index	US ODCE, Net	MSCI Global Listed Property	FTSE EPRA Nareit Developed	MSCI Global World Equities
Category	Compound Annual Return	2.60%	3.90%	2.50%	8.30%	10.20%	8.10%	5.90%	10.40%	10.40%	9.20%	10.30%
CPI	Global CPI	**1.00**	0.42	0.42	-0.11	0.44	-0.53	-0.66	0.41	-0.03	0.05	-0.22
Fixed Income	Bloomberg Barclays Global Treasury Index	0.42	**1.00**	0.97	-0.35	-0.16	-0.31	-0.03	-0.21	0.16	0.30	0.08
	Bloomberg Barclays Global Aggregate Index	0.42	0.97	**1.00**	-0.45	-0.20	-0.46	-0.18	-0.29	0.37	0.50	0.28
Property-Level Returns	MSCI Global Property Index	-0.11	-0.35	-0.45	**1.00**	0.76	0.80	0.40	0.74	-0.28	-0.34	0.40
	US NPI	0.44	-0.16	-0.20	0.76	**1.00**	0.27	-0.22	0.96	-0.13	-0.14	-0.34
Private Real Estate Funds	GREFI	-0.53	-0.31	-0.46	0.80	0.27	**1.00**	0.84	0.32	-0.42	-0.47	-0.34
	INREV Annual Fund Index	-0.66	-0.03	-0.18	0.40	-0.22	0.84	**1.00**	-0.16	-0.32	-0.32	-0.21
	US ODCE, Net	0.41	-0.22	-0.29	0.75	0.96	0.32	-0.16	**1.00**	-0.25	-0.25	-0.36
Listed Real Estate	MSCI Global Listed Property	-0.03	0.16	0.37	-0.28	-0.13	-0.42	-0.32	-0.25	**1.00**	0.98	0.70
	FTSE EPRA Nareit Developed	0.05	0.30	0.50	-0.34	-0.14	-0.47	-0.32	-0.24	0.98	**1.00**	**0.67**
Global Equities	MSCI Global World Equities	-0.22	0.08	0.28	-0.40	-0.34	-0.34	-0.21	-0.36	0.70	0.67	1.00

Note that the correlation between the global private property and global listed property indexes is negative, as are the correlations between the private property indexes and bonds. This indicates the potential for diversification benefits of adding private equity real estate investment to a stock and bond portfolio. When real estate is publicly traded, it tends to behave more like the stock market than the real estate market, at least in the short run. Several studies have shown that listed real estate does perform similarly to private real estate in the medium term after adjusting for leverage and the lagged and smoothed performance of private real estate. However, some argue that because private real estate indexes are based on infrequent appraisals or market transactions, their performance lags changes in the listed markets, which dampens price volatility and correlations with stock indexes. In fact, the correlation between the MSCI global listed real estate index and the global private property index rises above 0.50, in comparison to a negative correlation, when the 10-year period for the listed real estate begins one year prior to the private real estate index. This issue is discussed in more detail in Section 4.1 which covers the uses and shortcomings of appraisal-based indexes.

- *Tax benefits*: A final reason for investing in real estate, which may be more important to some investors in certain countries than others, is the preferential tax benefits that may result. Private real estate investments may receive favorable tax treatment compared with other investments. In other words, the same before-tax return may result in a higher after-tax return on real estate investments compared with the after-tax return on other possible investments. For example, the preferential tax treatment in the United States comes from the fact that real estate can be depreciated for tax purposes over a shorter period (i.e., faster) than the period over which the property actually deteriorates. Although some real estate investors, such as pension funds, do not normally pay taxes, they compete with taxable investors who might be willing to pay more for the same property because of depreciation tax benefits to which they may be entitled. In many countries, REIT structures also offer tax benefits because in those countries, REITs do not pay corporate income taxes on real estate income as long as the income is distributed to shareholders. Exhibit 5 shows the minimum profit distribution obligation for various markets.

Exhibit 5 Most REITs Required to Distribute at Least 90% of Income

Market	Minimum Profit Distribution Obligation
United States	90% of taxable ordinary income
Japan	90% of distributable profits
Hong Kong SAR	90% of net income
United Kingdom	90% of property rental profits
Germany	90% of annual net income
Australia	100% of trust income
Singapore	90% of tax transparent income
Canada	100% of income
Sweden	No REIT regime
France	100% of taxable profit
Netherlands	100% of taxable profit

Source: European Public Real Estate Association, "EPRA Global REIT Survey 2019" (2019).
Notes: Ordered by market capitalization based on the FTSE EPRA Nareit Global Developed Real Estate Index Series. As of 31 December 2019. Exhibit 5 represents a simple view of dividend distribution requirements. Rates may vary depending on the source of income, such as capital gains on property sales, income from real estate securities, and non-real estate income.

EXAMPLE 5

Motivations for Investing in Real Estate

Why would an investor want to include real estate equity investments in a portfolio that already has a diversified mixture of stocks and bonds?

Solution:

Real estate equity offers diversification benefits because it is less than perfectly correlated with stocks and bonds—for direct ownership (private equity investment) in particular. In other words, there are times when stocks and bonds may perform poorly while private equity real estate investments perform well, and vice versa. Thus, adding real estate equity investments may improve the risk-adjusted return of the portfolio.

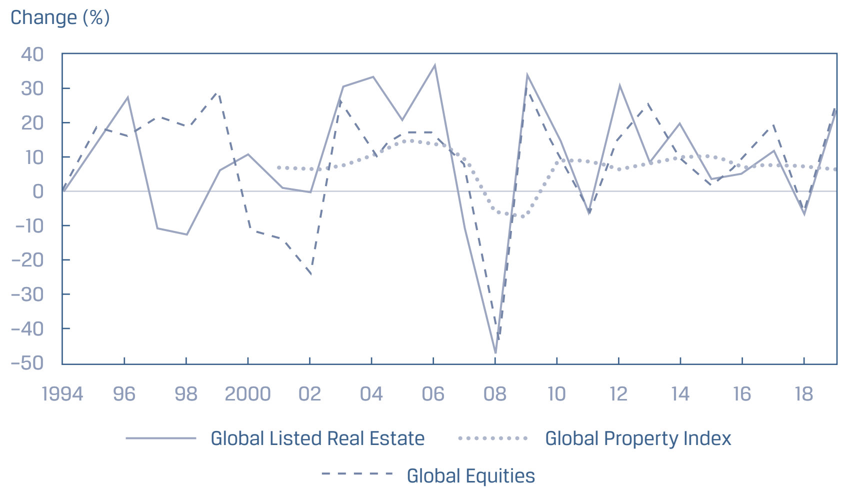
2.3 Real Estate Risk and Return Relative to Stocks and Bonds

Total returns from investing in real estate have proved attractive on an absolute basis. US real estate began the 1990s at depressed values, following significant overbuilding and a weak economy. Many private companies and partnerships embraced REIT IPOs as the best way to recapitalize with equity capital and reduce crushing debt burdens. Investors appreciated that REITs with strong balance sheets could buy real estate for well below replacement cost while rents were generally increasing, which, in turn, supported higher dividends.

Listed REITs fell out of favor toward the end of the 1990s as dot-com mania captured investors' attention. Following the collapse of the dot-com bubble beginning in early 2000, investors turned back to REITs for high dividend yields and cash flow growth. The backdrop of declining interest rates contributed to the attractiveness of REITs. The public and private real estate sectors continued to gain institutional acceptance as investors worldwide increased their allocation to real estate in the early to mid-2000s.

The REIT sector in many markets outperformed listed equities generally through the mid-2000s. The large run-up in real estate stock values reversed when the US housing bubble burst and contributed to the global financial crisis (see Exhibit 6). Most major banks in the United States and Europe dramatically curtailed lending while they focused on improving their own balance sheets, often with government assistance or intervention. Although there was significant overbuilding in the housing market that would take years to absorb, non-residential commercial real estate supply growth remained moderate. Real estate stock values declined in 2007 and collapsed during 2008, wiping out four years of gains, with private fund values turning negative following a slight lag. For the non-residential market, the financial crisis was a liquidity crisis that continued into 2009, making it very difficult to refinance maturing debt, but it was not a real estate overbuilding/excess supply crisis. Still, many listed real estate companies cut dividends and sold equity at depressed prices to refinance maturing debt and strengthen balance sheets through the worst of the crisis.

Exhibit 6 Global Listed Real Estate and Equities Plummeted in 2008



Notes: All indexes are in local currency. The MSCI World IMI Core Real Estate Index is a free-float-adjusted market capitalization index that consists of large-, mid-, and small-cap stocks across 23 developed markets engaged in the ownership, development, and management of specific core property-type real estate. The MSCI Global Annual Property Index (unfrozen; weighting: market size) measures unlevered total returns of directly held standing property investments from one valuation to the next. The index tracks the performance of 55,675 property investments, with a total capital value of USD1,775.5 billion as of December 2018. The MSCI ACWI Index (ACWI stands for "all country world index") captures large- and mid-cap global equities' representation across 23 developed markets and 26 emerging markets; with 2,849 constituents, it covers approximately 85% of the global investable equity opportunity set.
Sources: MSCI, MSCI Real Estate.

Equity markets generally began to move higher in 2009 as governments in many countries stepped forward to support banks and sharply reduce interest rates. While housing markets continued to languish, private commercial real estate and real estate stocks, including the residential rental sector, recovered strongly in response to economic stabilization, improved access to capital, and very low interest rates. Real government bond yields turned negative in some countries, which made dividend-paying stocks, such as REITs, that much more attractive. Real estate investments globally enjoyed as much as a 10-year run, from 2010 through 2019, during which property prices moved higher.

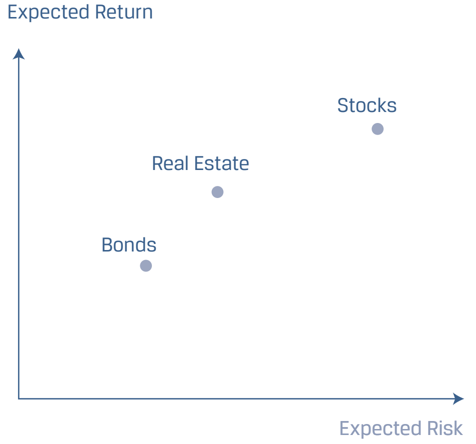
The run-up through 2019 came to an abrupt halt when the COVID-19 pandemic contributed to a collapse in stock values. The FTSE EPRA Nareit global REIT and real estate company index declined 43% from a peak in mid-February 2020 to a low on 23 March 2020 and ended the first quarter of 2020 with a 28% decline from year-end 2019. Through August 2020, the index recovered approximately two-thirds of the first-quarter decline as governments again lowered interest rates and increased fiscal spending.

Real estate investment has appealed to investors for providing income and being "better than bonds" by providing a higher current income with the possibility for income growth. The structure of leases, which are legal agreements requiring the tenant to make periodic payments to the space owner, and exposure to underlying tenant credits give real estate its bond-like characteristics. Like bond prices, real estate values are sensitive to changes in interest rates, inflation, and associated risk premiums.

At the end of the lease term, however, there will be uncertainty as to whether the tenant will renew the lease or the landlord will be required to find a new tenant and what the rental rate will be at that time. These issues depend on the availability of competing space and also on factors that affect the profitability of the companies leasing the space and the strength of the overall economy, in much the same way that stock prices are affected by the same factors. These factors give a stock market characteristic to the risk of real estate.

On balance, because of these two influences (bond-like and stock-like characteristics), core real estate, as an asset class, is expected to have a risk and return profile that falls between the risk and return profiles of stocks and bonds. By this, we mean the risk and return characteristics of a portfolio of real estate versus a portfolio of large-cap stocks and a portfolio of investment-grade bonds. Individual real estate investments or portfolios could certainly have risk that is greater or less than that of an individual stock or bond holdings or portfolios. The more aggressive real estate investment strategies, such as accepting high financial leverage or development risk, carry higher return expectations accompanied by higher volatility. Exhibit 7 illustrates the basic expected risk–return relationships of stocks, bonds, and core private real estate investments. In Exhibit 7, expected risk is measured by the standard deviation of expected returns. Given different correlations with stock and bond returns, it should be evident that adding real estate to a multi-asset-class portfolio expands the efficient frontier. Note that direct real estate won't have transparent (unambiguous) standard deviation and correlations with other, more liquid asset classes (stocks and bonds).

Exhibit 7 Expected Returns and Risks of Core Private Real Estate Compared with Stocks and Investment-Grade Bonds



EXAMPLE 6

Investment Risk

Which is a riskier investment: core private real estate or investment-grade bonds? Explain why.

Solution:

Historically, core private equity real estate with modest leverage is riskier than investment-grade bonds. Although real estate leases offer income streams somewhat like those of bonds, the income expected when leases renew can be uncertain and will depend on market conditions at that time—unlike the more certain face value of a bond at maturity.

2.4 Classifications

c discuss commercial property types, including their distinctive investment characteristics

Section 1.2 introduced several basic real estate classifications and property types. In this section, we expand on the distinctions and characteristics of each property type.

Residential properties include *single-family detached houses* and *multi-family properties*, such as apartments. In general, residential properties are properties that provide housing for individuals or families. Single-family properties may be owner-occupied or rental properties. Multi-family properties contain multiple residential units that can be on the same floor or stacked within a single building; these can include attached townhouses. The names given to the individual units, such as apartments or flats, vary by region. A multi-unit building may be owned by one investor, or each unit may be owned by a separate investor, who may occupy or rent the unit. Individually owned units are referred to as condominiums, or condos, in the United States, Canada and many other countries. In some countries, the units are known as apartments, flats, cooperative houses, and a variety of other names. Multi-family housing is usually differentiated by location (urban or suburban), structure height (high-rise, mid-rise, low-rise, or garden apartments or townhouses), and amenities (pool, balcony, exercise facilities, concierge services, etc.). Residential real estate properties, particularly multi-family properties, purchased with the intent to let, lease, or rent (in other words, produce income) are typically included in the category of **commercial real estate properties** (sometimes called income-producing real estate properties).

Non-residential properties include commercial properties other than multi-family properties, farmland, and timberland. Commercial real estate is by far the largest class of real estate for investment and is the focus of this reading. Commercial real estate properties are typically classified by end use. In addition to multi-family properties, core institutional commercial real estate properties include office, industrial and warehouse, and retail. Hospitality is sometimes included among the major commercial categories, but the higher cash flow volatility and the much greater importance of operations exclude it from being described as one of the core real estate sectors. Note, however, that the same *building* can serve more than one end use. For example, it can contain both office and retail space. In fact, the same building can contain residential as well as non-residential uses of space. A property that has a combination of end users is usually referred to as a *mixed-use development*. Thus, the classifications should be viewed mainly as a convenient way of categorizing the use of space for the purpose of analyzing the determinants of supply and demand and economic performance for each type of space.

- *Office* properties range from major multi-tenant office buildings found in the central business districts of most large cities to single-tenant office buildings. They are often built with the needs of specific tenants in mind (known in real estate terms as *build to suit* if it is for one occupant). Examples of properties developed and built considering the needs of prospective tenants would be a medical office building near a hospital or the office headquarters of a large company. At other times, new construction will begin after an anchor tenant has committed to occupy a large portion of the building and reduced the lease-up risk for the developer. Developments that are preleased to some or all of the tenants are easier to finance than "speculative" construction, which proceeds without tenant commitments. In some markets, speculative development is the norm. In general, speculative construction increases as the property cycle heats up. After a real estate bust, lenders may require preleasing as a condition for financing new development. As the cycle recovers, restrictions generally ease.

- *Industrial and warehouse* properties include wholesale and retail distribution centers, combination warehouse/showroom and office buildings, and light or heavy manufacturing facilities as well as associated warehouse space. Also included are special purpose buildings designed specifically for industrial use that would be difficult to convert to another use. Older buildings that originally had one use may be converted to another use. For example, office space may be converted to warehouse or light industrial space, and warehouse or light industrial space may be converted to residential or office space. Frequently, the conversion is based on the desirability of the area for the new use.
- *Retail* properties vary significantly in size and include the following: large regional shopping centers and malls with large department stores or big-box retailers as anchors and numerous smaller in-line stores between the anchors; neighborhood shopping centers with smaller anchor tenants and many in-line tenants; and stand-alone properties, such as a grocery store or restaurant. As indicated earlier, it is also common to find retail space combined with office space, particularly on the ground floor of office buildings in major cities, or residential space. Office tenants appreciate having restaurants, exercise facilities, and convenience stores in close proximity, and retailers benefit from the daily office-worker traffic.
- *Hospitality* properties vary considerably in size and available amenities. Motels and smaller hotels are used primarily as a place for business travelers and families to spend a night. These properties may have limited amenities and are often located very close to a major highway. Hotels designed for tourists who plan to stay longer usually have a restaurant, a swimming pool, and other amenities. They are also typically located near other attractions that tourists visit. Hotels at "destination resorts" provide the greatest number of amenities. These resorts are away from major cities, and the guests usually stay for several days or even several weeks. Facilities at these resort hotels can be quite luxurious, with several restaurants, swimming pools, nearby golf courses, and so on. Hotels that cater to convention business may be either in a popular destination resort or located near the center of a major city.
- *Other specialty types* of commercial real estate are almost too numerous to name. Investors' search for yield has led to the institutional acceptance of many property types, such as hospitals, medical office buildings, bioscience laboratories, self-storage, student housing, cell towers, and, recently, data centers. Investors can own parking facilities, restaurants, and recreational properties, such as country clubs, marinas, and sports complexes. Retail space that complements the recreational activity (such as gift and golf shops) is often associated with or part of these recreational real estate properties. Dining facilities and possibly hotel or residential facilities may also be present. The physical structure of a building intended for a specific use may be similar to the physical structure of buildings intended for other uses; for example, government office space is similar to other office space. In other cases, buildings intended for one use may not easily be adapted for other uses. For example, buildings used by universities and hospitals may not easily be adapted to other uses.

Some commercial property types are more management intensive than others. Of the main commercial property types, hotels require the most day-to-day management and are more like operating a business than multi-family, office, or retail space. Shopping centers (shopping malls) are also relatively management intensive because it is important for the owner to maintain the right tenant mix and promote the mall. Many of the "other" property types, such as recreational facilities, can also require significant management. Usually, investors consider properties that are more management intensive as riskier because of the operational risks. Therefore, investors typically require a higher rate of return on these management-intensive properties.

EXAMPLE 7

Commercial Real Estate Segments

Commercial real estate properties are *most likely* to include:

A residential, industrial, hospitality, retail, and office.

B multi-family, industrial, warehouse, retail, and office.

C multi-family, industrial, hospitality, retail, and timberland.

Solution:

B is correct. Commercial real estate properties include multi-family, industrial, warehouse, retail, and office, as well as hospitality and other. Residential properties include single-family, owner-occupied homes and income-producing (commercial) residential properties. Timberland is a unique category of real estate.

2.5 Investment Characteristics by Property Type

In this section, the main factors that influence property supply and demand and typical lease terms are discussed. It is important to discuss lease terms because they affect a property's value and the risk/return profile of the investment.

High-quality, well-leased office, retail, industrial/warehouse, and multi-family in strong markets are often considered the *core* property types used to create a portfolio that is relatively low risk. Hotels are usually considered riskier because there are no leases and their performance may be highly correlated with the business cycle—especially if they have a restaurant and depend on convention business. Specialty properties are excluded because the substitutability of the space is relatively low. It does not matter much what type of tenant occupies an office, retail, or distribution facility as long as the tenant blends well with the overall tenant mix and its credit quality is acceptable. Hospitals and cell towers have only one type of tenant, and the facilities are not easily converted to other uses.

For each property type, location is the critical factor in determining value. Properties with the highest value per unit of space are in the best locations and have modern features and functionality. Moderately valued properties are typically in adequate but not prime locations and/or have slightly outdated features. Properties with the lowest values per unit of space are in poor locations and have outdated features.

2.5.1 *Common Types of Leases*

An important consideration in commercial leases is whether the owner or tenant incurs the risk of operating expenses, such as utilities, increasing in the future. A **net lease** requires the tenant to be responsible for paying operating expenses, whereas a **gross lease** requires the owner to pay the operating expenses. Many apartment leases are gross leases, meaning the tenant pays one amount for use of the space and the property owner is responsible for operating expenses, including utilities and real estate taxes. It is also common to see residential tenants be responsible for their own energy (gas and electric) and telephone utilities, cable television costs, and internet access, in addition to the apartment's quoted rent.

Non-residential commercial properties that are net leased require tenants to pay a portion or all of the property's operating expenses in addition to the base or initial fixed rent. The amount varies by region and type of lease and is subject to negotiation. **Triple-net leases** (or NNN leases) are common in the United States and Canada and require each tenant to pay its share of the following three operating expenses: common area maintenance (CAM) and repair expenses; property taxes; and building insurance costs. Tenants are also responsible for insuring their own furnishings, equipment, systems, and so on, against fire, water damage, and other perils. Triple-net leases are among several types of leases common in the Netherlands. In the United Kingdom, lessees may pay a service charge that covers common area expenses, and landlords may negotiate for the tenants to contribute to building insurance costs. In Hong Kong SAR, lessees may pay a management fee, which covers a portion of the common area maintenance and repair expenses. The longer the lease, the more likely that tenants will be responsible for property expenses.

A long-term single-tenant net lease requires the tenant to pay all the operating expenses directly and a base rent to the property owner. This setup is common with a **sale-leaseback** and other types of long-term real estate financings. In a sale-leaseback, a company sells the building it owns and occupies to a real estate investor and the company then signs a long-term lease with the buyer to continue to occupy the building. The tenant is responsible for all aspects of property ownership while it leases the space, including major repairs, such as roof replacement. At the end of the lease, use of the property reverts to the landlord. The tenant is not responsible for normal property depreciation.

The base rent for net leases is lower than that for an equivalent gross lease because the tenant must bear the operating expenses and the risk of expenses being higher than expected. Alternatively, the landlord must charge a higher rent to earn a profit when it is responsible for all property operating expenses.

Not all leases are structured as net or gross leases. For example, a lease may be structured such that in the first year of the lease, the owner is responsible for paying the operating expenses; then, for every year of the lease after that, the owner pays for expenses up to the amount paid in the first year. Any increase in expenses above that amount is passed through to the tenant as an expense reimbursement. That is, the tenant bears the risk of any increase in expenses, although the owner benefits from any decline in expenses. In a multi-tenant building, the expenses may be prorated among the tenants on the basis of the amount of space they are leasing. Although having a small number of tenants can simplify managing a property, it increases risk. If one tenant gets into financial trouble or decides not to renew a lease, it can have a significant effect on cash flows.

EXAMPLE 8

Net and Gross Leases

What is the net rent equivalent for an office building where the gross rent is \$20 per square foot and operating expenses are currently \$8 per square foot?

Solution:

On a gross lease, the owner pays the operating expense, whereas on a net lease, the tenant pays. So, we might expect the rent on a net lease to be \$12 per square foot (or \$20 psf – \$8 psf). With the gross lease, the owner bears the risk of rising operating expenses, whereas the same is true of the tenant with net leases. If expenses decline, the benefit under a gross lease accrues to the owner through improved operating margins because the tenant still would pay \$20. With the net lease, the tenant would benefit from a decline in operating expenses by paying a lower amount to the landlord, thereby reducing the tenant's total cost of occupancy.

Medium- to long-term leases frequently include contractual increases in rents known as rent bumps, rent step-ups, or step-up rents. The lease may specify a given step-up each year, such as 1% of the prior-period rent; a period step-up of, for example, 3% of the prior-period rent every five years; or occasionally an adjustment to mark rents to then-prevailing market rates. It is common to specify step-ups tied to CPI, either annually or cumulatively after several years. The higher a country's inflation rate, the more likely the tenant will pay a base rent plus annual CPI.

Long-term leases provide greater cash flow stability than short-term leases, especially when market-level rents are changing. When market vacancy is low and rents are rising, property owners benefit more from short-term leases because they can raise rents more frequently. The marking to-market of rents hurts owners of properties with short-term leases when market rents decline. As above-market rent leases expire, tenants have greater ability to negotiate lower rates. Rental declines presumably occur when market vacancy increases, providing tenants with more space alternatives. When market rents decline, tenants cannot walk away from their leases in most countries unless they file for bankruptcy. When market rents increase, landlords cannot raise rents on existing leases (unless they negotiated a clause to reset rents from time to time on the basis of market conditions). Investors may prefer properties with long-term leases if they are risk averse or are concerned about declining market rents. Similarly, investors will expect greater returns from owning hotel properties, for example, than from a ground lease, at the other extreme, whereby a tenant has the right to develop and use the land property improvements for an extended period, with, say, 40 years remaining until lease expiration.

2.5.2 *Office*

The demand for office properties depends heavily on employment growth. The typical amount of space used per employee is also important because it tends to increase when the economy is strong and decline when the economy is weak. There also has been a tendency for the average amount of space per employee to decrease over time as technology has allowed more employees to spend more time working away from the office and less permanent space is needed.

The average length of an office building lease varies. Lease terms may depend on the desirability of the property and the financial strength of the tenant, as well as other terms in the lease, such as provisions for future rent changes and whether there are options to extend the lease. Smaller tenants tend to sign three- to five-year leases, whereas larger tenants more commonly sign five- to ten-year leases. Lease length, rent, renewal rights, and the tenant's ability to exit the lease vary by country, regulations, and culture and even through cycles and over time.

For example, in the United Kingdom since the early 1990s, lease terms have fallen. Rents are typically fixed for five years and then set at the higher of the then-market rent or contract rent upon review; these are known as upward-only rent reviews. Leases are typically on a full repairing and insuring (FRI) basis; the tenant is responsible for most costs. Therefore, detailed cost (expense) analysis is much less important in deriving net operating income—a critical measure in estimating the value of a commercial property—in the United Kingdom than in markets where operating costs are typically the responsibility of the owner.

2.5.3 *Industrial and Warehouse*

The demand for industrial and warehouse space heavily depends on the overall strength of the economy and economic growth. The demand for warehouse space also depends on import and export activity in the economy. Industrial leases are often long-term net leases, especially when the property is built specifically for the tenant, although gross leases or leases with expense reimbursements, as described for office properties, also exist.

Industrial and warehouse property values have shifted along with changing domestic and international trade routes. Developed economies outsourcing to low-cost manufacturing centers supported global trade centers. The opening of the wider Panama Canal in 2016, which can accommodate the much larger neo-Panamax container ships, has allowed US Gulf Coast and East Coast ports to accept more shipments from eastern Asia and take share from the US West Coast ports. In 2019 and 2020, various trade restrictions disrupted global trade and COVID-19 accelerated the penetration of online retail sales and e-commerce. The retail trend has increased demand for space closer to population centers. CBRE Research reported that in 2019 approximately 13.8 million square feet of retail shopping space had been or was in the process of being converted to 15.5 million square feet of warehouse space. If developed economies continue to increase domestic manufacturing, commodity and goods distribution routes will shift again.

2.5.4 *Retail*

The demand for retail space depends heavily on trends in consumer spending. Consumer spending, in turn, depends on the health of the economy, population growth, job growth, consumer confidence, and savings rates.

Retail lease terms, including length of leases and rental rates, vary not only on the basis of the quality of the property but also by the size and the importance of the tenant. For example, in the United States, the length of leases is typically shorter (three to five years) for the smaller tenants in a shopping center and is longer for larger, anchor tenants, such as department stores. Anchor tenants may be offered extremely favorable rental terms designed to attract them to the property. The quality of anchor tenants is often a key factor in attracting other tenants.

A unique aspect of many retail leases is the requirement that the tenants pay additional rent once their sales reach a certain level. This type of lease is referred to as a "percentage lease." The lease will typically specify a "minimum rent" or base rent that must be paid regardless of the tenant's sales and the basis for calculating percentage rent once the tenant's sales reach a certain level or breakpoint. For example, the lease may specify a minimum rent of $30 per square foot plus 10% of sales over $300 per square foot. Note that at the breakpoint of $300 per square foot in sales, we obtain the same rent per square foot based on either the minimum rent of $30 or 10% of $300. This is a typical way of structuring the breakpoint, and the sales level of $300 would be referred to as a "natural breakpoint."

EXAMPLE 9

Retail Rents

A retail lease specifies that the minimum rent is $40 per square foot plus 5% of sales revenue over $800 per square foot. What will the rent be if the tenant's sales are $1,000 per square foot?

Solution:

The rent per square foot will be $40 + 5% × ($1,000 – $800), or $40 + $10 = $50. We get the same answer by multiplying 5% × $1,000 (= $50) because $800 is the "natural breakpoint," meaning that 5% of $800 results in the minimum rent of $40. A lease may not have the breakpoint set at this natural level, in which case it is important that the lease clearly define how to calculate the rent.

2.5.5 *Multi-Family*

The demand for multi-family space depends on population growth, especially for the age segment most likely to rent apartments. In other words, population demographics are important. The relevant age segment can be very broad or very narrow, depending on the particular culture's propensity to rent. Homeownership rates vary from country to country. The relevant age segment for renters can also vary by type of property being rented or by locale. For example, the average age of a property renter in an area attractive to retirees may be higher.

Demand also depends on how the cost of renting compares with the cost of owning—that is, the ratio of home prices to rents. As home prices rise and become less affordable, more people will rent. Similarly, as home prices fall, there may be a shift from renting to owning. Mortgage markets also influence rental property and homeownership costs. Countries with well-developed or subsidized mortgage markets will see greater use of leverage. Home ownership usually receives greater subsidies and permits more leverage than investment properties. Higher interest rates will make homeownership more expensive: For owners that partially finance the purchase with debt, the financing cost will be higher, whereas for other homeowners, the opportunity cost of having funds tied up in a home will increase. This increase in the cost of ownership may cause a shift toward renting. If interest rates decrease, there may be a shift toward homeownership.

Multi-family rental properties typically have leases that range from six months to two years, with one year being most common. The tenant may or may not be responsible for paying expenses, such as utilities, depending on whether each unit has a separate meter. The owner is typically responsible for the upkeep of common property, insurance, repair, and maintenance of the property. The tenant is typically responsible for cleaning the space rented and for insurance on personal property.

EXAMPLE 10

Economic Value Determinants

1 The primary economic driver of the demand for office space is most likely:
   - A job growth.
   - B population growth.
   - C growth in savings rates.

2 The demand for which of the following types of real estate is likely most affected by population demographics?
   - A Office
   - B Multi-family
   - C Industrial and warehouse

Solution to 1:

A is correct. Job growth is the main economic driver of office demand, especially for jobs in industries that are heavy users of office space, such as finance and insurance. As the number of jobs increases, companies need to provide office space for the new employees. Population growth may indirectly affect the demand for office space because it affects demand and job growth. Growth in savings rates affects consumer spending and the demand for retail space.

Solution to 2:

B is correct. Population demographics are a primary determinant of the demand for multi-family space.

CONSIDERATIONS IN ANALYSIS AND DUE DILIGENCE

3

d explain the due diligence process for both private and public equity real estate investments

Direct real estate investors (individuals and companies), their advisers, appraisers, and lenders are well advised to perform thorough due diligence before acquiring or valuing properties or making secured loans. Property value will be based on the cash flow outlook, market conditions, and prices paid for recent properties. Therefore, due diligence should include an analysis of all the cash flow drivers, liabilities, and other qualitative factors, such as whether the seller/borrower has clear title to the property. Indirect investors should also perform due diligence on REITS and REOCs to determine whether the share valuation is appropriate before making buy/hold/sell decisions. Much of the information about public companies can be found in public filings of annual audited financial statements, quarterly reports, and investor presentations.

Property due diligence should include an examination of the following:

- Market review: Understand market trends, including local market population, job, and income growth; expected additions to supply; space absorption rates (how much net space is leased each year); tenant preferences; building amenities; market rents; and expense trends.
- Lease and rent review: Compare the tenant rents with market rent forecasts and lease length to determine how much rents will change when leases expire. A landlord who sets lease lengths from three to seven years on average may see approximately 20% of the leases expire in any given year. Or there may be some years with large lease expirations. Work into your review contractual step-ups and percentage rents, if any. Analyze the history of rental payments, late payments, and any defaults for the major tenants.
- Costs of re-leasing space: Look at lease renewal rates and incentives provided to both renewing and new tenants. Costs to lease space when a lease matures typically include brokerage commissions, allowances for tenants' improvements to their space, free rent, and downtime between leases. Such costs can be burdensome for landlords and are not usually included in annual operating income. Instead, these expenses are capitalized and amortized over the length of the lease.
- Seek underlying documentation. Get copies of bills for operating expenses, such as utility expenses and real estate taxes.

- Look at several years of audited financial statements. The seller's property-level cash flow statements should provide a perspective on operating expenses and revenue trends. Beware of underspending to boost operating income or free rent and other tenant incentives to lease vacant space and boost occupancy.
- Perform an environmental inspection to be sure there are no issues, such as a contaminant material at the site. Leaking fuel tanks can be a common problem.
- Have a physical/engineering inspection to be sure the property has no structural issues and to check the condition of the building systems, structures, and foundation and the adequacy of utilities.
- Have an attorney or appropriate party review the ownership history to be sure there are no issues related to the seller's ability to transfer free and clear title that is not subject to any previously unidentified liens.
- Review service and maintenance agreements to determine whether recurring problems exist.
- Have a property survey to determine whether the physical improvements are in the boundary lines of the site and to find out if any easements would affect the value.
- Verify that the property is compliant with zoning laws, environmental regulations, parking ratios, and so on.
- Verify that property taxes, insurance, special assessments, and so on, have been paid.

When differences in income, liabilities, property structural problems, permit issues, and the like are discovered, the investor should rework the cash flow forecast and include the cost to fix shortcomings into the valuation.

EXAMPLE 11

Due Diligence

What is the primary purpose of due diligence?

Solution:

Due diligence is done to identify legal, environmental, physical, and other unanticipated problems that have not been disclosed by the seller that could be costly to remediate or that could negatively affect value. If identified, an issue or issues could result in negotiating a lower price or allow the investor to walk away from the transaction.

A company analysis will require a similar review process. Instead of examining one property, the investor will be evaluating the company's portfolio and potential for growth. The larger the portfolio, the easier it is to make simplifying assumptions, such as applying expected market rent and occupancy growth to the company's properties in the market. Cash flow growth can also increase because of completion and stabilization of properties recently developed, new developments, and acquisitions. The analysis must include the source of capital to finance the growth, be it from new equity, borrowing, or selling assets. Again, the company's audited financials will provide a sense of rent, occupancy, expense, margin, and spending trends over at least two to three years. The financial statements and accompanying notes will describe other assets and liabilities.

The investor's goal is to determine the value of the company and its shares. Two commonly used approaches compare (1) the stock price with the net value per share and (2) cash flow multiples or dividend yields with market comparables, making adjustments for the quality of management, assets, the balance sheet, governance, the cash flow outlook, and so on.

INDEXES

4

e discuss real estate investment indexes, including their construction and potential biases

An investor will find a variety of real estate indexes to choose from and may find one that seems representative of the market of interest to them. There are indexes that measure property income performance, property total return, investment fund performance, and listed security returns. Investors should be aware, however, of how the index is constructed and the inherent limitations resulting from the construction method. Investors should also be aware that the apparent low correlation of real estate with other asset classes may be due to limitations in real estate index construction.

4.1 Appraisal-Based Indexes

Many indexes rely on appraisals to estimate how the value of a portfolio of properties or the real estate market in general is changing over time. Property and private real estate investment indexes often rely on appraisals to estimate values because there usually are not sufficient transactions of the same property to rely on transactions to indicate value. Even though real estate transactions may be occurring, they are not for the same property; differences in sale prices (transaction prices) can be due to changes in the market or differences in the characteristics of the property (size, age, location, and so on). Appraisal-based indexes combine valuation information from individual properties and provide a measure of market movements. (Section 5 discusses common valuation techniques that appraisers also rely on in estimating property values.)

For example, a well-known index that measures the change in values of real estate held by institutional investors in the United States is the NCREIF Property Index (NPI). Members of NCREIF (National Council of Real Estate Investment Fiduciaries) who are investment managers and pension fund plan sponsors contribute to NCREIF every quarter information on appraised values, along with net operating income (NOI), capital expenditures, and other information, such as occupancy. This information is then used to create an index that measures the performance of these properties quarterly. The return for all the properties is calculated as follows:

$$\text{Return} = \frac{\begin{array}{c}\text{NOI} - \text{Capital expenditures}\\ +(\text{Ending market value} - \text{Beginning market value})\end{array}}{\text{Beginning market value}}$$

In this calculation, the beginning and ending market values are based on the appraisals of the properties.

The return calculated with this formula is commonly known as the *holding period return* and is equivalent to a single-period internal rate of return, or IRR (the IRR if the property were purchased at the beginning of the quarter at its beginning market value and sold at the end of the quarter at its ending market value). A similar equation is used to calculate the returns on stocks and bonds, but in those cases, an actual transaction price is typically used. Because this is not possible for real estate, the appraised value is used.

Note that the income return is not the same as cash flow, because cash flow is calculated after capital expenditures. That is, the amount of cash flow available each quarter is **net operating income** (or rental revenue less property operating expenses, including property taxes and insurance) minus capital expenditures. Thus, we can also think of the total return in the formula as measuring the cash flow (NOI – Capital expenditures) plus the change in value (Ending market value – Beginning market value).

An index like the one described allows us to compare the performance of real estate with other asset classes, such as stocks and bonds. The quarterly returns are also important for measuring risk, which is often measured as the volatility or standard deviation of the quarterly returns. A major drawback, however, is that *the income component of real estate returns does not represent distributions to investors in real estate funds or REITs*. The total return for equities is based on capital appreciation plus dividends, not on the underlying company's operating income. The index does succeed, however, as a benchmark to compare returns among individual real estate funds.

NCREIF, which began aggregating data and reporting property index-level returns in 1978, produces a variety of US real estate indexes based on such factors as property type and location. The European Association for Investors in Non-Listed Real Estate Vehicles (INREV), launched in 2003 and performs similar functions. The Asian Association for Investors in Non-Listed Real Estate Vehicles (ANREV) was formed in 2007 as a sister organization to INREV. All three of these organizations contribute data to produce the Global Real Estate Fund Index (GREFI), a capitalization-weighted index incorporating local currency returns. Released quarterly, GREFI launched in 2014, with values going back to 2009.

Several other appraisal-based indexes are also available. MSCI publishes a wide range of property indexes that cover markets worldwide, including emerging markets. These MSCI indexes, formerly called the MSCI IPD (Investment Property Databank) indexes, are calculated in a manner similar to that of the NPI.

EXAMPLE 12

Appraisal-Based Indexes

Why are appraisals often used to create real estate performance indexes?

Solution:

Because properties do not transact very frequently, it is more difficult to create transaction-based indexes as is done for publicly traded securities. Appraisal-based indexes can be constructed even when there are no transactions by relying on quarterly or annual appraisals of the property. Of course, when no transactions occur, it is also difficult for appraisers to estimate value.

4.2 Transaction-Based Indexes

Some indexes are based on actual transactions rather than appraised values. These indexes have been made possible by companies that collect information on enough transactions to create an index based only on transactions. In fact, both NCREIF and MSCI have transaction information that can be used for this purpose. When creating a transaction-based index, the fact that the same property does not sell very frequently is still an issue. So, to develop an index that measures changes in value on a quarterly basis as discussed for appraisal indexes, the fact that different properties sell every quarter needs to be controlled for. Some econometric techniques, usually involving regression analysis, are used to address the issue and to create the index in two main ways. One is to create what is referred to as a **repeat sales index**, and the other is to create what is referred to as a **hedonic index**.

A repeat sales index, as the name implies, relies on repeat sales of the same property. A particular property may sell only twice during the entire period of the index. But if at least some properties have sold each quarter, the repeat sales regression methodology can use this information to create an index. Of course, the more sales, the more reliable the index. In general, the idea supporting this type of index is that because it is the same property that sold twice, the change in value between the two sale dates indicates how market conditions have changed over time. Property and tenant credit quality, the lease maturity schedule, and market conditions may have changed, depending on the amount of time between sales. The regression methodology allocates this change in value to each time period—that is, each quarter on the basis of the information from sales that occurred that quarter. The details of how the regression works are beyond the scope of this reading. An example of a repeat sales index for commercial real estate in the United States is the suite of RCA Commercial Property Price Indices (RCA CPPI).

A hedonic index does not require repeat sales of the same property. It requires only one sale. The way it controls for the fact that different properties are selling each quarter is to include variables in the regression that control for differences in the characteristics of the property, such as size, age, quality of construction, and location. These independent variables in the regression reflect how differences in characteristics cause values to differ so that they can be separated from the differences in value due to changes in market conditions from quarter to quarter. Again, the details of this regression are beyond the scope of this reading. The point is that indexes based only on transactions can be constructed. They require a lot of data and are usually most reliable at the national level for the major property types, but sometimes they are reliable at the regional level within a country if sufficient transactions are available.

EXAMPLE 13

Transaction-Based Indexes

Describe two main ways of creating transaction-based indexes.

Solution:

The two main ways are (1) a repeat sales index and (2) a hedonic index. A repeat sales index requires repeat sales of the same property; because it is the same property, controls for differences in property characteristics, such as its size and location, are not required. A hedonic index requires only one sale of a property and thus can usually include more properties than a repeat sales index; however, it must control for "hedonic" characteristics of the property, such as its size and location.

4.3 Advantages and Disadvantages of Appraisal-Based and Transaction-Based Indexes

All indexes, whether appraisal or transaction based, have advantages and disadvantages. Appraisal-based indexes are often criticized for having appraisal lag, which results from appraised values tending to lag when there are sudden shifts in the market. In an upward market, transaction prices usually start to rise first. Then as these higher prices are reflected in comparable sales and investor surveys, they are captured in appraised values. Thus, appraisal-based indexes may not capture the price increase until a quarter or more after it was reflected in transactions. The same lag would also occur in a down market, with appraised values not falling as soon as transaction prices. Another cause of appraisal lag is that all properties in an appraisal-based index may not be appraised every quarter. A manager may assume the value has stayed the same for several quarters until he or she goes through the appraisal process to estimate a new value. This situation causes a lag in the index. That being said, if the investment managers are all using appraised values to measure returns and if the index is based on appraised values, then it is an "apples to apples" comparison.

If the purpose of the index is for comparison with other asset classes that are publicly traded, however, appraisal lag is more of an issue. Appraisal lag tends to smooth the index, meaning that it has less volatility. It behaves somewhat like a moving average of what an index would look like if it were based on values obtained from transactions rather than appraisals. Thus, appraisal-based indexes may underestimate the volatility of real estate returns. Because of the lag in appraisal-based real estate indexes, they will also tend to have a lower correlation with other asset classes. (Exhibit 4 provides public and private real estate return correlations with fixed income and equities.) The smoothing effect will also overstate Sharpe ratios, which is problematic if the index is used in asset allocation models to determine how much of a portfolio should be allocated to real estate versus other asset classes. The appropriate allocation to and benefits from private real estate would likely be overestimated.

Appraisal lag can be adjusted for in two ways. The first is to "unsmooth" the appraisal-based index. Several techniques have been developed to do this, but they are beyond the scope of this reading. In general, the resulting unsmoothed index will have more volatility and more correlation with other asset classes. The second way of adjusting for appraisal lag is to use a transaction-based index when comparing real estate with other asset classes.

Transaction-based indexes tend to lead appraisal-based indexes for the reasons discussed, but they can be noisy (that is, they include random elements in the observations) because of the need to use statistical techniques to estimate the index. So, there may be upward or downward movements from quarter to quarter that are somewhat random even though in general (viewed over a year or longer) the index is capturing the correct movements in the market. The challenge for those creating these indexes is to try to keep the noise to a minimum through the use of appropriate statistical techniques and collecting as much data as possible.

EXAMPLE 14

Comparing Appraisal-Based and Transaction-Based Indexes

What are the main differences between the performance of appraisal-based and transaction-based indexes?

Solution:

An appraisal-based index will tend to have less volatility and lag a transaction-based index, resulting in a lower correlation with other asset classes being reported.

4.4 Real Estate Security Indexes

The major listed index providers, stock exchanges, and leading REIT trade organizations produce real estate equity security indexes. These include Bloomberg, FTSE Russell, MSCI, Nikkei, and S&P Dow Jones, among many others. There are also numerous real estate debt security indexes, such as the CMBX. FTSE and Nareit produce US REIT indexes; FTSE, EPRA, and Nareit publish European and Asia-Pacific listed real estate company indexes. FTSE, EPRA, and Nareit also provide a variety of global real estate securities indexes.

Depending on the split between REITs and REOCs, the indexes available from the various providers may contain equity REITs only, equity REITs and REOCs, or just REOCs. There are indexes based on market cap, country, property type, exchange listing, and major diversified index membership (e.g., S&P 500 REITs). In addition to total return, some indexes track capital appreciation and dividend yields.

SECTION B. INVESTMENTS IN REAL ESTATE THROUGH PRIVATE VEHICLES

Direct property ownership and investment through private vehicles has long been the preferred choice of institutional investors, including insurance companies, pension funds, sovereign wealth funds, foundations, endowments, and high-net-worth families and individuals. As mentioned previously, investors pursue private real estate for total return, income, tax benefits, and low correlation with other asset classes. Long-term investors expect to earn an illiquidity premium. Others like the control direct investments offer, which allows owners to decided where and when to invest and when to sell. Institutions may also seek private real estate for low volatility of returns stemming from appraisal-based valuations.

INTRODUCTION TO VALUATION APPROACHES

5

f discuss the income, cost, and sales comparison approaches to valuing real estate properties

In general, appraisers, or surveyors as they are known in the United Kingdom, use three different approaches to estimate real estate value: the **income approach**, the **cost approach**, and the **sales comparison approach**.

The income approach considers what price an investor would pay for a property based on forecasted cash flows discounted by an expected rate of return that is commensurate with the risk of the investment. It commonly relies on a discounted cash flow (DCF) analysis to calculate the present value of the expected future income from the property, including proceeds from resale at the end of a typical investment holding period, although there are other methods. The concept is that value depends on the expected rate of return that investors would require to invest in the property. The two commonly used income approaches are discussed in detail in Sections 6 and 7, respectively.

The cost approach considers what it would cost to reproduce or replicate the asset and deducts depreciation and other factors that reduce the value of the property. **Replacement cost** includes the expense of buying the land and constructing a new property on the site that has the same utility or functionality as the property being appraised (referred to as the subject property). Adjustments are made if the subject property is older or not of a modern design, if it is not feasible to construct a new property in the current market, or if the location of the property is not ideal for its current use. The concept is that you should not pay more for a property than the cost of buying vacant land and developing a comparable property. The development cost should include the developer's expected profit that would compensate the developer for development risk, including time and complexity, and the cost of financing development.

The sales comparison approach considers what similar or comparable properties (comparables) transacted for in the current market. It is also referred to as the market approach. Recent property sales serve as the basis for establishing market comparables (market comps), or units of comparison. Price per square meter or square foot of leasable area or total area is the most common measure. The UK-based RICS (Royal Institute of Chartered Surveyors), which promotes international valuation standards, identifies other common units of measurement, many of which are used to value businesses and securities, such as price per gross or net rent per square meter, price-to-revenue, and price-to-earnings before interest, taxes, depreciation, and amortization. Adjustments are made to reflect comparables' differences from the subject property, such as size, age, location, and condition of the property, and to adjust for differences in market conditions at the times of sale. The most recent transactions should carry more weight than prior-period sales. The concept is that you would not pay more than others are paying for similar properties.

EXAMPLE 15

London Office Property Valuation

You have been asked to appraise an office property in London. The following table provides some characteristics about the property and details on three other properties that sold in the past three months. How would you estimate the property value using the information given, and what is your estimate?

Property/Market	Target Property	Taller Towers/City of London	Fairview Ally/ Mayfair	Real Estate Road/ Knightsbridge
Size (square feet)	100,000	500,000	25,000	200,000
Occupancy	75%	85%	80%	80%
Market Net Rent	£80 psf	£65 psf	£90 psf	£75 psf
Property Net Rent	£72 psf	£75 psf	£95 psf	£65 psf
Annual Net Rent for Property	£5,400,000	£31,875,000	£1,900,000	£10,400,000
Annual Operating Income	£4,590,000	£28,687,500	£1,710,000	£8,840,000
Age	22 years	5 years	10 years	15 years
Quality	Class B	Class A	Class A	Class B
Market Rent Trend	Flat	+2%	+5%	Flat
Valuation Metrics				
Selling Price	?	£725,000,000	£40,000,000	£175,000,000
Price Per Square Foot		£1,150	£1,250	£735

Property/Market	Target Property	Taller Towers/City of London	Fairview Ally/ Mayfair	Real Estate Road/ Knightsbridge
Price/Rental Revenue		18.0×	16.3×	14.1×
Price/Operating Income		20.0×	18.0×	16.5×

Notes: Figures are rounded. Values are hypothetical.

Solution:

The target property has more in common with the other Class B property based on quality, age, and rents trailing the market average. In contrast, the two Class A, or Grade A, properties, rent at a premium to the local market, and Taller Towers has the highest occupancy. As the appraiser, you may come up with a range of values based on the property in its current condition, with the in-place tenant leases and occupancy, and what the property would be worth if occupancy and income were higher. Using the most comparable property valuation metrics without any adjustments, values would range from a low of £73.5 million, based on a purchase price of £735 per square foot, to £76.3 million using the same price-to-revenue multiple. A discount to these multiples may be warranted because the target property is older. Alternatively, if the target property's occupancy were to readily increase to 80%, the upper range of the valuation could move to £80.5 million. It is beyond the scope of this example to consider how much it would cost to raise occupancy by spending capital to improve the vacant space and pay broker leasing commissions, nor are we considering the property potential if larger amounts were invested in renovating the property. If you were to estimate the property value following such a renovation, you would subtract the cost of the renovation from the post-renovation value.

These approaches are unlikely to result in the same value because they rely on different assumptions and availability of data to estimate the value. The idea is to try to triangulate on the market value by approaching the estimate three different ways. The appraiser may have more confidence in one or two of the approaches depending on the availability of data for each approach. Part of the appraisal process is trying to reconcile the differences in the estimates of value from each approach and coming up with a final estimate of value for the subject property.

5.1 Highest and Best Use

Before we elaborate on the three approaches to estimating value, it is helpful to understand an important concept known as **highest and best use**. The highest and best use of a vacant site is the use that would result in the highest value for the land. Presumably, the developer that could earn the highest risk-adjusted profit based on time, effort, construction and development cost, leasing, and exit value would be the one to pay the highest price for the land.

Developers commonly back into the cost of land by estimating the expected exit price for the to-be-completed property, subtracting the all-in development costs, exit costs, and their minimum profit requirement. The amount that remains represents the most the developer would be willing to pay for the land. Developers would consider all possible uses given zoning constraints or incorporating the cost paid and time spent securing rezoning approval. Developers might consider the potential profit and risk associated with residential, office, retail, mixed-use, and other property types. Note that the highest and best use is not the use with the highest total value; it is the use that provides the developer with the highest profit based on the return required to compensate it for all the risks associated with the project.

The theory is that the land value is based on its highest and best use *as if vacant* even if an existing building is on the site. If an existing building is on the site that is not the highest and best use of the site, then the value of the building—not the land—will be lower. For example, suppose that a site with an old warehouse on it would sell for $1.5 million as a warehouse (land and building). If vacant, the land is worth $1 million. Thus, the value of the existing building (warehouse) is $500,000 (= $1,500,000 – $1,000,000). As long as the value under the existing use is more than the land value, the building should remain on the site. If another developer could build an office for an all-in cost of $2.5 million—*including its profit requirement*—and sell the property for $4.25 million, then the developer would be willing to pay as much as $2 million for the land. In that case, the value under the existing use falls below the land value ($1.5 million warehouse value minus $2.0 million for the land based on the highest and best use) and any buildings on the site will likely be demolished so the building that represents the highest and best use of the site can be constructed.

EXAMPLE 16

Highest and Best Use

Two uses have been identified for a property. One is an office building that would have a value after construction of $20 million. Development costs would be $16 million, including a profit for the developer. The second use is an apartment building that would have a value after construction of $25 million. Development costs, including a profit to the developer, would be $22 million. What is the highest and best use of the site and the implied land value?

Solution:

	Office	Apartment
Value on completion	$20,000,000	$25,000,000
Cost to construct building	(16,000,000)	(22,000,000)
Implied land value	$4,000,000	$3,000,000

An investor/developer could pay up to $4 million for the land to develop an office building but only $3 million for the land to develop an apartment building. The highest and best use of the site is an office building with a land value of $4 million. Of course, this answer assumes a competitive market with several potential developers who would bid for the land to develop an office building.

We will now discuss each of the approaches to estimating value in more detail and provide examples of each.

THE INCOME APPROACH TO VALUATION: DISCOUNT RATES AND THE DIRECT CAPITALIZATION OF NOI AND DCF METHODS

6

g compare the direct capitalization and discounted cash flow valuation methods

The **direct capitalization method** and the **discounted cash flow method** are income approaches used to appraise or estimate the value of a commercial (income-producing) property. The direct capitalization method estimates the value of an income-producing property based on the level and quality of its net operating income. The DCF method discounts future projected cash flows to arrive at a present value of the property. Net operating income, a measure of income and a *proxy for cash flow*, is a focus of both approaches.

6.1 Similarities in Approaches

Both income approaches focus on net operating income generated from a property. Recall, NOI is a measure of the income from the property after deducting operating expenses for such items as property taxes, insurance, maintenance, utilities, and repairs but before deducting any costs associated with financing (leverage) and before deducting federal income taxes. NOI in a real estate property context is similar to earnings before interest, taxes, depreciation, and amortization (EBITDA) in a financial reporting context. Note that neither property NOI nor EBITDA includes capital spending, financing costs, or taxes, and therefore, neither represents actual cash flow. That is not to say these considerations are unimportant. NOI is just the starting point.

Both income approaches consider growth and property quality. The first, the direct capitalization method, capitalizes the current NOI using an *implicit* growth **capitalization rate**, or cap rate. That is, properties expected to generate faster growth will use a capitalization rate that results in a higher value, and properties with slower growth will use a capitalization rate that results in a lower value. When the capitalization rate is applied to the forecasted first-year NOI for the property, the implicit assumption is that the first-year NOI is representative of what the typical first-year NOI would be for similar properties.

The second approach, the DCF method, applies an explicit growth rate to construct an NOI stream from which a present value can be derived. As we will see, there is some overlap because even for the second method, we generally estimate a terminal value by capitalizing NOI at some future date. Income can be projected either for the entire economic life of the property or for a typical holding period with the assumption that the property will be sold at the end of the holding period. The discount rate should reflect the risk characteristics of the property. It can be derived from market comparisons or from specific analysis; we will examine both cases.

6.2 The Direct Capitalization Method

The direct capitalization method capitalizes the current or expected NOI to calculate real estate value. If we think about the inverse of the cap rate as a multiplier, the approach is analogous to an income multiplier. The direct capitalization method differs from the DCF method, in which future operating income (a proxy for cash flow) is discounted at a discount rate to produce a present value.

6.2.1 *The Capitalization Rate and the Discount Rate*

The cap and discount rates are closely linked but are not the same. Briefly, the discount rate is the return required from an investment and comprises the risk-free rate plus a risk premium specific to the investment. The cap rate is lower than the discount rate because it is calculated using the current NOI. So, the cap rate is like a current yield for the property. The discount rate is applied to current and future NOI, which may be expected to grow. In general, when income and value are growing at a constant compound growth rate, we have

Cap rate = Discount rate − Growth rate. **(1)**

The growth rate is implicit in the cap rate in that the buyer incorporates cash flow growth into the value the buyer is willing to pay for the property, but we have to make it explicit for a DCF valuation. If the growth rate were negative, the cap rate would exceed the discount rate.

6.2.2 Defining the Capitalization Rate The capitalization rate is a very important measure for valuing income-producing real estate property. The cap rate is defined as follows:

Cap rate = NOI/Value, **(2)**

where NOI is usually based on what is expected during the current or first year of ownership of the property. Sometimes the term *going-in cap rate* is used to clarify that it is based on the first year of ownership when the investor is *going into* the deal. (Later, we will explain that the *terminal cap rate* is based on expected income for the year after the anticipated sale of the property.)

The value used in Equation 2 is an estimate of what the property is worth at the time of purchase. If we rearrange Equation 2 and solve for value, we get the following equation:

Value = NOI/Cap rate. **(3)**

So, if we know the appropriate cap rate, we can estimate the value of the property by dividing its first-year NOI by the cap rate.

Where does the cap rate come from? That will be an important part of our discussion. A simple answer is that it is based on observing what other similar or comparable properties are selling for. Assuming that the sale price for a comparable property is a good indication of the value of the subject property, we have

Cap rate of comparable = NOI of comparable/Sale price of comparable. **(4)**

We would not want to rely on the price for just one sale to indicate what the cap rate is. We want to observe several sales of similar properties before drawing conclusions about what cap rates investors are willing to accept for a property. As we will discuss later, there are also reasons why we would expect the cap rate to differ for different properties, such as what the future income potential is for the property—that is, how it is expected to change after the first year. This is important because the cap rate is only explicitly based on the first-year income. But the cap rate that investors are willing to accept depends on how they expect the income to change in the future and the risk of that income. These expectations are said to be implicit in the cap rate.

The cap rate is like a snapshot at a point in time of the relationship between NOI and value. It is somewhat analogous to the price–earnings multiple for a stock except that it is the reciprocal. The reciprocal of the cap rate is price divided by NOI. In the United Kingdom, the reciprocal of the cap rate is called the “years purchase” (YP). It is the number of years that it would take for income at the current level to be equal to the original purchase price. Just as stocks with greater earnings growth potential tend to have higher price–earnings multiples and, inversely, lower earnings yields, properties with greater income growth potential have higher ratios of price to current NOI and thus lower cap rates.

It is often necessary to make adjustments based on specific lease terms and characteristics of a market. For example, a similar approach is common in the United Kingdom, where the term "fully let property" is used to refer to a property that is leased at market rent because either it has a new tenant or the rent has just been reviewed. In such cases, the appraisal is undertaken by applying a capitalization rate to this rent rather than to NOI because leases usually require the tenant to pay all costs. The cap rate derived by dividing rent by the recent sale prices of comparables is often called the all-risks yield (ARY) and is shown in the following formula:

ARY = Rent/Recent sale prices of comparables. **(5)**

Note that the term "yield" in this case is used like a "current yield" based on first-year NOI. It is a cap rate and will differ from the total return that an investor might expect to get from future growth in NOI and value. If it is assumed, however, that the rent will be level in the foreseeable future (like a perpetuity), then the cap rate will be the same as the return and the all-risks yield will be an internal rate of return or yield to maturity.

In simple terms, the valuation is

Market value = Rent/ARY. **(6)**

Again, this valuation is essentially the same as dividing NOI by the cap rate as discussed earlier except the occupant is assumed to be responsible for all expenses, so the rent is divided by the ARY. In practice, management costs should also be considered—although operating costs falling on the landlord are typically much lower than in the United States. ARY is a cap rate and will differ from the required total return (the discount rate) an investor might expect to receive from future growth in NOI and value. When rents are expected to increase after every rent review, the investor's expected return will be higher than the cap rate. If rents are expected to increase at a constant compound rate, then the investor's expected return (discount rate) will equal the cap rate plus the growth rate, as was shown with Equation 1.

EXAMPLE 17

Capitalizing NOI

A property has just been let at an NOI of £250,000 for the first year, and the capitalization rate on comparable properties is 5%. What is the value of the property?

Solution:

Value = NOI/Cap rate = £250,000/0.05 = £5,000,000.

Suppose the rent review for the property in Example 17 occurs every year and rents are expected to increase 2% each year. *An approximation of the IRR would simply be the cap rate plus the growth rate*; in this case, a 5% cap rate plus 2% rent growth results in a 7% IRR. Of course, if the rent review were less frequent, as in the United Kingdom where it is typically every five years, then we could not simply add the growth rate to the cap rate to get the IRR. But the IRR would still be higher than the cap rate if rents were expected to increase.

6.2.3 *Stabilized NOI*

When the cap rate is applied to the forecasted first-year NOI for the property, the implicit assumption is that the first-year NOI is representative of what the typical first-year NOI would be for similar properties. In some cases, the appraiser might project an NOI to be used to estimate value that is different from what might actually be expected for the first year of ownership for the property if what is actually expected is not typical.

An example of this situation might be when a property is undergoing a renovation and has a temporarily higher-than-typical amount of vacancy until the renovation is complete. The purpose of the appraisal might be to estimate what the property will be worth once the renovation is complete. A cap rate will be used from properties that are not being renovated because they are more typical. Thus, the appraiser projects what is referred to as a **stabilized NOI**, which is what the NOI would be if the property were not being renovated—in other words, what the NOI will be once the renovation is complete. This NOI is used to estimate the value. Of course, if the property is being purchased before the renovation is complete, a slightly lower price will be paid because the purchaser has to wait for the renovation to be complete to get the higher NOI. Applying the cap rate to the lower NOI of the renovation period will understate the value of the property because it implicitly assumes that the lower NOI is expected to continue.

EXAMPLE 18

Value of a Property to Be Renovated

A property is being purchased that requires some renovation to be competitive with otherwise comparable properties. Renovations satisfactory to the purchaser will be completed by the seller at the seller's expense. If it were already renovated, it would have NOI of ¥9 million next year, which would be expected to increase by 3% per year thereafter. Investors would normally require a 12% IRR (discount rate) to purchase the property after it is renovated. Because of the renovation, the NOI will be only ¥4 million next year. But thereafter, the NOI is expected to be the same as it would be if it had already been renovated at the time of purchase. What is the value of or the price a typical investor is willing to pay for the property?

Solution:

If the property were already renovated (and the NOI had stabilized), the value would be as follows:

Value if renovated = ¥9,000,000/(0.12 – 0.03) = ¥100,000,000.

But because of the renovation, there is a loss in income of ¥5 million during the first year. If for simplicity we assume that this amount would have been received at the end of the year, then the present value of the lost income at a 12% discount rate is as follows:

Loss in value = ¥5,000,000/(1.12) = ¥4,464,286.

Thus, the value of the property is as follows:

Value if renovated	¥100,000,000
Less loss in value	– ¥4,464,286
= Value	¥95,535,714

An alternative approach is to get the present value of the first year's income and the value in a year when renovated:

{¥4,000,000 + [¥9,000,000(1.03)]/(0.12 – 0.03)]}/(1.12) = ¥95,535,714.

6.2.4 *Other Forms of the Income Approach*

Direct capitalization usually uses NOI and a cap rate. However, there are some alternatives to the use of NOI and a cap rate. For example, a *gross income multiplier* might be used in some situations. The gross income multiplier is the ratio of the sale price to the gross income expected from the property in the first year after sale. It may be obtained from comparable sales in a way similar to what was illustrated for cap rates. The problem with using a gross income multiplier is that it does not explicitly consider vacancy rates and operating expenses. Thus, it implicitly assumes that the ratio of vacancy and expenses to gross income is similar for the comparable and subject properties. But if, for example, expenses were expected to be lower on one property versus another because it was more energy efficient, an investor would pay more for the same rent. Thus, its gross income multiplier should be higher. The use of a gross rent multiplier is also considered a form of direct capitalization but is generally not considered as reliable as using a capitalization rate.

THE DCF METHOD, THE RELATIONSHIP BETWEEN DISCOUNT RATE AND CAP RATE, AND THE TERMINAL CAPITALIZATION RATE

7

h estimate and interpret the inputs (for example, net operating income, capitalization rate, and discount rate) to the direct capitalization and discounted cash flow valuation methods

i calculate the value of a property using the direct capitalization and discounted cash flow valuation methods

The direct capitalization method typically estimates value by capitalizing the first-year NOI at a cap rate derived from market evidence. The DCF method (sometimes referred to as a yield capitalization method) involves projecting income beyond the first year and discounting that income at a discount rate (yield rate). The terms *yield rate* and *discount rate* are used synonymously in this discussion, as are the terms *yield capitalization* and *discounted cash flow analysis*.

7.1 The Relationship between the Discount Rate and the Cap Rate

If the income and value for a property are expected to change over time at the same compound rate—for example, 3% per year—then the relationship between the cap rate and the discount rate is the same as in Equation 1:

Cap rate = Discount rate – Growth rate.

To see the intuition behind this, let us solve for the discount rate, which is the return that is required to invest in the property:

Discount rate = Cap rate + Growth rate.

Recall that the cap rate is based on first-year NOI. The growth rate captures how NOI will change in the future along with the property value. Thus, we can say that the investor's return (discount rate) comes from the return on first-year income (cap rate) plus the growth in income and value over time (growth rate). Although income and value may not always change at the same compound rate each year, this formula gives us insight into the relationship between the discount rate and the cap rate. Essentially, the difference between the discount and cap rates has to do with growth in income and value.

Intuitively, given that both methods start from the same NOI in the first year, you would pay more for an income stream that will grow than for one that will be constant. So, the price is higher and the cap rate is lower when the NOI is growing, which is what is meant by the growth being *implicit* in the cap rate. If the growth rate is constant, we can extend Equation 3 using Equation 1 to give

$$\text{Value of property} = \text{NOI}/(r - g), \quad (7)$$

where

r = the discount rate (required return)

g = the growth rate for income (given constant growth in income, value will grow at the same rate)

This equation is analogous to the dividend growth model applied to stocks. If NOI is not expected to grow at a constant rate, then NOIs are projected into the future and each period's NOI is discounted to arrive at a value of the property. Rather than project NOIs into infinity, NOIs typically are projected for a specified holding period, and a terminal value (estimated sale price) at the end of the holding period is estimated.

EXAMPLE 19

Growth Explicit Appraisal

NOI is expected to be \$100,000 the first year, and after that, NOI is expected to increase at 2% per year for the foreseeable future. The property value is also expected to increase by 2% per year. Investors expect to get a 12% IRR given the level of risk; therefore, the value is estimated using a 12% discount rate. What is the value of the property today (at the beginning of the first year)?

Solution:

$$V = \text{NOI}/(r - g)$$

$$= \$100{,}000/(0.12 - 0.02)$$

$$= \$100{,}000/0.10$$

$$= \$1{,}000{,}000.$$

The property value growth rate is not required to calculate the value of the property after the first year. However, it would be used in a DCF calculation to determine the property's terminal value.

7.2 The Terminal Capitalization Rate

When a DCF methodology is used to value a property, one of the important inputs is generally the estimated sale price of the property at the end of a typical holding period. This input is often referred to as the estimated terminal value. Estimating the terminal value of a property can be quite challenging in practice, especially given that the purpose of the analysis is to estimate the value of the property today. But if we do not know the value of the property today, how can we know what it will be worth in the future when sold to another investor? We must also use some method for estimating what the property will be worth when sold in the future.

In theory, this value is based on the present value of income to be received by the *next* investor. But we usually do not try to project NOI for another holding period beyond the initial one. Rather, we rely on the direct capitalization method using the NOI of the first year of ownership for the next investor and a cap rate. The cap rate used to estimate the resale price or terminal value is referred to as a *terminal cap rate* or *residual cap rate*. It is a cap rate that is selected at the time of valuation to be applied to the NOI earned in the first year after the property is expected to be sold to a new buyer.

Selecting a terminal cap rate is challenging. Recall that the cap rate equals the discount rate less the growth rate when income and value are growing constantly at the same rate. Whether constant growth is realistic or not, we know that the cap rate will be higher (lower) if the discount rate is higher (lower). Similarly, the cap rate will be lower if the growth rate is expected to be higher, and vice versa. These relationships also apply to the terminal cap rate and the going-in cap rate.

The terminal cap rate could be the same, higher, or lower than the going-in cap rate, depending on the expected discount and growth rates at the time of sale. If interest rates are expected to be higher in the future, pushing up discount rates, then terminal cap rates might be higher. The growth rate is often assumed to be a little lower because the property is older at the time of sale and may not be as competitive. This situation would result in a slightly higher terminal cap rate. Uncertainty about what the NOI will be in the future may also result in selecting a higher terminal cap rate. The point is that the terminal cap rate is not necessarily the same as the going-in cap rate at the time of the appraisal.

EXAMPLE 20

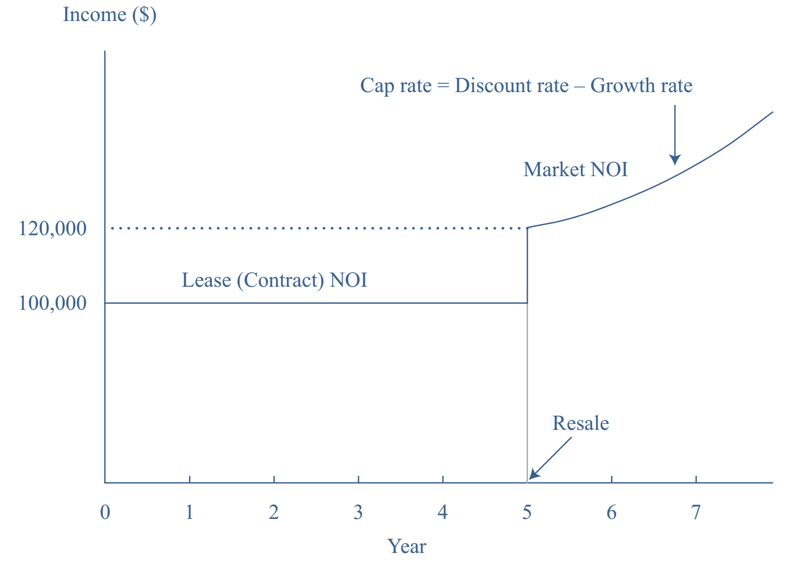
Appraisal with a Terminal Value

Net operating income is expected to be level at $100,000 per year for the next five years because of existing leases. Starting in Year 6, the NOI is expected to increase to $120,000 because of lease rollovers and increase at 2% per year thereafter. The property value is also expected to increase at 2% per year after Year 5. The investors in the property require a 12% return and expect to hold the property for five years. What is the current value of the property?

Solution:

Exhibit 8 shows the projected NOI for this example. Because NOI and property value are expected to grow at the same constant rate after Year 5, we can calculate the cap rate at that time based on the discount rate less the growth rate. That gives us a terminal cap rate that can be used to estimate the value that the property could be sold for at the end of Year 5 (based on the income a buyer would get after that). We can then discount this value along with the income for Years 1–5 to get the present value.

Exhibit 8 Projected Income



Step 1:

Estimate resale price after five years.

Resale (residual) or "terminal" cap rate = 12% – 2% = 10%.

Apply this to NOI in Year 6:

Resale = $120,000/0.10 = $1,200,000.

Note that the value that can be obtained by selling the property at some point in the future is often referred to as the "reversion" by real estate professionals.

Step 2:

Discount the level NOI for the first five years and the resale price.

PMT = $100,000.

FV = $1,200,000.

n = 5.

i = 12%.

Solving for PV, the current value of the property is estimated to be $1,041,390. Note that the implied going-in cap rate is $100,000/$1,041,390 = 9.60%.

In Example 20, the going-in cap rate is lower than the terminal cap rate. Investors would be willing to pay a higher price for the current NOI because they know that it will increase when the lease is renewed at market rents in five years. The expected rent jump on lease renewal is implicit in the cap rate.

As noted earlier, we often expect the terminal cap rate to be higher than the going-in cap rate because it is being applied to income that is more uncertain. Also, the property will be older and may have less growth potential or require increased spending. Finding a lower implied going-in cap rate in this example is consistent with this. At certain times, however, we would expect the terminal cap rate to be lower than the going-in cap rate—for example, if we thought that interest rates and thus discount rates would be lower when the property is sold in the future or we expected that markets would be a lot stronger in the future, with expectations for higher rental growth than in the current market. In general, the higher the risk, the higher the cap rate required to lift the investment return commensurately.

EXAMPLE 21

Appraisal with Level NOI

Suppose the NOI from a property is expected to be level at \$600,000 per year for a long period of time such that, for all practical purposes, it can be assumed to be a perpetuity. What is the value of the property assuming investors want a 12% rate of return?

Solution:

In this case, the growth rate is zero, so we have

Value = NOI/Discount rate.
Value = \$600,000/0.12 = \$5,000,000.

Note that in this case, the cap rate will be the same as the discount rate. This is true when there is no expected change in income and value over time.

EXAMPLE 22

Discounted Cash Flow Analysis

You work for a real estate investment firm that has been presented with the opportunity to purchase a 10-year-old warehouse distribution facility in Perth, Australia. The property contains 20,000 square meters (m^2) of leasable area that is 100% leased to three tenants. Two of the tenants have 10 years remaining on their respective leases. The third tenant has only three years left on its lease and has already indicated it will not renew the lease at expiration. One of the other tenants has indicated it would like to lease the space as soon as it becomes available as long as the property owner makes material improvements to the building.

Part 1:

Your manager has asked you to prepare a discounted cash flow analysis to determine whether the acquisition would meet the company's return targets. Based on the assumptions provided, calculate the first-year NOI, the purchase price, and the terminal value. Note that the rent presented is net of all operating expenses.

Assumptions			
Annual Net Rent per Square Meter	AUD 145.00	Initial Cap Rate	6.0%
Leasable Area (square meters)	20,000	Terminal Cap Rate	7.0%
Lease 1 (m^2)	6,897	Exit Year	7

(continued)

Assumptions			
Lease 2 (m^2)	6,897	NOI in Year 9 (millions)	AUD 3.79
Lease 3 (m^2)	6,207		

Solution:

First-year NOI = AUD145/m^2 × 20,000 m^2 of leasable area

= AUD2.9 million.

Purchase price = Year 1 NOI/Cap rate = AUD2.9 million/6.0%

= AUD48.3 million.

Terminal value at the end of Year 7 = AUD3.79 million/7.0%

= AUD54.1 million.

Part 2:

Why do you think your manager asked you to use a 6.0% acquisition cap rate and a 7.0% terminal cap rate?

Solution:

The 6.0% cap rate would be based on the cap rate from recent property sales adjusted for location, property age and quality, and tenant/lease composition. The 7.0% terminal cap rate is 100 bps higher than the initial cap rate because of a combination of the following: The property will be older and, besides the renovation of one building, will have continued to depreciate; the government benchmark rate, which is approximately 2.0%, could increase over the next seven years; and a measure of conservatism is advisable. A higher cap rate implies a lower *relative* price based on future earnings.

Part 3:

This part of the example illustrates the cash flow forecast, the DCF, and NPV/return analysis.

Net property income increases at a 3.0% annual rate based on lease terms, which is consistent with the market. Additional capital spending and other below-the-line expenses that are not reported on the income statement are estimated at 3.5% of NOI. The forecast includes the large renovation during Year 4 to prepare the space vacated by the tenant whose lease expires at the end of Year 3 for one of the remaining tenants. Year 4 NOI includes a 20% increase above the 3.0% trend, which was negotiated with the tenant for the new lease, offset by six months of forgone rent during the renovation as incentive for the expanding tenant. Capital spending is also assumed to be lower in the year immediately preceding and the year following the renovation before returning to the steady-state trend. Recall that the NOI forecast needs to extend one year past the exit year to calculate the forward income that the next property buyer would use to value the property and present its own bid.

Cash Flow Forecast (AUD thousands)

Property Net Operating Income:	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Lease 1		1,000.0	1,030.0	1,060.9	1,092.7	1,125.5	1,159.3	1,194.1	1,229.9
Lease 2		1,000.0	1,030.0	1,060.9	1,092.7	1,125.5	1,159.3	1,194.1	1,229.9
Lease 3		900.0	927.0	954.8	590.1	1,215.5	1,242.0	1,289.6	1,328.3
Net Property NOI		2,900.0	2,987.0	3,076.6	2,775.5	3,466.6	3,570.6	3,677.7	3,788.0
Capital Spending and Leasing Expense		(101.5)	(104.5)	(74.3)	(1,094.4)	(78.8)	(125.0)	(128.7)	(132.6)
Property Cash Flow		2,798.5	2,882.5	3,002.3	1,681.2	3,387.8	3,445.6	3,549.0	3,655.4

The next table presents the unlevered discounted cash flow. Your manager suggests using a 7.0% discount rate for the investment based on the attractive, lower-risk-location, 10-year leases; the likelihood that an existing tenant will expand into the to-be-vacated space; and strong market fundamentals. The present value of the future cash flows is then compared with the acquisition price. (It is only a coincidence that the discount rate equals the terminal cap rate.) If the net present value is positive, the investment would create some value over the hurdle rate of return represented by the discount rate.

Discounted Cash Flow Analysis (AUD thousands)

	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
Initial outlay	−48,333							
Annual property cash flow		2,799	2,882	3,002	1,681	3,388	3,446	3,549
Exit value								54,114
Total cash flow	(48,333)	2,799	2,882	3,002	1,681	3,388	3,446	57,663
Discount rate (input)	7.0%	$1/(1+0.07)^1$	$1/(1+0.07)^2$	$1/(1+0.07)^3$	$1/(1+0.07)^4$	$1/(1+0.07)^5$	$1/(1+0.07)^6$	$1/(1+0.07)^7$
Present value factor		0.9346	0.8734	0.8163	0.7629	0.7130	0.6663	0.6227
Present value of property cash flows		2,615	2,518	2,451	1,283	2,415	2,296	35,910

(continued)

(Continued)

NPV and Return Analysis	
Sum of present value of cash inflows	49,488
Cash outlay	(48,333)
Net present value of future CF	1,154
IRR (Output)	7.4%

The NPV of the property exceeds the acquisition price by AUD1.2 million. The positive net present value demonstrates that the investment would create value and excess return relative to the risk-adjusted cost of capital. The 7.4% expected IRR further confirms that the forecasted return would exceed the 7.0% discount rate.

Part 4:

This part of the example provides two sensitivity tables: (1) a sensitivity analysis on the entry cap rate and (2) a sensitivity analysis on the exit cap rate.

Sensitivity Tables (AUD thousands)

	Initial Capitalization Rate				
	7.0%	**6.5%**	**6.0%**	**5.5%**	**5.0%**
Purchase price	41,429	44,615	48,333	52,727	58,000
NPV	3,847	2,626	1,154	(647)	(2,896)

	Terminal Capitalization Rate				
	8.0%	**7.5%**	**7.0%**	**6.5%**	**6.0%**
Exit price	47,350	50,507	54,114	58,277	63,134
NPV	(3,058)	(1,092)	1,154	3,747	6,771

The preceding sensitivity tables present the range of net present values in thousands of Australian dollars for several entry and exit cap rates. The corresponding purchase and exit prices, also in thousands, are presented alongside the cap rates. The acquisition table is presented from lower asset price/higher cap rate to higher asset price/lower cap rate. The table indicates that the investor could pay more than AUD48.3 million, or the 6.0% cap rate on Year 1 NOI, and still exceed the cost of capital but not as much as AUD52.7 million, which equates to a 5.5% cap rate. In fact, the investor could pay approximately AUD51.1 million, the equivalent of a 5.67% cap rate, and still break even on the basis of the given discount rate. The breakeven purchase price represents the maximum value the investor could pay for the acquisition while holding all other factors constant without the net present value turning negative.

PRIVATE MARKET REAL ESTATE DEBT

8

j calculate and interpret financial ratios used to analyze and evaluate private real estate investments

Thus far, our focus has been on analyzing a property without considering whether debt financing would be on the property or whether it would be purchased on an all-cash basis. The reason is that the way a property is financed should not affect the property's value. This does not mean that the overall level of interest rates and the availability of debt in the market do not affect values. It means that for a given property, the investor paying all cash should be paying the same price as one who decides to use some debt financing. Of course, investors who do use debt financing will normally expect to earn a higher rate of return on their equity investment because they expect to earn a greater return on the property than what they will be paying the lender. Thus, there will be positive financial leverage. By borrowing money, the investor is taking on more risk in anticipation of a higher return on equity invested. The risk is higher because with debt comes more uncertainty as to what return the investor will actually earn on equity, since the investor gets what is left over after paying the lender. A small drop in property value can result in a large decrease in the investor's return if a high amount of debt was used to finance the property. When a property is valued without explicitly considering financing, the discount rate can be thought of as a weighted average of the rate of return an equity investor would want and the interest rate on the debt.

The maximum amount of debt that an investor can obtain on commercial real estate is usually limited by either the ratio of the loan to the appraised value of the property (loan to value, or LTV) or the debt service coverage ratio (DSCR), depending on which measure results in the lowest loan amount. The debt service coverage ratio is the ratio of the first-year NOI to the loan payment (referred to as debt service for commercial real estate). That is,

$$\text{DSCR} = \text{NOI/Debt service.} \tag{8}$$

The debt service includes both interest and principal payments on the mortgage. The principal payments are the portion of the loan payment that amortizes the loan over the loan term. An interest-only loan is one that has no principal payments, so the loan balance will remain constant over time. Interest-only loans typically either revert to amortizing loans at some point or have a specified maturity date. For example, an interest-only loan might be made that requires the entire balance of the loan to be repaid after 7–10 years (referred to as a "balloon payment"). Lenders typically require a DSCR of 1.25–1.5 depending on the property type to provide a margin of safety that the NOI from the property can cover the debt service.

EXAMPLE 23

Loans on Real Estate

A property has been appraised for $5 million and is expected to have NOI of $400,000 in the first year. The lender is willing to make an interest-only loan at an 8% interest rate as long as the loan-to-value ratio does not exceed 80% and the DSCR is at least 1.25. The balance of the loan will be due after seven years. How much of a loan can be obtained?

Solution:

Based on the loan-to-value ratio, the loan would be 80% of $5 million, or $4 million. With a DSCR of 1.25, the maximum debt service would be $400,000/1.25 = $320,000. This amount is the mortgage payment that would result in a 1.25 DSCR for an interest-only loan.

If the loan is interest only, then we can obtain the loan amount by simply dividing the mortgage payment by the interest rate. Therefore, the loan amount would be $320,000/0.08 = $4,000,000.

In this case, we obtain the same loan amount from either the LTV or the DSCR requirements of the lender. If one ratio had resulted in a lower loan amount, that would normally be the maximum that could be borrowed.

When financing is used on a property, equity investors often look at their first-year return on equity or the "equity dividend rate" as a measure of how much cash flow they are getting as a percentage of their equity investment. This measure is sometimes referred to as a *cash-on-cash return* because it measures how much cash they are receiving as a percentage of the cash equity they put into the investment.

EXAMPLE 24

Equity Dividend Rate

Using the information in Example 23, what is the equity dividend rate, or cash-on-cash return, assuming the property is purchased at its appraised value?

Solution:

The first-year cash flow is the NOI less the mortgage payment.

NOI	$400,000
Debt service	$320,000
Cash flow	$80,000

The amount of equity is the purchase price less the loan amount.

Price	$5,000,000
Mortgage	$4,000,000
Equity	$1,000,000

The equity yield rate is the cash flow divided by equity: $80,000/$1,000,000 = 8%. Keep in mind that this is not an IRR that would be earned over a holding period until the property is sold. The equity investor does not share any of the price appreciation in the value of the property with the lender.

For loans called "participation" loans, the lender might receive some of the price appreciation, but it would be in exchange for a lower interest rate on the loan.

EXAMPLE 25

Leveraged IRR

Refer to Examples 23 and 24. Suppose the same property is sold for $6 million after five years. What IRR will the equity investor receive on his or her investment?

Solution:

The cash flow received by the equity investor from the sale will be the sale price less the mortgage balance, or \$6 million − \$4 million = \$2 million. Using a financial calculator,

PV = −\$1,000,000 (using a calculator, this is input as a negative to indicate the negative cash flow at the beginning of the investment).
PMT = \$80,000.
n = 5.
FV = \$2,000,000.
Solving for i gives 21.14%.

This IRR is based on the equity invested in the property.

EXAMPLE 26

Unleveraged IRR

Refer to Examples 23, 24, and 25. What would the IRR be if the property were purchased on an all-cash basis (no loan)?

Solution:

Now the equity investor will receive all the cash flow from the sale (\$6 million) and the NOI (\$400,000). The initial investment will be \$5 million. Using a financial calculator,

PV = −\$5,000,000.
PMT = \$400,000.
n = 5.
FV = \$6,000,000.
Solving for i gives 11.20%.

This IRR is based on an unleveraged (all-cash) investment in the property. The difference between this IRR (11.20%) and the IRR the equity investor receives with the loan calculated in Example 25 (21.14%) reflects positive financial leverage. The property earns 11.20% before adding a loan, and the loan is at 8%. So, the investor is benefiting from the spread between 11.20% and 8%.

SECTION C. INVESTMENTS IN REAL ESTATE THROUGH PUBLICLY TRADED SECURITIES

Real estate investment trusts were initially conceived of as a way for small investors to gain exposure to a professionally managed, diversified real estate portfolio. REITs were viewed as a type of (closed-end) mutual fund and income passthrough vehicle through which the portfolio manager would acquire attractively valued properties, occasionally sell fully valued properties, and distribute property earnings to the trust's investors. Until legislation was passed in the United States in 1960 to authorize REITs, real estate investing was reserved for the wealthy and institutions. The Netherlands followed suit in 1969. The US model and other types of tax-advantaged real estate investment vehicles have been adopted worldwide. Subsequent liberalization of US REIT restrictions made it easier for pension funds to invest in REITs. The S&P 500 added REITs as a separate GICS sector in 2016.

Today, more than 35 countries have REITs or REIT-like structures, more are considering adopting similar vehicles, and REITs are held by individuals and institutions alike. The FTSE EPRA Nareit Developed Index had a market cap of $1.45 trillion as of March 2020, and 30 REITs are members of the S&P 500.

9 TYPES OF PUBLICLY TRADED REAL ESTATE SECURITIES

k discuss types of REITs

Publicly traded real estate securities allow investors to gain indirect exposure to real estate equity and debt, primarily mortgages, by purchasing shares of companies that own real estate, real estate loans, or both. The very definition of securitization—transforming an illiquid asset into a financial product—makes it possible for investors of all sizes to access an asset class that was once available only to the largest investors and the diversification of the underlying asset pool. Globally, the principal types of publicly traded real estate securities are real estate investment trusts, real estate operating companies, and mortgage-backed securities.

- **Real estate investment trusts** are companies or trusts that own, finance, and—to a limited extent—develop income-producing real estate property. REITs that own real estate are called **equity REITs**. Those that make or invest in loans secured by real estate are categorized as mortgage REITs. The companies' tax advantages result from being allowed to deduct dividends paid from income, which effectively exempts REITs from corporate income tax in many countries. In many jurisdictions, qualifying REITs are simply exempt from corporate income tax. Most REITs are required to distribute 90%–100% of their taxable income to shareholders.
- **Real estate operating companies** are ordinary taxable real estate ownership companies. Businesses are organized as REOCs, as opposed to REITs, if they are located in countries that do not have a tax-advantaged REIT regime in place, if they engage to a large extent in the development of for-sale real estate properties, or if they offer other non-qualifying services, such as brokerage and third-party property management. The primary cash inflows for merchant developers are from sales of developed or improved properties rather than from recurring lease or rental income. Other companies prefer the more flexible operating company structure, even when they develop, own, and operate qualifying rental properties, because the REIT prohibitions may be too restrictive or they may prefer to retain earnings for reinvestment.
- Mortgage-backed securities are asset-backed securitized debt obligations that represent rights to receive cash flows from portfolios of mortgage loans—mortgage loans on commercial properties in the case of commercial mortgage-backed securities (CMBS) and mortgage loans on residential properties in the case of residential mortgage-backed securities (RMBS). The market capitalization of publicly traded real estate equity securities is greatly exceeded by the market value of real estate debt securities—in particular, RMBS. Whereas

residential mortgage pools may contain thousands of loans, commercial mortgage pools may contain more than 100 loans or as few as 1 loan when the asset is very large.

In addition to publicly traded real estate securities, there are privately held real estate securities, including private REITs and REOCs, privately held mortgages, private debt issues, and bank debt. Many real estate private equity partnerships create private REITs to own income-producing properties.

9.1 REIT Structures

REITs are tax-efficient conduits for distributing earnings from rental income to shareholders. Most are structured as corporations or trusts. There are numerous requirements for a company to qualify as a REIT. In most countries, REITs are required to distribute 90%–100% of their otherwise taxable earnings, invest at least 75% of their assets in real estate, and derive at least 75% of income from real estate rental income or interest on mortgages. Countries may specify a minimum number of shareholders, maximum share ownership by a single shareholder, a minimum number of properties/maximum asset concentration, a maximum level of non-rental income, a maximum amount of development, and limits on leverage and types of loans. In the United States, a REIT must have at least 100 shareholders, and no fewer than 5 shareholders can own more than 50% of the shares (the five-or-fewer rule). There are numerous other requirements as well. The restrictions effectively bar an individual or small group from creating REITs to own individual assets.

Most REITs in the United States are self-managed and self-advised. Senior executives are company employees who report to trustees or the board of directors, who, in turn, are elected by shareholders. Fully integrated REITs generally have fewer conflicts than REITs that are externally advised or externally managed. Externally managed REITs pay asset management fees to the third-party adviser, which has an inherent incentive to increase the size of the REIT if fees are based on total assets. External managers may require REITs to pay for other services that are provided by affiliates of the manager, such as property management, acquisitions, and debt placement.

That is not to say all externally managed REITs should be avoided. Management quality, governance, alignment, reputation, and transparency clearly matter. Several services rate the quality of governance, transparency, and so forth, to assist in the investment valuation, for both externally managed and self-managed REITs.

In the United States, shareholder taxation of REIT dividends received may vary according to the underlying source of income. The portion of the distribution derived from ordinary rental income is classified as ordinary income and taxed at investors' top marginal tax rates. The portion of distributions in excess of a REIT's earnings is categorized as return of capital and is deducted from the investor's share cost basis for tax purposes (if an investor later sells shares at a higher price than the reduced cost basis and generates a long-term capital gain, the profit would qualify for lower capital-gains tax treatment). REIT distributions of capital gains on property sales may also be taxed at shareholders' capital gains tax rate. There are other classifications, such as recaptured depreciation, that are beyond the scope of this reading. In any country, foreign investors are likely to be subject to the host country's withholding tax rate.

9.2 Market Size

As of May 2019, the market value of publicly traded real estate investment trusts and real estate operating companies globally was approximately US$2.6 trillion for the developed markets, whereas the total face value of residential and commercial mortgage-related securities outstanding in the United States alone was approximately US$9.8 trillion. Details about the market's relative size by geographic area and security type are shown in Exhibit 9.

Exhibit 9 Relative Size and Composition of Publicly Traded Real Estate Equity Security Markets

A. Percentage of market value of publicly traded real estate equity securities (REITs and REOCs) in developed markets as of 31 May 2019

By Region (%)		By Market (%)	
North America	56.7	United States	53.9
Asia Pacific	26.4	Japan	11.3
Europe	16.8	Hong Kong SAR	7.9
Middle East, Africa	0.1	Australia	4.7
		Germany	4.7
		United Kingdom	4.6
		Canada	2.8
		Singapore	2.5
		Sweden	1.7
		Netherlands	1.6
		France	1.5
		Others	2.8

B. Percentage of market value of publicly traded equity real estate equity securities in developed markets by type of structure as of 29 March 2019

	Global	North America	Europe	Asia Pacific
REITs	68	96	43	42
Non-REITs, REOCs	32	4	57	58

Sources: www.ftserussell.com and www.epra.com. Based on data from the FTSE EPRA Nareit Developed Index.

As an investment asset class, income-producing real estate offers the advantages of stable income based on its contractual revenue from leases and a measure of long-term inflation protection because, over the long term, rents tend to rise with inflation. For example, in the United States over 1998–2018, the FTSE Nareit All Equity REITs Index achieved a compounded annual total return of 10.0%, compared with 5.6% for the S&P 500 and 4.5% for the Bloomberg Barclays US Aggregate Bond Index. The US Consumer Price Index increased by an average of 2.2% annually over the same period.

9.3 Benefits and Disadvantages of Investing in REITs

The benefits of investing in public real estate companies as compared with private real estate investments include the following:

1 *Liquidity*: Ability to buy and sell shares of almost any amount on major exchanges

2 *Transparency*: Readily available share prices and transaction histories

3 *Diversification*: By property type, geography, and underlying tenant credit when medium-sized and larger companies own several dozen properties, thousands of rental apartment units, and millions of square feet of leasable office space, retail space, industrial space, and so on

4 *High-quality portfolios*: Many companies own high-quality assets in leading markets

5 *Active professional management*: Most companies have strong executive management overseeing dedicated property management teams and achieve economies of scale and operating efficiencies

6 *High, stable income*: Well occupied properties subject to long-term leases generate predictable property income, sometimes with distributions occurring monthly

7 *Tax efficiency*: REIT and passthrough structures avoid corporate income taxation, leaving only the investor to pay taxes on dividends received

The largest disadvantage for REITs that seek to expand their portfolios is the lack of retained earnings. These companies need to access capital markets for equity and debt to fund growth. The faster the expansion, the more often the company must raise new capital. REITs are also constrained in the types of assets they own. Consequently, many REITs form **taxable REIT subsidiaries** (TRS), which pay income taxes on earnings from non-REIT-qualifying activities, such as merchant development or third-party property management.

EXAMPLE 27

Advantages of Publicly Traded Real Estate Investments

1 Which of the following assets requires the *most* expertise in real estate on the part of the investor?

  A A REOC share

  B An equity REIT share

  C A direct investment in a single property

2 Which of the following has the *most* operating and financial flexibility?

  A A REOC

  B An equity REIT

  C A direct investment in a single property

3 Investors seeking broad diversification would invest in the securities of which of the following companies?

  A A company that owns multi-family rental properties in Hong Kong SAR

  B A company that owns large office properties in New York City, San Francisco, Los Angeles, and Chicago

  C A company with a mix of office and retail properties in urban and suburban markets

Solution to 1:

C is correct. Direct investment in a single property requires a high level of real estate expertise. Investment in publicly traded equity investments (in REITs or REOCs) requires much less expertise because investors benefit from having their property interests actively managed on their behalf by professional managers and from having their business interests overseen and guided by boards of directors, as in the case of all public corporations.

Solution to 2:

A is correct. REOCs are free to invest in any kind of real estate or related activity without limitation. This freedom gives management the opportunity to create more value in development activity and in trading real estate and to retain as much of their income as they believe is appropriate. A wider range of capital structures and degrees of financial leverage may be used in the process. In contrast to REOCs, REITs face restrictions on the amount of income and assets accounted for by activities other than collecting rent and interest payments. Direct investment is less liquid and divisible than REOC and REIT shares, which limits the operational flexibility of such investment.

Solution to 3:

C is correct. It should be clear that a company with a mix of assets—office and retail—with exposure to urban and suburban markets offers the best diversification. A is incorrect because the company has only one type of asset, multi-family rentals, in one market, Hong Kong SAR. The systematic risk is high for that portfolio. B is incorrect because the company owns only one asset type, office properties, and the economic activity correlation may be high among urban cities with exposure to global trade and the financial sector.

Alternatively, investors looking for property and market diversification might, instead of the solutions provided, consider investing in fewer, large companies that own different asset types in multiple cities or several pure-play companies, each of which concentrates on a single asset type in its given region, as long as the companies' regions and product type do not overlap to a large extent.

EXAMPLE 28

Publicly Traded Real Estate Investments

Which of the following best represents an advantage of REITs over a direct investment in an income-producing property?

A Diversification

B Operating flexibility

C Income growth potential

Solution:

A is correct. REITs provide diversification of property holdings. B is incorrect because REITs do face restrictions on the amount of income and assets accounted for by activities other than collecting rent and interest payments; these restrictions can prevent a REIT from maximizing its returns. C is incorrect because the relatively low rates of income retention that are required to maintain a REIT's tax-free status can detract from income growth potential.

EXAMPLE 29

Investment Objectives

Two real estate investors are each choosing from among the following investment types: a REOC, an equity REIT, or a direct investment in an income-producing property. Investor A's primary objective is liquidity, and Investor B's primary objective is maximum growth/capital gain potential. State and explain which real estate investment type best suits:

1 Investor A.

2 Investor B.

Solution to 1:

For Investor A, with a liquidity objective, REOC and REIT investments are most appropriate because REOCs and REITs are traded on stock exchanges and are more liquid. Direct investments in income-producing property are generally less liquid.

Solution to 2:

For Investor B, with a maximum growth objective, REOCs and direct property investment are most appropriate because REOCs and direct investors are free to invest in any kind of real estate or related activity without limitation and to reinvest as much of their income as they believe is appropriate for their objectives. This freedom gives them the opportunity to create more value in development activity and in trading real estate. REITs' constraints prevent them from retaining earnings to reinvest, so their growth opportunities are more limited.

There are several caveats to note for each generalized solution. Shares of closely held listed companies with low market float that trade infrequently may not offer the desired liquidity. Management quality, corporate governance, balance-sheet capacity and leverage, attractive investment and reinvestment opportunities, and many other considerations matter greatly when it comes to selecting the vehicle and company that are best at delivering growth and value to shareholders. There are many REITs that offer dividend reinvestment programs (DRIPs), and there are some REOCs that continuously develop without selling assets when real estate valuations are high.

VALUATION: NET ASSET VALUE APPROACH 10

l. justify the use of net asset value per share (NAVPS) in REIT valuation and estimate NAVPS based on forecasted cash net operating income

The approaches analysts take in valuing equity include those based on asset value estimates, price multiple comparisons, and discounted cash flow.

Two possible measures of value that analysts might use are book value per share (BVPS) and net asset value per share (NAVPS) based on reported accounting values and market values for assets, respectively. In this reading, BVPS refers to depreciated real estate value rather than total shareholders' equity per share. NAVPS is the relevant market-based valuation measure for valuing REITs and REOCs.

NAVPS is a fundamental benchmark for the value of a REIT or REOC. In Europe and Asia, the price-to-NAV multiple is the primary measure that analysts use to value real estate companies. (US analysts more commonly report on price multiples of gross cash flow, as discussed in Section 11.) Real estate **net asset value** may be viewed as the largest component of the intrinsic value of a REIT or REOC. NAVPS should also include investors' assessments of the value of any non-asset-based income streams (e.g., fee or management income); the value of non–real estate assets, including cash, net of the value of any contingent liabilities; and the value added by management of the REIT or REOC.

Shares priced at discounts to NAVPS are interpreted as indications of potential undervaluation, and shares priced at premiums to NAVPS, in the absence of indications of positive future events, such as a successful property development completion or expected high value creation by a management team, suggest potential overvaluation. However, these assessments must be made in the context of the stock market's tendency to be forward looking in its valuations and at times to have different investment criteria from property markets. In addition, the stock price discount or premium to NAVPS may be explained by investors' view of management's added value, leverage, and company governance. REITs whose shares trade below NAVPS or have high leverage may have a more difficult time raising new capital to fund acquisitions and development, which, in turn, may limit long-term growth, in contrast to REITs that trade at or above NAVPS. Selling equity below NAVPS can be dilutive for investors.

10.1 Accounting for Investment Properties

If accounting is on a fair value basis, accounting values may be relevant for asset-based valuation. If historical cost values are used, however, accounting values are generally not relevant.

Under International Financial Reporting Standards (IFRS), companies are allowed to value investment properties using either a cost model or a fair value model. The cost model is identical to the cost model used for property, plant, and equipment. Under the fair value model, all changes in the asset's fair value affect net income. To use the fair value model, a company must be able to reliably determine the property's fair value on a continuing basis. In general, a company must consistently apply its chosen model (cost or fair value) to all of its investment property. If a company chooses the fair value model for its investment property, it must continue to use the fair value model until it disposes of the property or changes its use such that it is no longer considered investment property (e.g., it becomes owner-occupied property or part of inventory). The company must continue to use the fair value model for that property even if transactions on comparable properties, used to estimate fair value, become less frequent.

Investment property appears as a separate line item on the balance sheet. Companies are required to disclose whether they use the fair value model or the cost model for their investment property. If the company uses the fair value model, it must make additional disclosures about how it determines fair value and must provide reconciliation between the beginning and ending carrying amounts of the investment property. If the company uses the cost model, it must make additional disclosures—for example, the depreciation method and useful lives must be disclosed. In addition, if the company uses the cost model, it must also disclose the fair value of investment property.

In contrast to IFRS, under US GAAP, most US real estate owners use the historical cost accounting model, which values an asset at its original purchase price plus capital investment less historical depreciation. This model does not accurately represent the economic values of assets and liabilities in environments of significant operating income and asset price changes or long-term inflation. Net asset values can be written down when there is a permanent impairment in economic value, but they can be written up only under exceptional circumstances, such as mergers, acquisitions, or reorganizations. US GAAP historical cost accounting practices tend to distort the measure of economic income and asset value by (1) understating carrying values on long-held property assets that are often appreciating in value because of general price inflation or other property-specific reasons and (2) overstating depreciation when companies use accelerated depreciation.

10.2 Net Asset Value per Share: Calculation

As a result of shortcomings in accounting reported values, investment analysts and investors use estimates of **net asset value per share**. NAVPS is the difference between a real estate company's assets and its liabilities, *all taken at current market values instead of accounting book values*, divided by the number of shares outstanding. NAVPS is a superior measure of a company's net worth compared with historical book value per share.

In valuing a REIT's or REOC's real estate portfolio, analysts will look for the results of existing appraisals if they are available (such as those provided by companies reporting under IFRS). If such appraisals are unavailable or if they disagree with the assumptions or methodology of those appraisals, analysts will often capitalize the rental streams—represented by net operating income—produced by a REIT's or REOC's properties, using a market-required rate of return. NOI is defined as gross rental revenue minus operating costs (which include estimated vacancy and collection losses, insurance costs, taxes, utilities, and repair and maintenance expenses) before deducting depreciation, general and administrative (G&A) expenses, and interest expense. After deducting G&A expenses from NOI, the figure obtained is analogous to earnings before interest, depreciation, and amortization (EBITDA). Recall that this approach is similar to the valuation of private real estate covered in Sections 5 –7. These estimated asset values will be substituted for the book values of the properties on the balance sheet and adjustments made to any related accounting assets, such as capitalized leases, to avoid double counting.

Generally, goodwill, deferred financing expenses, and deferred tax assets will be excluded to arrive at a "hard" economic value for total assets. Liabilities will be similarly adjusted to replace the face value of debt with market values if these are significantly different (e.g., as a result of changes in interest rates), and any such "soft" liabilities as deferred tax liabilities will be removed. The revised net worth of the company divided by the number of shares outstanding is the NAV. Although this figure is calculated before provision for any income or capital gains taxes that might be payable on liquidation, the inability to predict how the company or its assets might be sold and the prospect that it might be kept intact in an acquisition cause investors to look to the pre-tax asset value as their primary net worth benchmark. If a company has held its assets for many years and has a very low remaining depreciable value for its assets for tax purposes, it can affect investors' perspectives on valuation. Quantifying the effects of a low adjusted cost base, however, is impeded by lack of knowledge of the tax circumstances and strategies of a would-be acquirer.

Exhibit 10 provides an example of the calculations involved in estimating NAV based on capitalizing rental streams. Because the book values of assets are based on historical costs, the analyst estimates NAVPS. First, by capitalizing NOI with certain adjustments, the analyst obtains an estimate of the value of rental properties; then, the value of other tangible assets is added and the total is netted of liabilities. This net amount, NAV, is then divided by the number of shares outstanding to obtain NAVPS.

Exhibit 10 Analyst Adjustments to REIT Financials to Obtain NAVPS

Last-12-month real estate NOI	$270,432
Less: Non-cash rent	7,667

(continued)

Exhibit 10 (Continued)

Plus: Adjustment for full impact of acquisitions (1)	4,534
Pro forma cash NOI for last 12 months	$267,299
Plus: Next-12-month growth in NOI (2)	$4,009
Estimated next-12-month cash NOI	$271,308
Assumed cap rate (3)	7.00%
Estimated value of operating real estate	$3,875,829
Plus: Cash and equivalents	65,554
Plus: Land held for future development	34,566
Plus: Accounts receivable	45,667
Plus: Prepaid/other assets (4)	23,456
Estimated gross asset value	$4,045,072
Less: Total debt	1,010,988
Less: Other liabilities	119,886
Net asset value	$2,914,198
Shares outstanding	55,689

1 An incremental 50% of the annual expected return on acquisitions that were completed midway through 2020.

2 Growth is estimated at 1.5%.

3 Cap rate is based on recent comparable transactions in the property market.

4 This figure does not include intangible assets.

NAVPS is calculated to be $2,914,198 divided by 55,689 shares, which equals $52.33 per share.

The second line in Exhibit 10 shows the adjustment to remove **non-cash rent**; these are the result of the accounting practice of "straight lining" the rental revenue from long-term leases with contractual step-ups. When the real estate company reports the average contractual rent it expects to receive over the course of each lease, rent received from the tenant is less than the average revenue booked during the early years of the lease, and the tenant pays more rent than the company reports during the latter years of the lease term. (The amount of this deduction is the difference between the average contractual rent over the leases' terms and the cash rent actually paid.) NOI is also increased to reflect a full year's rent for properties acquired during the course of the year, resulting in pro forma "cash NOI" for the previous 12 months of $267,299,000. This amount is then increased to include expected growth for the next 12 months at 1.5%, resulting in expected next-12-month cash NOI of $271,308,000.

An appropriate capitalization rate is then estimated on the basis of recent transactions for comparable properties in the property market. An estimated value for the REIT's operating real estate is obtained by dividing expected next-12-month cash NOI by the decimalized capitalization rate (in this case, 0.07). The book values of the REIT's other tangible assets, including cash, accounts receivable, land for future development, and prepaid expenses, are added to obtain estimated gross asset value. (Land is sometimes taken at market value if this amount can be determined reliably; but because land is often difficult to value and of low liquidity, analysts tend to use book values.) From this figure, debt and other liabilities (but not deferred taxes, because this item is an accounting provision rather than an economic liability) are subtracted to obtain net asset value. Division by the number of shares outstanding produces NAVPS.

10.3 Net Asset Value per Share: Application

NAVPS can be reasonably estimated when there are ample market transactions to provide property comparables. Investors can make observations about how such properties trade on the basis of the capitalization rate (the rate obtained by dividing net operating income by total value) or on the basis of price per square foot. Broker reports and private real estate research companies also track rental rates by property and other tenant incentives, such as free rent or capital to improve the space, and then apply these valuations to the assets of a public company. In the United States, close to 15% of commercial real estate is held by publicly traded REITs, according to EPRA (www.epra.com). In Europe, only 7% of the commercial real estate market is owned by listed real estate companies (REITs and REOCs), and in Singapore, 34% of the commercial market is owned by listed real estate companies

Over time, REITs and REOCs in the United States and globally have at times traded at premiums to NAV of more than 25% and at other times at discounts to NAV exceeding 25%. Thus, if the NAV of a REIT were $20 per share, the stock might trade as low as $15 per share or as high as $25 per share, depending on a range of factors.

10.3.1 *Important Considerations in a NAV-Based Approach to Valuing REITs*

Although NAV estimates provide investors with a specific value, a number of important considerations should be taken into account when using this approach to value REITs and REOCs. First, investors must understand the implications of using a private market valuation tool on a publicly traded security. In this context, it is useful to examine how NAVs are calculated.

The methods most commonly used to calculate NAV are (1) using the cap rate approach to valuing the NOI of a property or portfolio of properties, (2) applying value per square foot (or unit) to a property or portfolio of properties, and (3) using appraised values disclosed in the company's financial statements. An analyst may adjust these appraised values reported by the company if she does not agree with the underlying assumptions and if there is sufficient information to do so. In the first two instances, the cap rates and values per square foot are derived from observing transactions that have occurred in the marketplace. In contrast, most sophisticated direct purchasers of commercial real estate arrive at a purchase price after performing detailed forecasting of the cash flows they expect to achieve from owning and operating a specific property over their investment time horizon. These cash flows are then discounted to a present value or purchase price. Whatever that present value or purchase price is, an analyst can estimate value by dividing an estimate of NOI by the cap rate—essentially, the required rate of current return for income streams of that risk. In addition, an analyst can take the present value or purchase price and divide by the property's rentable area for a value per square foot. The point is that cap rates and values per square foot result from a more detailed analysis and discounted cash flow process. The discount rate used by a private owner/operator of commercial real estate could differ from the discount rate used by investors purchasing shares of REITs.

Real estate stocks trade at premiums and discounts to NAV. The price-to-NAV ratio will vary by market, sector, outlook, and perceived quality of management and governance. Private property investors may or may not value individual assets the same way public equity investors value listed real estate companies. Property buyers frequently consider the long-term prospects and valuation for an asset when making an investment. Appraisal-based NAV estimates, however, often lag changes in market conditions. Stock investors tend to focus more on the near-term projected outlook for changes in income and asset value. These factors help explain why share valuation may differ from NAV. As alluded to earlier, it is possible that REITs and REOCs can trade at some premium or discount to NAV until the premium/discount becomes wide enough for market forces to close the arbitrage gap.

Another factor to consider when using a NAV approach to REIT/REOC valuation is that NAV implicitly treats a company as an individual asset or static pool of assets. In reality, such treatment is not consistent with a going-concern assumption. Management teams have different track records and abilities to produce value over time, assets can be purchased and sold, and capital market decisions can add or subtract value. An investor must thus consider how much value a management team can add to (or subtract from) current NAV. For instance, an investor may be willing to purchase REIT A trading at a 10% premium to NAV versus REIT B trading at a small discount to NAV because the management team of REIT A has a stronger track record and better opportunities to grow the NAV compared with REIT B, thus justifying the premium at which REIT A trades relative to REIT B.

NAV estimates can also become quite subjective when property markets become illiquid and few transactions are observable or when REITs and REOCs own hundreds of properties, making it difficult for an investor to estimate exactly how much the portfolio would be worth if the assets were sold individually. There may also be a large-portfolio premium in good economic environments when prospective strategic purchasers may be willing to pay a premium to acquire a large amount of desired property at once or a large-portfolio discount when there are few buyers for the kind of property in question. In addition, such assets as undeveloped land, very large properties with few comparable assets, properties with specific uses, service businesses, and joint ventures complicate the process of estimating NAV with accuracy and confidence.

10.3.2 *Further Observations on NAV*

Among institutional investors, the most common view is that if REIT management is performing well in the sense of creating value, REITs and REOCs should trade at premiums to underlying NAVPS. This rationale is based on the following:

1. Investors in the stocks have liquidity on a day-to-day basis, whereas a private investor in real estate does not, thus warranting a lower required return rate (higher value) in the public market than in the private market for the same assets.
2. The competitive nature of the public markets and the size of the organizations should attract above-average management teams, which should produce better real estate operating performance and lead to better investment decisions than the average private real estate concern.

In conclusion, although NAV is by its nature an absolute valuation metric, in practice it is often more useful as a relative valuation tool. If all REITs are trading above or below NAV, selecting individual REITs could become a relative exercise—that is, purchasing the REIT stock trading at the smallest premium to NAV when REITs are trading above NAV or selling the REIT trading at the smallest discount to NAV when REITs are all trading at a discount to NAV. In practice, NAV is also used as a relative metric by investors looking at implied cap rates. To calculate the implied cap rate of a REIT or REOC, the current price is used in a NAV model to work backward and solve for the cap rate. By doing so, an investor looking at two similar portfolios of real estate could ascertain whether the market is valuing these portfolios differently on the basis of the implied cap rates.

VALUATION: RELATIVE VALUE (PRICE MULTIPLE) APPROACH

11

m describe the use of funds from operations (FFO) and adjusted funds from operations (AFFO) in REIT valuation

Conventional equity valuation approaches, including market-based or relative value approaches, are used with some adaptations to value REITs and REOCs. Such multiples as the price-to-funds from operations ratio (P/FFO), the price-to-adjusted funds from operations ratio (P/AFFO), and the enterprise value-to-EBITDA ratio (EV/EBITDA) are used for valuing shares of REITs and REOCs in much the same way as for valuing shares in other industries.

11.1 Relative Value Approach to Valuing REIT Stocks

REIT analysts and investors make extensive use of two measures of operating performance that are specific to REITs. **Funds from operations (FFO)** is defined as net income plus depreciation and amortization less gains or losses on the sale of real property. It is one of the most commonly used metrics in the United States. (In Europe and Asia, NAVPS is more commonly used, as discussed in Section 10.) **Adjusted funds from operations (AFFO)** subtracts recurring capital expenditure and the difference between reported rents and cash rents (i.e., rent straight lining). AFFO better approximates a company's sustainable dividend-paying capacity. These definitions are discussed in greater detail in Section 11.2.

The relative value measures most frequently used in valuing REIT shares are P/FFO and P/AFFO. The ratio EV/EBITDA is used to a lesser extent. Similar to the P/E and P/CF multiples used for valuing equities in other industries, P/FFO and P/AFFO multiples allow investors to quickly ascertain the value of a given REIT's shares compared with that of other REIT shares or to compare the current valuation level of a REIT's shares with historical levels. Within the REIT sector, P/FFO and P/AFFO multiples are also often compared with the average multiple of companies owning similar properties—for example, comparing the P/FFO multiple of a REIT that owns office properties with the average P/FFO multiple for all REITs owning office properties. These multiples are typically calculated using current stock prices and year-ahead estimated FFO or AFFO.

FFO and AFFO are based on net income available to equity and thus represent levered income. P/FFO multiples are generally lower for companies with higher leverage, all things equal. EBITDA, by definition, measures income before the leveraging effect of debt. Not only do EV/EBITDA multiples facilitate like-for-like valuation comparisons; they also better approximate how investors evaluate real estate. Recall that the inverse of the multiple, EBITDA/EV, closely approximates the real estate capitalization rate formula (NOI/market value).

There are three main drivers that differentiate P/FFO, P/AFFO, and EV/EBITDA multiples among most REITs and REOCs:

1 *Expectation for growth in FFO and AFFO*: The higher the expected growth, the higher the multiple or relative valuation. Growth can be driven by business model (e.g., REITs and REOCs successful in real estate development often generate above-average FFO and AFFO growth over time); geography (e.g., having a concentration of properties in primary, supply-constrained markets, such as New York City or London, can give landlords more pricing power and higher cash flow growth than can be obtained in secondary markets); and other factors (e.g., management skill or lease structure).

2 *Risk associated with the underlying real estate*: Cash flow volatility related to asset type, quality, and age; market conditions; lease types; and submarket location also affect valuation. For example, owning apartments is viewed as having less cash flow variability than owning hotels. As such, apartment-focused REITs have tended to trade at relatively high multiples compared with hotel REITs. Likewise, shares of companies with younger, well-maintained portfolios generally trade at higher multiples than stocks of companies with older or out-of-date properties with deferred maintenance that will require higher capital expenditures to sustain rent growth.

3 *Risks associated with the company's capital structures and access to capital*: As financial leverage increases, equities' FFO and AFFO multiples decrease because required return increases as risk increases. Higher leverage constrains a company's incremental borrowing capacity and may create a stock overhang if investors avoid buying shares in anticipation of future equity offerings.

There are many other factors that affect valuation, as with any investment, including investor perceptions of management, asset types or markets being in or out of favor, complexity, quality of financial disclosure, transparency, and governance.

FFO has some shortcomings, but because it is the most standardized measure of a REIT's or REOC's earning power, P/FFO is the most frequently used multiple in analyzing the sector. It is, in essence, the REIT sector equivalent of P/E. Investors can derive a quick "cash flow" multiple by looking at P/AFFO because AFFO makes a variety of adjustments to FFO that result in an approximation of cash earnings.

11.2 Funds from Operations and Adjusted Funds from Operations

FFO has long been the standard measure of REIT performance. The National Association of Real Estate Investment Trusts (Nareit) took steps to standardize and promote the definition. FFO is an SEC-accepted non-GAAP financial measure (as is EBITDA), which, according to the SEC and as specified in updated guidance from Nareit (2018), must be reconciled with GAAP net income. The SEC also recommends that companies that report adjustments to FFO reconcile those figures with the Nareit-defined FFO, sometimes referred to as Nareit FFO.

FFO attempts to approximate continuing operating performance. The more complete definition of FFO is as follows: net income (computed in accordance with GAAP) plus losses (minus gains) from sales of properties, plus depreciation and amortization related to real estate, plus real estate impairments and write-downs unrelated to depreciation.

Why is depreciation added back to net income? Investors believe that real estate maintains its value to a greater extent than other business assets, often appreciating in value over the long term, and that depreciation deductions under IFRS and US GAAP do not represent economic reality. A taxable REOC that uses a moderate degree of leverage and regularly chooses to reinvest most of its income in its business usually will be able to defer a large part of its annual tax liability—that is, its cash income taxes will be low as a result of the accelerated depreciation rates for tax purposes permitted in most countries, and reinvesting continues to add to the depreciable real estate base. (Section 2.2 highlights the tax benefits derived from investing in real estate.)

Net income is adjusted for gains and losses from sales of previously depreciated operating properties on the grounds that they do not represent sustainable, normal income. The amortization add-back includes amortization of leasing commissions, tenant improvements, and tenant allowances.

Similar to cash flow from operations, FFO is not a measure of cash flow. It does not include investment and spending necessary to sustain cash flow growth or cash flow related to financing activities. FFO also includes FFO from unconsolidated businesses.

Adjusted funds from operations, also known as **funds available for distribution (FAD)** or **cash available for distribution**, is a refinement of FFO that is designed to be a more accurate measure of current economic income. AFFO is most often defined as FFO adjusted to remove any non-cash rent and to subtract maintenance-type capital expenditures and leasing costs (including leasing agents' commissions and tenants' improvement allowances). So-called **straight-line rent** is the average contractual rent over a lease term, and this figure is recognized as revenue under IFRS and US GAAP. The difference between this figure and the cash rent paid during the period is the amount of the non-cash rent or **straight-line rent adjustment**. Because most long-term leases contain escalating rental rates, this difference in rental revenue recognition can be significant. Also, deductions from FFO for capital expenditures related to maintenance and for leasing the space in properties reflect costs that need to be incurred to maintain the value of properties.

The purpose of the adjustments to net earnings made in computing FFO and AFFO is to obtain a more tangible, cash-focused measure of sustainable economic income that reduces reliance on non-cash accounting estimates and excludes non-economic, non-cash charges.

AFFO is superior to FFO as a measure of economic income and thus economic return because it takes into account the capital expenditures necessary to maintain the economic income of a property portfolio. AFFO is also more reflective of a REIT's dividend-paying ability than FFO. It is open, however, to more variation and error in estimation than FFO. The precise annual provision required to maintain and lease the space in a property is difficult to predict, and the actual expense in any single year may be significantly more or less than the norm because of the timing of capital expenditure programs and the uneven expiration schedule of leases. Consequently, estimates of FFO are more frequently referenced measures, although analysts and investors will tend to base their investment judgments to a significant degree on their AFFO estimates. Although many REITs and REOCs compute and refer to AFFO in their disclosures, their methods of computation and their assumptions vary. Firms that compile statistics and estimates of publicly traded enterprises for publications, such as Bloomberg and Refinitiv, tend not to gather AFFO estimates because of the absence of a universally accepted methodology for computing AFFO and inconsistent corporate reporting of actual AFFO figures, which hinders corroboration of analysts' estimates. An example of FFO data compiled for US equity REITs by property classification is shown in Exhibit 11.

Exhibit 11 FFO of All Listed US Equity REITs

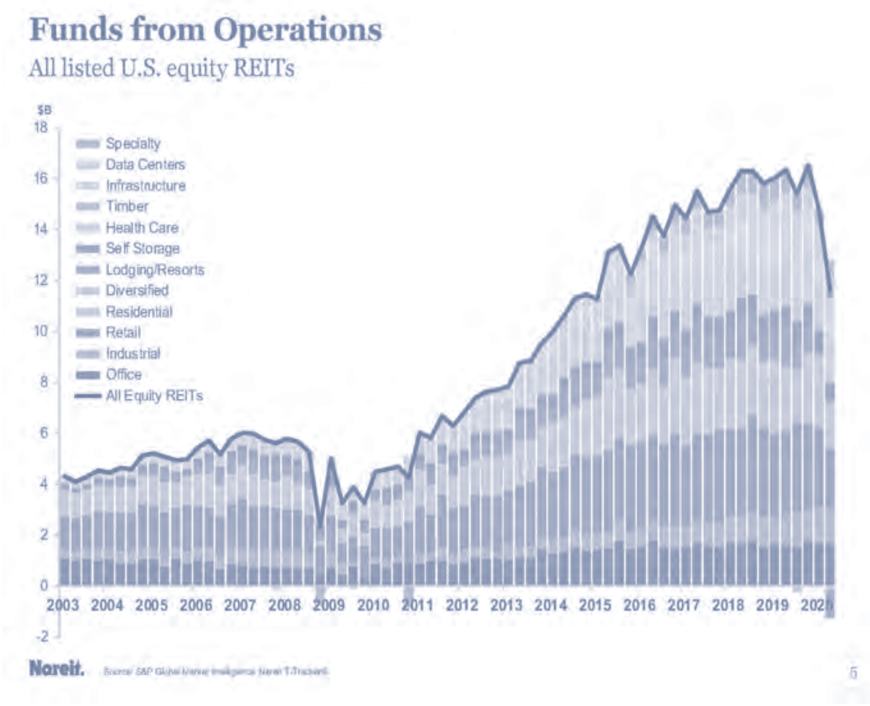


Exhibit 12 illustrates the most straightforward, convenient way of calculating FFO and AFFO for a hypothetical firm, Office Equity REIT Inc.

Exhibit 12 Calculation of FFO and AFFO for Office Equity REIT Inc. (SGD thousands, except per-share data)

A. Calculation of funds from operations	
Net income	160,638
Add: Depreciation and amortization	76,100
Add: (Gains)/losses from sale of depreciable real estate	25,000
Funds from operations	261,738
FFO per share (55,689 shares outstanding)	4.70

B. Calculation of adjusted funds from operations	
Funds from operations	261,738
Less: Non-cash (straight-line) rent adjustment	21,103
Less: Recurring maintenance-type capital expenditures and leasing commissions	55,765
Adjusted funds from operations	184,870
AFFO per share (55,689 shares outstanding)	3.32

EXAMPLE 30

Analyst Adjustments (I)

1 Which of the following is the *best* measure of a REIT's current economic return to shareholders?

A FFO

B AFFO

C Net income

2 An analyst gathers the following information for a REIT:

Net operating income	\$115 million
USD	\$1,005 million
Market value of debt outstanding	\$505 million
Market cap rate	7%
Shares outstanding	100 million
Book value per share	\$5.00

The REIT's NAV per share is *closest* to:

A \$10.05.

B \$11.38.

C \$16.42.

3 All else equal, estimated NAV per share will decrease with an increase in the:

A capitalization rate.

B estimated growth rate.

C deferred tax liabilities.

Solution to 1:

B is correct. AFFO is calculated from FFO by deducting non-cash rent, capital expenditures for maintenance, and leasing costs.

A is incorrect because it does not account for non-cash rent, capital expenditures for maintenance, and leasing costs. C is incorrect because it includes non-cash depreciation and amortization expense and does not account for non-cash rent, capital expenditures, and capitalized leasing costs, which are appropriate adjustments to net income in calculating current economic return.

Solution to 2:

B is correct. NAVPS estimates real estate values by capitalizing NOI. Valuing \$115 million of NOI with a capitalization rate of 7% yields a value for the properties of \$1,642,857,000. After deducting \$505 million of debt at market value, NAV is \$1,137,857,000; NAVPS equals NAV divided by 100 million shares outstanding, or \$11.38.

A is incorrect because it is the book value of the assets (not the net assets) per share: \$1,005 million divided by 100 million shares = \$10.05 per share. It does not take into account the market value of the assets and does not deduct debt. C is incorrect because it is the market value of the real estate—that is, NOI capitalized at 7%, divided by 100 million shares: \$1,642,857,000/100,000,000 = \$16.42. This calculation excludes the liabilities of the entity.

Solution to 3:

A is correct. The capitalization rate is used to calculate the estimated value of operating real estate because it is the NOI as a percentage of the value of operating real estate: NOI/Capitalization rate = Estimated value. As the capitalization rate increases, the estimated value of operating real estate and thus NAV will decrease.

B is incorrect because an increase in the estimated growth rate would increase the estimated NOI and the estimated value of operating income. C is incorrect because deferred liabilities are not counted as "hard" liabilities and are not subtracted from the NAV.

EXAMPLE 31

Analyst Adjustments (II)

1 An increase in the capitalization rate will *most likely* decrease a REIT's:

A cost of debt.

B estimated NOI.

C estimated NAV.

2 An analyst gathers the following information for a REIT:

Non-cash (straight-line) rent	€207,430
Depreciation	€611,900
Recurring maintenance-type capital expenditures and leasing commissions	€550,750
Adjusted funds from operations	€3,320,000
AFFO per share	€3.32

The REIT's FFO per share is *closest* to:

A €3.93.

B €4.08.

C €4.48.

3 Which of the following estimates is *least likely* to be compiled by firms that publish REIT analysts' estimates?

A FFO

B AFFO

C NAV

Solution to 1:

C is correct. The capitalization rate is used to estimate the market value of real estate, which is then used to calculate NAV.

A is incorrect because a higher capitalization rate does not decrease the REIT's cost of debt. B is incorrect because the estimated NOI is based on income growth, not the capitalization rate.

Solution to 2:

B is correct. FFO = AFFO + Non-cash (straight-line) rent + Recurring maintenance-type capital expenditures and leasing commissions = 3,320,000 + 550,750 + 207,430 = €4,078,180. The number of shares outstanding = 3,320,000/3.32 = 1,000,000. FFO per share = 4,078,180/1,000,000 ≈ €4.08.

A is incorrect because it adds depreciation to AFFO (3,320,000 + 611,900 = €3,931,900; 3,931,900/1,000,000 ≈ €3.93 per share). C is incorrect because it also adds depreciation to AFFO + Non-cash (straight-line) rent + Recurring maintenance-type capital expenditures and leasing commissions.

Solution to 3:

B is correct. Firms that compile statistics and estimates of REITs tend not to gather AFFO estimates because of the absence of a universally accepted methodology for computing AFFO and inconsistent corporate reporting of actual AFFO figures. FFO is commonly tracked in the United States, and NAV is the standard measure in Europe and Asia.

11.3 P/FFO and P/AFFO Multiples: Advantages and Drawbacks

The key benefits of using P/FFO and P/AFFO multiples in the valuation of REITs and REOCs are as follows:

1 Multiples of earnings measures of this kind are widely accepted in evaluating shares across global stock markets and industries.

2 In light of this acceptance, portfolio managers can put the valuation of REITs and REOCs into context with other investment alternatives.

3 FFO estimates are readily available through market data providers, such as Bloomberg and Refinitiv, which facilitates calculating P/FFO multiples.

4 Multiples can be used in conjunction with such items as expected growth and leverage levels to deepen the relative analysis among REITs and REOCs. Because FFO and AFFO do not take into account differences in leverage, leverage ratios can be used to adjust for leverage differences among REITs when using these multiples to compare valuations.

There are also drawbacks. Multiples are not a perfect basis for valuation because of the following:

1 Applying a multiple to FFO or AFFO may not capture the intrinsic value of all real estate assets held by the REIT or REOC, such as non-income-producing assets (for example, land held for development, vacant buildings, and properties under development), underused assets (current use may not represent highest and best use), or assets with below-market rents.

2 P/FFO does not adjust for the impact of recurring capital expenditures needed to keep properties operating smoothly. Although P/AFFO should do so, wide variations in estimates and assumptions are incorporated into the calculation of AFFO.

3 An increased level of such one-time items as gains and accounting charges, as well as new revenue recognition rules, has affected the income statement, thus making P/FFO and P/AFFO more difficult to compute and complicating comparisons between companies.

12 REIT MINI CASE STUDY: EXAMPLE OF DISCLOSURES AND VALUATION ANALYSIS

- calculate and interpret the value of a REIT share using the net asset value, relative value (price-to-FFO and price-to-AFFO), and discounted cash flow approaches

In this section, we undertake the valuation of a REIT by using the previously outlined approaches for valuation. The REIT in our example is Capitol Shopping Center REIT Inc. (CRE), a fictitious company that owns and operates retail shopping centers primarily in the Washington, DC, metropolitan area. The following are CRE's income statements, balance sheets, and cash flow statements for 2019 and 2020.

Exhibit 13 Capitol Shopping Center REIT Inc. Financial Statements (USD thousands, except per-share data)

A. Income statements

	Three Months Ending 31 December		Year Ending 31 December	
	2020	2019	2020	2019
Rental revenue	133,700	130,300	517,546	501,600
Other property income	3,600	2,100	14,850	13,450
Total property revenue	137,300	132,400	532,396	515,050
Rental expenses	29,813	28,725	112,571	109,775
Property taxes	15,050	14,850	57,418	55,375
Total property expenses	44,863	43,575	169,989	165,150
Property net operating income	92,437	88,825	362,407	349,900
Other income	450	385	1,840	1,675
General and administrative expenses	6,150	7,280	23,860	26,415
EBITDA	86,737	81,930	340,387	325,160
Depreciation and amortization	28,460	27,316	115,110	111,020
Net interest expense	25,867	25,015	100,823	99,173
Net income available to common	32,410	29,599	124,454	114,967
Weighted average common shares	61,100	60,100	60,600	60,100
Earnings per share	0.53	0.49	2.05	1.91

B. Balance sheets

	Year Ending 31 December	
	2020	2019
Assets		
Real estate, at cost		
Operating real estate	3,627,576	3,496,370
Land held for future development	133,785	133,785
	3,761,361	3,630,155
Less accumulated depreciation	(938,097)	(822,987)
Net real estate	2,823,264	2,807,168
Cash and equivalents	85,736	23,856

Exhibit 13 (Continued)

B. Balance sheets

	Year Ending 31 December	
	2020	**2019**
Accounts receivable, net	72,191	73,699
Deferred rent receivable, net	38,165	33,053
Prepaid expenses and other assets	106,913	101,604
Total assets	3,126,269	3,039,380
Liabilities and shareholders' equity		
Liabilities		
Mortgages payable	701,884	647,253
Notes payable	1,090,745	1,090,745
Accounts payable and other liabilities	219,498	200,439
Total liabilities	2,012,127	1,938,437
Common shares and equity	1,114,142	1,100,943
Total liabilities and shareholders' equity	3,126,269	3,039,380

C. Cash flow statements

	Year Ending 31 December	
	2020	**2019**
Operating activities		
Net income	124,454	114,967
Depreciation and amortization	115,110	111,020
Change in accounts receivable	1,508	452
Change in deferred rents	(5,112)	(4,981)
Change in prepaid expenses and other assets	(5,309)	1,237
Change in accounts payable and other liabilities	19,059	(11,584)
Net cash provided by operating activities	249,710	211,111
Investing activities		
Acquisition of real estate	(111,200)	(22,846)
Capital expenditures on operating real estate	(20,006)	(18,965)
Net cash used in investing activities	(131,206)	(41,811)
Financing activities		
Issuance of mortgages	54,631	14,213
Issuance of common shares	58,425	0
Dividends paid to common shareholders	(169,680)	(165,275)
Net cash used in financing activities	(56,624)	(151,062)
Increase (decrease) in cash and equivalents	61,880	18,238
Cash and cash equivalents, beginning of year	23,856	5,618
Cash and cash equivalents, end of year	85,736	23,856

CRE also publishes a supplemental investor packet that provides further disclosures used by the investment community to analyze the company. Exhibit 14 shows its adjustments to arrive at FFO and AFFO, as well as its calculation of dividend payouts based on dividends paid.

Exhibit 14 Capitol Shopping Center REIT Inc. FFO, AFFO, and Dividend Payouts (USD thousands, except per-share data)

	Three Months Ending 31 December		Year Ending 31 December	
	2020	2019	2020	2019
Funds from operations				
Net income	32,410	29,599	124,454	114,967
Depreciation and amortization	28,460	27,316	115,110	111,020
Funds from operations	60,870	56,915	239,564	225,987
FFO/share	1.00	0.95	3.95	3.76
Adjusted funds from operations				
Funds from operations	60,870	56,915	239,564	225,987
Less non-cash rents (1)	(1,469)	(1,325)	(5,112)	(4,981)
Less recurring capital expenditures (2)	(5,638)	(5,101)	(20,006)	(18,965)
Adjusted funds from operations	53,763	50,489	214,446	202,041
AFFO/share	0.88	0.84	3.54	3.36
Dividends/share	0.70	0.69	2.80	2.75
Dividend payout ratios				
On FFO	70.0%	72.6%	70.9%	73.1%
On AFFO	79.6%	82.1%	79.1%	81.8%
Weighted average common shares	61,100	60,100	60,600	60,100

1 Non-cash rents include the impact of straight lining contractual rent increases in leases, per accounting rules. The change in deferred rents can often provide the impact of this accounting on rental revenues.

2 Recurring capital expenditures include those costs needed to maintain the revenue-producing ability of existing assets, such as leasing commissions to keep or attract new tenants, such maintenance items as roofs and parking lot repairs, and basic buildouts of space as an inducement to attract tenants.

The historical stock price and the company's financial statements, including disclosures, are used to complete a simple analysis of the balance sheet, as shown in Exhibit 15.

Exhibit 15 Capitol Shopping Center REIT Inc. Balance Sheet Analysis (USD thousands, except per-share data)

	Year Ending 31 December	
	2020	2019
Ending debt	1,792,629	1,737,998
Ending stock price	72.36	61.50
Ending shares	61,100	60,100
Ending market capitalization	4,421,196	3,696,150

Exhibit 15 (Continued)

	Year Ending 31 December	
	2020	2019
Debt/total market capitalization	*40.5%*	*47.0%*
Peer group debt/total market capitalization	47.1%	56.7%
All REITs debt/total market capitalization	42.8%	49.6%
EBITDA	340,387	325,160
Interest expense	100,823	99,173
Interest coverage	*3.38x*	*3.28x*
Peer group interest coverage	2.35×	2.16×
All REITs interest coverage	2.58×	2.27×
Ending net debt	1,706,893	1,714,142
EBITDA	340,387	325,160
Net debt-to-EBITDA	*5.01×*	*5.27×*
Peer group net debt-to-EBITDA	7.10×	8.60×
All REITs net debt-to-EBITDA	6.70×	7.80×
Ending net debt	1,706,893	1,714,142
Ending gross real estate	3,761,361	3,630,155
Net debt/gross real estate (book)	*45.4%*	*47.2%*
Peer group net debt/gross real estate (book)	52.8%	55.1%
All REITs net debt/gross real estate (book)	49.6%	52.6%

The exhibits provide a historical picture of CRE's financial performance and balance sheet. Some key points about the company's properties, operations, dividend policy, recent business activity, and historical trading attributes follow.

- CRE owns properties that are generally considered defensive in the commercial real estate sector because many of its properties are tenanted by basic necessity goods retailers such as grocery stores and drug stores.
- CRE's location in the Washington, DC, metropolitan area is generally viewed as favorable for two key reasons: (1) Washington, DC, is the capital of the United States, and the government is the largest driver of employment and has historically provided more stability compared with the private sector; and (2) the city is a fairly dense area with strict zoning restrictions that make new construction of shopping centers difficult, which limits competing new supply.
- CRE has been able to increase its rents and net operating income by 2%–3% each year, on average, in the past decade.
- The past two reported years (2019 and 2020) were difficult for the broader commercial real estate markets. CRE was able to achieve positive growth while many of its peers saw FFO and AFFO decline. Because forecasts now call for improving fundamental property-level conditions, CRE's portfolio may not have as much "upside" because it did not experience the decline in occupancy and rents that other REITs did.
- In the middle of 2020, the company purchased a portfolio of three shopping centers from a local developer for a total price of $111.2 million. The return on these assets in the first year is an estimated 6.75%. The company was able to achieve a better going-in cap rate on this acquisition than the market averages

of 6.0%–6.25% because of its strong relationships and reputation with tenants, commercial property brokers, and competitors, as well as its ability to act quickly because of its strong balance sheet. In addition, the property is not fully leased, leaving the potential to increase net operating income if CRE can attract additional tenants. CRE funded the purchase with a $54.6 million mortgage at a 6% interest rate and cash from a common stock offering of 1 million shares and from cash on hand.

- The company intends to make additional acquisitions in the future as part of its growth plan. It intends to use a combination of debt, common equity, and internally generated cash to make these purchases. It typically requires the properties it acquires to generate an unleveraged internal rate of return of 9.5% in the form of current yield and capital appreciation over time.
- CRE's balance sheet strategy is to operate at less than 50% debt/market capitalization, with a preference for leverage to be closer to 40%. At year-end 2018, CRE's debt/market capitalization was 40.5% and its interest coverage was 3.38×. The company's current in-place average debt cost is 5.7%. In comparison, CRE's peers operate at an average leverage level of 47.1% and have an interest coverage ratio of 2.35×.
- CRE's board has chosen a dividend policy that provides an approximate 80% payout of cash flow, or AFFO. This level allows the company to pay an attractive dividend to shareholders, retain some cash flow, provide a cushion in the event of a downturn, and remain in compliance with REIT payout requirements in the United States. It is easily able to meet these REIT payout requirements because the requirements are based on taxable net income, which is calculated after deducting depreciation. In fact, CRE's dividend level has run well in excess of taxable net income, according to comments made by its management.
- Over the last decade, CRE has traded between 9× and 19× FFO, while its peers have traded between 8× and 18×, and all REITs have traded between 7× and 20×. On an AFFO basis, CRE's historical multiple has been 10×–21×, with its peers trading between 9×–19× and all REITs being in the 9×–24× range.
- Currently, shopping center REITs are estimated to be trading at 7.6% above analyst estimates of NAV. The overall REIT sector is estimated to be trading at a 14.8% premium to estimated NAV.
- CRE's historical beta to the broader equity market is 0.80. The current risk-free rate of return is 4.0%, and the market risk premium is estimated at 5.0%.

Investors and analysts who cover CRE have published estimates for its FFO per share, AFFO per share, and dividends per share for the next three years. Putting the average, or "consensus," of these estimates together with the company's reported results reveals the FFO/AFFO and dividend snapshot shown in Exhibit 16.

Exhibit 16 Capitol Shopping Center REIT Inc. Historical and Forecast Earnings and Dividends (all amounts are per share)

	Year Ending 31 December				
	2019A	**2020A**	**2021E**	**2022E**	**2023E**
CRE's FFO/share	$3.76	$3.95	$4.23	$4.59	$4.80
Growth	—	5.1%	7.1%	8.5%	4.6%
Peer group FFO/share growth	—	3.4%	6.8%	8.2%	4.2%
All REITs FFO/share growth	—	1.2%	7.9%	9.8%	10.2%
CRE's AFFO/share	$3.36	$3.54	$3.76	$4.09	$4.31

Exhibit 16 (Continued)

	Year Ending 31 December				
	2019A	2020A	2021E	2022E	2023E
Growth	—	5.4%	6.2%	8.8%	5.4%
Peer group AFFO/share growth	—	−1.0%	6.2%	9.1%	4.8%
All REITs AFFO/share growth	—	−3.0%	8.1%	9.7%	10.8%
CRE's dividends/share	$2.75	$2.80	$2.98	$3.25	$3.40
Growth	—	1.8%	6.4%	9.1%	4.6%
Peer group dividends/share growth	—	−2.0%	5.6%	7.9%	5.1%
All REITs dividends/share growth	—	−5.0%	7.8%	8.9%	6.0%
CRE's dividend payout on AFFO	81.8%	79.1%	79.3%	79.5%	78.9%

Taking the recent stock price of $69.85 per share and focusing on the next two years (as most analysts looking at multiples do), we can determine comparative FFO and AFFO multiples for CRE. Exhibit 17 also includes the multiples of its direct peers and the entire REIT industry.

Exhibit 17 Comparative Multiple Analysis

	P/FFO		P/AFFO	
	2021E	2022E	2021E	2022E
Capitol Shopping Center REIT Inc. (CRE)[a]	16.5×	15.2×	18.6×	17.1×
Shopping center–oriented REITs	14.5×	13.3×	16.1×	14.5×
All REITs	14.2×	12.8×	16.5×	14.6×
CRE's premium/(discount) to:				
Shopping center REITs	13.8%	14.3%	15.5%	17.9%
All REITs	16.2%	18.8%	12.7%	17.1%

[a]Based on a current stock price of $69.85.

12.1 Selection of Valuation Methods

As this discussion demonstrates, different valuation methods can yield different results. Under such circumstances, an analyst should re-examine the assumptions made to investigate why the approaches are generating such different results. The methods selected by an analyst may depend on which ones the analyst believes use the most reliable assumptions, which ones the analyst believes will be used by other investors, or which ones best reflect the analyst's own investment philosophy or view of value. The analyst may choose to use a single valuation approach, a midpoint in the range of values obtained by using several approaches, or a weighted average of the values obtained based on the analyst's view of the relative reliability of the models used to arrive at the values.

EXAMPLE 32

Valuation (I)

1 If the outlook for economic growth turns negative and property market transaction volumes decline, it is *least likely* that CRE's:

A P/FFO and P/AFFO would be lower.

B relative P/FFO and P/AFFO multiples would be higher than those of peers.

C NAV becomes the most useful valuation method.

2 If other REITs have no land on their balance sheets, how is CRE's "Land held for future development" *best* factored into a relative P/FFO or P/AFFO multiple valuation?

A There should be no impact on multiples as a result of land value.

B CRE would warrant lower multiples to account for land value.

C CRE would warrant higher multiples to account for land value.

3 An analyst speaks with private market real estate investors and learns that because interest rates have just increased 200 bps, buyers will require future property acquisitions to have going-in cap rates that are 100 bps to 200 bps higher than those on recent property market transactions. The analyst's estimate of NAV for CRE *most likely*:

A increases as cap rates are higher.

B decreases as cap rates are higher.

C remains the same unless CRE has debt maturing in the near term.

4 An analyst determines that CRE purchased its "Land held for future development" 15 years ago and that on average, land values at that time were one-third of what they are today. Which of the following *best* adjusts NAV to reflect this consideration?

A The cap rate on operating assets should be changed.

B Land value and thus NAV should be adjusted higher to reflect today's valuations.

C NAV is still mainly a representation of book values; thus, there should be no adjustments.

5 Zoning in CRE's real estate markets is changed to allow more new space in the future, dampening CRE's long-term FFO growth by about 0.5%. The effect on CRE's valuation using a dividend discount model is *most likely* that the present value of the dividend stream:

A decreases because of lower growth.

B remains the same.

C increases because of the new supply.

Solution to 1:

C is correct. NAV becomes more subjective in a negative and less liquid market with fewer observable transactions, and thus this basis of valuation becomes less useful and reliable.

A and B are incorrect because P/FFO and P/AFFO are likely to fall in a negative economic environment, but investors may be willing to pay a relative premium for CRE's stock based on its superior stability in economically challenging times. Thus, P/FFO and P/AFFO are likely to be higher than those of peers.

Solution to 2:

C is correct. Although it may not produce income that contributes to FFO or AFFO, the land has value and represents a source of greater internal growth potential. For that reason, A and B are incorrect.

Solution to 3:

B is correct. Estimated real estate value decreases as the cap rate increases. Because NAV is derived directly from estimated real estate value, it also decreases. For this reason, A is incorrect. C is incorrect because an increase in cap rates decreases asset values. The fact that CRE has debt maturing in the near term is not a key factor influencing NAV.

Solution to 4:

B is correct. An analyst tries to attribute market values to real property owned.

A is incorrect because the cap rate used by analysts in calculating NAVs represents the return on only the income-producing asset portfolio and does not relate to land holdings that are not currently producing any income. C is incorrect because NAV is not a representation of book values, which rely on accounting methodology rather than market values.

Solution to 5:

A is correct. Lower growth affects the projected dividend stream, decreasing its present value. For that reason, B and C are incorrect.

EXAMPLE 33

Valuation (II)

1 An analyst gathers the following information for two REITs:

	Price/NAV	Capitalization Rate Used in NAV
REIT A	100%	6%
REIT B	99%	8%

If the REITs have similar property portfolio values, interest expense, and corporate overhead, which REIT *most likely* has the higher price/FFO?

A REIT A

B REIT B

C They will have similar P/FFO because their ratios of price to NAV are almost identical.

Solution to 1:

A is correct. If both companies have similar portfolio values as indicated in the text and by the similar P/NAV, then the company with the lower capitalization rate is more expensive, which results in lower FFO and hence a higher P/FFO. If each company were worth ¥100, then REIT A, which is valued at a 6% cap rate, would have ¥6 of NOI and REIT B would have ¥8 of NOI. Because interest expense and overhead are similar for both companies, REIT A would also have lower FFO and a correspondingly higher P/FFO multiple.

B is incorrect because A has a lower capitalization rate, implying a lower FFO and hence a higher P/FFO if P/NAV for each company is similar, which is the case here.

C is incorrect because it neglects the effect of the lower capitalization rate of REIT A.

13 PRIVATE VS. PUBLIC: A COMPARISON

- explain advantages and disadvantages of investing in real estate through publicly traded securities compared to private vehicles

Large institutional and high-net-worth investors have historically pursued private real estate investments through direct ownership, joint ventures, and private fund investments, whereas individual investors, without the resources to invest directly, typically invested in listed property companies. As more real estate companies went public and continued to issue equity to fund acquisitions, developments, and mergers, the market cap of the publicly listed real estate sector rose significantly. This larger market float and liquidity permitted institutional investors to add to their real estate exposure by creating allocations to public real estate companies

Should investors with the ability to pursue both public *and* private real estate investments choose one over the other? The answer depends on investor objectives, including total return requirements, volatility (risk) tolerance, diversification goals, and the expected returns from each investment. Many institutional investors such as pension funds and endowments have chosen to allocate to both.

Both public and private real estate equity investments provide exposure to real estate properties, potentially hedge inflation, deliver attractive risk-adjusted returns, and provide some diversification benefits to stock and bond portfolios.

Listed real estate can play a complementary role in private real estate. Listed real estate's liquidity makes it easier to express a short-term view, such as when markets become too negative on retail and drive shares of public companies below net asset value. When there are sustained valuation differences between public and private real estate, fund and company managers can capture opportunities. If public companies trade well below net asset value, the public companies may choose to go private or sell to private real estate funds. When real estate values are high, public companies can sell real estate to realize gains and private funds may seek exits through the IPO market.

Private real estate investors have the ability to pursue a variety of strategies, such as merchant (for sale) development, which is highly restricted for REITs. In some countries, REITs were early movers in specialty sectors, such as self-storage and data centers. Investors wanting exposure to some of these niches had to seek out listed company exposure until the private funds moved into these sectors, often in the search for higher yield.

Private and public real estate investments both have something to offer investors, and each has its drawbacks. Exhibit 18 summarizes some of the key differences, advantages, and disadvantages of public and private real estate investing.

Exhibit 18 Advantages and Disadvantages of Private and Public Real Estate

	Private Real Estate (Direct Investment)	Public Real Estate (Equity REITs and Real Estate Operating Companies)
Advantages	■ Direct exposure to real estate fundamentals ■ Stable returns/low volatility ■ Income and capital appreciation ■ Property performance drives returns ■ Low correlations with other asset classes ■ Potential inflation hedge ■ Control (direct real estate and separate accounts) ■ Potential to earn illiquidity premium ■ Wide variety of strategies/few restrictions ■ Tax benefits (e.g., accelerated depreciation, deferred taxes in some markets when sales are reinvested in other real estate)	■ Tracks real estate fundamentals over the long term ■ Liquidity ■ Access to professional management ■ Potential inflation hedge ■ Potential for strong alignment of interests ■ Tax-efficient structure avoids double taxation (REITs only) ■ Potential for exposure to diversified portfolios ■ Access to diverse sectors, including data centers, medical offices, and self-storage ■ Low investment requirements ■ Low entry/exit costs ■ No special investor qualifications beyond equity investing generally ■ Limited liability ■ Greater regulation and investor protections ■ High transparency
Disadvantages	■ Low liquidity ■ Difficult-to-exit funds' redemption activity is high ■ High fees and expenses ■ Appraisal valuations commonly lag changes in market conditions ■ Fewer regulations to protect investors ■ Some managers focus on asset gathering over high profitability ■ High investment minimums and high net-worth requirements ■ Low transparency ■ High returns often derived from leverage	■ High volatility (compared with private real estate) ■ Equity market correlation is high in short term ■ REIT structure limits possible activities ■ Stock prices may not reflect underlying property values (i.e., trade at discount to NAV) ■ Dividends taxed at high current income tax rates ■ Regulatory compliance costs are prohibitive for small companies ■ Poor governance/mis-aligned interests can penalize stock performance ■ Equity markets often penalize companies with high leverage

SUMMARY

Real estate property is an asset class that plays a significant role in many investment portfolios and is an attractive source of current income. Investor allocations to public and private real estate have increased significantly over the last 20 years. Because of the unique characteristics of real estate property, real estate investments tend to behave differently from other asset classes—such as stocks, bonds, and commodities—and thus have different risks and diversification benefits. Private real estate investments are further differentiated because the investments are not publicly traded and require analytic techniques different from those of publicly traded assets. Because of the lack of transactions, the appraisal process is required to value real estate property. Many of the indexes and benchmarks used for private real estate also rely on appraisals, and because of this characteristic, they behave differently from indexes for publicly traded equities, such as the S&P 500, MSCI Europe, FTSE Asia Pacific, and many other regional and global indexes.

General Characteristics of Real Estate

- Real estate investments make up a significant portion of the portfolios of many investors.
- Real estate investments can occur in four basic forms: private equity (direct ownership), publicly traded equity (indirect ownership claim), private debt (direct mortgage lending), and publicly traded debt (securitized mortgages).
- Each of the basic forms of real estate investment has its own risks, expected returns, regulations, legal structures, and market structures.
- Equity investors generally expect a higher rate of return than lenders (debt investors) because they take on more risk. The returns to equity real estate investors have two components: an income stream and capital appreciation.
- Many motivations exist for investing in real estate income property. The key ones are current income, price appreciation, inflation hedge, diversification, and tax benefits.
- Adding equity real estate investments to a traditional portfolio will potentially have diversification benefits because of the less-than-perfect correlation of equity real estate returns with returns to stocks and bonds.
- If the income stream can be adjusted for inflation and real estate prices increase with inflation, then equity real estate investments may provide an inflation hedge.
- Debt investors in real estate expect to receive their return from promised cash flows and typically do not participate in any appreciation in value of the underlying real estate. Thus, debt investments in real estate are similar to other fixed-income investments, such as bonds.
- Regardless of the form of real estate investment, the value of the underlying real estate property can affect the performance of the investment. Location is a critical factor in determining the value of a real estate property.
- Real estate property has some unique characteristics compared with other investment asset classes. These characteristics include heterogeneity and fixed location, high unit value, management intensiveness, high transaction costs, depreciation, sensitivity to the credit market, illiquidity, and difficulty of value and price determination.

- There are many different types of real estate properties in which to invest. The main commercial (income-producing) real estate property types are office, industrial and warehouse, retail, and multi-family. Other types of commercial properties are typically classified by their specific use.
- Certain risk factors are common to commercial property, but each property type is likely to have a different susceptibility to these factors. The key risk factors that can affect commercial real estate include business conditions, lead time for new development, excess supply, cost and availability of capital, unexpected inflation, demographics, lack of liquidity, environmental issues, availability of information, management expertise, and leverage.
- Location, lease structures, and economic factors—such as economic growth, population growth, employment growth, and consumer spending—affect the value of each property type.
- An understanding of the lease structure is important when analyzing a real estate investment.
- Appraisals estimate the value of real estate income property. Definitions of value include market value, investment value, value in use, and mortgage lending value.
- Due diligence investigates factors that might affect the value of a property prior to making or closing on an investment. These factors include leases and lease history, operating expenses, environmental issues, structural integrity, lien/proof of ownership, property tax history, and compliance with relevant laws and regulations.
- Appraisal-based and transaction-based indexes are used to track the performance of private real estate. Appraisal-based indexes tend to lag transaction-based indexes and appear to have lower volatility and lower correlation with other asset classes than transaction-based indexes.

Private Equity Real Estate

- Generally, three different approaches are used by appraisers to estimate value: income, cost, and sales comparison.
- The income approach includes direct capitalization and discounted cash flow methods. Both methods focus on net operating income as an input to the value of a property and indirectly or directly factor in expected growth.
- The cost approach estimates the value of a property based on adjusted replacement cost. This approach is typically used for unusual properties for which market comparables are difficult to obtain.
- The sales comparison approach estimates the value of a property based on what price comparable properties are selling for in the current market.
- When debt financing is used to purchase a property, additional ratios and returns calculated and interpreted by debt and equity investors include the loan-to-value ratio, the debt service coverage ratio, and leveraged and unleveraged internal rates of return.

Publicly Traded Real Estate Securities

- The principal types of publicly traded real estate securities available globally are real estate investment trusts, real estate operating companies, and residential and commercial mortgage-backed securities.
- Publicly traded equity real estate securities offer investors participation in the returns from investment real estate with the advantages of superior liquidity; greater potential for diversification by property, geography, and property type; access to a larger range of properties; the benefit of management services; limited liability; protection accorded by corporate governance, disclosure, and other securities regulations; and in the case of REITs, exemption from corporate income taxation within the REIT if prescribed requirements are met.
- Disadvantages include the costs of maintaining a publicly traded corporate structure and complying with regulatory filings, pricing determined by the stock market and returns that can be volatile, the potential for structural conflicts of interest, and tax differences compared with direct ownership of property that can be disadvantageous under some circumstances.
- Compared with other publicly traded shares, REITs offer higher-than-average yields and greater stability of income and returns. They are amenable to a net asset value approach to valuation because of the existence of active private markets for their real estate assets. Compared with REOCs, REITs offer higher yields and income tax exemptions but have less operating flexibility to invest in a broad range of real estate activities and less potential for growth from reinvesting their operating cash flows because of their high income-to-payout ratios.
- In assessing the investment merits of REITs, investors analyze the effects of trends in general economic activity, retail sales, job creation, population growth, and new supply and demand for specific types of space. They also pay particular attention to occupancies, leasing activity, rental rates, remaining lease terms, in-place rents compared with market rents, costs to maintain space and re-lease space, tenants' financial health and tenant concentration in the portfolio, financial leverage, debt maturities and costs, and the quality of management and governance.
- Analysts make adjustments to the historical cost-based financial statements of REITs and REOCs to obtain better measures of current income and net worth. The three principal figures they calculate and use are (1) funds from operations or accounting net earnings, excluding depreciation, deferred tax charges, and gains or losses on sales of property and debt restructuring; (2) adjusted funds from operations, or funds from operations adjusted to remove straight-line rent and to provide for maintenance-type capital expenditures and leasing costs, including leasing agents' commissions and tenants' improvement allowances; and (3) net asset value or the difference between a real estate company's asset and liability ranking prior to shareholders' equity, all valued at market values instead of accounting book values.
- REITs and some REOCs generally return a significant portion of their income to their investors and, as a result, tend to pay high dividends. Thus, dividend discount or discounted cash flow models for valuation are also applicable. These valuation approaches are applied in the same manner as they are for shares in other industries. Usually, investors use two- or three-step dividend discount models with near-term, intermediate-term, and/or long-term growth

assumptions. In discounted cash flow models, investors often use intermediate-term cash flow projections and a terminal value based on historical cash flow multiples.

REFERENCES

Huerta, Daniel, Thanh Ngo, and Mark Pyles. 2019. "*The Role of Institutional Ownership in REIT Acquisitions.*" Working paper (August): 3. www.fmaconferences.org/NewOrleans/Papers/TheRoleofInstitutionalOwnershipinREITAcquisitions.pdf.

Nareit. 2018. "Nareit Funds from Operations White Paper—2018 Restatement." White paper (December). www.reit.com/sites/default/files/2018-FFO-white-paper-(11-27-18).pdf.

PRACTICE PROBLEMS

The following information relates to Questions 1–12

Amanda Rodriguez is an alternative investment analyst for a US investment management firm, Delphinus Brothers. Delphinus's chief investment officer, Michael Tang, has informed Rodriguez that he wants to reduce the amount invested in traditional asset classes and gain exposure to the real estate sector by acquiring commercial property in the United States. Rodriguez is asked to analyze potential commercial real estate investments for Delphinus Brothers. Selected data on three commercial real estate properties are presented in Exhibit 1.

Exhibit 1 Selected Property Data

	Property 1	Property 2	Property 3
Property Type	**Downtown Office Building**	**Grocery-Anchored Retail Center**	**Multi-Family Building**
Location	New York City	Miami	Boston
Occupancy	90.00%	93.00%	95.00%
Square Feet or Number of Units	100,000 sf	205,000 sf	300 units
Gross Potential Rent	$4,750,000	$1,800,000	$3,100,000
Expense Reimbursement Revenue	$333,333	$426,248	$0
Other Income (includes % rent)	$560,000	$15,000	$45,000
Potential Gross Income	$5,643,333	$2,241,248	$3,145,000
Vacancy Loss	($564,333)	($156,887)	($157,250)
Effective Gross Income	$5,079,000	$2,084,361	$2,987,750
Property Management Fees	($203,160)	($83,374)	($119,510)
Other Operating Expenses	($2,100,000)	($342,874)	($1,175,000)
Net Operating Income	$2,775,840	$1,658,113	$1,693,240

Rodriguez reviews the three properties with Tang, who indicates that he would like her to focus on Property 1 because of his prediction of robust job growth in New York City over the next 10 years. To complete her analysis, Rodriquez assembles additional data on Property 1, which is presented in Exhibits 2, 3, and 4.

As part of the review, Tang asks Rodriguez to evaluate financing alternatives to determine whether it would be better to use debt financing or to make an all-cash purchase. Tang directs Rodriguez to inquire about terms with Richmond Life Insurance Company, a publicly traded company that is an active lender on commercial real estate property. Rodriquez obtains the following information from Richmond Life for a loan on Property 1: loan term of five years, interest rate of 5.75% interest only, maximum loan to value of 75%, and minimum debt service coverage ratio of 1.5×. Data on Property 1 are provided in Exhibit 2, Exhibit 3, and Exhibit 4.

© 2021 CFA Institute. All rights reserved.

Exhibit 2 Six-Year Net Operating Income (NOI) and DCF Assumptions for Property 1

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
NOI	$2,775,840	$2,859,119	$2,944,889	$3,033,235	$3,124,232	$3,217,959
DCF Assumptions						
Investment Hold Period			5 years			
Going-In Cap Rate			5.25%			
Terminal Cap Rate			6.00%			
Discount Rate			7.25%			
Income/Value Growth Rate			Constant			

Exhibit 3 Sales Comparison Data for Property 1

Variable	Property 1	Sales Comp A	Sales Comp B	Sales Comp C
Age (years)	10	5	12	25
Condition	Good	Excellent	Good	Average
Location	Prime	Secondary	Secondary	Prime
Sale Price psf		$415 psf	$395 psf	$400 psf
Adjustments				
Age (years)		–10%	2%	10%
Condition		–10%	0%	10%
Location		15%	15%	0%
Total Adjustments		**–5%**	**17%**	**20%**

Exhibit 4 Other Selected Data for Property 1

Land Value	$7,000,000
Replacement Cost	$59,000,000
Total Depreciation	$5,000,000

After reviewing her research materials, Rodriguez formulates the following two conclusions:

Conclusion 1 Benefits of private equity real estate investments include the owners' ability to attain diversification benefits, to earn current income, and to achieve tax benefits.

Conclusion 2 Risk factors of private equity real estate investments include business conditions, demographics, the cost of debt and equity capital, and financial leverage.

1 Which of the following is *most likely* accurate regarding Property 2, described in Exhibit 1?

A Operating expense risk is borne by the owner.

B The lease term for the largest tenant is longer than three years.

C A significant amount of percentage rent is linked to sales levels.

2 Based on Exhibits 2, 3, and 4, which of the following statements is *most* accurate regarding the valuation of Property 1?

A The cost approach valuation is $71 million.

B The adjusted price psf for Sales Comp B is $423 psf.

C The terminal value at the end of Year 5 in the income approach is $53,632,650.

3 Based on Exhibit 2, the growth rate of Property 1 is *closest* to:

A 0.75%.

B 1.25%.

C 2.00%.

4 Based on Exhibit 2, the value of Property 1 using the discounted cash flow method is *closest* to:

A $48,650,100.

B $49,750,900.

C $55,150,300.

5 Based on Exhibit 2, relative to the estimated value of Property 1 under the discounted cash flow method, the estimated value of Property 1 using the direct capitalization method is:

A the same.

B lower.

C higher.

6 Based on Exhibits 1 and 3, the estimated value of Property 1 using the sales comparison approach (assigning equal weight to each comparable) is *closest* to:

A 40,050,000.

B 40,300,000.

C 44,500,000.

7 In the event that Delphinus purchases Property 2, the due diligence process would *most likely* require a review of:

A all tenant leases.

B tenant sales data.

C the grocery anchor lease.

8 Compared with an all-cash purchase, a mortgage on Property 1 through Richmond Life would *most likely* result in Delphinus earning:

A a lower return on equity.

B a higher return on equity.

C the same return on equity.

9 Assuming an appraised value of $48 million, Richmond Life Insurance Company's maximum loan amount on Property 1 would be *closest* to:

A $32 million.

B $36 million.

C $45 million.

10 Rodriguez's Conclusion 1 is:

A correct.

B incorrect, because tax benefits do not apply to tax-exempt entities.

C incorrect, because private real estate is highly correlated with stocks.

11 Rodriguez's Conclusion 2 is:

A correct.

B incorrect, because inflation is not a risk factor.

C incorrect, because the cost of equity capital is not a risk factor.

12 Richmond Life Insurance Company's potential investment would *most likely* be described as:

A private real estate debt.

B private real estate equity.

C publicly traded real estate debt.

The following information relates to Questions 13–28

First Life Insurance Company, Ltd., a life insurance company located in the United Kingdom, maintains a stock and bond portfolio and also invests in all four quadrants of the real estate market: private equity, public equity, private debt, and public debt. Each of the four real estate quadrants has a manager assigned to it. First Life intends to increase its allocation to real estate. The chief investment officer (CIO) has scheduled a meeting with the four real estate managers to discuss the allocation to real estate and to each real estate quadrant. Leslie Green, who manages the private equity quadrant, believes her quadrant offers the greatest potential and has identified three investment properties to consider for acquisition. Selected information for the three properties is presented in Exhibit 1.

Exhibit 1 Selected Information on Potential Private Equity Real Estate Investments

	Property		
	A	B	C
Property description	**Single-Tenant Office**	**Shopping Center**	**Warehouse**
Size (square meters)	3,000	5,000	9,000
Lease type	Net	Gross	Net
Expected loan-to-value ratio	70%	75%	80%
Total economic life	50 years	30 years	50 years
Remaining economic life	30 years	23 years	20 years
Rental income (at full occupancy)	£575,000	£610,000	£590,000
Other income	£27,000	£183,000	£29,500
Vacancy and collection loss	£0	£61,000	£59,000
Property management fee	£21,500	£35,000	£22,000

(continued)

Exhibit 1 (Continued)

	Property		
	A	B	C
Property description	**Single-Tenant Office**	**Shopping Center**	**Warehouse**
Other operating expenses	£0	£234,000	£0
Discount rate	11.5%	9.25%	11.25%
Growth rate	2.0%	See Assumption 2	3.0%
Terminal cap rate		11.00%	
Market value of land	£1,500,000	£1,750,000	£4,000,000
Replacement costs			
■ Building costs	£8,725,000	£4,500,000	£12,500,000
■ Developer's profit	£410,000	£210,000	£585,000
Deterioration—curable and incurable	£4,104,000	£1,329,000	£8,021,000
Obsolescence			
■ Functional	£250,000	£50,000	£750,000
■ Locational	£500,000	£200,000	£1,000,000
■ Economic	£500,000	£100,000	£1,000,000
Comparable adjusted price per square meter			
■ Comparable Property 1	£1,750	£950	£730
■ Comparable Property 2	£1,825	£1,090	£680
■ Comparable Property 3	£1,675	£875	£725

To prepare for the upcoming meeting, Green has asked her research analyst, Ian Cook, for a valuation of each of these properties under the income, cost, and sales comparison approaches using the information provided in Exhibit 1 and the following two assumptions:

Assumption 1 The holding period for each property is expected to be five years.

Assumption 2 Property B is expected to have the same net operating income for the holding period because of existing leases and a one-time 20% increase in Year 6 because of lease rollovers. No further growth is assumed thereafter.

In reviewing Exhibit 1, Green notes the disproportionate estimated obsolescence charges for Property C relative to the other properties and asks Cook to verify the reasonableness of these estimates. Green also reminds Cook that they will need to conduct proper due diligence. In that regard, Green indicates that she is concerned whether a covered parking lot that was added to Property A encroaches (is partially located) on adjoining properties. Green would like for Cook to identify an expert and present documentation to address her concerns regarding the parking lot.

In addition to discussing the new allocation, the CIO informs Green that she wants to discuss the appropriate real estate index for the private equity real estate quadrant at the upcoming meeting. The CIO believes that the current index may result in over-allocating resources to the private equity real estate quadrant.

13 The *most* effective justification that Green could present for directing the increased allocation to her quadrant would be that relative to the other quadrants, her quadrant of real estate investments:

- **A** provides greater liquidity.
- **B** requires less professional management.
- **C** enables greater decision-making control.

14 Relative to the expected correlation between First Life's portfolio of public REIT holdings and its stock and bond portfolio, the expected correlation between First Life's private equity real estate portfolio and its stock and bond portfolio is *most likely* to be:

- **A** lower.
- **B** higher.
- **C** the same.

15 Which of the properties in Exhibit 1 exposes the owner to the greatest risk related to operating expenses?

- **A** Property A
- **B** Property B
- **C** Property C

16 Which property in Exhibit 1 is *most likely* to be affected by import and export activity?

- **A** Property A
- **B** Property B
- **C** Property C

17 Which property in Exhibit 1 would *most likely* require the greatest amount of active management?

- **A** Property A
- **B** Property B
- **C** Property C

18 Which property in Exhibit 1 is *most likely* to have a percentage lease?

- **A** Property A
- **B** Property B
- **C** Property C

19 The disproportionate charges for Property C noted by Green are *least likely* to explicitly factor into the estimate of property value using the:

- **A** cost approach.
- **B** income approach.
- **C** sales comparison approach.

20 Based on Exhibit 1, which of the following statements regarding Property A is *most* accurate?

- **A** The going-in capitalization rate is 13.5%.
- **B** It appears to be the riskiest of the three properties.
- **C** The net operating income in the first year is £298,000.

21 Based on Exhibit 1, the value of Property C using the direct capitalization method is *closest* to:

A £3,778,900.

B £4,786,700.

C £6,527,300.

22 Based on Exhibit 1 and Assumptions 1 and 2, the value of Property B using the discounted cash flow method, assuming a five-year holding period, is *closest* to:

A £4,708,700.

B £5,034,600.

C £5,050,900.

23 Which method under the income approach is *least likely* to provide a realistic valuation for Property B?

A Layer method

B Direct capitalization method

C Discounted cash flow method

24 Based on Exhibit 1, the value of Property A using the cost approach is *closest* to:

A £5,281,000.

B £6,531,000.

C £9,385,000.

25 Based on Exhibit 1, the value of Property B using the sales comparison approach is *closest* to:

A £4,781,000.

B £4,858,000.

C £6,110,000.

26 Which due diligence item would be *most* useful in addressing Green's concerns regarding Property A?

A Property survey

B Engineering inspection

C Environmental inspection

27 The real estate index currently being used by First Life to evaluate private equity real estate investments is *most likely:*

A an appraisal-based index.

B a transaction-based index.

C the NCREIF Property Index.

28 Based on Exhibit 1, the property expected to be most highly leveraged is:

A Property A.

B Property B.

C Property C.

The following information relates to Questions 29–34

Hui Lin, CFA, is an investment manager looking to diversify his portfolio by adding equity real estate investments. Lin and his investment analyst, Maria Nowak, are discussing whether they should invest in publicly traded real estate investment trusts or public real estate operating companies. Nowak expresses a strong preference for investing in public REITs in taxable accounts.

Lin schedules a meeting to discuss this matter, and for the meeting, Lin asks Nowak to gather data on three specific REITs and come prepared to explain her preference for public REITs over public REOCs. At the meeting, Lin asks Nowak,

> "Why do you prefer to invest in public REITs over public REOCs for taxable accounts?"

Nowak provides Lin with an explanation for her preference of public REITs and provides Lin with data on the three REITs shown in Exhibits 1 and 2.

The meeting concludes with Lin directing Nowak to identify the key investment characteristics along with the principal risks of each REIT and to investigate the valuation of the three REITs. Specifically, Lin asks Nowak to value each REIT using four different methodologies:

Method 1 Net asset value

Method 2 Discounted cash flow valuation using a two-step dividend model

Method 3 Relative valuation using property subsector average P/FFO multiple

Method 4 Relative valuation using property subsector average P/AFFO multiple

Exhibit 1 Select REIT Financial Information

	REIT A	REIT B	REIT C
Property subsector	**Office**	**Storage**	**Health Care**
Estimated 12-month cash net operating income	$350,000	$267,000	$425,000
Funds from operations	$316,965	$290,612	$368,007
Cash and equivalents	$308,700	$230,850	$341,000
Accounts receivable	$205,800	$282,150	$279,000
Debt and other liabilities	$2,014,000	$2,013,500	$2,010,000
Non-cash rents	$25,991	$24,702	$29,808
Recurring maintenance-type capital expenditures	$63,769	$60,852	$80,961
Shares outstanding	56,100	67,900	72,300

Exhibit 2 REIT Dividend Forecasts and Average Price Multiples

	REIT A	REIT B	REIT C
Expected annual dividend next year	$3.80	$2.25	$4.00
Dividend growth rate in Years 2 and 3	4.0%	5.0%	4.5%

(continued)

Exhibit 2 (Continued)

	REIT A	REIT B	REIT C
Dividend growth rate (after Year 3 into perpetuity)	3.5%	4.5%	4.0%
Assumed cap rate	7.0%	6.25%	6.5%
Property subsector average P/FFO multiple	14.4×	13.5×	15.1×
Property subsector average P/AFFO multiple	18.3×	17.1×	18.9×

Note: Nowak estimates an 8% cost of equity capital for all REITs and a risk-free rate of 4.0%.

29 Nowak's *most likely* response to Lin's question is that the type of real estate security she prefers:

- **A** offers a high degree of operating flexibility.
- **B** provides dividend income that is exempt from double taxation.
- **C** has below-average correlations with overall stock market returns.

30 Based on Exhibits 1 and 2, the value per share for REIT A using valuation Method 1 is *closest* to:

- **A** $51.26.
- **B** $62.40.
- **C** $98.30.

31 Based on Exhibits 1 and 2, the value per share of REIT B using valuation Method 3 is *closest* to:

- **A** $40.77.
- **B** $57.78.
- **C** $73.19.

32 Based on Exhibit 2, the value per share of REIT C using valuation Method 2 is *closest* to:

- **A** $55.83.
- **B** $97.57.
- **C** $100.91.

33 Based on Exhibits 1 and 2, the value per share of REIT A using valuation Method 4 is *closest* to:

- **A** $58.32.
- **B** $74.12.
- **C** $103.40.

34 The risk factor *most likely* to adversely affect an investment in REIT B is:

- **A** new competitive facilities.
- **B** tenants' sales per square foot.
- **C** obsolescence of existing space.

The following information relates to Questions 35–40

Tim Wang is a financial adviser specializing in commercial real estate investing. He is meeting with Mark Caudill, a new client who is looking to diversify his investment portfolio by adding real estate investments. Caudill has heard about various investment vehicles related to real estate from his friends and is seeking a more in-depth understanding of these investments from Wang.

Wang begins the meeting by advising Caudill of the many options that are available when investing in real estate, including the following:

Option 1	Direct ownership in real estate
Option 2	Publicly traded real estate investment trusts
Option 3	Publicly traded real estate operating companies
Option 4	Publicly traded residential mortgage-backed securities

Wang next asks Caudill about his investment preferences. Caudill responds by telling Wang that he prefers to invest in equity securities that are highly liquid, provide high income, and are not subject to double taxation.

Caudill asks Wang how the economic performance of REITs and REOCs is evaluated and how their shares are valued. Wang advises Caudill there are multiple measures of economic performance for REITs and REOCs, including the following:

Measure 1	Net operating income
Measure 2	Funds from operations
Measure 3	Adjusted funds from operations

In response, Caudill asks Wang,

> "Which of the three measures is the best measure of a REIT's current economic return to shareholders?"

To help Caudill's understanding of valuation, Wang presents Caudill with data on Baldwin, a health care REIT that primarily invests in independent and assisted senior housing communities in large cities across the United States. Selected financial data on Baldwin for the past two years are provided in Exhibit 1.

Exhibit 1 Baldwin REIT Summarized Income Statement (USD thousands, except per-share data)

	Year Ending 31 December	
	2019	2018
Rental income	339,009	296,777
Other property income	6,112	4,033
Total income	345,121	300,810
Rental expenses		
Property operating expenses	19,195	14,273
Property taxes	3,610	3,327
Total property expenses	22,805	17,600
Net operating income	322,316	283,210
Other income (gains on sale of properties)	2,162	1,003
General and administrative expenses	21,865	19,899

(continued)

Exhibit 1 (Continued)

	Year Ending 31 December	
	2019	**2018**
Depreciation and amortization	90,409	78,583
Net interest expenses	70,017	56,404
Net income	142,187	129,327
Weighted average shares outstanding	121,944	121,863
Earnings per share	1.17	1.06
Dividend per share	0.93	0.85
Price/FFO, based on year-end stock price	11.5×	12.7×

Before the meeting, Wang had put together some valuation assumptions for Baldwin in anticipation of discussing valuation with Caudill. Wang explains the process of valuing a REIT share using discounted cash flow analysis, and he proceeds to estimate the value of Baldwin on a per-share basis using a two-step dividend discount model using the data provided in Exhibit 2.

Exhibit 2 Baldwin Valuation Projections and Assumptions

Current risk-free rate	4.0%
Baldwin beta	0.90
Market risk premium	5.0%
Appropriate discount rate (CAPM)	8.5%
Expected dividend per share, 1 year from today	$1.00
Expected dividend per share, 2 years from today	$1.06
Long-term growth rate in dividends, starting in Year 3	5.0%

35 Based on Caudill's investment preferences, the type of real estate investment Wang is *most likely* to recommend to Caudill is:

- **A** Option 2.
- **B** Option 3.
- **C** Option 4.

36 Relative to Option 2 and Option 3, an advantage of investing in Option 1 is:

- **A** greater liquidity.
- **B** lower investment requirements.
- **C** greater control over property-level investment decisions.

37 The Baldwin REIT is *least likely* to experience long-run negative effects from:

- **A** an economic recession.
- **B** an unfavorable change in population demographics.
- **C** a major reduction in government funding of health care.

38 The *most appropriate* response to Caudill's question is:

- **A** Measure 1.

- **B** Measure 2.
- **C** Measure 3.

39 Based on Exhibit 1, the 2019 year-end share price of Baldwin was *closest* to:

- **A** \$13.23.
- **B** \$21.73.
- **C** \$30.36.

40 Based on Exhibit 2, the intrinsic value of the Baldwin REIT on a per share basis using the two-step dividend discount model is *closest* to:

- **A** \$26.72.
- **B** \$27.59.
- **C** \$28.76.

SOLUTIONS

1 B is correct. The lease term for the anchor tenant is typically longer than the usual three- to five-year term for smaller tenants. The data in Exhibit 1 suggest that the operating expenses are passed on to the tenant; the sum of property management fees and other operating expenses equals the expense reimbursement revenue. Also, other income is only $15,000, suggesting that a minimal amount of percentage rent is linked to sales thresholds.

2 C is correct. The terminal value using the income approach is $53,632,650 (= Year 6 NOI/terminal cap rate = $3,217,959/0.06). The value of the property using the cost approach is $61 million (= Land value + Building replacement cost – Total depreciation = $7,000,000 + $59,000,000 – $5,000,000). The adjusted sales price per square foot for Sales Comp B is $462 psf (= $395 × 1.17).

3 C is correct. There is a constant growth rate in income and value: Growth rate = Discount rate (7.25%) – Going-in cap rate (5.25%) = 2.00%.

4 B is correct. The value of Property 1 using the discounted cash flow method is $49,750,931, or $49,750,900 rounded, calculated as follows:

		Discount Period	Discounted Value[a]
Year 1 NOI	$2,775,840	1	$2,588,196
Year 2 NOI	$2,859,119	2	$2,485,637
Year 3 NOI	$2,944,889	3	$2,387,135
Year 4 NOI	$3,033,235	4	$2,292,540
Year 5 NOI	$3,124,232	5	$2,201,693
Terminal Value[b]	$53,632,650	5	$37,795,731
Property 1 DCF Value			$49,750,932

[a]Discount rate = 7.25%.
[b]Terminal value = Year 6 NOI/Terminal cap rate = $3,217,959/0.06 = $53,632,650.

5 C is correct. The direct capitalization method estimate of value for Property 1 is $52,873,143 (= Year 1 NOI/Going-in cap rate = $2,775,840/0.0525), which is greater than the estimated DCF value of $49,750,932.

The value of Property 1 using the discounted cash flow method can be calculated from the following table:

		Discount Period	Discounted Value[a]
Year 1 NOI	$2,775,840	1	$2,588,196
Year 2 NOI	$2,859,119	2	$2,485,637
Year 3 NOI	$2,944,889	3	$2,387,135
Year 4 NOI	$3,033,235	4	$2,292,540
Year 5 NOI	$3,124,232	5	$2,201,693
Terminal Value[b]	$53,632,650	5	$37,795,731
Property 1 DCF Value			$49,750,932

[a]Discount rate = 7.25%.
[b]Terminal value = Year 6 NOI/Terminal cap rate = $3,217,959/0.06 = $53,632,650.

6 C is correct. The estimate of the value of Property 1 using the sales comparison approach can be calculated using the following table:

	Unadjusted psf	Adjusted psf
Sales Comp 1	\$415	\$394 (= \$415 × 0.95)
Sales Comp 2	\$395	\$462 (= \$395 × 1.17)
Sales Comp 3	\$400	\$480 (= \$400 × 1.20)
Average	\$403	\$445

Estimated value of Property 1 = \$44,500,000 (= \$445 psf × 100,000 sf).

7 C is correct. The due diligence process includes a review of leases for major tenants, which would include the grocery anchor tenant. Typically, only major tenant leases will be reviewed in the due diligence process; smaller tenant leases will likely not be reviewed. Also, the fact that other income is only \$15,000 suggests that percentage rent linked to sales levels is minimal and has not been underwritten in the valuation and acquisition process.

8 B is correct. Delphinus will expect to earn a higher return on equity with the use of a mortgage to finance a portion of the purchase. The quoted mortgage interest rate of 5.75% is less than the discount rate of 7.25%.

9 A is correct. The maximum amount of debt that an investor can obtain on commercial real estate is usually limited by either the ratio of the loan to the appraised value of the property (loan-to-value, or LTV, ratio) or the debt service coverage ratio (DSCR), depending on which measure results in the lowest loan amount. The maximum LTV ratio is 75% of the appraised value of \$48 million, or \$36,000,000. The loan amount based on the minimum DSCR would be \$32,183,652, determined as follows:

Maximum debt service = Year 1 NOI/DSCR = \$2,775,840/1.5 = \$1,850,560.

Loan amount (interest-only loan) = Maximum debt service/Mortgage rate = \$1,850,560/0.0575 = \$32,183,652 (rounded to \$32,000,000).

10 A is correct. Benefits of private equity real estate investments include the owners' ability to attain diversification benefits, to earn current income, and to achieve tax benefits.

11 A is correct. Business conditions, demographics, the cost of debt and equity capital, and financial leverage are characteristic sources of risk for real estate investments.

12 A is correct. Richmond Life's investment would be a mortgage that falls under private debt in the four quadrants.

13 C is correct. Private equity investments in real estate enable greater decision-making control relative to real estate investments in the other three quadrants. A private real estate equity investor or direct owner of real estate has responsibility for the management of the real estate, including maintaining the properties, negotiating leases, and collecting rents. These responsibilities increase the investor's control in the decision-making process. Investors in publicly traded REITs or real estate debt instruments would not typically have significant influence over these decisions.

14 A is correct. Evidence suggests that private equity real estate investments have a lower correlation with stocks and bonds than publicly traded REITs. When real estate is publicly traded, it tends to behave more like the rest of the stock market than the real estate market.

15 B is correct. Property B is a gross lease, which requires the owner to pay the operating expenses. Accordingly, the owner, First Life, incurs the risk of Property B's operating expenses, such as utilities, increasing in the future.

16 C is correct. Property C is a warehouse and is most likely affected by import and export activity in the economy. Property A (office) and Property B (retail) would typically be less dependent on import and export activity compared with a warehouse property.

17 B is correct. Property B is a shopping center and would most likely require more active management than a single-tenant office (Property A) or a warehouse (Property C); the owner would need to maintain the right tenant mix and promote the facility.

18 B is correct. Property B is a shopping center, a type of retail property. A percentage lease is a unique aspect of many retail leases, which requires the tenant to pay additional rent once its sales reach a certain level. The lease will typically specify a "minimum rent" that must be paid regardless of the tenant's sales. Percentage rent may be paid by the tenant once the tenant's sales reach a certain level or breakpoint.

19 B is correct. Obsolescence charges reduce the value of a property using the cost approach and are factored into the sales comparison approach by adjustments, including condition and location, to the price per square meter. The cash flows to the property should reflect obsolescence: Less rent is received if the property is not of an appropriate design for the intended use, if it is in a poor location, or if economic conditions are poor. Therefore, obsolescence is implicitly, not explicitly, factored into the estimate of property value using the income approach.

20 B is correct. Property A has been assigned the highest discount rate (11.5%) and thus is considered to be the riskiest investment of the three alternatives. This may be the result of the reliance on a single tenant. The going-in capitalization rate is 9.5% (Cap rate = Discount rate – Growth rate). The net operating income is £580,500 (= Rental income + Other income – Property management fee = £575,000 + £27,000 – £21,500).

21 C is correct. Under the direct capitalization method, the value of the property = NOI/(r – g).

Calculate net operating income (NOI):

NOI = Rental income + Other income – Vacancy and collection loss – Property management costs.
NOI = £590,000 + £29,500 – £59,000 – £22,000 = £538,500.

Then, value the property using the cap rate:

Value of property = £538,500/(11.25% – 3.0%) = £6,527,273, rounded to £6,527,300.

22 B is correct. The value of Property B using the discounted cash flow method is £5,034,600.

The value using the discounted cash flow method is based on the present value of the net operating income and the estimated property resale price.

Calculate NOI (constant during the five-year holding period from Assumption 2):

NOI = Rental income (at full occupancy) + Other income – Vacancy and collection loss – Property management fee – Other operating expenses.

NOI = £610,000 + £183,000 – £61,000 – £35,000 – £234,000 = £463,000.

Estimate property value at the end of five years:

NOI starting in Year 6 is 20% higher because of lease rollovers (from Assumption 2).

NOI starting in Year 6 = £463,000 × 1.20 = £555,600.

Terminal cap rate (given) = 11%.

Applying the terminal cap rate yields a property value of £5,050,909 (= £555,600/0.11).

Find the present value of the expected annual NOI and the estimated property resale value using the given discount rate of 9.25%:

N = 5.
FV = £5,050,909.
PMT = £463,000.
I = 9.25.

Solving for PV, the current value of the property is estimated to be £5,034,643, or £5,034,600 rounded.

23 B is correct. The net operating income for Property B is expected to be level for the next five years, because of existing leases, and grow 20% in Year 6. A direct capitalization method would not be appropriate because of the multiple growth rates. A discounted cash flow method that assigns a terminal value, or a layer method, should be used.

24 A is correct. The value of Property A using the cost method is equal to the replacement cost, adjusted for the different types of depreciation (loss in value):

Value of Property A = Land value + (Replacement building cost + Developer's profit) – Deterioration – Functional obsolescence – Locational obsolescence – Economic obsolescence
= £1,500,000 + (£8,725,000 + £410,000) – £4,104,000 – £250,000 – £500,000 – £500,000
= £5,281,000.

25 B is correct. The value of a property using the sales comparison approach equals the adjusted price per square meter using comparable properties times property size. The value of Property B using the sales comparison approach is calculated as follows:

Average adjusted price per square meter of Comparable Properties 1, 2, and 3 for Property B = (£950 + £1,090 + £875)/3 = £971.67.

Applying the £971.67 average adjusted price per square meter to Property B gives a value of £4,858,300 (= £971.67 × 5,000 square meters = £4,858,350, or £4,858,000 rounded).

26 A is correct. A property survey can determine whether the physical improvements, such as the covered parking lot, are in the boundary lines of the site and whether any easements would affect the value of the property.

27 A is correct. An appraisal-based index is most likely to result in the over-allocation mentioned by the CIO due to the appraisal lag. The appraisal lag tends to "smooth" the index, meaning that it has less volatility. It behaves somewhat like a moving average of what an index would look like if it were based on values obtained from transactions rather than appraisals. Thus, appraisal-based indexes may underestimate the volatility of real estate returns. Because of the lag in the index, appraisal-based real estate indexes will also tend to have a lower correlation with other asset classes. This situation is problematic if the index is used in asset allocation models; the amount allocated to the asset class that appears to have lower correlation with other asset classes and less volatility will be greater than it should be.

28 C is correct. Property C has an expected loan-to-value ratio of 80%, which is higher than the loan-to-value ratio for Property A (70%) or Property B (75%).

29 B is correct. REITs are tax-advantaged entities, whereas REOC securities are not typically tax-advantaged entities. More specifically, REITs are typically exempt from the double taxation of income that comes from taxes being due at the corporate level and again when dividends or distributions are made to shareholders in some jurisdictions, such as the United States.

30 B is correct. The NAV is $62.40.

Estimated cash NOI	350,000
Assumed cap rate	0.07
Estimated value of operating real estate (350,000/0.07)	5,000,000
Plus: Cash + accounts receivable	514,500
Less: Debt and other liabilities	2,014,000
Net asset value	3,500,500
Shares outstanding	56,100
NAV/share	**$62.40**

31 B is correct. The value per share is $57.78, calculated as follows:

Funds from operations = $290,612.

Shares outstanding = 67,900 shares.

FFO/share = $290,612/67,900 shares = $4.28.

Applying the property subsector average P/FFO multiple of 13.5× yields a value per share of

$4.28 × 13.5 = $57.78.

32 C is correct. The value per share for REIT C is $100.91.

	Step 1			Step 2
	Year 1	**Year 2**	**Year 3**	**Year 4**
Dividends per share:	$4.00	$4.18	$4.37	$4.54
Value of stock at end of 2013:[a]			$113.57	
			$117.94	

	Step 1			Step 2
	Year 1	Year 2	Year 3	Year 4
Discount rate: 8.00%				
Net present value of all dividends:[b] $100.91				

[a]Calculated as $4.54/(0.08 – 0.04) = $113.57.
[b]Calculated as $4.00/(1.08) + $4.18/$(1.08)^2$ + $117.94/$(1.08)^3$ = $100.91.

33 B is correct. The value per share is $74.11, calculated as follows:

Funds from operations (FFO) = $316,965.

Less: Non-cash rents: $25,991

Less: Recurring maintenance-type capital expenditures: $63,769

Equals: AFFO: $227,205

Shares outstanding = 56,100 shares.

AFFO/share = $227,205/56,100 shares = $4.05.

Applying the property subsector average P/AFFO multiple of 18.3× yields a value per share of

$4.05 × 18.3 = $74.12.

34 A is correct. As a storage REIT, this investment faces competitive pressures because the ease of entry into the field of self-storage properties can lead to periods of overbuilding.

35 A is correct. Option 2 (publicly traded REITs) best satisfies Caudill's investment preferences. REITs are equity investments that, in general, are income tax exempt at the corporate/trust level, so there is no double income taxation. To qualify for the income tax exemption, REITs are legally obligated to pay out a high percentage of income to their shareholders, which typically results in relatively high income for investors. Lastly, public REITs are generally liquid because they are traded in stock exchanges.

36 C is correct. Direct property ownership offers greater control over property-level investment decisions compared with the level of control exhibited by shareholders in REITs and REOCs.

37 A is correct. Baldwin, a health care REIT, is largely resistant to economic recessions but is exposed to changes in population demographics and changes in government funding for health care.

38 C is correct. Measure 3, adjusted funds from operations, is a refinement of FFO that is designed to be a more accurate measure of current economic income. In essence, FFO is adjusted to remove any non-cash rent and to include a provision for maintenance-type capital expenditures and leasing costs. Maintenance expenses are required for a business to continue as a going concern.

39 B is correct. Baldwin's FFO per share in 2019 was $1.89, and the resulting share price was $21.73. First, calculate FFO per share in 2019, and then apply the year-end P/FFO multiple of 11.5×.

FFO = accounting net earnings, excluding (a) depreciation charges on real estate, (b) deferred tax charges, and (c) gains or losses from sales of property and debt restructuring.

2019 accounting net income: \$142,187

2019 depreciation charges: \$90,409

2019 deferred tax charges: na

2019 gains on sale of properties (other income): \$2,162

2019 shares outstanding: 121,944

2019 year-end price/FFO = 11.5×.

2019 Baldwin FFO per share = (\$142,187 + \$90,409 – \$2,162)/121,944 shares = \$1.89. At the given 2019 year-end price/FFO multiple of 11.5×, this results in a share price for Baldwin of \$1.89 × 11.5 = \$21.73.

40 C is correct. The estimated value per share for the Baldwin REIT using a two-step dividend discount model is \$28.76, calculated as follows:

	Step 1		Step 2
	Year 1	**Year 2**	**Year 3**
Dividends per share:	\$1.00	\$1.06	\$1.11
Value of stock at end of Year 2:[a]		\$31.71	
		\$32.77	
Discount rate: 8.50%			
Net present value of all dividends:[b] \$28.83			

[a]Calculated as \$1.11/(0.085 – 0.05) = \$31.71.
[b]Calculated as \$1.00/(1.085) + \$32.77/$(1.085)^2$ = \$28.76.

READING

36

Private Equity Investments

by Yves Courtois, CMT, MRICS, CFA, and Tim Jenkinson, PhD

Yves Courtois, CMT, MRICS, CFA, is at KPMG (Luxembourg). Tim Jenkinson, PhD, is at Saïd Business School, Oxford University (United Kingdom).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** explain sources of value creation in private equity;
☐	**b.** explain how private equity firms align their interests with those of the managers of portfolio companies;
☐	**c.** compare and contrast characteristics of buyout and venture capital investments;
☐	**d.** interpret LBO model and VC method output;
☐	**e.** explain alternative exit routes in private equity and their impact on value;
☐	**f.** explain risks and costs of investing in private equity;
☐	**g.** explain private equity fund structures, terms, due diligence, and valuation in the context of an analysis of private equity fund returns;
☐	**h.** interpret and compare financial performance of private equity funds from the perspective of an investor;
☐	**i.** calculate management fees, carried interest, net asset value, distributed to paid in (DPI), residual value to paid in (RVPI), and total value to paid in (TVPI) of a private equity fund.

INTRODUCTION

1

Private equity's shift from a niche activity to a critical component of the financial system is evident from investors' financial commitment: around $2.8 trillion globally as of mid-2018. And that's just the equity portion. The use of debt means transaction value is often two or three times the actual equity raised. Blackstone, Carlyle, and KKR are household names and publicly traded companies of significant size. Private

© 2020 CFA Institute. All rights reserved.

equity funds may account for 15%–18% of the value of all mergers and acquisitions, and the market capitalization of Alibaba, Amazon, Facebook, and Google has raised the profile of venture capital investing.

We take two approaches to illuminate our subject: In Sections 2–6 the perspective is primarily that of the private equity firm evaluating potential investments. Valuing acquisitions is particularly complex; except for public-to-private transactions, there will be no market prices to refer to, and the challenges are considerable. In Sections 7–9 we take the perspective of an outside investor investing in a fund sponsored by the private equity firm.

Definitions of private equity differ, but here we include the entire asset class of equity investments that are not quoted on stock markets. Private equity stretches from venture capital (VC)—working with early-stage companies that may be without revenues but that possess good ideas or technology—to growth equity, providing capital to expand established private businesses often by taking a minority interest, all the way to large buyouts (leveraged buyouts, or LBOs), in which the private equity firm buys the entire company. When the target is publicly traded, the private equity fund performs a public-to-private transaction, removing the target from the stock market. But buyout transactions usually involve private companies and very often a particular division of an existing company.

Some exclude venture capital from the private equity universe because of the higher risk profile of backing new companies as opposed to mature ones. For this reading, we refer simply to *venture capital* and *buyouts* as the two main forms of private equity.

Many classifications of private equity are available. Classifications proposed by the European and Private Equity Venture Capital Association (EVCA) are displayed in Exhibit 1.

Exhibit 1 Classification of Private Equity in Terms of Stage and Type of Financing of Portfolio Companies

Broad Category	Subcategory	Brief Description
Venture capital	Seed stage	Financing provided to research business ideas, develop prototype products, or conduct market research
	Start-up stage	Financing to recently created companies with well-articulated business and marketing plans
	Later (expansion) stage	Financing to companies that have started their selling effort and may already be covering costs: Financing may serve to expand production capacity, product development, or provide working capital.
	Replacement capital	Financing provided to purchase shares from other existing venture capital investors or to reduce financial leverage.
Growth	Expansion capital	Financing to established and mature companies in exchange for equity, often a minority stake, to expand into new markets and/or improve operations
Buyout	Acquisition capital	Financing in the form of debt, equity, or quasi-equity provided to a company to acquire another company
	Leveraged buyout	Financing provided by an LBO firm to acquire a company
	Management buyout	Financing provided to the management to acquire a company, specific product line, or division (carve-out)

Exhibit 1 (Continued)

Broad Category	Subcategory	Brief Description
Special situations	Mezzanine finance	Financing generally provided in the form of subordinated debt and an equity kicker (warrants, equity, etc.) frequently in the context of LBO transactions
	Distressed/turnaround	Financing of companies in need of restructuring or facing financial distress
	One-time opportunities	Financing in relation to changing industry trends and new government regulations
	Other	Other forms of private equity financing are also possible—for example, activist investing, funds of funds, and secondaries.

Private equity funds may also be classified geographically, by sector, or both. Certain specialists target real asset classes, such as real estate, infrastructure, energy, and timber, or they seek out emerging or niche sectors, such as agribusiness or royalties in pharmaceuticals, music, film, or TV.

US private equity enjoyed a far larger market size historically than private equity in other regions, with few restrictions on hostile takeovers. Buyouts subsequently expanded to Europe and then Asia as friendly deals became commonplace. In broad terms, around four-fifths of the money has been flowing into buyout, growth, and other types of private equity in both the United States and Europe, with buyout amounts far exceeding other types. The sheer scale of buyouts means that an individual deal can absorb billions of dollars in capital. Buyout funds have benefited from increased allocations given their ability to absorb far higher capital amounts and to deliver historically higher-than-average returns.

Venture capital deals, in contrast, tend to drip, providing small amounts of feed money. Still, advances in technology and communications are causing the number of venture capital funds and the availability of start-up capital to grow. Investor attention started to shift to China in 2015, an especially active year for raising capital. VC funds targeting Asia had more than US$200 billion in 2017, up from US$50 billion in 2010.

Most private equity money comes from institutional investors, such as pension funds, sovereign wealth funds, endowments, and insurance companies, although many family offices and high-net-worth individuals also invest directly or through fund-of-funds intermediaries. Venture capital investors include government agencies and corporations seeking to promote regional investment or gain insight into, and possibly control of, emerging businesses and technologies.

Private equity investment is characterized by a buy-to-sell orientation: Investors typically expect their money to be returned, with a handsome profit, within 10 years of committing their funds. The economic incentives of the funds are aligned with this goal.

2 INTRODUCTION TO VALUATION TECHNIQUES IN PRIVATE EQUITY TRANSACTIONS

a explain sources of value creation in private equity;

b explain how private equity firms align their interests with those of the managers of portfolio companies;

This reading is not intended to be a comprehensive review of valuation techniques. Instead, we highlight some essential considerations specific to private equity. Private equity firms serve as a rich laboratory for applying the principles of asset and equity valuation.

First and foremost, we must distinguish between the price paid for a private equity stake and the valuation of that same private equity stake. The price paid for a private equity stake is the outcome of a negotiation process in which each party may assign a different value to the same stake. Whereas public company shares are traded on a regulated market and their prices are transparent, the buyers and sellers of private equity interests generally make greater efforts to uncover their value. Private equity valuation is thus time-bound and dependent on the respective motives of buyers and sellers.

Selecting a valuation methodology for a private equity (PE) portfolio company depends largely on its stage of development. Common methodologies appear in Exhibit 2, along with the stages in which they may apply.

Exhibit 2 Overview of Selected Valuation Methodologies and Their Possible Application in Private Equity

Valuation Technique	Brief Description	Application
Income approach: discounted cash flow (DCF)	Value is obtained by discounting expected future cash flows at an appropriate cost of capital.	Generally applies across the broad spectrum of company stages. Given the emphasis on expected cash flows, DCF provides the most relevant results when applied to companies with a sufficient operating history. It is most applicable to companies operating from the expansion phase up to the maturity phase.
Relative value: earnings multiples	Application of an earnings multiple to the earnings of a portfolio company. The earnings multiple is frequently obtained from the average of a group of public companies operating in a similar business and of comparable size. Commonly used multiples include price/earnings (P/E), enterprise value/EBITDA, enterprise value/sales.	Generally applies to companies with a significant operating history and a predictable stream of cash flows. May also apply (with caution) to companies operating at the expansion stage. Rarely applies to early-stage or start-up companies.
Real option	The right to undertake a business decision (call or put option). Requires judgmental assumptions about key option parameters.	Generally applies to situations in which the management or shareholders have significant flexibility in making radically different strategic decisions (i.e., option to undertake or abandon a high-risk, high-return project). Therefore, generally applies to some companies operating at the seed or start-up phase.
Replacement cost	Estimated cost to recreate the business as it stands as of the valuation date.	Generally applies to early-stage (seed and start-up) companies, companies operating at the development stage and generating negative cash flows, or asset-rich companies. Rarely applies to mature companies because it is difficult to estimate the cost to recreate a company with a long operating history. For example, it would be difficult to estimate the cost to recreate a long-established brand, such as Coca-Cola, whereas the replacement cost methodology may be used to estimate the brand value for a recently launched beverage (R&D expenses, marketing costs, etc.).

In a vibrant and booming private equity market, there is a natural tendency among participants to focus on the earnings approach to determine value. The benchmark value it offers is perceived as corresponding best to the market's present state, but given the lack of liquidity of private equity investments, the concurrent use of other metrics is strongly recommended.

In most transactions, private equity investors are faced with a set of investment decisions that are based on an assessment of prospective returns and associated probabilities. Private equity firms are confronted generally with a large flow of information arising from detailed due diligence investigations and from complex financial models. It is essential to understand the potential upside and downside impact of internal and external factors on the business, net income, and cash flows. The interplay between exogenous factors (such as favorable and unfavorable macroeconomic conditions, interest rates, and exchange rates) and value drivers for the business (such as sales margins and required investments) should also be considered carefully. For example, what will be the sales growth if competition increases or if competing new technologies are introduced?

When building financial forecasts, variables in the financial projections should be linked to key business drivers with assigned subjective probabilities. The use of Monte Carlo simulation can further enhance the analysis and identify significant financial upsides and downsides to the business. In a Monte Carlo simulation, the analyst must model the fundamental value drivers of the portfolio company, which are in turn linked to a valuation model. The objective is to ensure that the simulation is as close as possible to the realities of the business and encompasses the range of possible outcomes, including base case, worst case, and best case scenarios (sometimes called a triangular approach).

Other key considerations when evaluating a private equity transaction include the value of control, the impact of illiquidity, and the extent of any country risk. Estimating the discount for illiquidity and marketability and a premium for control are among the most subjective decisions in private equity valuation. The control premium is an incremental value associated with a bloc of shares that will be instrumental in gaining control of a company. In most buyouts, the entire equity capital is acquired by the private equity purchasers. But in venture capital deals, investors often acquire minority positions. In this case, the control premium (if any) largely depends on the relative strength and alignment of interest of shareholders willing to gain control. For example, in a situation with only a limited number of investors able to acquire control, the control premium is likely to be much more significant relative to a situation with a dominant controlling shareholder invested along with a large number of much smaller shareholders.

The distinction between marketability and liquidity is more subtle. The cost of illiquidity may be defined as the cost of finding prospective buyers and represents the speed of conversion of the assets to cash, whereas the cost of marketability is closely related to the right to sell the assets. In practice, the marketability and liquidity discounts are frequently lumped together.

The cost for illiquidity and premium for control may be closely related because illiquidity may be more acute when there is a fierce battle for control. But there are many dimensions to illiquidity. The size of the illiquidity discount may be influenced by such factors as the shareholding structure, the level of profitability and its expected sustainability, the possibility of an initial public offering (IPO) in the near future, and the size of the private company. Because determining the relative importance of each factor may be difficult, the illiquidity discount is frequently assessed overall on a judgmental basis. In practice, the discount for illiquidity and premium for control are both adjustments to the preliminary value estimate instead of being factored into the cost of capital.

When valuing private equity portfolio companies in emerging markets, country and currency risk represent additional sources of risk frequently added to a modified version of the standard CAPM. Estimating the appropriate country risk premium represents a significant challenge in emerging market private equity valuation, and numerous estimation approaches exist.

All of this is to say that PE valuation is highly challenging and valuation does not simply involve a net present value calculation based on a static set of future profit projections. Using a combination of valuation methodologies, supplemented by stress testing and scenario analysis, provides the strongest support to estimating value. One of the key ways private equity firms add value is by challenging the way businesses are run. Should the PE firm manage to improve the business's finances, operations, management, or marketing, we can expect additional value.

2.1 How Is Value Created in Private Equity?

How private equity funds actually create value has been the subject of much debate. Rather than ownership and control being separate, as in most publicly quoted companies, private equity concentrates ownership and control. Many view the combining of ownership and control as a fundamental source of the returns earned by the best private equity funds. The survival of the private equity governance model depends on economic advantages it may have over the public equity governance model, including (1) the ability to re-engineer the private firm to generate superior returns, (2) financial leverage and the ability to access credit markets on favorable terms, and (3) a better alignment of interests between private equity firm owners and the managers of the firms they control.

Do private equity houses have a superior ability to re-engineer companies and therefore generate superior returns? Some of the largest private equity organizations, such as Ardian, Blackstone Group, Carlyle Group, CVC Capital Partners, KKR, and Partners Group, have developed high-end consulting capabilities supported frequently by seasoned industry veterans, such as former CEOs, chief financial officers, and senior advisers. They have proven their ability to execute deals on a global basis. Irrespective of their size, some of the very best firms have developed effective re-engineering capabilities to add value to their investments. But it is hard to believe that this factor, all else being equal, is the main driver of added value. Assuming that private equity houses have a superior ability to re-engineer companies, this would mean that public companies have inherently less ability to do so. Many public companies, however, such as Apple, Berkshire Hathaway, Samsung, Tencent Holdings, Toyota, and Unilever, have long track records of creating value through organizational changes and re-engineering. Only a portion of the value added by private equity houses may be explained by superior capabilities in this sphere.

Is financial leverage the main driver of private equity returns in buyouts? In private equity, target companies are rarely purchased using only the equity of the buyout company. Relative to comparable publicly quoted companies, there is a much greater use of debt in a typical buyout transaction. The use of debt is central to the structure and feasibility of buyouts, and private equity firms use significant proportions of debt to finance each deal. The leverage increases equity returns and the number of transactions a particular fund can make. A private equity firm may invest equity representing 30% of the buyout purchase price and raise the rest in the debt markets. It may use a combination of bank loans—often called leveraged loans because of the prominent proportion of the company's capital structure they represent—and high-yield bonds.

Leveraged Loan Covenants

To protect investors, leveraged loans often carry covenants that may require or restrict certain actions. For instance, the covenants may require the company (1) to maintain specified financial ratios, (2) within certain limits, to submit information regularly so that the bank can monitor performance, or (3) to operate within certain parameters. The covenants may restrict the company from further borrowing (in other words, no additional bonds can be issued and no additional funds can be borrowed from banks or other sources), or they may impose limits on paying dividends or even making certain operating decisions.

Similarly, bond terms may include covenants intended to protect the bondholders. One of the key differences between leveraged loans and high-yield bonds, however, is that leveraged loans are generally senior secured debt whereas the bonds are unsecured in the case of bankruptcy. Even given covenants on the bonds, the bonds issued to finance an LBO are usually high-yield bonds that receive low-quality ratings and must offer high coupons to attract investors because of the amount of leverage used.

The ample availability of credit at favorable terms—think low credit spreads and fewer covenants—before the 2007 global financial crisis (GFC) and a resumption of covenant-lite terms combined with low interest rates during much of the 2010s contributed to a significant increase in available leverage for buyouts. Borrowing six to eight times EBITDA has been common for large buyout transactions. Note that in private equity, leverage is typically measured as a multiple of EBITDA instead of equity.

When considering the impact of leverage on value, we should naturally turn to one of the foundations of modern finance: the Modigliani–Miller (1958) theorem. This theorem, in its basic form, states that in the absence of taxes, asymmetric information, and bankruptcy costs and assuming efficient markets, the value of a firm is not affected by how the firm is financed. In other words, it should not matter if the firm is financed by equity or debt as far as firm value is concerned. The relaxing of the no-tax assumption raises interesting questions in leveraged buyouts, as the tax shield on the acquisition debt creates value because of the tax-deductibility of interest. One would also expect that the financial leverage of a firm would be set at a level where bankruptcy costs do not outweigh these tax benefits. Private equity firms may have a better ability than public companies to raise high levels of debt as a result of their better control over management but also as a result of their reputation for having raised and repaid such high levels of debt in previous transactions.

Such debt financing is raised initially from the syndicated loan market but then is frequently repackaged via sophisticated structured products, such as collateralized loan obligations (CLOs), which typically consist of a portfolio of secured floating-rate leveraged loans issued by non-investment-grade companies. In some cases, the private equity funds issue high-yield bonds as a way of financing the portfolio company, and these often are sold to funds that create collateralized debt obligations (CDOs).

This raises the question of whether a massive transfer of risk to the credit markets is taking place in private equity. If the answer to this question is yes, then one would expect that it will self-correct during the next economic downturn. During early 2008, the CDO and CLO markets were undergoing a significant slowdown as a result of the credit market turmoil that started in the summer of 2007, triggered by the subprime mortgage crisis. As a result, the LBO market for very large transactions ("mega-buyouts") was affected by a lack of financing. Global equity market declines beginning in mid- to late 2015, which accompanied or responded to the Brexit referendum, the

Asian currency weakness, and changing regulatory capital rules, further interrupted LBO and leveraged debt activity. The pause ended up being short-lived; market volumes resumed their upward growth in 2017.

Additional leverage is also gained by means of equity-like instruments at the acquisition vehicle level, which are frequently located in a favorable jurisdictions, such as Luxembourg, the Channel Islands, the Cayman Islands, the British Virgin Islands, or Malta. Acquisitions by large buyout private equity firms are generally held by a top holding company in a favorable tax jurisdiction. The top holding company's share capital and equity-like instruments are held in turn by investment funds run by a general partner who is controlled by the private equity buyout firm. These instruments are treated as debt for tax purposes in certain jurisdictions.

The effect of leverage may be analyzed through Jensen's (1986, 1989) free cash flow hypothesis. According to Jensen, low-growth companies generating high free cash flows tend to invest in projects that destroy value (i.e., with a negative net present value) instead of distributing excess cash to shareholders. This is a possible explanation for why an LBO transaction generates value, because excess cash is used to repay the senior debt tranche, effectively removing management's discretionary use of cash. Here, too, financial leverage may explain part of the value added by private equity investment.

Examining the value created via the re-engineering of private firms and the availability and effects of leverage is informative. Now we turn to the alignment of economic interests between private equity owners and the managers of the companies they control to ensure the latter's efforts to achieve the ambitious milestones set by the former. Results-driven pay packages and contractual clauses ensure that managers are incentivized to reach their targets and that they will not be left behind after the private equity house exits the investment. One common clause stipulates that any offer made by a future acquirer of the company be extended to all shareholders, including company management.

Consider the managers of public companies subsequently acquired by private equity groups. Empirical evidence shows that managers tend to acknowledge an increased level of directness and intensity of input after the takeover, which enables them to conduct higher-value-added projects. Crucially, these projects can be implemented over a longer time frame after the buyout; this situation is in contrast to the short-termism that prevailed during their public market period. This short-termism is mostly driven by shareholders' expectations, the analyst community, and market participants more broadly who place significant emphasis on meeting quarterly earnings targets. Private equity firms have a longer time horizon, so they attract talented managers with the ability to implement sometimes profound restructuring plans, isolated against short-term market consequences.

Private equity firms are not, however, the sole catalysts of change at large companies. Some large organizations, such as Google, SAP, and Tencent, have proven their ability to inspire entrepreneurship at all levels within their ranks while generating substantial value over the long term.

A balance of rights and obligations between the private equity firm and the management team requires effective structuring. The following matters are covered by the contractual clauses that private equity firms use to ensure that the management team is focused on achieving the business plan. If the agreed objectives are not met, the control and equity allocation held by the private equity firm may increase.

- *Corporate board seats*: A seat ensures some degree of private equity control in the case of major corporate events, such as a company sale, takeover, restructuring, IPO, bankruptcy, or liquidation.
- *Noncompete clause*: This is generally imposed on founders, preventing them from restarting the same activity during a predefined period of time.

- *Preferred dividends and liquidation preference*: Private equity firms generally come first when distributions take place, and they may be guaranteed a minimum multiple of their original investment before other shareholders receive their returns.
- *Reserved matters*: Some domains of strategic importance (such as changes in the business plan, acquisitions, or divestitures) are subject to approval or veto by the private equity firm.
- *Earnouts (mostly in venture capital)*: These are agreements that the acquisition price paid by the private equity firm is contingent on company management achieving predefined financial performance over a specified future time period (e.g., over one or two years). Earnouts are not specific to private equity.

How the PE firm structures the investment contract can have a major bearing on the returns. Venture capital firms, in particular, whose investee companies face considerable uncertainties, can set terms that increase their level of control over time or can even seize control if too many targets are missed.

2.2 Using Market Data in Valuation

With the exception of public-to-private transactions, there is no direct market evidence of company valuation with most private equity deals. But virtually all valuation techniques use evidence from the market at different stages in the calculation rather than relying entirely on accounting data and management forecasts.

The two most important ways in which market data are used to infer the value of the entity being acquired are (1) by analyzing publicly traded comparison companies and (2) by considering the valuations that are implied by recent transactions involving similar entities. Typically, these techniques involve trading or acquisition multiples. Suppose we need a valuation for a privately owned company in the food-retail sector. The comparison-company approach would look at the trading multiples—such as enterprise value to EBITDA—of comparable public food-retail companies and use this multiple to value the target. Similarly, the transaction multiples that were paid in recent food-retail M&A transactions can inform the market value of our target. It is very important, of course, to make sure that the comparisons are appropriate, and this is simply not always possible, especially for niche businesses or targets that are pioneering products and services.

Market data come into play for DCF approaches, in particular when estimating the discount rate. The same weighted average cost of capital (WACC) formula we use for public companies is used to establish the cost of capital for private companies. We face a serious challenge, however, in assessing the cost of equity in PE settings: the lack of public historical data on share prices and returns. Therefore, beta (β), which represents the relative exposure of company shares to the market, must be estimated by means of a proxy. Typically the proxy is the result of estimating the beta for comparable companies and then adjusting it for financial and operating leverage. This benchmark exercise calls for analyst judgment: To what extent are the comparable public firms genuinely comparable to the target firm? Should outlying companies be excluded? What is the target debt-to-equity ratio of the target firm versus the industry average? What comparable public companies are appropriate if the target firm operates in several business segments?

Given that forecasts of future financial performance are usually only available for a few years ahead, when it comes to DCF valuation it is almost always necessary to estimate the terminal value of the company beyond this forecasting horizon. It is possible to apply a perpetual growth rate assumption, although small changes in the assumed growth rate, which itself is very difficult to predict, can have a significant impact on the valuation. An alternative is to use a trading multiple that exists (or is predicted to exist) in public markets and apply this to the final-year forecasted values. For instance, if the average enterprise-value-to-EBITDA ratio for comparable publicly quoted companies is 10, then this might be applied to the private target's final forecast EBITDA value as a way of estimating the terminal value.

3 CONTRASTING VENTURE CAPITAL AND BUYOUT INVESTMENTS

c compare and contrast characteristics of buyout and venture capital investments;

Our two main categories of private equity investments, buyout and venture capital funds, dominate in terms of number of funds and invested amounts. Whereas a VC firm may have a specialized industry focus—seeking the next rising star in technology or life sciences—LBO firms generally invest in a portfolio of firms with more predictable cash flow. VC firms seek revenue growth from new enterprise and technology; buyout firms focus more on EBIT or EBITDA growth by established companies. Valuation is thus fundamentally different, and Exhibit 3 presents certain key distinctions.

Exhibit 3 Characteristics of Buyout and Venture Capital Investments

Buyout Investments:	Venture Capital Investments:
■ Steady and predictable cash flows	■ Low cash flow predictability; cash flow projections may not be realistic
■ Excellent market position (can be a niche player)	■ Lack of market history; new market and possibly an unproven future market (early-stage venture)
■ Significant asset base (may serve as a basis for collateral lending)	■ Weak asset base
■ Strong and experienced management team	■ Newly formed management team with strong individual track record as entrepreneurs
■ Extensive use of leverage consisting of a large proportion of senior debt and a significant layer of junior and/or mezzanine debt	■ Primarily equity funded; the use of leverage is rare and very limited
■ Risk is measurable; investments are in mature businesses with long operating histories	■ The assessment of risk is difficult because of new technologies, new markets, and a lack of operating history
■ Predictable exit (secondary buyout, sale to a strategic buyer, IPO)	■ Exits are difficult to anticipate (secondary venture sale, sale to strategic/financial buyer, IPO)
■ Established products	■ Technological breakthrough but the route to market is yet unproven
■ Potential for restructuring and cost reduction	■ Significant cash burn rate required to ensure company development and commercial viability
■ Low working capital requirement	■ Expanding capital requirement if in the growth phase
■ Buyout firms typically conduct full-blown due diligence before investing in the target firm (financial, strategic, commercial, legal, tax, environmental)	■ VC firms tend to conduct technology and commercial due diligence before investing; financial due diligence is limited as portfolio companies have no or very little operating history
■ Buyout firms monitor cash flow management and strategic and business planning	■ VC firms monitor the achievement of milestones defined in the business plan

Exhibit 3 (Continued)

Buyout Investments:	Venture Capital Investments:
■ Investment portfolio returns are generally characterized by a lower variance across returns from underlying investments; bankruptcies are rare	■ Investment portfolio returns are generally characterized by very high returns from a limited number of highly successful investments and a significant number of write-offs from poor-performing investments or failures
■ Large buyout firms are generally significant players in the capital markets	■ VC firms tend to be much less active in the capital markets
■ Most transactions are auctions involving multiple potential acquirers	■ Many transactions are "proprietary," arising from relationships between venture capitalists and entrepreneurs
■ High-performing buyout firms tend to have a better ability to secure larger pools of financing given their track record	■ VC firms tend to be less scalable relative to buyout firms; the increase in size of subsequent funds tends to be less significant
■ Variable revenue to the general partner (GP) at buyout firms generally comes in the form of carried interest, transaction fees, and monitoring fees	■ Carried interest (participation in profits) is generally the main source of variable revenue to the general partner at VC firms; transaction and monitoring fees are rare in practice

LBO MODEL FOR VALUATION OF BUYOUT TRANSACTIONS 4

d interpret LBO model and VC method output;

When the buyer in a private equity transaction acquires from the seller a controlling stake in the equity capital of a target company, it is called a buyout. The generic term "buyout" refers explicitly to the notion of acquiring control. It denotes a wide range of techniques, including but not limited to management buyouts (MBOs), leveraged buyouts (LBOs), and takeovers. In this reading, we focus on LBOs: using borrowed money to finance a significant portion of the acquisition price.

Given their target sector, private equity firms look for characteristics that make a company particularly attractive as an LBO target:

Undervalued/depressed stock price. The private equity firm perceives that the company's intrinsic value exceeds its market price. Firms are therefore willing to pay a premium to the market price in order to secure approval by the seller's shareholders. In other circumstances, firms see a chance to make an acquisition cheaply, and the stock prices of out-of-favor public companies may make them attractive.

Willing management and shareholders. Company management is looking for a deal. They may have opportunities to increase value, but they lack the resources to make investments in the processes, personnel, equipment, and so on, that would drive long-term growth. Company shareholders may have insufficient access to capital and so welcome a private equity partner. Family business owners may want to cash out. PE firms can provide the time and capital to expand a company or turn it around.

Inefficient companies. Private equity firms seek to generate attractive returns by identifying companies that are inefficiently managed and have the potential to perform well if managed better.

Strong and sustainable cash flow. Companies that generate strong cash flow are attractive because LBO transactions have the target company taking on significant debt. Cash flow is necessary to make interest payments.

Low leverage. Private equity firms focus on target companies that have no significant debt on their balance sheets because then it's easier to use debt to finance a large portion of the purchase price.

Assets. Private equity managers like companies that have a significant amount of unencumbered physical assets. These physical assets can be used as security, and secured debt is cheaper than unsecured debt.

Earlier we considered a typical LBO capital structure that entailed 30% equity along with leveraged loans and high-yield bonds to make up the rest of the purchase price. Leveraged loans are often the source of a larger amount of capital than either equity or high-yield bonds. As an alternative to high-yield bonds, mezzanine financing may be used. Mezzanine financing refers to debt or preferred shares with a relationship to common equity that results from a feature such as attached warrants or conversion options. Being subordinate to both senior and high-yield debt, mezzanine financing typically pays a higher coupon rate. In addition to interest or dividends, this type of financing offers a potential return based on increases in the value of common equity and is generally customized to fit the specific requirements of the transaction in question.

4.1 The LBO Model

The LBO model is not a separate valuation technique but, rather, a way of determining the impact of the capital structure, purchase price, and other parameters on the returns expected by the private equity firm from the deal.

The LBO model has three main input parameters: the cash flow forecasts of the target company, the return that the providers of financing (equity, senior debt, high-yield bonds, mezzanine) are expecting, and the amount of financing available for the transaction. The free cash flow forecasts of the target company are generally prepared by its management and are subject to an extensive due diligence process (strategic, commercial, financial, legal, and environmental) to determine their reliability. The forecasts assume an explicit horizon that generally corresponds to the expected holding period (i.e., investment period) of the private equity firm.

The exit year is typically considered to determine the expected IRR sensitivity of the equity capital around the anticipated exit date. The exit value is determined most frequently by reference to an expected range of exit multiples determined on the basis of a peer group of comparable companies (enterprise value-to-EBITDA ratio).

Given the significant predictability of cash flows in buyout transactions, the income-based approach (discounted cash flows, adjusted present value, LBO model, target IRR) is frequently used as a primary method to determine the value of equity, considering the expected change in leverage until the time of exit of the investment. The initial high and declining financial leverage is the main technical valuation issue that needs to be adequately factored into the income approach when applied to a buyout valuation. The value is also frequently corroborated by an analysis of the peer group of comparable publicly traded companies.

On the basis of the input parameters, the LBO model provides the maximum price that can be paid to the seller while satisfying the target returns for the providers of financing. This is why the LBO model is not a valuation methodology per se. It is a negotiation tool that helps develop a range of acceptable prices to conclude the transaction.

Exhibit 4 is a value-creation chart that illustrates the sources of the additional value between the original cost and the exit value. Value creation comes from a combination of factors: earnings growth arising from operational improvements and enhanced corporate governance, multiple expansion depending on pre-identified potential exits, and the optimization of financial leverage and repayment of part of the debt with operational cash flows before the exit. Each component of the value creation chart should be carefully considered and backed by supporting analyses, which frequently come from the lengthy due diligence process (especially commercial, tax, and financial analysis) and also from a strategic review that quantifies the range of plausible value creation.

Exhibit 4 Typical Leveraged Buyout Value-Creation Chart

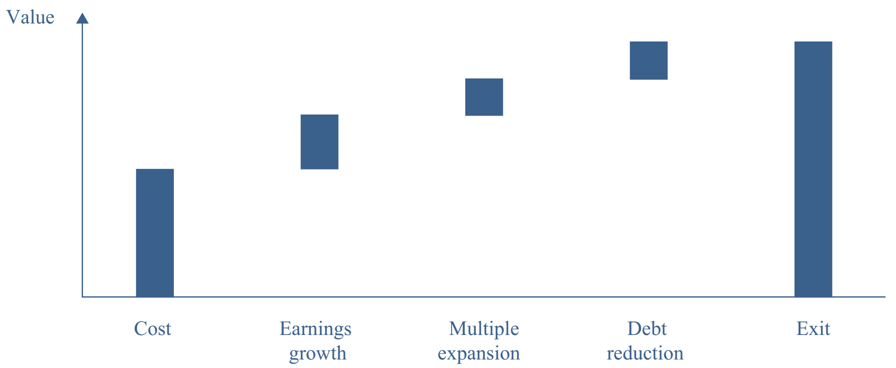
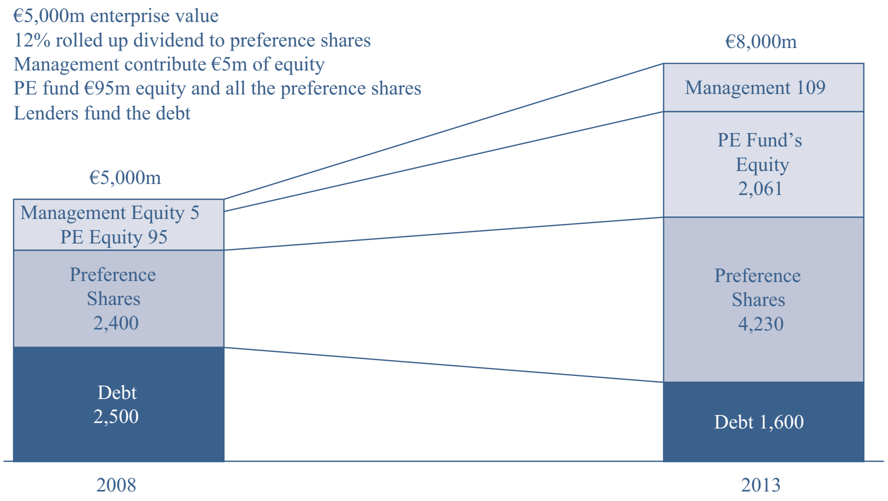


Exhibit 5 provides an example of a €5,000 (amounts in millions) investment in a private equity transaction. The transaction is financed with 50% debt and 50% equity. The €2,500 equity investment is further broken into €2,400 of preference shares owned by the private equity fund, €95 of equity owned by the private equity fund, and €5 of management equity. The preference shares are promised a 12% annual return (paid at exit). The private equity fund's equity is promised 95% of the residual value of the firm after creditors and preference shares are paid, and management equity holders are promised the remaining 5%.

Exhibit 5 Stakeholder Payoffs

	Invested	Proceeds	Multiple	IRR
Management	€5m	€109m	21.8x	85%
PE fund	€2,495m	€6,291m	2.5x	20%



Assume that the exit value, five years after investment, is 1.6 times the original cost. The initial investment of €5,000 has an exit value of €8,000. The specific payoffs for the four claimants are as follows:

- Senior debt has been partially retired with operational cash flows, reducing debt from €2,500 to €1,600. So debtholders get €1,600.
- Preference shares are paid a 12% return for five years, so they receive €2,400 × $(1.12)^5$ = €4,230.
- PE fund equity receives 95% of the terminal equity value, or 0.95 × [8,000 – (4,230 + 1,600)] = €2,061.
- Management equity receives 5% of the terminal equity value, or 0.05 × [8,000 – (4,230 + 1,600)] = €109.

As you can see, preference shares increase in value over time as a result of their preferred dividend being capitalized, and the equity held by the PE fund and by the management is expected to increase significantly depending on the total enterprise value upon exit. Both the equity sold to managers, frequently known as the management equity program (MEP), and the equity held by the private equity firm are most sensitive to the level of the exit. The larger the exit multiple, the larger the upside potential for both the MEP and the equity held by the private equity firm. In the example, assuming that an exit of 1.6 times cash may be achieved at the anticipated exit date (five years from investment), the management would realize an IRR of 85% per annum on its investment and the private equity fund equity holders an IRR of 20% per annum. The private equity firm also earns 12% per annum on its preference shares.

This chart also demonstrates the critical importance of leverage in buyout transactions. A reduction in financial leverage over time is instrumental in magnifying the returns available to shareholders. Note that the bulk of financial leverage in LBOs consists of senior debt, much of which will be amortizing. Therefore, the reduction in financial leverage gradually increases over time as a proportion of principal is paid back to senior lenders on an annual or semi-annual basis depending on the terms of the senior debt. As a result of the gradual repayment of senior debt over time, a larger proportion of operating cash flows becomes available to equity holders. Of course, this mechanism works well as long as no significant adverse economic factors impact the business of the target LBO company and also as long as a successful exit can be secured in the foreseeable future. It should be remembered that these high levels of debt increase the risks borne by the equity investors significantly. Such risks should be accounted for when comparing the expected returns with alternative investment classes, such as investments in the stock market.

Typically, a series of scenarios with varying levels of cash exits, growth assumptions, and debt levels are engineered with the use of an LBO model, using as inputs the required rate of return from each stakeholder (equity, mezzanine, senior debtholders) to gain a sound understanding of the buyout firm's flexibility in conducting the deal.

VC METHOD FOR VALUATION OF VENTURE CAPITAL TRANSACTIONS[1] 5

d interpret LBO model and VC method output;

The primary difficulty of venture capital investing is the substantial uncertainty about the company's future prospects. Traditional valuation tools, such as discounted cash flow, earnings multiples, and the LBO model, are not usually practical in the VC context. Moreover, due to the uncertainty, VC financing is done in stages—financing rounds referred to as Series A, Series B, and so on—and it is essential to understand the effects of ownership dilution. So, we turn to the VC method.

Imagine a start-up company is worth \$16 million, and its shareholders own 100% of the equity. The company decides to raise \$4 million in additional equity capital from a venture capital firm. The value of the start-up company after the financing is \$20 million (= \$16 million + \$4 million).

According to the VC method, \$16 million is known as the pre-money (i.e., pre-financing) valuation, \$4 million is the amount of new equity Series A investment, and \$20 million is known as the post-money (i.e., post-financing) valuation. That is,

Pre-money valuation + New equity investment = Post-money valuation.

What is the total equity stake owned by the VC firm if it were to make the \$4 million investment? After the transaction, the VC firm will own 20% (= \$4 million/\$20 million) of the equity stake or, in VC terms, will have 20% fractional ownership of the start-up. The original shareholder equity stake gets diluted to 80%. This basic example is central to understanding the VC Method.

The central challenge of the VC method is determining the pre-money valuation. The closer you are to the inception date of the company, the harder it is to assess its pre-money valuation. Indeed, in early-stage venture capital, the investor might be backing an idea with no tangible proof of commercial viability. At the heart of a successful VC investment is a fundamentally novel and uncertain idea; otherwise, existing firms would easily replicate it. It is hard to envision a comparable company in the public domain or even comparable VC investments when you are early in the process. So the VC Method works backwards to come up with the pre-money valuation, taking as inputs the expected exit valuation and the required return on investment.

1 Authored by Victoria Ivashina, PhD, Harvard Business School.

5.1 Expected Exit Valuation

VC investments are typically financed through a closed-end fund with a finite life horizon; 10 years is very common in private equity. The average holding period for individual investments within the fund is between five and eight years. Performance metrics for the firm, measuring its progress toward commercial success, are identified at the end of the expected investment period. The performance metric could be, for example, revenues or number of users. For early-stage VC investments, it might be ambitious to expect that the company becomes profitable, but if that is the expectation, one could also use earnings per share or EBITDA. Choosing the right multiple to apply depends on the specifics of the industry, and the general idea is that, whatever the performance metrics, the value of the VC investment at exit can be determined based on projected performance metrics at anticipated exit and an appropriate industry multiple.

5.2 Required Rate of Return

VC firms target an expected hurdle return. The failure rate of VC investments, especially early-stage VC investments, is much higher than that of growth equity or buyouts. Thus, the target return on investment tends to be 10× to 30×, compared with 2.0× to 2.5× in the buyout space.

Armed with the hurdle return and the value of the equity at exit, we can compute the post-money valuation:

$$\text{ROI} = \frac{\text{Value of equity at exit}}{\text{Post-money valuation}},$$

so,

$$\text{Post-money valuation} = \frac{\text{Value of equity at exit}}{\text{ROI}}.$$

Given the capital needs of the company, we know that the pre-money valuation is as follows:

Pre-money valuation = Post-money valuation − New equity injection.

Take an entrepreneur who is looking to raise \$500,000. Given the size of its market and the industry, the entrepreneur's company expects to reach sales of \$80 million over the investment horizon. A typical revenue multiple for a revenue-generating business in its industry is 2×. If the VC firm's ROI is 20× and the entrepreneur's company has no debt, then what is the pre-money valuation? And what is the VC firm's fractional ownership?

$$\text{Pre-money valuation} = \frac{\$80 \text{ million} \times 2}{20} - \$0.5 \text{ million} = \$7.5 \text{ million}.$$

$$\text{VC fractional ownership} = \frac{\$0.5 \text{ million}}{\$7.5 \text{ million} + \$0.5 \text{ million}} = 6.25\%.$$

These numbers are sensitive to the revenue figure. The required ROI will be dictated by the competition and the cost of capital. Although ROI is not sensitive to the length of the holding period, an expected holding period of five to eight years is generally the assumption in the VC fund structure. The previous example could also be reframed in terms of required internal rate of return (IRR) to equity. Over a five-year period, 20× ROI is equivalent to an 82% IRR. In general,

$$(1 + \text{IRR})^t = \text{ROI} = \frac{\text{Value of equity at exit}}{\text{Post-money valuation}},$$

or in this case,

$$(1 + \mathrm{IRR})^5 = 20 = \frac{\$80 \text{ million} \times 2}{\$8 \text{ million}}.$$

5.3 Option Pools

Returning to the first example, if the start-up were owned by the founders, they held a total of 10 million shares, and there is no option pool, what is number of new shares that the company needs to issue? And at what price?

We know that

$$\frac{N(\text{new shares})}{N(\text{total shares after financing})} = \frac{N(\text{new shares})}{N(\text{old shares}) + N(\text{new shares})} = 20\%.$$

This means that 2.5 million new shares [= (0.2 × 10 million)/0.8] will have to be issued. Based on the $4 million of new equity raised, the shares would be issued at $4 million/2.5 million = $1.60 per share.

Scenarios without an outstanding option pool, however, are unrealistic. To attract and incentivize employees, start-ups grant their employees the option to purchase shares. When these options are exercised, they will naturally have a dilutive effect, so VC firms tend to calculate the per-share price on a fully diluted basis.

The basis of the dilution is contractually defined. The central question is about who assumes the effect of future dilution: original shareholders or new investors? There is tension between what benefits the VC investors and what benefits the founders, which is why this is important. By calculating the share price on a fully diluted basis, VC investors are effectively left untouched by the dilution effect. Instead, the original shareholders, our founders, absorb the effects of dilution. We will return to this concept when we examine follow-up financing series.

Back to our start-up example, let's say that in addition to the founders holding 10 million shares, there is an outstanding option pool of 2 million shares. How many shares need to be issued? And at what price per share? (Recall that the default is to make calculations on a fully diluted basis.) What is the starting equity stake of the VC firm? What is the VC firm's equity stake after dilution?

- The number of existing or original shares on a fully diluted basis now is 12 million. This means that 3 million new shares [= (0.2 × 12 million)/0.8] will have to be issued.
- The new price per share is $4 million/$3 million = $1.33.
- The VC firm's starting ownership is 3 million/13 million = 23.08%, but as options are exercised, this ownership will be diluted to 20%.
- The VC fractional ownership on a post-dilution basis is not affected and is equal to 20% = $4 million/$20 million. The ownership structure should not affect the value of the assets—the key insight of the Modigliani–Miller theorem. If the pre-money valuation is $4 million, it should remain at $4 million regardless of how many options the firm issues.

5.4 Stage Financing

We've established that VC financing is usually executed in stages, largely due to the uncertainty that surrounds VC investments.

As an example, let's take Facebook and its initial financing rounds. The first external investment in the company took place in 2004, shortly after the platform launched, in the amount of $500,000. Just a year later, Facebook raised $12.7 million. In 2006, it raised \$27.5 million, followed by \$300 million in 2007. Behind this astonishing growth in fundraising sits an even more impressive rise in company valuation. So what changed for Facebook in the three years between 2004 and 2007? Why were the initial rounds so small by comparison? The answer is uncertainty about the company's future, and this illustrates that even one year out, company valuation can fluctuate dramatically. This is despite the fact that Facebook was a sought-after company with multiple veteran VC investors pursuing it. As the Facebook user base grew, so did their investment and investors' visibility into the company's future commercial success. Stage financing, thus, is a key mitigator of the risk that is fundamental to venture capital: significant uncertainty about growth and profitability prospects.

How do the different series of financing relate to each other? Because the earlier-stage investors take on higher risk, the return for those investors has to be higher. Valuations, specifically pre-money valuations, at which later rounds of financing are raised, provide insight into the performance of an otherwise illiquid asset class.

Let's assume that our earlier example—a VC firm raising \$500,000 for a start-up company in exchange for a 6.25% stake—describes a Series A financing. Imagine that one year later the firm raises \$2 million in a Series B financing at 10× ROI. The exit of all investors is expected to occur simultaneously, and Series B investors were projecting an exit valuation of \$300 million. We can compute the ownership structure in each of the financing rounds along with the implied ROI for the Series A and B financings.

For Series B investors,

$$\text{Post-money valuation} = \frac{\$300\text{ million}}{10} = \$30\text{ million}.$$

$$\text{Pre-money valuation} = \frac{\$300\text{ million}}{10} - \$2\text{ million} = \$28\text{ million}.$$

$$\text{VC fractional ownership} = \frac{\$2\text{ million}}{\$28\text{ million} + \$2\text{ million}} = 6.67\%.$$

$$\text{ROI} = (1 + \text{IRR})^5 = 10 = \frac{\$300\text{ million}}{\$30\text{ million}}.$$

These figures are summarized in Exhibit 6.

Exhibit 6 Stage Financing Example

(in thousands)	Series A	Series B
Required ROI	20.0	10.0
Investment	500	2,000
Exit valuation	160,000	300,000
Post-money valuation	8,000	30,000
Pre-money valuation	7,500	28,000
Fractional ownership required	6.25%	6.67%
Ownership:		
Entrepreneurs	93.75%	87.50%
Series A investors	6.25%	5.83%
Series B investors	—	6.67%
Total	100.00%	100.00%

Exhibit 6 (Continued)

(in thousands)	Series A	Series B
Implied ROI		
Series A investors	20.0	35.0
Series B investors	—	10.0

This example illustrates that the pre-money valuation implied by Series B indicates a substantial appreciation of the Series A investment. Instead of 20× ROI, Series A now has an implied ROI of 35×, although no exits occurred, and despite dilution of the Series A stake from 6.25% to 5.83%.

For Series A investors, the implied ROI, which was originally 20×, or

$$\text{ROI} = 20 = \frac{\$160 \text{ million} \times 6.25\%}{\$0.5 \text{ million}},$$

increases to 35× at the time of the Series B financing one year later, or

$$\text{ROI} = 35 = \frac{\$300 \text{ million} \times 5.83\%}{\$0.5 \text{ million}}.$$

The entry of new investors with Series B dilutes ownership for both Series A investors and the entrepreneurs on a proportionate basis. That is,

Entrepreneurs' ownership = [(1 – 6.67%) 93.75%] = 87.50%.

Series A investors' ownership = [(1 – 6.67%) 6.25%] = 5.83%.

Venture capital investments tend to be minority stake investments. This is partly because the founders might not be willing to give up control but also because entrepreneurs are essential in the initial stages of business development. So, the dilution of initial investors through the subsequent financing rounds is common. In the previous example, the dilution from Series B was absorbed pro rata by the entrepreneurs and the Series A investors. Because control is not essential, this arrangement is typical. If an alternative economic arrangement is sought, however—as in the earlier example of employees' stock options—designing an arrangement whereby the dilution is absorbed disproportionately by entrepreneurs is possible.

Whereas our example treated Series A and Series B shares as common stock, it is typical to use convertible preferred equity in later-stage financing. The capital that comes in later stages is less risky than earlier-stage financing. In addition, to mitigate risk further, later-stage capital tends to have a preferred dividend. Series B shares could entitle shareholders to a preferred dividend of 5%, for example. On an investment of $2 million over three years, the cumulative dividend is $315,250. If the investee company performance is as expected and the returns are high, the preference shares will be irrelevant. However, if things do not go as planned, the accumulated dividend is treated as junior debt, diminishing the value held by earlier equity investors while preserving the value for Series B. Clearly, this makes Series B more valuable than if it had been just common equity. Importantly, the value comes at the expense of earlier investors. These adjustments are rarely accounted for in practice, however.

6 EXIT ROUTES: RETURN CASH TO INVESTORS

e explain alternative exit routes in private equity and their impact on value;

The exit is a critical mechanism for unlocking value in private equity. Most private equity firms consider their exit options prior to investing, and they factor their assessment of the exit outcome into their IRR analysis.

Generally, PE investors have access to the following four exit routes:

- *Initial public offering (IPO)*: Going public offers significant advantages, including higher valuation multiples as a result of enhanced liquidity, access to large amounts of capital, and the possibility of attracting higher-caliber managers. But the process is cumbersome and less flexible and entails significant costs. Therefore, an IPO is an appropriate exit route for private companies that are of a sufficient size with an established operating history and excellent growth prospects. Timing is important and heavily dependent on public equity market conditions. IPO markets have shut down for long periods following major events, such as the internet bubble collapse that began in March 2000 and the GFC, starting in 2007. Regional economic concerns and regulatory issues, such as uncertainty around Britain's plan to exit the European Union, have had negative ramifications for equity and IPO markets worldwide. In fact, any extended market downturn can limit the ability of new companies to come to market. IPO exits are more common for VC-backed companies, but they are not the largest divestment alternative.
- *Secondary market*: The sale of an investor's stake to other financial investors or to strategic investors (think companies that operate in the same sector or are keen to try). As private equity has become increasingly segmented, secondary market transactions tend to occur within each segment—that is, buyout firms tend to sell to other buyout firms (secondary buyouts) and venture firms to other venture firms (secondary venture capital transactions). These secondary market transactions account for a significant proportion of exits, especially in the buyout segment. Venture capital exits by means of a buyout are also possible but are rare in practice, because buyout firms are reluctant to finance development-stage companies with a significant amount of leverage. The main advantages of secondary market transactions are (1) the possibility of achieving the highest valuation multiples in the absence of an IPO and (2) the fact that given the segmentation of private equity firms, specialized firms have the skill to bring their portfolio companies to the next level—say, through a restructuring or merger or by bringing them to a new market—and then to sell, either to a strategic investor seeking to exploit synergies or to another private equity firm with another set of skills to further add value to the portfolio company.
- *Management buyout (MBO)*: These takeovers by company management use significant amounts of leverage to finance the acquisition. Alignment of interest between management and investors is optimal under this exit scenario, but it may come at the expense of excessive leverage that significantly reduces the company's flexibility.
- *Liquidation*: Controlling shareholders have the power to liquidate the company if it is no longer viable. This exit mechanism generally results in a floor value for the company but may come at a cost of very negative publicity for the private equity firm if the company is large and the employee count is significant.

Timing the exit and determining the optimal exit route are important. Even carefully planned exits face the unexpected, however, and may be delayed or accelerated depending on market conditions or purely opportunistic circumstances.

Suppose, for example, that an LBO firm is planning to exit one of its portfolio companies, but the public market and economic conditions have collapsed, rendering any exit via a trade sale or an IPO unprofitable. Instead, the LBO firm exploits the depressed pricing environment to conduct another acquisition and merges the target with the original portfolio company in order to strengthen its market position and product range. Then it waits for better market conditions before conducting the sale. Such flexibility is critical for private equity firms during hard times and underlines the importance of PE firms maintaining sufficient financial strength.

There seems to be no boundary to the size of buyout transactions, as expectations have consistently been exceeded. Three of the largest buyout transactions—TXU Energy ($32.1 billion), First Data ($25.7 billion), and Alltel ($25.1 billion)—all took place in 2007 immediately prior to the GFC. Strength in capital markets combined with the increasing prevalence of megafunds suggest even larger transactions may be in store for the biggest buyout firms. Private equity firms appear to be moving into uncharted territory by managing exits at such levels. The central question about these mega-buyout transactions is how the exits will take place given that the possibilities are much more limited relative to smaller deals. IPOs, for example, raise significantly more challenges, restricting sellers to a gradual exit (only a single block of shares can be sold initially) and proving excessively risky when market conditions are suboptimal. And a real challenge exists for large, unified companies for which a single exit is the only way out. This is in contrast to the type of large companies that may be viewed as holding companies for a portfolio of real assets, which can be sold in tranches.

When an exit is anticipated within one or two years, the multiples observed from comparable publicly quoted firms provide good guidance for an expected exit multiple, and stress tests on that value may be conducted for small incremental changes and based on market knowledge. When the exit horizon is much further out, these multiples are less reliable, and stress tests may be performed on valuation model inputs, such as discount factor and terminal growth rates, and on financial forecasts, such as sales growth and operating margins.

6.1 Exit Routes: Summary

Valuation is the most critical aspect of private equity transactions. The investment decision-making process typically flows from the screening of investment opportunities to preparing a proposal, appraising the investment, and structuring the deal and finally to the negotiating phase. Along with the various due diligence investigations (commercial or strategic, financial, legal, tax, environmental) that are generally conducted on private equity investment opportunities, valuation serves to assess a company's ability to generate superior cash flows from a distinctive competitive advantage and as a benchmark for negotiations with the seller. Because of the difficulties in valuing private companies, a variety of alternative valuation methods are typically used to provide guidance. Private equity valuation is a process that starts as a support for decision making at the transaction phase but also serves as a monitoring tool to capture new opportunities, create value, or protect from losses during the investment period and as a performance reporting tool for investors.

7 RISKS AND COSTS OF INVESTING IN PRIVATE EQUITY

f explain risks and costs of investing in private equity;

We turn now to the perspective of a private equity fund investor.

7.1 What Are the Risks and Costs of Investing in Private Equity?

Most jurisdictions restrict private equity investing to "qualified investors"—typically, institutions and high-net-worth individuals who meet certain criteria. These restrictions are a product of the high levels of risk associated with private equity investing, which are generally subject to disclosure in the private equity fund prospectus. Such risks may be categorized as general private equity risk factors, investment strategy–specific risk factors (buyout, venture capital, mezzanine), industry-specific risk factors, risk factors specific to the investment vehicle, or regional or emerging market risks when applicable.

The following are some general private equity risk factors:

- *Illiquidity of investments*: Because private equity investments are generally not traded on any securities market, the exit of investments may not end up being conducted on a timely basis.
- *Unquoted investments*: Investing in unquoted securities may be risky relative to investing in securities quoted on a regulated securities exchange.
- *Competition for attractive investment opportunities*: Competition for investment opportunities on attractive terms may be high.
- *Reliance on the management of investee companies (agency risk)*: There is no assurance that the management of the investee companies will run the company in the best interests of the private equity firm, particularly in earlier-stage deals in which the management retains a controlling stake in the company and enjoys certain private benefits of control.
- *Loss of capital*: High business and financial risks may result in a substantial loss of capital.
- *Government regulations*: Investee companies' products and services may be subject to changes in government regulations that adversely affect their business model.
- *Taxation risk*: The tax treatment of capital gains, dividends, or limited partnerships may change over time.
- *Valuation of investments*: The valuation of private equity investments is subject to significant judgment. When valuations are not conducted by an independent party, they may be subject to bias.
- *Lack of investment capital*: Investee companies may require additional future financing that may not be available.

- *Lack of diversification*: Highly concentrated investment portfolios may be exposed to significant losses. Instead, private equity investors should consider a mix of funds of different vintage, portfolio companies at different stages of development, and investing across private equity strategies, such as large and mid-market buyouts, venture capital, mezzanine finance, and restructuring.
- *Market risk*: Changes in general market conditions (interest rates, currency exchange rates) may adversely affect private equity investments. The impact of market risk is, however, long-term in nature given the long-term horizon of private equity funds. Temporary short-term market fluctuations are generally irrelevant.

The costs associated with private equity are substantially higher than the costs of public market investing. We break them down here:

- *Transaction fees*: These arise from due diligence work, bank financing costs, the legal fees of arranging an acquisition, and the costs of arranging the sale of an investee company.
- *Fund setup costs*: These are mainly the legal costs of setting up the investment vehicle, and they are typically amortized over the life of the investment vehicle.
- *Administrative costs*: Custodian, accounting, and transfer-agent costs are generally charged yearly as a fraction of the investment vehicle's net asset value.
- *Audit costs*: This is a fixed annual fee.
- *Management and performance fees*: 2% management fees and 20% performance fees, which are generally more significant than the fees charged by regular investment funds.
- *Dilution*: A more subtle cost—from our preceding examples we know that dilution comes from the stock options granted to management and the PE firm itself, as well as from additional rounds of financing.
- *Placement fees*: Fundraising fees may be charged up front—2% is not uncommon in private equity—or by means of a trailer fee. A trailer fee is generally charged annually and figures as a fraction of the amount invested by limited partners as long as these amounts remain invested in the investment vehicle.

PRIVATE EQUITY FUND STRUCTURES AND TERMS 8

g explain private equity fund structures, terms, due diligence, and valuation in the context of an analysis of private equity fund returns;

When analyzing an investment in a private equity fund, a solid understanding of PE fund structures, terms of investment, due diligence, and PE fund valuation are an absolute prerequisite for investors. When interpreting financial performance, private equity raises many more challenges than public equities do. In addition to the structure and terms, two of the main differentiating characteristics relate to the nature of the subscriptions investors make in private equity structures and to the J-curve effect. Investors initially commit a certain amount to the private equity fund that is subsequently drawn by the fund as the fund's capital is deployed to portfolio companies. This contrasts with public market investing in which investment orders are typically fully disbursed at the time the orders are settled on the markets. The J-curve effect refers to the typical profile of reported PE fund returns, whereby low or negative returns are reported in the early years of a private equity fund (in large part as a result of the fees' impact on net returns), followed by increased returns thereafter as the private equity firm manages portfolio companies toward the exit.

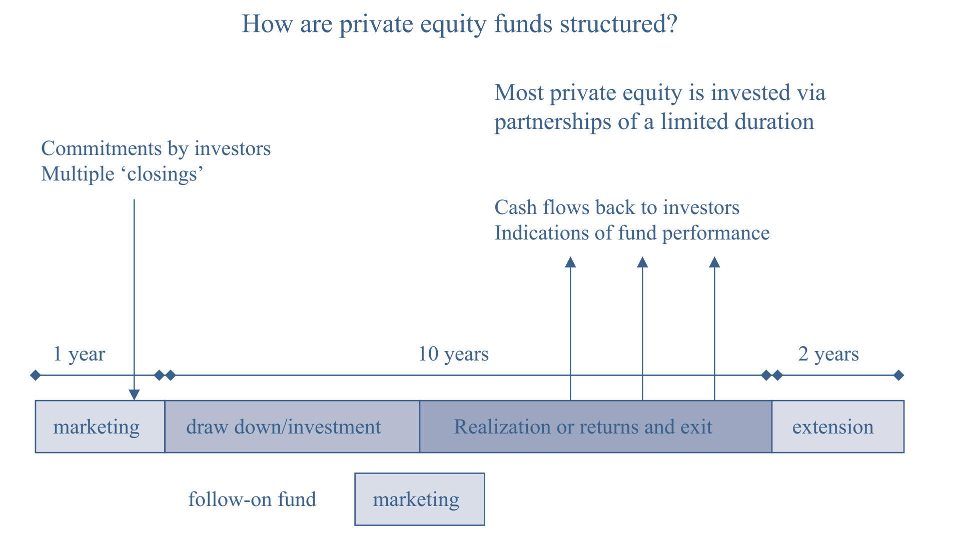
The limited partnership has emerged as the dominant form in most jurisdictions. Funds that are structured as limited partnerships are governed by a limited partnership agreement between the fund manager, called the general partner (GP), and the fund's investors, called limited partners (LPs). Whereas the GP has management control over the fund and is jointly liable for all debts, LPs have limited liability; that is, they do not risk more than the amount of their investment in the fund.

The main alternative to the limited partnership is a corporate structure called a company limited by shares, which mirrors in its functioning the limited partnership but offers better legal protection to the GP and to some extent the LPs, depending on the jurisdiction. Some fund structures are subject to light regulatory oversight, which offers enhanced protection to LPs.

The vast majority of these private equity fund structures are closed-end, meaning they restrict existing investors from redeeming their shares over the lifetime of the fund and limit new investors to entering the fund at predefined time periods, at the discretion of the GP.

Private equity firms operate effectively in two spheres: the business of managing private equity investments and the business of raising funds. Their marketing efforts tend to be planned well in advance of the launch of their funds to ensure that the announced target fund size will be met once the fund is effectively started. The marketing phase of a private equity fund, depending on whether it is a first fund or a following fund, may take between one and two years. Once investors have committed their investment to the fund, private equity managers draw on investors' commitments—this action is called a drawdown or capital call—as the fund is being deployed and invested in portfolio companies. After commitment, private equity funds tend to have a duration of 10–12 years, generally extendable to an additional 2–3 years. Exhibit 7 illustrates structuring and stages in the life cycle for a PE fund.

Exhibit 7 Structuring and Life Cycle Stages for a Private Equity Fund



Fund terms are contractually defined in a fund prospectus or limited partnership agreement, which is available to qualified prospective investors. The definition of qualified investors depends on the jurisdiction. Typically, wealth criteria (e.g., exceeding US$1 million) and/or a minimum subscription threshold (e.g., a minimum of €125,000) apply. The terms are frequently the result of the balance of negotiation power between GPs and LPs. Although the balance of negotiation power shifted in favor of LPs during and immediately following the GFC, it turned back in favor of GPs as the largest sponsor continued to increase in size and gain market share. Any significant downturn in private equity's fortunes may change the balance of power in favor of LPs once again. In any event, the negotiation of terms informs an alignment of interests between the GP and LPs and defines GP incentives, such as transaction fees and profit sharing. The most significant fund terms may be categorized into economic and corporate governance terms.

8.1 Economic Terms

- *Management fees* represent an annual percentage of committed capital, which is paid quarterly to the GP during the fund investment period. Fees of 1.5%–2.5% are common, and some of the most successful funds charge even more. After the investment period, fees may decline somewhat and be calculated based on invested capital. Less frequently, management fees may continue to be calculated based on capital commitment or net asset value.
- *Transaction fees* are fees paid to GPs in their advisory capacity when they provide investment banking services for transactions (mergers and acquisitions, IPOs) that benefit the fund. These fees may be subject to sharing agreements with LPs, typically a 50/50 split. When such fee-sharing agreements apply, they generally come as a deduction to the management fees.
- *Carried interest* represents the general partner's share of the profits generated by a private equity fund. Carried interest is frequently in the range of 20% of the fund's profits (after management fees).
- *Ratchet* is a mechanism that determines the allocation of equity between shareholders and the management team of the PE-controlled company. A ratchet enables the management team to increase its equity allocation depending on the company's actual performance and the return achieved by the PE firm.
- *Hurdle rate* is the internal rate of return that a private equity fund must achieve before the GP receives any carried interest. The hurdle rate is typically in the range of 7%–10%. The objective is to align the interests of the GP with those of LPs by giving additional incentives to the GP to outperform traditional investment benchmarks and to protect against LP downside.

EXAMPLE 1

Calculation of Carried Interest

Suppose that a LBO fund has committed capital of US$100 million, carried interest of 20%, and a hurdle rate of 8%. The fund called 75% of its commitments from investors at the beginning of Year 1, which was invested at the beginning of Year 1 in target company A for $40 million and target company B for $35 million. Suppose that at the end of Year 2, a profit of $5 million has been realized by the GP upon exit of the investment in company A, and the value of the investment in company B has remained unchanged. Suppose also that the GP is entitled to carried interest on a deal-by-deal basis; that is, the IRR used to calculate carried interest is calculated for each investment upon exit. A theoretical carried interest of $1 million (20% of $5 million) could be granted to the GP, but the IRR upon exit of investment in company A is only 6.1%. Unless the IRR exceeds the hurdle rate, no carried interest may be paid to the GP.

- *Target fund size* is expressed as an absolute amount in the fund prospectus or information memorandum (also called the private placement memorandum, or PPM; offering memorandum, or OM; or offering circular, or OC). Target fund size is critical investor information because it signals the GP's capacity to manage a portfolio of a predefined size and also the GP's ability to raise funds. A fund that closed with a significantly lower size relative to the target size is perceived as a negative signal.
- *Vintage year* is the year the private equity fund is launched. Reference to the vintage year allows performance comparison of funds operating at the same stage and under the same market conditions.
- *Term of the fund* is typically 10 years, which is extendable for additional shorter periods by agreement with the investors. Although infrequently observed, funds can also be of unlimited duration, and in this case they are often quoted on stock markets—for example, investment trusts.

8.2 Corporate Governance Terms

- *Key man clause*. Under the key man clause, a certain number of key named executives are expected to play an active role in the management of the fund. In case of the departure of such a key executive or insufficient time spent in the management of the fund, the clause provides that the GP may be prohibited from making any new investments until a new key executive is appointed.
- *Disclosure and confidentiality*. Private equity firms have no obligations to publicly disclose their financial performance. Following a 2002 court ruling requiring the California Public Employees' Retirement System (CalPERS) to publicly report its returns on private equity investments, the Freedom of Information Act (FOIA) in the United States and similar legislation in some European countries led public pension funds to report on their private equity investments. Disclosable information relates to the financial performance of the underlying funds but does not extend to information on the companies in which the funds invest, which is not typically disclosed. The reporting by CalPERS is a prominent example of the application of this clause. Some PE fund terms may be more restrictive on confidentiality and disclosure, subject to FOIA.
- *Distribution waterfall*. This is a mechanism that delineates how distributions are allocated to LPs and GPs. The predominant mechanisms are deal-by-deal waterfalls, which allow earlier distribution of carried interest to the GP after each individual deal (also known as an American waterfall), and total return waterfalls, which result in earlier distributions to LPs because carried interest is calculated on the profits of the entire portfolio (also known as a European waterfall). Two alternatives for calculating carried interest exist under the total return method: In the first alternative, the GP receives carried interest only after the fund has returned the entire committed capital to LPs; in the second alternative, the GP receives carried interest on any distribution as long as the

value of the investment portfolio exceeds a certain threshold above invested capital, usually 20%. The European waterfall has become prevalent among large funds marketed to institutional investors.

- *Clawback provision*. A clawback provision requires the GP to return a portion or all of the carried interest to LPs if it turns out the GP has received more than its share of profits. This provision ensures that when an LP exits a highly profitable investment early in the fund's life and subsequent exits are less profitable, the GP pays back capital contributions, fees, expenses, and carried interest profits to the LPs in order to ensure that the profit split is in line with the fund's prospectus. The clawback is normally due on termination of the fund but may be subject to an annual reconciliation (or true-up).

EXAMPLE 2

Distribution Waterfalls

Suppose a private equity fund has committed capital of £300 million and a carried interest of 20%. After a first investment of £30 million, the fund exits the investment nine months later with a £15 million profit. Under the deal-by-deal method, the GP would be entitled to 20% of the profit—that is, £3 million. In the first alternative for calculating carried interest under the total return method, the LPs are entitled to the entire proceeds of the sale—that is, £45 million—and the GP is entitled to nothing (yet). Under the second alternative, the exit value of £45 million exceeds the invested value of £30 million by more than 20%. The GP would thus be entitled to £3 million.

Continuing this example with a clawback provision with an annual true-up, suppose that the deal-by-deal method applies and that a second investment of £25 million is concluded with a loss of £5 million one year later. At the annual true-up, the GP would have to pay back £1 million to LPs. In practice, an escrow account is used to regulate these fluctuations until termination of the fund.

- *Tag-along, drag along rights* are contractual provisions in share-purchase agreements. Tag-along rights ensure that minority shareholders have the right to join in a sale entered into by a majority shareholder at the same terms offered to the majority shareholder. Essentially the buyer cannot acquire control without extending its offer to all shareholders, including the management of the company. Drag-along rights allow majority shareholders who have negotiated an exit to require the minority investors to participate in the sale at the same terms, preventing minority investors from vetoing a sale.
- *No-fault divorce*. A GP may be removed both with and without cause provided that a supermajority (generally above 75%) of LPs approve the removal. In practice it is unusual for investors to succeed in removing the GP.
- *Removal for cause* is a clause that allows for removal of the GP or an earlier termination of the fund for "cause." Cause may include gross negligence on the part of the GP, a key person event, the felony conviction of a key person, bankruptcy of the GP, or a material breach of the fund prospectus. It is difficult for LPs to remove the GP for cause because when there is an allegation of wrongdoing, the GP will often agree to an out-of-court settlement and pay a fine without having to admit guilt. Moreover, it may be many years until a final court hearing takes place.

- *Investment restrictions* generally impose a minimum level of diversification on the fund's investments, a geographic and/or sector focus, or limits on borrowing.
- *Co-investment*. LPs generally have a first right of co-investing along with the GP. This can be advantageous for the LPs as fees and profit share are likely to be lower (or zero) on co-invested capital. The GP and affiliated parties are also typically restricted in their co-investments to prevent conflicts of interest with their LPs. Crossover co-investments are a classic example of a conflict of interest. A crossover co-investment occurs when a subsequent fund launched by the same GP invests in a portfolio company that has received funding from a previous fund.

8.3 Due Diligence Investigations by Potential Investors

Prior to investing in a private equity fund, prospective investors generally conduct thorough due diligence. Outlining several fundamental characteristics of PE funds will underline how important the due diligence process is:

- Private equity funds tend to exhibit a strong persistence of returns over time. This means that, typically, top-performing funds continue to outperform and poorly performing funds continue to perform poorly or disappear.
- The performance range between funds varies widely. For example, the difference between top-quartile and third-quartile fund IRRs can be 20 percentage points.
- Liquidity in private equity is typically very limited, and thus LPs are locked in for the long term. However, when private equity funds exit an investment, they return the cash to the investors immediately. Therefore, the duration of an investment in private equity is typically shorter than the maximum life of the fund.

Standard due diligence questionnaires (DDQs) have been developed by numerous international and country trade organizations. The Institutional Limited Partners Association (ILPA) publishes and makes available a DDQ on its website. These guides are not a substitute for the due diligence process that is conducted by LPs before investing in a venture capital or private equity fund.

8.4 Private Equity Fund Valuation

The description of private equity valuation in a fund prospectus is generally associated with the fund's calculation of net asset value (NAV), which itself is generally defined as the value of the fund assets less liabilities corresponding to the accrued expenses of the fund. The fund's assets are frequently valued by GPs, depending on their valuation policies, in the following ways:

1 At cost with significant adjustments for subsequent financing events or deterioration

2 At the lower of cost or market value

3 By the revaluation of a portfolio company whenever a new financing round involving new investors takes place

4 At cost with no interim adjustment until the exit

5 With a discount for restricted securities—for example, Reg. 144 securities

6 More rarely, marked to market by reference to a peer group of public comparables and applying illiquidity discounts

Private equity valuation standards, such as those originally produced by British, French, and European industry associations, have been adopted by funds operating in numerous jurisdictions.

Because the fund's valuation is adjusted with each new round of financing, the NAV may be more stale in down markets when there is a long gap between funding rounds. This mechanism is similar to the valuation of investment funds of publicly quoted securities. There is thus a fundamental implicit break-up assumption whereby the fund may be broken up at any time, the funds underlying investments may be liquidated individually and immediately, and the proceeds returned to LPs. Whereas this fundamental break-up assumption may hold for publicly traded securities, which are marked to market, this assumption may be more questionable for private equity investment portfolios typically held over a long period of time. At what value should investments in portfolio companies be reported prior to the private equity fund exiting the investment and returning the proceeds to the LPs? There is no clear answer to that question, because there is no market for securities issued by private equity companies.

Undrawn LP commitments represent legal obligations to meet capital calls in the future. They are not accounted for in the NAV calculation and should be viewed as unfunded liabilities by each LP for as long as they are callable.

Given that PE funds have different investment strategies, an understanding of their respective valuation policies will prevent biases. Whereas an early-stage VC fund may record its investments at cost, a late-stage VC fund may mark its portfolio companies to market by reference to public market comparables. When market bubbles form, as they did in the year leading up to the GFC, public market comparables may distort the valuation of portfolio companies and thus reported fund returns.

Private equity valuations are mostly performed by GPs. Under pressure from LPs, an increasing number of annual and semi-annual valuations are performed by independent valuers that are mandated by GPs. Although auditors sign off on annual results, their responsibility does not extend much beyond testing the reasonableness and allocation of the GP's model for illiquid investment values.

EVALUATING FUND PERFORMANCE AND CONCEPT IN ACTION: EVALUATING A PRIVATE EQUITY FUND 9

h interpret and compare financial performance of private equity funds from the perspective of an investor;

i calculate management fees, carried interest, net asset value, distributed to paid in (DPI), residual value to paid in (RVPI), and total value to paid in (TVPI) of a private equity fund.

Because each private equity fund is unique, assessing financial performance depends on good knowledge of a fund's objectives, structure, terms, and valuation policies. Typically an analysis of a private equity fund's financial performance includes a detailed examination of each investment's and the fund's return using IRR and multiples.

9.1 Analysis of IRR since Inception

Here, *net of fees* means net of management fees, carried interest, or any other financial arrangements that accrue to the GP. The IRR, a cash flow–weighted rate of return, is deemed the most appropriate measure of private equity performance by the Global Investment Performance Standards (GIPS), by the International Private Equity and Venture Capital Valuation Guidelines (2018 Guidelines), and by other venture capital and private equity standards. The interpretation of IRR in private equity is subject to caution, however, because an implicit assumption behind the IRR calculation is that the fund is fully liquid, whereas a significant portion of the fund's underlying investments is illiquid during much of a private equity fund's life. Therefore, valuation of portfolio companies according to industry standards is important to ensure the quality of the IRR figures.

Gross IRR is a function of the cash flows between the portfolio companies and the PE fund and is often considered a good measure of the PE firm's track record in creating value. Net IRR is a function of the cash flows between the PE fund and LPs, capturing the returns to investors. Fees and profit sharing create significant deviations between gross and net IRR. IRR analysis is often combined with a benchmark IRR analysis—that is, the median IRR for the relevant peer group of comparable private equity funds operating with a similar investment strategy and vintage year. This is particularly important because there are clear trends in private equity returns, with some vintage years producing much higher returns than others.

Despite the widespread prevalence of IRR as a performance measure and a hurdle for GP profit participation, it has its drawbacks. One drawback is that it can be easily manipulated. Imagine that a GP delays investor capital calls by using a fund's line of credit or bridge financing to make an initial investment instead or that the GP refinances an investment in order to return capital to investors prior to an exit. Both strategies reduce the time that investors' capital is outstanding, which boosts IRR independently of the investment holding period. In some cases, the only reason the GP may be entitled to its carried interest in the end is the manipulation and timing of leverage. Another problem with IRR is that it provides no information about the size of the return.

9.2 Analysis of Return Multiples

Return multiples simply measure the total return to investors relative to the total sum invested. Although multiples ignore the time value of money, their ease of calculation and their ability to differentiate between "realized" actual proceeds from divestments and the "unrealized" portfolio subject to the GP's valuation make these ratios very popular among LPs. The return multiples used most frequently by LPs and also defined by GIPS that provide additional information about fund performance are as follows:

- PIC (paid-in capital): The ratio of paid-in capital, which is the proportion of the LPs' total committed capital that the GP has so far deployed following any capital calls, to total committed capital.
- DPI (distributed to paid-in): The ratio of the cumulative distributions, which is the amount of cash and stock that has already been paid out to LPs from the fund, to the paid-in capital. This ratio indicates the fund's realized return on investment and is often called the cash-on-cash return. DPI is presented net of management fees and carried interest.
- RVPI (residual value to paid-in): This is the value of LPs' shareholding held with the private equity fund as a proportion of the cumulative invested capital (i.e., the paid-in capital). The numerator is the value that the GP assigns to the remaining portfolio companies. This ratio is a measure of the private equity fund's unrealized return on investment. RVPI is presented net of management fees and carried interest.
- TVPI (total value to paid-in): This ratio is the portfolio companies' distributed (or realized) and undistributed (or unrealized) value as a proportion of the cumulative invested capital. TVPI is the sum of DPI and RVPI and is presented net of management fees and carried interest.

In addition to quantitative measures of return, an analysis of fund financial performance includes

- an analysis of realized investments since inception, with comments on all successes and failures;
- an analysis of unrealized investments, highlighting all red flags in the portfolio and the expected time to exit for each portfolio company;
- a cash flow forecast for each portfolio company and for the aggregate portfolio; and
- an analysis of portfolio valuation, audited financial statements, and the NAV.

EXAMPLE 3

Calculating and Interpreting Private Equity Fund Performance

Suppose that a private equity fund has a DPI of 0.07 and an RVPI of 0.62 after five years. The IRR is –17%. The fund follows a venture capital strategy in high technology and has a vintage year of 2006 and a term of 10 years. A DPI of 7% indicates that few successful exits were made. An RVPI of 62% points to an extended J-curve effect for the fund, as TVPI amounts to 69% at the midlife of the fund. A vintage year of 2006 hints that the fund was started before the 2007 global financial crisis and that the routes to exit for portfolio companies have been dramatically changed. During the financial crisis, the investment portfolio probably suffered a number of complete write-offs. LPs should thus consider the state of the portfolio and examine the number of write-offs and other signals of ailing companies. The risk of not recovering the invested amount at termination of the fund is significant. The GP's compliance with valuation policies should also be closely monitored by LPs in order to ensure that the GP's expectations are not excessive given the state of the portfolio.

With increased allocations to private equity, performance comparisons across asset classes are often misinterpreted. The IRR, a standard measure of private equity returns, is cash flow weighted, whereas the performance of most other asset classes is measured in terms of a time-weighted rate of return.

There have been ongoing attempts to solve performance-comparison issues. The Public Market Equivalent (PME) compares a fund's IRR to a public market index (e.g., the MSCI World, the S&P 500 Index, and the FTSE All-Share Index) by assuming the fund's cash flows were invested and disinvested in the public index at the same amount and time. The PME, sometimes referred to as the Long Nickels PME for its developers Austin Long and Craig Nickels or as the Index Comparison Method (ICM), often calculates the IRR based on public index values. The PME facilitates a direct comparison of the fund's IRR to public markets, but it breaks down when the fund significantly outperforms the public index, resulting in a negative NAV and IRR. The PME+ (developed by Christophe Rouvinez in 2003), the modified PME (created by Cambridge Associates in 2013), and several other approaches attempt to address the shortcomings of the Long Nickels PME.

9.3 Concept in Action: Evaluating Private Equity Fund Performance

Michael Hornsby, CFA, is a senior investment officer at Icarus, a UK-based institutional investor in private equity. He is contemplating an investment in Europa Venture Partners III, a new late-stage technology venture capital fund, after thorough due diligence was performed both on the fund and on the GP. Icarus has been an investor in Europa Venture Partners' (EVPs') previous two funds, EVP I and EVP II, and has been satisfied with performance so far. Icarus is seeking to further expand its relationship with this GP because it sees it as a niche venture capital firm operating in a less crowded segment of the pan-European technology markets. In light of its past success, EVP is increasing its carried interest for the third fund to 25% from 20% for the previous two funds. Hornsby has received information about the fund's financial performance and is seeking assistance in calculating and interpreting financial performance for a number of specific queries as outlined below.

Europa Venture Partners (EVP)

General Partner: Europa Venture Partners (EVP) was established to provide equity financing to later-stage European technology companies in need of development capital. The GP seeks to provide strategic support to seasoned entrepreneurial teams and to bring proven new technologies to the market. The GP targets investment in portfolio companies between €2 million and €10 million.

	Established in 2012		Type: Development Capital					
Fund	**Vintage**	**Actual Fund Size (€ Millions)**	**Capital Called (%)**	**Mgmt. Fees (%)**	**Carried Interest (%)**	**Hurdle Rate (%)**	**Term**	**Report Date**
EVP I	2014	125	92	2	20	8	2012	31 Dec 2019
EVP II	2016	360	48	2	20	8	2025	31 Dec 2019

The financial performance for Icarus' investments in EVP funds follows.

Fund	Committed Capital (€ Millions)	Capital Called Down (€ Millions)	Gross IRR (%)	Net IRR (%)	DPI (×)	RVPI (×)	TVPI (×)	Quartile
EVP I	10	9.2	16.1	11.3	1.26	1.29	2.55	1
EVP II	25	12.0	1.6	(0.4)	0.35	1.13	1.48	2

Hornsby is also interested in verifying management fees, carried interest, and the NAV of EVP I. He has the following information about yearly capital calls (assumed to occur on 1 January of the given year), operating results, and annual distributions (as of 31 December of the given year).

Calls, Operating Results, and Distributions (€ Millions)

	2014	2015	2016	2017	2018	2019
Called down	50	15	10	25	10	5
Realized results	0	0	10	35	40	80

(Continued)

	2014	2015	2016	2017	2018	2019
Unrealized results	–5	–15	15	10	15	25
Distributions	—	—	—	25	45	75

Operating results are the sum of realized results from exiting portfolio companies and unrealized results from the revaluation of investments presently held in portfolio companies. In addition to the information available on EVP I, Hornsby also knows from the fund prospectus that the distribution waterfall is calculated according to the total return method, in which the GP receives carried interest only after the fund has returned the entire committed capital to LPs. Management fees are calculated on the basis of the paid-in capital. Hornsby also wants to calculate DPI, RVPI, and TVPI of EVP I for 2019, and he is interested in understanding how to calculate gross and net IRRs.

1 Interpret and compare the financial performance of EVP I and EVP II.

2 Calculate the management fees, the carried interest, and the NAV of EVP I. Also calculate DPI, RVPI, and TVPI of EVP I for 2019. Explain on the basis of EVP I how gross and net IRRs are calculated.

Solution to 1

In the table above, the first venture capital fund (EVP I) made its initial capital call in 2014 and returned to LPs €1.26 (all amounts in millions) for every €1 that had been drawn down two years ahead of the termination of the fund. EVP I residual value remains high, at 1.29 times capital drawn down, which is a good signal about the profitability of the fund at termination. The fund ranks in the first quartile, which means it belongs to the best-performing funds of that category and vintage year. Gross IRR of 16.1% after six years of operations and 11.3% net of fees represents good performance.

The second fund exhibits very modest performance to date in terms of gross and net IRR, which indicates that the fund is still experiencing the J-curve effect. EVP II has returned to LPs 35% of capital drawn down and a residual value of 113% of the capital drawn down, which indicates that despite the fund being in its early years, the GP has already managed a number of profitable exits and increased the value of the investment portfolio halfway through the termination of the fund. Actual fund size significantly exceeds previous fund size and is an indication that the GP is gaining momentum in terms of fundraising, probably partly attributable to the strong performance of the first fund.

Solution to 2

Cash Flows and Distributions (€ Millions)

Year	Called Down (1)	Paid-In Capital (2)	Mgmt. Fees (3)	Operating Results (4)	NAV before Distributions (5)	Carried Interest (6)	Distributions (7)	NAV after Distributions (8)
2014	50	50	1.0	–5	44.0			44.0
2015	15	65	1.3	–15	42.7			42.7

(continued)

(Continued)

Year	Called Down (1)	Paid-In Capital (2)	Mgmt. Fees (3)	Operating Results (4)	NAV before Distributions (5)	Carried Interest (6)	Distributions (7)	NAV after Distributions (8)
2016	10	75	1.5	25	76.2			76.2
2017	25	100	2.0	45	144.2	3.8	25	115.4
2018	10	110	2.2	55	178.2	6.8	45	126.4
2019	5	115	2.3	105	234.1	11.2	75	147.9

Based on this table, the calculations of DPI, RVPI, and TVPI can be derived as follows:

- Paid-in capital = Cumulative capital called down (Column 2).
- Management fees = (2%) × (Column 2).
- Carried interest: The first year that NAV is higher than committed capital (€125 million), carried interest is 20% of the excess, or (20%) [(NAV in Column 5) − €125 million]. Thereafter, provided that NAV before distribution exceeds committed capital, carried interest is (20%) × (Increase in NAV before distributions). Carried interest in 2019 is calculated as follows: (20%) × (€234.1 million − €178.2 million).
- NAV before distributions = NAV after distributions$_{t-1}$ + (Column 1) − (Column 3) + (Column 4).
- NAV after distributions = (Column 5) − (Column 6) − (Column 7).
- DPI = (€25 + €45 + €75)/€115, or 1.26×.
- RVPI = €147.9/€115, or 1.29×.
- TVPI = 1.26× + 1.29× = 2.55×.

The IRRs may be developed as follows:

- Gross IRRs are estimated by calculating the internal rate of return for the following cash flows: called-down capital at the beginning of period (Column 1) and the previous year's operating results (Column 4).
- Net IRRs are estimated by calculating the internal rate of return for the following cash flows: called-down capital at the beginning of period (Column 1) and the previous year's operating results (Column 4), net of management fees (Column 3) and carried interest (Column 6). Cash flows for gross and net IRRs are shown in the following table. Gross IRR and net IRR are shown in the bottom row.

Year End	Called Down	Operating Results	Cash Flows Gross IRR	Mgmt. Fees	Carried Interest	Cash Flows Net IRR
2013	50		−50			−50.0
2014	15	−5	−20	1.0		−21.0
2015	10	−15	−25	1.3		−26.3
2016	25	25	0	1.5		−1.5
2017	10	45	35	2.0	3.8	29.2
2018	5	55	50	2.2	6.8	41.0

Year End	Called Down	Operating Results	Cash Flows Gross IRR	Mgmt. Fees	Carried Interest	Cash Flows Net IRR
2019		105	105	2.3	11.2	91.5
IRR			16.1%			11.3%

SUMMARY

- Private equity funds seek to add value by various means, including optimizing financial structures, incentivizing management, and creating operational improvements.
- Private equity can be thought of as an alternative system of governance for corporations: Rather than ownership and control being separated as in most publicly quoted companies, private equity concentrates ownership and control. Many view the combination of ownership and control as a fundamental source of the returns earned by the best private equity funds.
- A critical role for the GP is valuation of potential investments. But because these investments are usually privately owned, valuation encounters many challenges.
- Valuation techniques differ according to the nature of the investment. Early-stage ventures require very different techniques than leveraged buyouts. Private equity professionals tend to use multiple techniques when performing a valuation, and they explore many different scenarios for the future development of the business.
- In buyouts, the availability of debt financing can have a big impact on the scale of private equity activity, and it seems to impact valuations observed in the market.
- Because private equity funds are incentivized to acquire, add value, and then exit within the lifetime of the fund, they are considered buy-to-sell investors. Planning the exit route for the investment is a critical role for the GP, and a well-timed and well-executed investment can be a significant source of realized value.
- In addition to the problems encountered by the private equity funds in valuing potential portfolio investments, challenges exist in valuing the investment portfolio on an ongoing basis. This is because the investments have no easily observed market value and there is a large element of judgment involved in valuing each of the portfolio companies prior to their sale by the fund.
- The two main metrics for measuring the ongoing and ultimate performance of private equity funds are IRR and multiples. Comparisons of PE returns across funds and with other assets are demanding because it is important to control for the timing of cash flows, differences in risk and portfolio composition, and vintage-year effects.

REFERENCES

Jensen, M. 1986. "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review* 76 (2): 323–29.

Jensen, M. 1989. "Eclipse of the Public Corporation." *Harvard Business Review* 67 (September–October).

Modigliani, F., and M. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review* 48 (3): 261–97.

Rouvinez, C. 2003. "*Private Equity Benchmarking with PME+*." Venture Capital Journal (1 August).

PRACTICE PROBLEMS

1 Jo Ann Ng is a senior analyst at SING INVEST, a large regional mid-market buyout manager in Singapore. She is considering the exit possibilities for an existing investment in a mature automotive parts manufacturer that was acquired three years ago at a multiple of 7.5 times EBITDA. SING INVEST originally anticipated exiting its investment in China Auto Parts, Inc., within three to six years. Ng noted that market conditions have deteriorated and that companies operating in a similar business trade at an average multiple of 5.5 times EBITDA. She expects, however, based on analyst reports and industry knowledge, that the market will recover strongly within the next two years because of the fast-increasing demand for cars in emerging markets. Upon review of market opportunities, Ng also noted that China Gear Box, Inc., a smaller Chinese automotive parts manufacturer that presents strong potential synergies with China Auto Parts, Inc., is available for sale at an EBITDA multiple of 4.5. Exits by means of an IPO or a trade sale to a financial or strategic (company) buyer are possible in China. How would you advise Ng to enhance value upon exit of China Auto Parts?

2 Wenda Lee, CFA, is a portfolio manager at a UK-based private equity institutional investor. She is considering an investment in a mid-market European buyout fund to achieve better diversification for her firm's private equity portfolio. She short-listed two funds that she sees as having similar risk–return profiles. Before deciding which one to invest in, she is carefully reviewing and comparing the terms of each fund.

	Mid-Market Fund A	Mid-Market Fund B
Management fees	2.5%	1.5%
Transaction fees	100% to the GP	50–50% split
Carried interest	15%	20%
Hurdle rate	6%	9%
Clawback provision	No	Yes
Distribution waterfall	Deal-by-deal	Total return

Based on the analysis of terms, which fund would you recommend to Lee?

3 Jean-Pierre Dupont is the chief investment officer (CIO) of a French pension fund that allocates a substantial portion of its assets to private equity. The fund's PE portfolio comprises mainly large buyout funds and mezzanine funds with a limited allocation to a special situations fund. A decision has been made to increase allocations to European venture capital. The investment committee of the pension fund requested that Dupont present an analysis of five key investment characteristics specific to venture capital relative to buyout investing. Can you assist Dupont in this request?

4 Discuss the ways that private equity funds can create value.

© 2018 CFA Institute. All rights reserved.

5 What problems are encountered when using comparable publicly traded companies to value private acquisition targets?

6 What are the main ways that the performance of private equity limited partnerships can be measured (A) during the life of the fund and (B) once all investments have been exited?

The following information relates to Questions 7–12

Martha Brady is the CIO of the Upper Darby County (UDC) public employees' pension system. Brady is considering an allocation of the pension system's assets to private equity. She has asked two of her analysts, Jennifer Chau, CFA, and Matthew Hermansky, to provide more information about the workings of the private equity market.

Brady recognizes that the private equity asset class covers a broad spectrum of equity investments that are not traded in public markets. She asks Chau to describe the major differences between assets within this asset class. Chau notes that private equity ranges from venture capital financing of early-stage companies to complete buyouts of large publicly traded or even privately held companies. Chau describes some of the characteristics of venture capital and buyout investments.

Chau mentions that private equity firms take care to align the economic interests of the managers of the investments they control with their own. Various contractual clauses are inserted in the compensation contracts of the management team in order to reward or punish managers who meet or do not meet agreed-upon target objectives.

One concern Chau highlights is the illiquidity of private equity investments over time. Some funds are returned to investors, however, over the life of the fund because a number of investment opportunities are exited early. Provisions in a fund's prospectus describe the distribution of returns to investors, some of which favor the limited partners. One such provision is the distribution waterfall mechanism that provides distributions to limited partners (LPs) before the general partner (GP) receives the carried interest. This distribution mechanism is called the total return waterfall.

Chau prepares the following data to illustrate the distribution waterfall mechanism and the funds provided to limited partners when a private equity fund with a zero hurdle rate exits from its first three projects during a three-year period.

Exhibit 1 Investment Returns and Distribution Waterfalls

Private equity committed capital	$400 million
Carried interest	20%
First project investment capital	$20 million
Second project investment capital	$45 million
Third project investment capital	$50 million
Proceeds from first project	$25 million
Proceeds from second project	$35 million
Proceeds from third project	$65 million

Chau cautions that investors must understand the terminology used to describe the performance of private equity funds. Interpretation of performance numbers should be made with the awareness that much of the fund assets are illiquid during a substantial part of the fund's life. She provides the latest data in Exhibit 2 for the Alpha, Beta, and Gamma Funds, diversified high-technology venture capital funds formed five years ago, each with five years remaining to termination.

Exhibit 2 Financial Performance of Alpha, Beta, and Gamma Funds

Fund	PIC	DPI	RVPI
Alpha	0.30	0.10	0.65
Beta	0.85	0.10	1.25
Gamma	0.85	1.25	0.75

Chau studies the data and comments,

Of the three funds, the Alpha Fund has the best chance to outperform over the remaining life. First, it's because the management has earned such a relatively high residual value on capital and will be able to earn a high return on the remaining funds called down. At termination, the RVPI will be double the 0.65 value when the rest of the funds are called down. Second, its cash-on-cash return as measured by DPI is already as high as that of the Beta Fund. The PIC (or paid-in capital) ratio indicates the proportion of capital already called by the GP. The PIC of Alpha is relatively low relative to Beta and Gamma.

Hermansky notes that a private equity fund's ability to properly plan and execute its exit from an investment is vital for the fund's success. Venture funds, such as Alpha, Beta, and Gamma, take special care to plan their exits.

Brady then asks the analysts what procedures private equity firms would use to value investments in their portfolios as well as investments that are added later. She is concerned about buying into a fund with existing assets that do not have public market prices that can be used to ascertain value. In such cases, she worries, what if a GP overvalues the assets and new investors in the fund pay more for the fund assets than they are worth?

Hermansky makes three statements regarding the valuation methods used in private equity transactions during the early stages of selling a fund to investors.

Statement 1 For venture capital investment in the early stages of analysis, emphasis is placed on the discounted cash flow approach to valuation.

Statement 2 For buyout investments, income-based approaches are used frequently as a primary method of valuation.

Statement 3 If a comparable group of companies exist, multiples of revenues or earnings are used frequently to derive a value for venture capital investments.

7 The characteristic that is *most likely* common to both the venture capital and buyout private equity investment is:

A measurable and assessable risk.

B the extensive use of financial leverage.

C the strength of the individual track record and ability of members of management.

8 The contractual term enabling management of the private equity–controlled company to be rewarded with increased equity ownership as a result of meeting performance targets is called:

A a ratchet.

B the tag-along right.

C the clawback provision.

9 For the projects described in Exhibit 1, under a deal-by-deal method with a clawback provision and true-up every three years, the cumulative dollar amount the GP receives by the end of the three years is equal to:

A 1 million.

B 2 million.

C 3 million.

10 Are Chau's two reasons for interpreting Alpha Fund as the best-performing fund over the remaining life correct?

A No

B Yes

C The first reason is correct, but the second reason is incorrect.

11 The exit route for a venture capital investment is *least likely* to be in the form of a(n):

A initial public offering (IPO).

B sale to other venture funds targeting the same sector.

C buyout by the management of the venture investment.

12 Which statement by Hermansky is the *least* valid?

A Statement 1

B Statement 2

C Statement 3

The following information relates to questions 13–18

Daniel Collin is a junior analyst at JRR Equity Partners (JRR), a private equity firm. Collin is assigned to work with Susan Tseng, a senior portfolio manager. Tseng and Collin meet to discuss existing and potential investments.

Tseng starts the meeting with a discussion of LBO firms and VC firms. Collin tells Tseng,

> LBO firms tend to invest in companies with predictable cash flows and experienced management teams, whereas VC firms tend to invest in companies with high EBITDA or EBIT growth and where an exit is fairly predictable.

Tseng and Collin next analyze a potential investment in the leveraged buyout of Stoneham Industries. Specifically, they assess the expected gain if they elect to purchase all of the preference shares and 90% of the common equity through the LBO. Details of the LBO include the following:

- The buyout requires an initial investment of $10 million.
- Financing for the deal includes $6 million in debt, $3.6 million in preference shares that promise a 15% annual return paid at exit, and $0.4 million in common equity.

The expected exit value in six years is $15 million, with an estimated reduction in debt of $2.8 million over the six years prior to exit.

Tseng and Collin next discuss JRR's investment in Venture Holdings, a private equity fund. Selected details on the Venture Holdings fund include the following:

- Total committed capital is $115 million.
- The distribution waterfall follows the deal-by-deal method, and carried interest is 20%.
- On its first exit event a few years ago, the fund generated a $10 million profit.
- At the end of the most recent year, cumulative paid-in capital was $98 million, cumulative distributions paid out to LPs were $28 million, and the year-end NAV, before and after distributions, was $170.52 million and $131.42 million, respectively.
- Tseng and Collin estimate that the fund's NAV before distributions will be $242.32 million at the end of next year.

Finally, Tseng and Collin evaluate two venture capital funds for potential investment: the Squire Fund and the Treble Fund. Both funds are in Year 7 of an estimated 10-year term. Selected data for the two funds are presented in Exhibit 1.

Exhibit 1 Selected Data for the Squire Fund and the Treble Fund

	Squire Fund	Treble Fund
DPI	0.11	0.55
RVPI	0.95	0.51
Gross IRR	–11%	10%
Net IRR	–20%	8%

After reviewing the performance data in Exhibit 1, Collin draws the following conclusions:

Conclusion 1 The unrealized return on investment for the Squire Fund is greater than the unrealized return on investment for the Treble Fund.

Conclusion 2 The TVPI for the Treble Fund is higher than the TVPI for the Squire Fund because the Treble Fund has a higher gross IRR.

13 Is Collin's statement about LBO firms and VC firms correct?

A Yes

B No, because he is wrong with respect to VC firms

C No, because he is wrong with respect to LBO firms

14 The multiple of expected proceeds at exit to invested funds for JRR's Stoneham LBO investment is *closest* to:

A 2.77.

B 2.89.

C 2.98.

15 The distribution available to the limited partners of the Venture Holdings fund from the first exit is *closest* to:

A \$2 million.

B \$8 million.

C \$10 million.

16 At the end of the most recent year, the ratio of total value to paid-in capital (TVPI) for the Venture Holdings fund was *closest* to:

A 0.29.

B 1.34.

C 1.63.

17 Based on Tseng and Collin's estimate of NAV next year, the estimate of carried interest next year is *closest* to:

A \$14.36 million.

B \$22 million.

C \$25.46 million.

18 Which of Collin's conclusions regarding the Squire Fund and the Treble Fund is correct?

A Only Conclusion 1

B Only Conclusion 2

C Both Conclusion 1 and Conclusion 2

SOLUTIONS

1 The exit strategies available to SING INVEST to divest their holding in China Auto Parts, Inc., will largely depend on the following two factors:

- Time remaining until the fund's term expires: If the time remaining is sufficiently long, the fund's manager has more flexibility to work out an exit at more favorable market circumstances and terms.
- Amount of undrawn commitments from LPs in the fund: If sufficient LP commitments can be drawn, the fund manager may take advantage of current investment opportunities at depressed market prices to enhance returns upon exit in a more favorable market environment.

In the case of China Auto Parts, Inc., depending on an analysis of the factors discussed, Ng could offer an opinion to support the acquisition of China Gear Box, Inc., subject to an in-depth analysis of potential synergies with China Auto Parts, Inc. The objective here may thus be twofold: to benefit from short-term market conditions and to enhance the value of existing investments by reinforcing their market potential with a strategic merger.

2 Assuming that both funds have similar risk–return characteristics, a closer analysis of economic and corporate governance terms should be instrumental in determining which fund to select.

In economic terms, Mid-Market Fund B has a higher carried interest than Mid-Market Fund A, but Mid-Market Fund B has a fee structure that is better aligned with the interests of LPs. A larger proportion of Mid-Market Fund B's fees (through the carried interest) will come from achieving successful exits, whereas Mid-Market Fund A will earn relatively larger fees on running the fund (management fees and transaction fees) without necessarily achieving high performance. In addition, the 9% hurdle rate of Mid-Market Fund B is indicative of confidence in the fund manager's ability to achieve a minimum compounded 9% return to LPs for which no carried interest will be paid.

In corporate governance terms, Mid-Market Fund B is far better aligned with the interests of LPs as a result of a clawback provision and a more favorable distribution waterfall that will allow payment of carried interest on a total return basis instead of deal by deal.

The conclusion is that Mid-Market Fund B appears better aligned with the interests of LPs.

3

Venture Capital	Buyout
Primarily equity funded. Use of leverage is rare and very limited.	Extensive use of leverage consisting of a large proportion of senior debt and a significant layer of junior and/or mezzanine debt.
Returns of investment portfolios are generally characterized by very high returns from a limited number of highly successful investments and a significant number of write-offs from low performing investments or failures.	Returns of investment portfolios are generally characterized by lower variance across returns from underlying investments. Bankruptcies are rare events.

(continued)

Venture Capital	Buyout
Venture capital firm monitors achievement of milestones defined in business plan and growth management.	Buyout firm monitors cash flow management and strategic and business planning.
Expanding capital requirement if in the growth phase.	Low working capital requirement.
Assessment of risk is difficult because of new technologies, new markets, and lack of operating history.	Risk is measurable (e.g., mature businesses, long operating history, etc.).

4 The main ways that private equity funds can create value include the following:

- Operational improvements and clearly defined strategies: In the case of later-stage companies and buyouts, private equity owners can often create value by focusing the business on its most profitable opportunities and providing strategic direction for the business. In the case of venture capital deals, the private equity funds provide valuable business experience, mentor management, and offer access to their network of contacts and other portfolio companies.
- Creating incentives for managers and aligning their goals with the investors: This is often achieved by providing significant monetary rewards to management if the private equity fund secures a profitable exit. In the case of buyouts, the free cash flow available to management is minimized by taking on significant amounts of debt financing.
- Optimizing the financial structure of the company: In the case of buyouts, the use of debt can reduce the tax payments made by the company and also reduce the cost of capital. There may also be opportunities in certain market conditions to take advantage of any mispricing of risk by lenders, which can allow the private equity funds to take advantage of interest rates that do not fully reflect the risks being carried by the lenders. Many would point to various periods from 2015 to 2019 when government interest rates were low, debt spreads were tight, and/or lender covenants were loose as examples of such prevailing conditions.

5 There are many complexities in using comparable companies to value private targets, including the following:

- The lack of public comparison companies operating in the same business, facing the same risks, and at the same stage of development. It is often possible to identify "approximate" comparisons but very rare to find an exact match. It is essential, therefore, to use judgment when using comparison company information, rather than just taking the average multiples derived from a sample of disparate companies.
- Comparison companies may have different capital structures, so estimated beta coefficients and some financial ratios should be adjusted accordingly.
- Reported accounting numbers for earnings must be chosen carefully and adjusted for any exceptional items, atypical revenues, and costs in the reference year. Care must also be taken to decide which earnings figures to compare; the main choices are trailing earnings (the last 12 months), earnings from the last audited accounts, or prospective year-ahead earnings.

6 In the early years of a fund, all measures of return are of little relevance because fees drag down the reported returns and investments are initially valued at cost. This produces the J-curve effect. After a few years (or longer in the case of venture capital investments), performance measures become more meaningful,

and the two main measures used by investors are IRR and return multiples (of the initial sum invested). During the life of the fund, it is necessary to value the non-exited investments and add them to the realized returns. The former inevitably involves an element of judgment on the part of the General Partner, especially when it is difficult to estimate the likely market value of the investment. Once all the investments have been exited, the multiples and IRR can be estimated easily, taking account of the exact timing of the cash flows into and out of the fund. The most relevant measures for investors are computed net of management fees and any carried interest earned by the general partner.

7 C is correct. Members of both the firm being bought out and the venture capital investment usually have strong individual management track records. Extensive financial leverage is common in buyouts but not venture capital investments, whereas measurable risk is more common in buyouts than in venture capital situations.

8 A is correct.

9 B is correct. On a cumulative basis for three years, the fund earns \$10 million, of which \$2 million goes to the GP. The \$2 million earned by the GP corresponds to 20% of the difference between total three-year proceeds and three-year invested capital, or $0.2 \times [(25 + 35 + 65) - (20 + 45 + 50)]$.

10 A is correct. Chau misinterprets DPI, RVPI, and PIC. The returns earned to date are for each dollar of invested capital—that which has been drawn down—not total returns. Chau mistakenly believes (assuming the same management skill) the result for Alpha Fund at termination will be on the order of $3 \times 0.65 = 1.95$, instead of 0.65. In both cases, Alpha Fund has underperformed relative to the other two funds.

11 C is correct. Leverage needed to finance a management buyout is not readily available to firms with limited history.

12 A is correct. Statement 1 is the least likely to be valid.

13 B is correct. LBO firms generally invest in firms with a predictable cash flow pattern (EBITDA or EBIT growth) and experienced management teams. In contrast, venture capital firms tend to invest in new firms and new technologies with high revenue growth. Also, VC investments tend to be characterized as having exits that are difficult to anticipate.

14 B is correct. The investment exit value is \$15 million. The expected payoff to JRR is calculated as follows (all amounts in millions):

Expected exit value:	\$15.00
Debt remaining at exit: (\$6.0 – 2.8)	3.20
Preference shares: [$\$3.60 \times (1.15)^6$]	8.33
Common equity: (\$15 exit – 3.2 debt – 8.33 preference)	3.47

Initial investment: \$3.6 (preference) + 0.9 × \$0.4 (common) = \$3.96.

Proceeds at exit: \$8.33 (preference) + 0.9 × \$3.47 (common) = \$11.45.

Multiple of expected proceeds to invested funds: \$11.45 exit value/\$3.96 initial investment = 2.89×.

15 B is correct. The distribution waterfall for the Venture Holdings fund follows the deal-by-deal method. The investment generated a profit of \$10 million, and with carried interest of 20%, the general partner would receive \$2 million (\$10 million × 20%), leaving \$8 million for the limited partners.

16 C is correct. Total value to paid-in capital (TVPI) represents the fund's distributed value and undistributed value as a proportion of the cumulative invested capital. TVPI is the sum of distributed to paid-in capital (DPI) and residual value to paid-in capital (RVPI):

$$\text{DPI} = \frac{\text{Cumulative distributions}}{\text{Cumulative invested capital}} = \frac{\$28\text{ million}}{\$98\text{ million}} = 0.29\times$$

$$\text{RVPI} = \frac{\text{NAV (after distributions)}}{\text{Cumulative invested capital}} = \frac{\$131.42\text{ million}}{\$98\text{ million}} = 1.34\times$$

$$\text{TVPI} = \frac{\text{Cumulative distribution} + \text{NAV (after distributions)}}{\text{Cumulative invested capital}}$$

$$= \frac{\$28\text{ million} + 131.42\text{ million}}{\$98\text{ million}} = 1.63\times$$

17 A is correct. Provided that NAV before distribution exceeds committed capital, the general partner is entitled to carried interest, calculated as the 20% multiplied by the increase in NAV before distributions. So, the carried interest is calculated as follows:

Carried interest = 20% × (\$242.32 – \$170.52) = \$14.36 million.

18 A is correct. DPI provides an indication of a fund's realized return, whereas RVPI provides an indication of a fund's unrealized return. The Squire Fund has a higher RVPI (0.95) than the Treble Fund (0.51). TVPI, which is the sum of DPI and RVPI, is the same for both funds: 0.11 + 0.95 = 1.06 for the Squire Fund and 0.55 + 0.51 = 1.06 for the Treble Fund.

READING

37

Introduction to Commodities and Commodity Derivatives

by David Burkart, CFA, and James Alan Finnegan, CAIA, RMA, CFA

David Burkart, CFA, is at Coloma Capital Futures, LLC (USA). James Alan Finnegan, CAIA, RMA, CFA at American Century Investments (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** compare characteristics of commodity sectors;
☐	**b.** compare the life cycle of commodity sectors from production through trading or consumption;
☐	**c.** contrast the valuation of commodities with the valuation of equities and bonds;
☐	**d.** describe types of participants in commodity futures markets;
☐	**e.** analyze the relationship between spot prices and futures prices in markets in contango and markets in backwardation;
☐	**f.** compare theories of commodity futures returns;
☐	**g.** describe, calculate, and interpret the components of total return for a fully collateralized commodity futures contract;
☐	**h.** contrast roll return in markets in contango and markets in backwardation;
☐	**i.** describe how commodity swaps are used to obtain or modify exposure to commodities;
☐	**j.** describe how the construction of commodity indexes affects index returns.

INTRODUCTION

1

In the upcoming sections, we present the characteristics and valuation of commodities and commodity derivatives. Given that investment in commodities is conducted primarily through futures markets, the concepts and theories behind commodity

© 2020 CFA Institute. All rights reserved.

futures is a primary focus of the reading. In particular, the relationship between spot and futures prices, as well as the underlying components of futures returns, are key analytical considerations.

What do we mean when we talk about investing in commodities? A basic economic definition is that a commodity is a physical good attributable to a natural resource that is tradable and supplied without substantial differentiation by the general public.

Commodities trade in physical (spot) markets and in futures and forward markets. Spot markets involve the physical transfer of goods between buyers and sellers; prices in these markets reflect current (or very near term) supply and demand conditions. Global commodity futures markets constitute financial exchanges of standardized futures contracts in which a price is established in the market today for the sale of some defined quantity and quality of a commodity at a future date of delivery; completion of the contract may permit cash settlement or require physical delivery.

Commodity futures exchanges allow for risk transfer and provide a valuable price discovery mechanism that reflects the collective views of all market participants with regard to the future supply and demand prospects of a commodity. Given the financial (versus physical) nature of their contract execution, commodity exchanges allow important parties beyond traditional suppliers and buyers—speculators, arbitrageurs, private equity, endowments, and other institutional investors—to participate in these price discovery and risk transfer processes. Standardized contracts and organized exchanges also offer liquidity (i.e., trading volumes) to facilitate closing, reducing, expanding, or opening new hedges or exposures as circumstances change on a daily basis.

Forward markets exist alongside futures markets in certain commodities for use by entities that require customization in contract terms. Forwards are largely outside the scope of this reading and are discussed only briefly. Exposure to commodities is also traded in the swap markets for both speculative and hedging purposes. Investment managers may want to establish swap positions to match certain portfolio needs, whereas producers may want to more precisely adjust their commodity risk (e.g., the origin of their cattle or the chemical specifications of their crude oil).

Commodities offer the potential for diversification benefits in a multi-asset class portfolio because of historically low average return correlation with stocks and bonds. In addition, certain academic studies (e.g., Gorton and Rouwenhorst 2006; Erb and Harvey 2006) demonstrate that some commodities have historically had inflation hedging qualities.

Our coverage of the commodities topic is organized as follows: We provide an overview of physical commodity markets, including the major sectors, their life cycles, and their valuation. We then describe futures market participants, commodity futures pricing, and the analysis of commodity returns, including the concepts of contango and backwardation. The subsequent section reviews the use of swap instruments rather than futures to gain exposure to commodities. We then review the various commodity indexes given their importance as benchmarks for the asset class and investment vehicles. Finally, we conclude with a summary of the major points.

2 COMMODITY SECTORS

a compare characteristics of commodity sectors;

Commodities are an asset class inherently different from traditional financial assets, such as equities and bonds. These latter assets are securities that are claims on productive capital assets and/or financial assets and thus are expected to generate cash flows for their owners. The intrinsic value of these securities is the present discounted value of their expected future cash flows. Commodities are valued differently. Commodities' value derives from either their use as consumables or as inputs to the production of goods and services. Because a number of commodities need to be processed or have a limited life before spoiling or decaying, an astute analyst will take into account the growth and extraction patterns of the various commodities as well as the logistics associated with transporting these physical goods. Therefore, commodities, while seemingly familiar from everyday life, offer distinct sets of risk exposures for investors.

Fundamental analysis of commodities relies on analyzing supply and demand for each of the products as well as estimating the reaction to the inevitable shocks to their equilibrium or underlying direction. For example, a growing world population demands more crude oil or related products as transportation of goods and people increases. However, technological improvements (e.g., shale drilling or electric vehicles) can disrupt that trend and in the case of armed conflict or adverse weather, for example, may alter it on very short notice! This means that the quantitative analysis of commodities is often imperfect because of high degrees of non-normalcy and shifting correlations. Furthermore, the coefficients to underlying variables are often non-stationary; for example, much corn today is genetically modified to resist heat, rendering drought impact estimates derived from history less predictive. Much of the raw data are held off market by private firms engaged in the commodity industry (such as oil or agricultural companies), which also hinders a purely quantitative approach. Therefore, the framework offered here will be at a high level. We will later provide a breakdown of individual areas for the investor to apply discretionary or quantitative techniques, as circumstances allow. Because the framework can be applied to both supply and demand, we shall set that distinction aside until we focus on individual sectors and commodities. The tools and considerations in fundamental analysis are as follows:

a Direct announcements: Various government agencies and private companies broadcast production and inventory data that can be used to infer demand, which is often unobservable. Possible public sources include the USDA (US Department of Agriculture), OPEC (Organization of the Petroleum Exporting Countries), the NBS (National Bureau of Statistics of China), and the IEA (International Energy Agency). Setting aside questions of reliability, sometimes estimating current conditions is as straightforward as monitoring official announcements, even with a lag.

b Component analysis: The more diligent analyst will attempt to break down high-level supply and demand into various components. Applying a stock and flow approach is a logical method. The stock or potential production or demand attempts to set boundaries around what is actually produced or wanted. This can be as general as the amount of arable land in all of Europe or as specific as the current capacity of the Ghawar oil field in Saudi Arabia. The flow considers the utilization of that stock of raw material. Examples include understanding the oil tanker traffic heading to China, estimating the historical yields of US cotton (the amount of fiber per unit of land) in various weather conditions, and estimating the number of piglets per mother hog in Canada.

These examples lend themselves to historical quantitative or conditional analysis. However, care needs to be taken regarding the qualitative aspects of supply and demand; a new policy such as stricter emissions standards can affect both supply (higher standards often strand lower-quality materials) and demand (not all consumers may be properly equipped to utilize a changing standard). Political unrest may not touch an isolated farm but may disrupt consumption.

c Timing considerations: Stocks and flows from (b) can be further affected by timing issues—such as seasonality and logistics—and, therefore, price reaction. A shock, by definition, is a sudden timing switch; an earthquake that destroys a pipeline does not affect the stock, but it does halt the flow. A more common consideration is seasonality, such as the growing period for crops and people's demand for winter heat generated from natural gas. This last aspect in particular feeds into the shape of the commodity futures curve, as discussed later.

d Money flow: Short-term and long-term prices can be affected by sentiment and macro monetary conditions, such as inflation. If investor risk tolerance is particularly high or low, then expecting exaggerated price movements would be rational as fundamental conditions are hyped up or beaten down. Alternatively, capital availability from low interest rates can help trigger the building of new mines and affect future supply. Government subsidies of substitute technologies can limit commodity price appreciation (e.g., available funds for electric cars indirectly affect the price of gasoline).

In summary, although the casual investor can perhaps focus solely on public summary statements, the engaged researcher will apply a framework of examining the stock and flow components and their related timing to better understand and weigh the pressures leading to higher or lower prices.

2.1 Commodity Sectors

The world of commodities is relatively broad but can be defined and separated in a reasonable manner. Although there are several ways to segment the asset class by sector, here we use the approach that is the basis for the Bloomberg Commodity Index: energy, grains, industrial (base) metals, livestock, precious metals, and softs (cash crops). This segmentation is more granular than some other indexes but is reasonably consistent with the breakdown in the specialties of most market participants. As noted previously, each sector has a number of individual characteristics that are important in determining the supply and demand for each commodity. A key concept is how easily and cost-effectively the commodity can be produced and stored, as well as such related issues as frequency/timing of consumption, spoilage, insurance, and ease of transportation to consumers. Note that many commodities, such as uranium or water, are traded only in thin, private markets. They are really just individual transactions, as opposed to the markets we are discussing. For the purposes of our coverage, we have to constrain ourselves to primary commodities, recognizing that there are many others that may offer investment opportunities or require hedging. Exhibit 1 reviews each sector and its main characteristics and influences.

Exhibit 1 A Description of Commodity Sectors and Factors

	Energy: Fuel transportation, industrial production, and electrical generation. Primary commodities include crude oil, natural gas, coal, and refined products, such as gasoline and heating oil.	
Primary Influences	Stocks: Discovery and depletion of new fields, economic and political costs/certainty of access to those fields, refinery technology and maintenance, power plant type and construction, economic (GDP) size	Flows: Pipeline and tanker reliability, seasonality (summer/winter), adverse weather (cold, hurricanes), automobile/truck sales, geopolitical instability, environmental requirements, economic (GDP) growth
	Grains: Provide human and animal sustenance but also can be distilled into fuel (e.g., ethanol). Primary commodities include corn, soy, wheat, and rice.	

Exhibit 1 (Continued)

Primary Influences	Stocks: Arable farmland, storage/port facilities (infrastructure), human and animal population size	Flows: Weather (moisture, temperature), disease, consumer preferences, genetic modification, biofuel substitution, population growth
	Industrial/Base Metals: Materials for durable consumer goods, industry, and construction. Primary commodities include copper, aluminum, nickel, zinc, lead, tin, and iron.	
Primary Influences	Stocks: Mined acreage, smelter capacity, economic (GDP) stage of industrial/consumer development	Flows: Government industrial and environmental policies, economic (GDP) growth, automobile/truck sales, infrastructure investment
	Livestock: Animals raised for human consumption. Primary commodities include hogs, cattle, sheep, and poultry.	
Primary Influences	Stocks: Herd size, processing plant capacity, consumer preferences, feed availability/cost	Flows: Speed of maturation to slaughter weight, economic (GDP) growth/consumer income, disease, adverse weather
	Precious Metals: Certain metals that act as monetary stores of value (as well as industrial uses). Primary commodities include gold, silver, and platinum.	
Primary Influences	Stocks: Mined acreage, smelter capacity, fiat money supply/banking development	Flows: Central bank monetary policy, geopolitics, economic (GDP) growth
	Softs (Cash Crops): Crops sold for income—as opposed to consumed for subsistence—and often originally seen as luxuries. Primary commodities include cotton, cocoa, sugar, and coffee.	
Primary Influences	Stocks: Arable farmland, storage/port facilities (infrastructure), economic (GDP) size	Flows: Weather (moisture, temperature), disease, consumer preferences, biofuel substitution, economic (GDP) growth/consumer income

As noted in this section, each commodity sector is unique in its fundamental drivers but with the overlapping context of economic and monetary data. With this context in mind, we will now examine the life cycle of the sectors from production to consumption—and their interaction—in more detail.

EXAMPLE 1

Commodity Sector Demand

Industrial activity *most likely* affects the demand for which of the following commodities?

A Copper

B Natural gas

C Softs (e.g., cotton, coffee, sugar and cocoa)

Solution:

A is correct. Copper is used for construction, infrastructure development, and the manufacture of durable goods, all of which are economically sensitive. B is incorrect because demand for natural gas is driven primarily by weather conditions (heating or cooling) and only secondarily by industrial activity. C is incorrect because demand for softs is driven primarily by global income.

EXAMPLE 2

Commodity Sector Risks

Which of the following commodity sectors are *least* affected in the short term by weather-related risks?

A Energy

B Livestock

C Precious metals

Solution:

C is correct. Weather has very little impact on the availability of precious metals given their ease of storage. Inflation expectations, fund flows, and industrial production are more important factors. A is incorrect because energy demand is strongly influenced by weather (e.g., heating demand in the winter or transportation demand in the summer). B is incorrect because the health of livestock is vulnerable to unfavorable weather conditions increasing the risks of death and disease by extreme cold, wet, and heat.

3 LIFE CYCLE OF COMMODITIES

b compare the life cycle of commodity sectors from production through trading or consumption;

The life cycle of commodities varies considerably depending on the economic, technical, and structural (i.e., industry, value chain) profile of each commodity, as well as the sector. Conceptually, the commodity production life cycle reflects and amplifies the changes in storage, weather, and political/economic events that shift supply and demand. Recall from the earlier discussion that timing/seasonality is, in effect, an overlay on top of the underlying supply/demand factors. A short life cycle allows for relatively rapid adjustment to outside events, whereas a long life cycle generally limits the ability of supply or demand to react to new conditions. These shifts, in turn, feed into the economics for the valuation and shape of the commodity supply and demand curves, plus their respective price elasticities of demand and supply. Understanding the life cycle builds understanding of, and ideally ability to forecast, what drives market actions and commodity returns.

Among the food commodities, agriculture and livestock have well-defined seasons and growth cycles that are specific to geographic regions. For example, by March of each year, corn planting may be finished in the southern United States but not yet started in Canada. Meanwhile, the corn harvest may be underway in Brazil and Argentina given their reverse seasonal cycle in the Southern Hemisphere. Each geographic location also represents local markets that have different domestic and export demand. These differences affect the nature (level and reliability) of demand and the power of buyers to extend or contract the life cycle.

In comparison, commodities in the energy and metals sectors are extracted all year round. Their life cycle changes are generally at the margin of a continuous process, as opposed to being centered at a discrete time or season. But the products from crude oil and metal ore have seasonal demands depending on weather (e.g., gasoline demand in the summer and heating oil demand in the winter) that affect the life cycle and usage of the underlying commodity. And with all the differences between the varieties even within the same sector, the life cycles depicted have to be representative and selective. The life cycles of several key commodity sectors are as follows.

3.1 Energy

For an example of the differences within a sector, one need look no further than energy. Natural gas can be consumed almost immediately after extraction from the ground. Crude oil, in contrast, has to be transformed into something else; crude is useless in its innate form. The refined products (e.g., gasoline and heating oil), in turn, have a number of potential processing steps depending on the quality of crude oil input and the relative demand for the various products. The steps for the energy complex can be summarized as shown in Exhibit 2.

Exhibit 2 Steps for the Energy Complex

Step	Title	Description
1.	Extraction	A drilling location is selected after surveys, and the well is dug. Enough underground pressure for the hydrocarbons to come out naturally may exist, or water or other tools may be required to create such pressure. Water is also used for the fracturing process known as "fracking," which breaks up shale formations to allow for oil or gas to be extracted.
2.	Storage	After extraction, crude oil is commercially stored for a few months on average in the United States, Singapore, and northern Europe and is strategically stored by many countries. In addition, oil may temporarily be stored on tanker ships. Natural gas may be delivered directly to the end consumer. Summer-extracted natural gas is often injected into storage for the winter months.
3.	Consumption Stage	Only natural gas is consumed at this stage because it does not need to be refined. Crude oil requires further processing.
4.	Refining	Crude oil is distilled into its component parts via a process called "cracking." Heat is used to successively boil off the components that are, in turn, cooled down and collected (e.g., gasoline, kerosene), until only the remnants (e.g., asphalt) are left.
5.	Consumption Stage	The distilled products are separated and shipped to their various locations—by ship, pipe, train, or truck—for use by the end consumer.

Sources: Based on information from www.eia.gov/energyexplained/index.php?page=oil_refining#tab1, https://en.wikipedia.org/wiki/Petroleum_refining_processes (accessed 23 April 2019), and authors' research.

Refineries are extraordinarily expensive to build—typically costing several billion US dollars—depending on the processes required to purify and distill the oil. Part of the cost depends on the expected specifications of the crude oil input. Generally speaking, a low-grade, high sulfur source would require more investment than one with an assured lighter, "sweeter" source. Pipelines are also very costly: For example, the Keystone XL pipeline expansion between Canada and the United States was originally estimated to cost $5 billion in 2010, but the estimate was doubled to

$10 billion in 2014. Even in countries dealing with violent insurrections (e.g., Libya, Iraq, Nigeria), damage to refineries has been generally modest because of their value to all parties. Pipelines, however, are often destroyed or cut off. Although these costs may appear staggering, they actually pale in comparison with the costs (and risks) of oil exploration, especially in deep offshore locations or geographically remote (or geopolitically risky) regions.

The crude oil market has a number of futures contracts and indexes that follow local grades and origins, but the two most commonly traded set of contracts follow the US-based crude oil (West Texas Intermediate, or WTI, crude oil) and the UK-located Brent crude oil from the North Sea. Likewise, there are futures for natural gas, gasoil, gasoline, and heating oil. Each has different delivery locations and standards, but the WTI and Brent contracts represent a high-quality refinery input that exploration and production companies can use as a hedging device.

EXAMPLE 3

Energy Life Cycle

Which of the following is a primary difference in the production life cycle between crude oil and natural gas?

A Only crude oil needs to be stored.

B European companies are the only ones that store crude oil.

C Natural gas requires very little additional processing after extraction compared with crude oil.

Solution:

C is correct. Natural gas can be used after it is extracted from the ground upon delivery, but crude oil must first be processed for later use. A is incorrect because both oil and natural gas are stored before usage. B is incorrect because many countries around the world store crude oil, both commercially and strategically.

3.2 Industrial/Precious Metals

The life cycle of both precious and industrial metals is probably the most flexible because the ore, as well as the finished products, can be stored for months (if not years) given the relative resistance to spoilage of metals (assuming proper storage). Otherwise, the life cycle parallels the energy one outlined previously, as shown in Exhibit 3.

Exhibit 3 Copper Purification Process

	Step Name	Description
1.	Extracting and Preparing	Ore (raw earth with ~2% metal content) is removed via a mine or open pit. Ore is then ground into powder and concentrated to roughly 25% purity.
2.	Smelting	The purified ore is heated, and more impurities are removed as slag, increasing the metal content to 60%. Further processes increase the concentration to 99.99%.
3.	Storage/Logistics	The purified metal is held typically in a bonded warehouse until it is shipped to an end user.

Sources: Based on information from http://resources.schoolscience.co.uk/CDA/14-16/cumining/copch2pg1.html (accessed 23 April 2019), www.madehow.com/Volume-4/Copper.html (accessed 23 April 2019), and authors' research.

Similar to refining crude oil, creating the economies of scale involved in the smelter and ore processing plants is critical. These are huge facilities for which marginal costs (i.e., the cost to convert the last pound or kilogram of processed ore into a useful metal) decline substantially with both the scale of the facility and its utilization (output as a percentage of capacity). As a result, when supply exceeds demand for a given industrial metal, it is difficult for suppliers to either cut back production or halt it entirely. Overproduction often continues until smaller or financially weaker competitors are forced to shut down. Because demand for industrial metals fluctuates with overall economic growth, as was discussed previously, there are substantial incentives for metals producers to invest in new capacity when their utilization (and profit) is high but huge economic and financial penalties for operating these facilities when demand falls off during an economic downturn. Ironically, given the typical economic cycle and the time lag involved after deciding to expand capacity, new supply often arrives just as demand is declining—which exacerbates pricing and profit declines.

With the lack of annual seasonality in the production of metals and ease of storage without spoilage, much of time variability comes from the demand side of the equation (e.g., construction and economic growth).

EXAMPLE 4

Industrial Metals Life Cycle

Because of large economies of scale for processing industrial metals, producers:

A immediately shut down new capacity when supply exceeds demand.

B have an incentive to maintain maximum operating production levels when demand declines.

C find it difficult to cut back production or capacity even when supply exceeds demand or demand slows.

Solution:

C is correct. Given the sizable facilities in which metals are produced and their capital requirements, reducing capacity is difficult when demand slows. A is incorrect because of the time lag involved in responding to reduced demand conditions. B is incorrect because producers would face financial losses if they maintained maximum production levels when there is a decline in demand.

3.3 Livestock

Livestock grows year round, but good weather and access to high-quality pasture and feed accelerate weight gain. As a result, there is fluctuation in the availability of animals ready for slaughter. The timing to maturity typically increases with size, with poultry maturing in a matter of weeks, hogs in months, and cattle in a few years. Taking the example of a hog, the life cycle begins with a sow (female hog) giving birth. Normally it takes about six months to raise a piglet to slaughter weight, and during that time it can be fed almost anything to get it up to proper bulk. In mass-scale production, soymeal and cornmeal are the most common foods. In contrast, cattle take longer to raise. For mass-scale breeding, the first one to two years are spent as "feeder cattle," first eating a grass diet in pasture. The next phase covers an additional 6–12 months whereby cattle are in a feed lot being fattened to slaughter weight, generally on a corn-based diet. Note that the various types of feed for these animals are other traded commodities.

The livestock industry in the United States has historically been among the least export-oriented of all the commodities because of the high risk of spoilage once an animal is slaughtered. However, advances in cryogenics (freezing) technologies with regard to chicken, beef, and pork mean that increasingly these products are moving from one part of the world to another in response to differences in production costs and demand. And as emerging and frontier market countries develop middle class consumers capable of purchasing meat protein as a regular part of their diet, there has been increased investment in the livestock and meatpacking industries in such countries as the United States and Brazil. These industries combine low-cost sources of animal feed, large grazing acreage, and strong domestic demand (leading to facilities with substantial economies of scale) as key export points to supply global demand.

Ranchers and slaughterhouses trade hog and cattle futures to hedge against their commitments. Ranchers can hedge both young cattle that are still in pasture (called feeder cattle) and animals being fattened for butchering (called live cattle).

EXAMPLE 5

Livestock Life Cycle

The US livestock sector has been among the least export-oriented commodity sectors because of:

A low technological innovation in the sector.

B high risk of spoilage once animals are slaughtered.

C little or no demand for US livestock from outside the United States.

Solution:

B is correct. Livestock incur a high risk of spoilage once they are slaughtered unless the meat is frozen. A is incorrect because advances in cryogenics have improved the ability to export from the United States. C is incorrect because demand for US livestock has expanded internationally, particularly in emerging market countries that are experiencing economic growth.

3.4 Grains

Grains in the Northern Hemisphere follow a similar growth cycle, with an analogous but opposite growth cycle in the Southern Hemisphere. Plants mature according to the following steps: (1) planting (placing the seeds in the ground after preparation/

fertilization work); (2) growth (the emerging of the seedling to full height); (3) pod/ear/head formation (the food grain is created by the plant); and (4) harvest (the collection of the grain by the farmer). The timing in North America is shown in Exhibit 4 to illustrate the time it takes to grow each crop.

Exhibit 4 Timing for Grain Production in North America

	Corn	Soybeans	Wheat*
Planting	April–May	May–June	Sep.–Oct.
Growth	June–Aug.	July–Aug.	Nov.–March
Pod/Ear/Head Formation	Aug.–Sep.	Sep.	April–May
Harvest	Sep.–Nov.	Sep.–Oct.	June–July

* The hard winter wheat variety, which has a higher protein content, is used here.
Source: Authors' research.

Because demand for grains is year round, they are regularly stored in silos and warehouses globally. Some countries have a central purchasing bureau, and others depend on local or international trading companies to maintain stockpiles. Poor hygienic standards and logistics can result in a substantial loss of value to grains due to mold or insect/animal infestation. Monitoring the purchasing patterns of these government tenders can assist a research analyst in determining grain demand.

Farmers and consumers can trade futures to hedge their exposure to the crop in question, and the contract delivery months reflect the different times of the growing cycle outlined earlier. Ranchers also can use grain futures to hedge against the cost of feeding an animal.

3.5 Softs

Coffee, cocoa, cotton, and sugar are very different soft commodities in this sector, so we will focus on one that is grown and enjoyed broadly—coffee. Coffee is harvested somewhere all year round in the various countries that circle the Equator. After the coffee cherries are picked (still often by hand, to ensure that only ripe ones are taken), the husk and fruit are removed and the remaining bean dried. More than half of coffee uses the dry method in which the harvested cherries are laid out in the sun for two to three weeks. The wet method uses fresh water to soak the cherries, the soft pulp is removed, the bean is fermented for 12–48 hours, and then the bean is dried. The "green" beans are then hulled, sorted, and bagged for their final markets. With most of the consumption in faraway foreign markets, ships are commonly used to transport the beans to their buyer, which may store them in a bonded warehouse. The local buyer roasts the beans and ships them to the retail location (e.g., coffee house or supermarket) for purchase or brewing.

Coffee comes in two main varieties, robusta and arabica, although there are many others. Generally speaking, robusta beans are lower quality with less flavor than the arabica. There are two futures contracts associated with coffee: The robusta variety is traded in London, and the arabica variety is traded in New York. Note that the contracts are for the unroasted or "green" beans. The physical delivery aspect of these contracts allows for sellers to deliver the beans to an authorized bonded warehouse as fulfillment of the contract at expiration. Therefore, farmers and distributors can sell futures contracts to hedge the sales price of production, and coffee roasters can buy futures contracts to hedge coffee bean purchase costs; contract maturities can be selected by each to match their product delivery schedules.

4 VALUATION OF COMMODITIES

c contrast the valuation of commodities with the valuation of equities and bonds;

The valuation of commodities compared with that of equities and bonds can be summarized by the fact that stocks and bonds represent financial assets and are claims on the economic output of a business, a government, or an individual. Commodities, however, are almost always physical assets. We say "almost always" because some newer classes of commodities, such as electricity or weather, are not physical assets in the sense that you can touch or store them.

Commodities are typically tangible items with an intrinsic (but variable) economic value (e.g., a nugget of gold, a pile of coal, a bushel of corn). They do not generate future cash flows beyond what can be realized through their purchase and sale. In addition, the standard financial instruments that are based on commodities are not financial assets (like a stock or bond) but are derivative contracts with finite lifetimes, such as futures contracts. As with other types of derivatives, commodity derivative contracts can and do have value, but they are contingent on some other factors, such as the price of the underlying commodity. Hence, the valuation of commodities is based not on the estimation of future profitability and cash flows but on a discounted forecast of future possible prices based on such factors as the supply and demand of the physical item or the expected volatility of future prices. On the one hand, this forecast may be quite formal and elaborately estimated by a producer or consumer. One can imagine the detailed inputs available to an oil company based on the labor and capital expenses needed to extract oil, refine it, and transport it to final sale as gasoline in your automobile. On the other hand, this forecast may be instinctively made by a floor trader with little fundamental analysis but instead with professional judgment based on years of experience and perhaps some technical analysis.

As opposed to a stock or bond that receives periodic income, owning a commodity incurs transportation and storage costs. These ongoing expenditures affect the shape of the forward price curve of the commodity derivative contracts with different expiration dates. If storage and transportation costs are substantial, the prices for a commodity futures contract will likely be incrementally higher as one looks farther into the future. However, sometimes the current demand for the commodity can move the spot price higher than the futures price. The spot price reflects the fact that, instead of going long a futures contract, one could buy the commodity today and store it until a future date for use. The expenditure would be the outlay/investment at today's spot price for the commodity along with (or net of) the future costs one would incur to store and hold it. This time element of commodity storage and supply and demand can generate "roll return" and affect investment returns. These and other factors figure into the assessment of futures pricing, which we will cover later.

Some commodity contracts require actual delivery of the physical commodity at the end of the contract versus settlement in a cash payment (based on the difference between the contract futures price and the spot price prevailing at the time of contract expiration). The force of arbitrage—which reflects the law of one price—may not be entirely enforced by arbitrageurs because some participants do not have the ability to make or take delivery of the physical commodity. In these situations, the relationships that link spot and futures prices are not an equality but are a range that only indicates the limit or boundary of value differences that can occur.

There is an important additional consideration concerning the link between spot and futures prices in commodities. Some of the largest users of commodity futures are businesses seeking to hedge price risk when that price is a critical source of either revenue or cost in their business operations. For example, the airline industry is very dependent on the cost of jet fuel for operating planes. The highly competitive nature of the industry results in tremendous price pressure on airfares, with a need for airlines to fill each flight with as many passengers as possible. The futures and swap markets for jet fuel allow airlines to lower the risk of higher fuel costs by hedging the price of future fuel purchases (particularly against surprise shocks in oil prices).

In addition, the price discovery process of the commodity futures markets provides airlines with insights about future fuel prices that help determine what prices to offer their customers for future flights while still making a profit. In fact, airline ticket sales are—in effect—selling a contract at a price set today for future delivery of a service—namely, a plane flight. In this case, the airlines will typically hedge their price risk and uncertainty about future fuel costs by purchasing ("going long") energy futures contracts.

EXAMPLE 6

Commodities versus Stocks and Bonds

In contrast to financial assets, such as stocks and bonds:

A commodities are always physical goods.

B commodities generate periodic cash flows.

C commodity investment is primarily via derivatives.

Solution:

C is correct. The most common way to invest in commodities is via derivatives. A is incorrect because although most commodities are physical goods, certain newer classes, such as electricity or weather, are not tangible. B is incorrect because commodities may incur, rather than generate, periodic cash flow through transportation and storage costs (when the commodities are physically owned).

EXAMPLE 7

Spot Commodity Valuation

What is a key distinction between the valuation of commodities compared with the valuation of stocks and bonds?

A Valuation of commodities cannot be conducted using technical analysis.

B Valuation of commodities focuses on supply and demand, whereas valuation of stocks and bonds focuses on discounted cash flows.

C Valuation of stocks and bonds focuses on future supply and demand, whereas commodity valuation focuses on future profit margins and cash flow.

Solution:

B is correct. The valuation of commodities is based on a forecast of future prices based on supply and demand factors, as well as expected price volatility. In contrast, the valuation of stocks and bonds is based on estimating future profitability and/or cash flow. A is incorrect because technical analysis is sometimes applied to valuing commodities. C is incorrect for the reasons stated for choice B.

5 COMMODITIES FUTURES MARKETS: PARTICIPANTS

d describe types of participants in commodity futures markets;

Public commodity markets are structured as futures markets—that is, as a central exchange where participants trade standardized contracts to make and take delivery at a specified place at a specified future time. As mentioned, futures contracts are derivatives because the value of the contract is derived from another asset. Both futures and forward contracts are binding agreements that establish a price today for delivery of a commodity in the future (or settlement of the contract in cash at expiration). As mentioned at the beginning of the reading, the focus of this reading is on futures, with forwards discussed only briefly.

5.1 Futures Market Participants

The key differences between futures and forward contracts is that futures contracts are standardized agreements traded on public exchanges, such as the Chicago Mercantile Exchange (CME), Intercontinental Exchange (ICE), and the Shanghai Futures Exchange (SHFE), and gains/losses are marked to market every day. Standardization allows a participant to enter into a contract without ever knowing who the counterparty is. In addition, the exchange oversees trading and margin requirements and provides some degree of self-imposed regulatory oversight. In contrast, forward contracts are commonly bilateral agreements between a known party that wants to go long and one that wants to go short. Because of their bilateral nature, forwards are considered to be OTC (over the counter) contracts with less regulatory oversight and much more customization to the specific needs of the hedging (or speculating) party. Often, the counterparty for a forward contract is a financial institution that is providing liquidity or customization in exchange for a fee. Although futures markets require that daily cash movements in the futures price be paid from the losing positions to the winning positions, forward contracts are usually only settled upon expiration or with some custom frequency dictated by the contract.

Early commodity exchanges operated as forward markets, but too often participants would go bankrupt when unrealized losses became realized at the end of the contract. The futures process was introduced to minimize this risk, with the exchange acting as payment guarantor. The first modern organized futures exchange was the Dojima Rice Exchange in Osaka, Japan, which was founded in 1710, although futures contracts were traded in England during the 16th century. The structure of futures markets is important to understand as a way of understanding the goals and roles of the various participants. When we consider any commodity, for every producer of that commodity there is a consumer. Thus, for participants who are long the physical commodity and want to sell it, there are also participants who are short the physical commodity and want to buy it. Therefore, for fairness between the two sets of participants, longs and shorts need to operate on an equal basis. As a coincident observation, the commodity markets are net zero in terms of aggregate futures positions (futures contract longs equal futures contract shorts). In contrast, in markets for stocks and bonds, there is a net long position because the issued stocks' and bonds' market values are equal to the net aggregate positions at the end of each day. Shorting an equity is constrained by the short seller's need to locate shares to short, the requirement to reimburse dividends on borrowed shares, and requirements to post and pay interest on margin that generally exceeds the margin required for long equity positions (as in the United States under Regulation T). In contrast, shorting commodity futures is much simpler, with short investors selling to long investors directly, and thus short investors post the same margin required of long investors.

There are a number of participants in commodity futures markets. First are *hedgers*, who trade in the markets to hedge their exposures related to the commodity. The second are long-term and short-term *traders* and *investors* (including index investors), who speculate on market direction or volatility and provide liquidity and price discovery for the markets in exchange for the expectation of making a profit. Third are the *exchanges* (or clearing houses), which set trading rules and provide the infrastructure of transmitting prices and payments. Fourth are *analysts*, who use the exchange information for non-trading purposes, such as evaluating commodity businesses, creating products that are based on commodity futures (e.g., exchange-traded funds, swaps, and notes), and making public policy decisions. Analysts also include brokers and other financial intermediaries who participate in the markets but do not take a position. Finally, *regulators* of both the exchange and traders exist to monitor and police the markets, investigate malfeasance, and provide a venue for complaints.

5.1.1 *Commodity Hedgers*

Hedgers tend to be knowledgeable market participants: One would expect that a company that drills for oil knows something about the supply and demand for oil and related forms of energy (at least in the long run). However, hedgers may not be accurate predictors of the future supply and demand for their product. Consider a baker who buys wheat for future delivery and benefits from a surprise drought (has locked in a low price in a supply-constrained market). However, the baker is hurt if the weather is beneficial (has effectively overpaid during a bumper crop). Given that a hedger can make delivery (if short the futures contract) or take delivery (if long the futures contract), he or she is generally motivated by risk mitigation with regard to cash flow, so the risk is more of an opportunity cost than an actual one.

It is important to keep in mind that hedging and speculating are not synonymous with being (respectively) long or short. As Exhibit 5 illustrates with some examples, both long and short positions can be associated with either hedging or speculating.

Exhibit 5 Examples of Hedging and Speculating Positions

	Long Position	Short Position
Hedging	Food manufacturer seeking to hedge the price of corn needed for snack chips	Gold mining company seeking to hedge the future price of gold against potential declines
Speculating	Integrated oil company seeking to capitalize on its knowledge of physical oil markets by making bets on future price movements	Commodity trading adviser (CTA) seeking to earn a profit for clients via a macro-commodity investment fund

Note also that hedgers tend to speculate based on their perceived unique insight into market conditions and determine the amount of hedging that is appropriate. From a regulatory standpoint in the United States, the difficulty in clearly distinguishing between hedging and speculating, therefore, has resulted in the separation of commodity producers and consumers from other trading participants regardless of whether commercial participants are actually speculating.

5.1.2 *Commodity Traders and Investors*

The commodity trading community, like other groups of traders, consists of three primary types: (1) informed investors, (2) liquidity providers, and (3) arbitrageurs. Informed investors largely represent the aforementioned hedgers and speculators, including index and institutional investors. With regard to the hedger, as mentioned previously, a company that drills for oil clearly is familiar with the supply and demand for oil and related forms of energy (at least in the long run). But hedgers may not be accurate predictors of the *future* supply and demand for their product.

Speculators, who believe that they have an information advantage, seek to outperform the hedger by buying or selling futures contracts in conjunction with—or opposite from—the hedger. This trading may be on a micro-second time scale or a multi-month perspective. For example, if a speculator has a superior weather prediction process, he or she has an information advantage and will trade accordingly. Alternatively, a speculator may be willing to act as a liquidity provider, knowing that producers and consumers may not be in the market at the same time. By buying when the producer wants to sell and selling when the consumer is ready to buy, speculators may be able to make a profit. In this sense, speculators are willing to step in, under the right pricing circumstances, to provide insurance to hedgers in return for an expected (albeit not guaranteed) profit.

Finally, arbitrageurs who have the ability to inventory physical commodities can attempt to capitalize on mispricing between the commodity (along with related storage and financing cost) and the futures price. They may own the storage facilities (bonded warehouses, grain silos, feedlots) and work to manage that inventory in conjunction with the futures prices to attempt to make arbitrage-style profits.

5.1.3 *Commodity Exchanges*

Commodity futures markets are found throughout the world. The CME and ICE are the primary US markets, having consolidated the bulk of the various specialist exchanges. Elsewhere in the Americas, the primary commodity exchange is in Brazil, where B3 trades softs, grains, and livestock. In Europe, the London Metal Exchange (owned by Hong Kong Exchanges and Clearing Limited (HKEX) is the main industrial metals location globally. Energy and shipping are also traded out of London. In Asia, major commodity exchanges include China's Dalian Commodity Exchange and Shanghai Futures Exchange and Japan's Tokyo Commodity Exchange, among others. Finally, Indonesia (palm oil), Singapore (rubber), and Australia (energy, grains, wool) have supplementary commodity futures markets. Given that people all over the world need food, energy, and materials, exchanges have formed globally to meet those needs.

5.1.4 *Commodity Market Analysts*

Non-market participants use the exchange information to perform research and conduct policy as well as to facilitate market participation. Their activities affect market behavior, albeit in an indirect manner. Research may be commercially based. For example, a manufacturer may want to project and forecast the energy cost of a new process or product as part of an academic study comparing one market structure with another. Commodity prices are a key component in understanding sources of inflation and are used in other indexes that indicate quality of life for consumers and households. Governments that control natural resource extraction (e.g., nationalized oil companies) or tax commodity extraction by private entities are also interested in understanding futures markets to promote or discourage investment and/or raise revenue.

5.1.5 *Commodity Regulators*

Finally, various regulatory bodies monitor the global commodity markets. In the United States, commodity and futures regulation falls under the Commodity Futures Trading Commission (CFTC), which is a regulatory body separate from the better-known Securities and Exchange Commission. The CFTC delegates much of the direct monitoring to the National Futures Association (NFA)—a self-regulatory body—whose members are the authorized direct participants in the markets with customer responsibilities (e.g., clearing firms, brokers, advisers).

Outside the United States, most other countries have a unified regulatory structure. For example, the China Securities Regulatory Commission regulates both futures and securities (i.e., stocks and bonds). In Europe, most legislation in the area of financial services is initiated at the European Union (EU) level primarily through the European Securities and Markets Authority (ESMA). The Markets in Financial Instruments Directive (MiFID, and subsequently MiFID II), which first came into force in 2007, was a key element of EU financial market integration that focused largely on deregulation (MiFID II took effect in January 2018). Since 2009, existing legislative instruments, particularly for commodity derivative markets, have been revised and new regulations have been introduced with the aim to strengthen oversight and regulation, and they are subject to G–20 commitments. Harmonizing these different regulatory bodies is the International Organization of Securities Commissions (IOSCO), which is the international association of the world's securities and futures markets.

In all regions, the interests of the financial sector strongly influence debates and legislation on financial market regulation, including that of commodities.

EXAMPLE 8

Commodity Market Participants

Commodity traders that often provide insurance to hedgers are *best* described as:

A arbitrageurs.

B liquidity providers.

C informed investors.

Solution:

B is correct. Liquidity providers often play the role of providing an insurance service to hedgers who need to unload and transfer price risk by entering into futures contracts. A is incorrect because arbitrageurs typically seek to capitalize and profit on mispricing due to a lack of information in the marketplace. C is incorrect because informed investors predominantly keep commodity futures markets efficient by capitalizing on mispricing attributable to a lack of information in the marketplace.

COMMODITY SPOT AND FUTURES PRICING 6

e analyze the relationship between spot prices and futures prices in markets in contango and markets in backwardation;

Commodity prices are typically represented by (1) spot prices in the physical markets and (2) futures prices for later delivery. The **spot price** is simply the current price to deliver a physical commodity to a specific location or purchase it and transport it away from a designated location. Examples of a spot price may be the price quoted at a grain silo, a natural gas pipeline, an oil storage tank, or a sugar refinery.

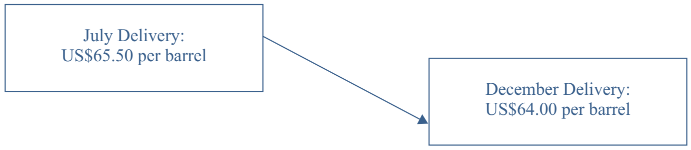
A **futures price** is a price agreed on to deliver or receive a defined quantity (and often quality) of a commodity at a future date. Although a producer and a consumer can enter into a bilateral contract to exchange a commodity for money in the future, there are (conveniently) many standardized contracts that trade on exchanges for buyers and sellers to use. Recall that a bilateral agreement is a forward contract, compared with a futures contract that is standardized and trades on a futures exchange. One benefit of futures markets is that information regarding contracts (number, price, etc.) is publicly available. In this way, the price discovery process that brings buyers and sellers into agreement is shared broadly and efficiently (in real time) with a global marketplace among the aforementioned market participants. The longest-maturity futures contract outstanding can have maturity extending from about a year (e.g., livestock) to several years (e.g., crude oil).

The difference between spot and futures prices is generally called the **basis**. Depending on the specified commodity and its current circumstances (e.g., supply and demand outlook), the spot price may be higher or lower than the futures price. When the spot price exceeds the futures price, the situation is called **backwardation**, and the opposite case is called **contango**. The origin of the word "contango" is a bit murky, but one theory is that it came from the word "continuation" used in the context of the London Stock Exchange in the mid-1800s. During this period, contango was a fee paid by the buyer to the seller to defer settlement of a trade (hence the near-term price would be less expensive than the longer-term price). The term "backwardation" describes the same arrangement if it were "backward," or reversed (i.e., payment to defer settlement was made by the seller to the buyer).

Backwardation and contango are also used to describe the relationship between two futures contracts of the same commodity. When the near-term (i.e., closer to expiration) futures contract price is higher than the longer-term futures contract price, the futures market for the commodity is in backwardation. In contrast, when the near-term futures contract price is lower than the longer-term futures contract price, the futures market for the commodity is in contango. The price difference (whether in backwardation or contango) is called the calendar spread. Generally speaking and assuming stable spot prices, the producer is willing to take a price in the future that is lower than the current spot price because it provides a level of certainty for the producer's business. The seller of that insurance on the other side of the trade profits because the lower futures price converges to the higher spot price over time. This relationship occurs when future commodity prices are expected to be higher because of a variety of reasons related to economic growth, weather, geopolitical risks, supply disruptions, and so on. As a long owner of a futures contract in contango, value will erode over time as the contract pricing moves closer to the spot price, assuming all else is unchanged. This relationship can be very costly for long holders of contracts if they roll futures positions over time. Although backwardation is "normal" for some contracts, there are other commodities that often trade in contango.

Exhibit 6 is a stylized representation of backwardation in West Texas Intermediate crude oil on CME Group's New York Mercantile Exchange (NYMEX).

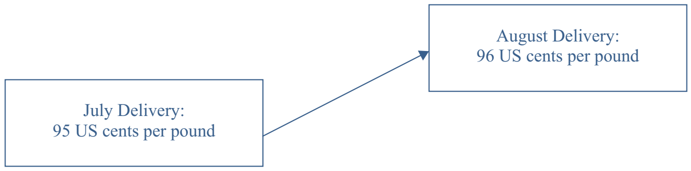
Exhibit 6 Backwardation



For contracts in a single (common) commodity, such as lean hogs or crude oil, the price differences may be traded as a spread rather than individually.

Exhibit 7 is a stylized representation of contango in lean hogs on the CME.

Exhibit 7 Contango



From these examples, the lean hogs July–August calendar spread is –1.0 cent per pound (95 – 96) and the crude oil July–December calendar spread is $1.50 per barrel (65.50 – 64.00).

A positive calendar spread is associated with futures markets that are in backwardation, whereas a negative calendar spread in commodities is associated with futures markets that are in contango. These calendar spreads are traded with their own bid–ask prices, trading range, and order book, similar to the single-month (i.e., nearest to expiration) futures contracts. Note that from this one trade, two contracts (one for each side, or "leg", of the spread) appear on an exchange's trading account and use their respective closing prices to determine profit or loss. Therefore, in the end, all trades and positions are valued at the close-of-day prices.

Commodity futures are settled by either cash or physical delivery. Cash-settled contracts, such as feeder cattle traded on the CME, have no value after the maturity date. Cash settlement is an important innovation in the evolution and development of commodity futures markets. To a certain extent, cash settlement enabled more involvement of two key participants in today's futures markets: speculators and arbitrageurs. It also introduced an entirely new way that hedgers (long or short) could participate in the market to transfer the future price risk of having to sell or buy a commodity without the complications associated with requiring physical delivery. Physical-settled commodity futures contracts require that the title of the actual commodity be transferred by the seller of the futures contract to the buyer at a particular place, on or by a particular date, and of a particular quality specification. For example, under a futures contract with West Texas Intermediate crude oil as the underlying physical commodity, crude oil meeting minimum specifications must be delivered to a particular set of tanks at Cushing, Oklahoma, in the United States. Meanwhile, a similar futures contract with Brent crude oil as the underlying physical commodity has delivery points in the North Sea off the coast of the United Kingdom and Norway. Supply and demand differences at these two faraway geographic locations can cause price divergences despite otherwise similar specifications.

Physical delivery also ensures a convergence of the futures and spot markets, which may not necessarily occur in a cash-settled market. Note that this statement does not imply market manipulation in cash-settled markets, because trading costs or other factors may limit complete convergence. The emergence of central exchanges for trading commodity futures facilitated this convergence with standardized contracts. In addition, these exchanges provided centrally established, publicly available pricing, which quickly replaced private pricing that was dependent on both contract terms and the location where transactions occurred.

Physical delivery can become complicated by such factors as quality or variety differences in the commodity. For example, robusta coffee (traded in the United Kingdom) cannot be delivered for arabica coffee (traded in the United States) because it is a different variety of coffee with a different venue for delivery. Likewise, raw (or unprocessed) sugar that is traded in the United States cannot be delivered for white processed sugar that is traded in the United Kingdom. Futures markets can address some of these peculiarities involving quality or differences in supply. When physical delivery is required, some futures contracts require a premium or discount associated with specifications. For example, arabica coffee prices are automatically adjusted based on the country of origin and the location of the warehouse where delivery is made.

In summary, spot prices are highly localized and associated with physical delivery, limiting the degree to which interested participants can seek to hedge or speculate on their future direction. In contrast, futures prices can be global (and if not, at least regional or national) in scope. They also are standardized for trading on exchanges to promote liquidity; act as a reference price point for customized (i.e., forward) contracts; and generate widely available, minimally biased data for market participants and governments to judge supply and demand and to make planning decisions.

In this manner, futures can be used to allocate risk and generate returns for market participants. On the surface, futures trading may seem muddled and chaotic on a micro level but serves as an overall social benefit by sending signals to producers and consumers for hedging and inventory-sizing purposes and to governments for the potential impact of policy decisions.

EXAMPLE 9

Spot and Futures Pricing (1)

The current price of the futures contract nearest to expiration for West Texas Intermediate (WTI) crude oil is $65.00 per barrel, whereas the six-month futures contract for WTI is priced at $60.75 per barrel. Based on this information:

A the futures market for WTI crude oil is currently in a state of contango.

B the futures market for WTI crude oil is currently in a state of backwardation.

C the shipping and delivery cost of WTI crude oil for a futures contract expiring in six months with physical delivery to Cushing, Texas, is $4.25 per barrel.

Solution:

B is correct. Commodity futures markets are in a state of backwardation when the spot price is greater than the price of near-term (i.e., nearest to expiration) futures contracts, and correspondingly, the price of near-term futures contracts is greater than longer-term contracts. A is incorrect because the market would be in contango only if the deferred futures price exceeded that of the nearby futures price. C is incorrect because the shipping and delivery costs associated with physical delivery of a commodity are only one component in determining a commodity futures contract price. Geopolitical, seasonal, and other factors also influence the difference in delivery months.

EXAMPLE 10

Spot and Futures Pricing (2)

An important distinction between spot and futures prices for commodities is that:

A spot prices are universal across regions, but futures prices vary by location.

B futures prices do not reflect differences in quality or composition for a commodity.

C spot prices vary across region based on quality/composition and local supply and demand factors.

Solution:

C is correct. Spot prices of commodities vary across regions, reflecting logistical constraints and supply and demand imbalances that hinder the movement of materials. A is incorrect because spot prices tend to vary by region while futures are purposely standardized to facilitate trading. B is incorrect because while futures contracts are based on standardized specifications, composition and quality can be assigned premiums or discounts for delivery.

EXAMPLE 11

Spot and Futures Pricing (3)

An arbitrageur has two active positions in the commodity futures markets—one for lean hogs and the other for natural gas. The calendar spread on the lean hogs contract is quoted at –50 cents per pound, and the calendar spread on the natural gas contract is +$1.10 per million BTU (British thermal units). Based on this information, we can say that:

A only the spreads of these commodities, and not the individual prices, can be traded in commodity markets.

B the lean hogs futures market is in a state of backwardation and the natural gas futures market is in a state of contango.

C the lean hogs futures market is in a state of contango and the natural gas futures market is in a state of backwardation.

Solution:

C is correct. The spread is the difference between the current spot price for a commodity and the futures contract price. Because futures markets in a state of contango will have futures prices that exceed the spot price, the spread for these markets is negative. Conversely, in a state of backwardation, the spread is positive. A is incorrect because either the individual contract prices or the combined spreads can be traded. B is incorrect because, as mentioned earlier, the negative sign of the spread of lean hogs futures indicates a state of contango, whereas the positive sign of the spread of natural gas futures indicates a state of backwardation.

EXAMPLE 12

Spot and Futures Pricing (4)

A futures price curve for commodities in backwardation:

A always remains in backwardation in the long term.

B can fluctuate between contango and backwardation in the long term.

C reflects structural long-term industry factors, as opposed to dynamic market supply and demand pressures.

Solution:

B is correct. During periods of market stress or fundamental structural change in market conditions, some commodity futures price curves can rapidly shift from contango to backwardation or vice versa. A is incorrect because futures price curves can vacillate between contango and backwardation. C is incorrect because the shape of a commodity futures price curve reflects both long-term industry factors as well as market expectations of future supply and demand of the underlying commodity(ies).

7 THEORIES OF FUTURES RETURNS

f compare theories of commodity futures returns;

Commodity futures markets have a reputation for volatility, but similar to other asset classes, there are theoretical bases for their long-run behavior. The original purpose of futures markets is for producers and consumers to hedge physical raw materials. In this section, we will discuss the underpinning theories of commodity futures returns, deconstruct the components of futures returns (i.e., at an index level), and close with thoughts on term structure (i.e., contango versus backwardation and implications of rolling futures contracts).

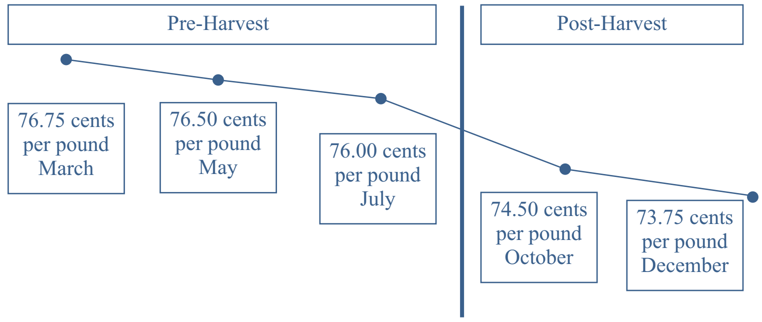
7.1 Theories of Futures Returns

Several theories have been proposed to explain the shape of the futures price curve, which has a dramatic impact on commodity futures returns. This reading covers three of the most important theories: (1) insurance theory, (2) hedging pressure hypothesis, and (3) theory of storage.

7.1.1 *Insurance Theory*

Keynes (1930), the noted economist and market speculator, proposed one of the earliest known theories on the shape of a commodity futures price curve. Also known as his theory of "normal backwardation," Keynes, in his 1930 tome *A Treatise on Money*, proposed that producers use commodity futures markets for insurance by locking in prices and thus make their revenues more predictable. A commodity producer is long the physical good and thus would be motivated to sell the commodity for future delivery to hedge its sales price. Imagine a farmer who thinks that next year she will grow a certain amount of soybeans on her land. She can sell a portion of her crop today that will be harvested months later to lock in those prices. She can then spend money on fertilizer and seed with more confidence about her budget. She may not be locking in a profit, but she would better understand her financial condition. Keynes's theory assumes that the futures curve is in backwardation "normally" because our farmer would persistently sell forward, pushing down prices in the future. Alternatively, this theory posits that the futures price has to be lower than the current spot price as a form of payment or remuneration to the speculator who takes on the price risk and provides price insurance to the commodity seller. The concept of normal backwardation is illustrated in Exhibit 8, using cotton prices pre- and post-harvest.

Exhibit 8 Normal Backwardation



In terms of returns, if the front price is stable (in our example, 76.75 cents), then an investor can buy a further-dated contract (e.g., October) at 74.50 cents and wait for that contract to become the current contract. As the month of October approaches (and assuming no change in front prices), the October contract will reach 76.75 cents at maturity, and the speculator will make a profit of 2.25 cents per pound (note that a contract is 50,000 pounds, so that is a total profit of $1,125 per contract). Even if the contract does not fully converge, this theory holds that there should be positive excess returns (sometimes referred to as the risk premium) via this process to induce buying. As noted earlier, this process acts as a type of insurance for the farmer as well as a return for the investor providing such insurance.

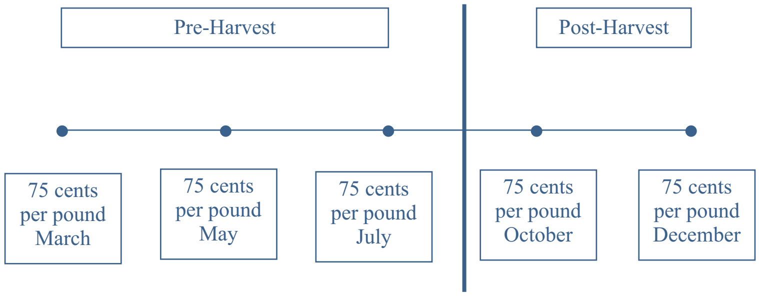
Looking at the evidence, however, markets failed to match Keynes's hypothesis. Kolb (1992) looked at 29 futures contracts and concluded (with some humor) that "normal backwardation is not normal." That is, the presence of backwardation does not necessarily generate positive returns in a statistically significant fashion for the investor (or that contango leads to negative returns, for that matter). This result confirmed other studies, including one by Fama and French (1987). Therefore, a more sophisticated view developed to explain futures markets in contango (i.e., when the shape of the futures price curve is upward sloping with more distant contract dates), recognizing that certain commodity futures markets often show persistently higher prices in the future as opposed to the backwardation outlined by Keynes. This view is called the hedging pressure hypothesis.

7.1.2 *Hedging Pressure Hypothesis*

This perspective stemmed from multiple works, most notably outlined by De Roon, Nijman, and Veld (2000), who drew from Cootner (1960). Their research analyzed 20 futures markets from 1986 to 1994 and concluded that hedging pressure plays an important role in explaining futures returns. Hedging pressure occurs when both producers and consumers seek to protect themselves from commodity market price volatility by entering into price hedges to stabilize their projected profits and cash flow. Producers of commodities will tend or want to sell commodities forward and thus sell commodity futures. On the other side, consumers of commodities want to lock in prices of their commodity purchases and buy commodity futures. This theory applies to the aforementioned farmer selling a portion of next year's crop today. It can also apply to a central bank that wants to buy gold during each of the next 12 months as part of its monetary operations or a refinery that may want to lock in the price of its oil purchases and, conversely, the prices of its gasoline and heating oil production.

If the two forces of producers and consumers both seeking price protection are equal in weight, then one can envision a flat commodity curve, such as Exhibit 9 illustrates. In this idealized situation, the natural needs for price insurance by commodity buyers and sellers offset each other. There is no discount on the commodity futures price required to induce speculators to accept the commodity price risk because the hedging needs of both the buyer and seller complement and offset each other.

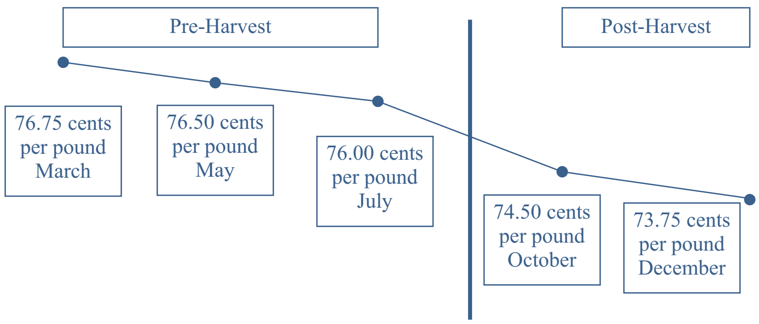
Exhibit 9 Balanced Hedging between Producers and Consumers



To use a different example, consider the problem of snowfall in the New England region of the United States. On one hand, small municipalities in Vermont, New Hampshire, or Maine may experience high levels of annual snowfall that are a risk to their snow removal budgets. On the other hand, ski resorts in New England have an opposite risk challenge: Low snowfall creates skiing revenue shortfalls (or adds to costs because of the need for man-made snow), whereas high snowfall winters are a potential bonanza for both higher revenue and lower operating costs. This situation is another example of when the hedging needs of two parties can offset each other and create a mutually beneficial outcome.

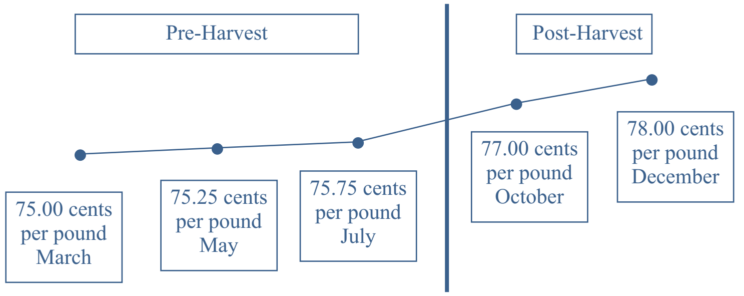
If commodity producers as a group are more interested in selling forward (seeking price insurance) than commodity consumers (as per the concept of normal backwardation), then the relative imbalance in demand for price protection will lead to the need for speculators to complete the market. But speculators will only do so when futures prices trade at a sufficient discount to compensate for the price risk they will take on. In this case, the shape and structure of the futures price curve can be illustrated as backwardation, as shown in Exhibit 10, which is consistent with Keynes's insurance theory.

Exhibit 10 Commodity Producers Exceed Consumers (Backwardation)



Finally, if the buyers of soybeans (as a group) are especially worried about the availability of the crop in the next harvest but producers of soybeans are less concerned about crop prices, there would be an imbalance in the demand for price insurance away from producers and toward buyers. This situation would lead to a futures price curve that represents a market in contango, as illustrated in Exhibit 11. In this case, the additional demand for price insurance among buyers (versus sellers) of the commodity will lead them to bid up the futures price to induce speculators to take on this price uncertainty risk.

Exhibit 11 Commodity Consumers Exceed Producers (Contango)



Although this theory is more robust than the Keynes's insurance theory, it is still incomplete. One issue is that producers generally have greater exposure to commodity price risk than consumers do (Hicks 1939). There are companies (as well as countries) that are almost entirely dependent on commodity production and thus are very concentrated in one sector, such as energy (e.g., British Petroleum, ExxonMobil), grains (e.g., Cargill, Louis Dreyfus), and metals (e.g., BHP Billiton, Vale, Rio Tinto, Shenhua).

Commodity consumers, in contrast, are very diffuse and often have other priorities (i.e., few if any individual people hedge their meat consumption or gasoline spending). Companies that purchase and use commodities in their products have a mixed record of price hedging, depending on the importance of the commodities in their cost structure. Clothing companies (e.g., Gap) generally do not hedge cotton because the spending is only a few percentage points of their expense base. Marketing and store experience (seen in rent, occupancy, and depreciation expenses) are much more important. But fast food companies hedge a wide variety of commodity inputs (e.g., livestock, grains, energy) because of the high degree of competition for prepared food at a low price point (e.g., McDonald's, Burger King, Wendy's).

In addition, both producers and consumers speculate on commodity prices, whether it is intended or unintended. Corporate treasury departments that serve as profit centers may adjust their hedges based on their views of the commodity markets. Their primary function may be to hedge, but a profit incentive can lead them to speculate. Individual farmers may not be overly aware of the commodity markets and thus have an inconsistent hedging approach. Trading companies actively trade the futures and physical markets in energy, metals, and grains. The very nature of trading companies is to know what is happening at all times along the value chain of any commodity market and profit from that informational advantage while bringing together buyers and sellers. In their case, profit maximization does not come from the production of commodities but trading around that production. In all of these examples, attempts to hedge may result instead in unintended speculative positions in which a company is not transferring price risk away but instead taking on more risk. The collapse in 1993 of Metallgesellschaft AG, one of Germany's largest industrial conglomerates at the time, from a poorly constructed gasoline, fuel oil, and heating oil hedge is a defining example of flawed commercial hedging.

In summary, despite its intuitive logic, applying the hedging pressure hypothesis remains a challenge because measuring the asymmetry in hedging pressure between buyers and sellers of a commodity is very difficult.

7.1.3 *Theory of Storage*

This theory, originally postulated by Kaldor (1939), focuses on how the level of commodity inventories helps shape commodity futures price curves. The key issue this theory attempts to address is whether supply or demand of the commodity dominates in terms of its price economics. Recall that commodities are physical assets, not virtual assets like stocks and bonds. Physical assets have to be stored, and storage incurs costs (rent, insurance, inspections, spoilage, etc.). Therefore, a commodity that is regularly stored should have a higher price in the future (contango) to account for those storage costs. In other words, supply dominates demand. In contrast, a commodity that is consumed along a value chain that allows for just-in-time delivery and use (i.e., minimal inventories and storage) can avoid these costs. In this situation, demand dominates supply and current prices are higher than futures prices (i.e., backwardation).

In theoretical terms, available inventory generates a benefit called a convenience yield. Having a physical supply of the commodity available is convenient for consumers of the commodity (e.g., individuals, bread companies, meat processors, refiners) because it acts as a buffer to a potential supply disruption that could otherwise force a shutdown of their operations. Because this type of risk/concern is inversely related to the inventory size and the general availability of the commodity (and confidence in its continued availability), the convenience yield is low when stock is abundant. However, the yield rises as inventories diminish and concerns regarding future availability of the commodity increase.

As a result, the theory of storage states that futures prices can be written this way:

Futures price = Spot price of the physical commodity + Direct storage costs (such as rent and insurance) − Convenience yield.

This equation indicates that price returns and the shape of the curve can move in conjunction with the changes in the available inventory as well as actual and expected supply and demand. For example, when civil war broke out in Libya in 2011, the production of that country's high-quality crude oil was placed in jeopardy, constricting supply. In reaction, the spot price for high-quality crude oil increased. At the same time, the convenience yield increased in the futures contracts closer to expiration because there was a scramble to tap into alternative oil supplies for European refiners. The high quality of Libyan crude oil also restricted which substitute crude oil supplies could be used to replace production from the blocked oil fields and how soon these replacements could be available. The real-world constraints and complications imposed by geography and the logistics of the oil industry resulted in a multi-month delay for replacement supplies. As a result, in the further-out (i.e., longer time to expiration) futures contracts, the reaction was muted as traders assumed that such replacement supplies would be available. Thus the convenience yield remained lower in the deferred months. For this and other reasons, crude oil was pressured to trade in backwardation during 2011.

Unfortunately, while all these theories are reasonable and attractive, they have components that are unobservable or highly volatile and, therefore, not reliably calculable. Commodity producers and consumers regard storage costs as proprietary information. Events (weather, war, technology) can radically adjust convenience yield in a short time with unknown magnitude. Corn suitable for feed may not be suitable for human consumption, so defining inventories is tricky. In the end, we have frameworks and theories, but they are not easily applied and require judgment and analysis by a trader or a valuation system.

EXAMPLE 13

Theories of Commodity Futures Returns (1)

Which of the following *best* describes the insurance theory of futures returns?

A Speculators will not provide insurance unless the futures price exceeds the spot price.

B Producers of a commodity will accept a lower future price (versus the spot price) in exchange for the certainty of locking in that price.

C Commodity futures markets result in a state of contango because of speculators insisting on a risk premium in exchange for accepting price risk.

Solution:

B is correct. Under the insurance theory of futures returns, Keynes stated that producers of a commodity would prefer to accept a discount on the potential future spot price in return for the certainty of knowing the future selling price in advance. A is incorrect because the futures price must be below the spot price (normal backwardation) under the insurance theory of futures returns. C is incorrect because the insurance theory of futures returns implies markets are in backwardation, not contango.

EXAMPLE 14

Theories of Commodity Futures Returns (2)

Under the hedging pressure hypothesis, when hedging activity of commodity futures buyers exceeds that of commodity futures sellers, that futures market is *most likely*:

A flat.

B in contango.

C in backwardation.

Solution:

B is correct. Under the hedging pressure hypothesis, a market in contango typically results when excess demand for price insurance among commodity futures buyers drives up the futures price to induce speculators to take on price uncertainty risk. A is incorrect because a flat market would likely exist if futures demand activity largely equaled that of supply. C is incorrect because under this scenario, the futures market would be in contango, not backwardation.

EXAMPLE 15

Theories of Commodity Futures Returns (3)

Under the theory of storage, the convenience yield is:

A not affected by the supply of a commodity.

B typically low when the supply of a commodity is scarce.

C typically high when the supply of a commodity is scarce.

Solution:

C is correct. Under the theory of storage, the convenience yield of a commodity increases as supply (inventories) diminish and concerns about the future availability increase. A is incorrect because supply levels have a discernible effect on the convenience yield, as mentioned. B is incorrect because the convenience yield would likely be high, as opposed to low, when supply is limited.

EXAMPLE 16

Theories of Commodity Futures Returns (4)

Which of the following represents the formula for a futures price according to the theory of storage?

A Futures price = Spot price of the physical commodity + Direct storage costs − Convenience yield.

B Futures price = Spot price of the physical commodity + Direct storage costs + Convenience yield.

C Futures price = Spot price of the physical commodity − Direct storage costs + Convenience yield.

Solution:

A is correct. According to the theory of storage, the futures price reflects the current spot price as well as costs incurred in actually holding the commodity until its delivery. Such costs include direct storage, such as inventory and insurance costs. Finally, because there is a convenience yield (or benefit) to owning a commodity as a form of insurance against potential supply disruptions, this term is subtracted from the current price of the commodity.

8 COMPONENTS OF FUTURES RETURNS

g describe, calculate, and interpret the components of total return for a fully collateralized commodity futures contract;

h contrast roll return in markets in contango and markets in backwardation;

The total return on a commodity investment in futures is different from a total return on the physical assets. So, why do investors tend to use futures to gain their exposure to commodities? Building on the previous section, one can see that physical commodities need to be stored, fed, or perhaps treated against spoilage. Each commodity can be very different in its maintenance requirements; sustaining a hog in Mexico would be very different from storing crude oil in Nigeria.

The total return on commodity futures is traditionally broken into three components:

- the price return (or spot yield),

- the roll return (or roll yield), and
- the collateral return (or collateral yield).

The price return is the change in commodity futures prices, generally the front month contract. Note that this change is different from the change in the price of the physical commodity because lack of standardization of the physical markets makes that a difficult task. Calculating the price return is straightforward, as shown in the following equation:

Price return = (Current price − Previous price)/Previous price.

In addition, as investors move from futures contract to futures contract, they must "roll" that exposure by selling the current contract as it approaches expiration and buying the next contract (assuming a long position). Depending on the shape of the futures curve, there is likely a difference between the two prices. Thus, a portfolio may require buying more far contracts than the near contracts being sold. Investors can observe this scenario if backwardation is driving the shape of the commodity futures price curve.

Example (stylized): Assume an investor has £110 of exposure in wheat futures and the near contract is worth £10 of exposure (so, the investor has £110 exposure divided by £10 per contract, or 11 contracts), but the far (i.e., longer expiration date) contract is worth only £9 of exposure. Therefore, for the investor to roll forward his contracts and maintain a constant level of exposure, he needs to roll the 11 contracts forward and also buy an additional 1 contract to keep the post-roll exposure close to the pre-roll exposure (£110 exposure divided by £9 per contract equals 12.2, or 12 contracts rounded).

In the opposite case, if the futures price curve shape is being driven by contango—with a higher futures price in the far contract—this scenario will require the purchase of fewer commodity contracts than in the near position.

Example: Assume an investor has £108 of exposure in regular unleaded gasoline (or petrol) futures and the near contract is worth £9 of exposure (so, the investor has £108 exposure divided by £9 per contract, or 12 contracts), but the far contract is worth £10 of exposure. Therefore, for the investor to roll forward her contracts and maintain a constant level of exposure, she needs to roll only 11 contracts and sell the extra 1 near contract to keep the post-roll exposure close to the pre-roll exposure (£108 exposure divided by £10 per contract equals 10.8, or 11 contracts rounded).

Note that this roll return is not a return in the sense that it can be independently captured; investors cannot construct a portfolio consisting of only roll returns. Instead, **roll return** is an accounting calculation used to replicate a portion of the total return for a fully collateralized (i.e., with no leverage) commodity index. As defined, the roll return is effectively the accounting difference (in percentage terms) between the near-term commodity futures contract price and the farther-term commodity futures contract price (note that roll return is sometimes defined in monetary terms rather than as a percentage):

Roll return = [(Near-term futures contract closing price − Farther-term futures contract closing price)/Near-term futures contract closing price] × Percentage of the position in the futures contract being rolled.

As an example, consider the roll from the March contract to the April contract for WTI crude oil on 7 February 2019 using the S&P GSCI methodology, which rolls its positions over a five-day period (so 1/5 = 20% per day):

March contract closing price: $52.64/barrel

April contract closing price: $53.00/barrel

($52.64 − $53.00)/$52.64 = −0.68% gross roll return × 20% rollover portion

= −0.13% net roll return (note the negative return in contango).

Note that different indexes use different periods and/or weights in their "rolling methodology." In Section 11, we will further discuss the rolling methodology of various indexes.

In his book *Expected Returns*, Ilmanen (2011) made the argument (challenged by others) that roll return is approximately equal to a risk premium. This concept relates back to Keynes and his theory of "normal backwardation." Keynes proposed that speculators take the other side of the transaction from commodity producers—who sell forward to lock in their cash flows—in an attempt to earn an excess return as compensation for providing price insurance to producers. Ilmanen attempted to demonstrate that positive long-run average returns are associated with positive roll return (i.e., in commodities for which futures prices are in backwardation) and negative long-run average returns are associated with negative roll return. However, because 40% of the commodities examined by Ilmanen (p. 255) had negative roll returns but positive total returns, one cannot directly conclude that backwardation earns a positive total return.

The **collateral return** is the yield (e.g., interest rate) for the bonds or cash used to maintain the investor's futures position(s). The minimum amount of funds is called the initial margin. If an investor has less cash than required by the exchange to maintain the position, the broker who acts as custodian will require more funds (a margin call) or close the position (buying to cover a short position or selling to eliminate a long position). Collateral thus acts as insurance for the exchange that the investor can pay for losses.

For return calculations on indexed investments, the amount of cash would be considered equal to the notional value of the futures. This approach means no leverage. For expected returns, commonly, investors should use a risk-free government bond that most closely matches the term projected. Most commodity indexes use short-term US Treasury bills, but if one is forecasting 10-year returns, then for collateral return purposes, a 10-year constant maturity government bond would have a more appropriate term.

Although indexes will be discussed more fully later in the reading, to illustrate the commodity return elements just discussed, one can use an index—in this case, the aforementioned S&P GSCI, which has one of the longest backtested and live history of the investable commodity indexes. Exhibit 12 shows the disaggregation of its return components.

Exhibit 12 Average Annual Return Components of the S&P GSCI, January 1970–March 2019

S&P GSCI Return	Total Return	Spot Return	Roll Return[1]	Collateral Return[1]
Return[2]	6.8%	3.0%	−1.3%	5.0%
Risk[3]	19.8%	19.8%	4.2%	1.1%
Correlation[4]		0.97	−0.11	−0.14

[1] Roll return is defined as the excess return on the S&P GSCI minus the spot of the S&P GSCI. Collateral return is defined as the total return on the S&P GSCI minus the excess return of the S&P GSCI. The excess return measures the returns accrued from investing in uncollateralized nearby commodity futures.
[2] Monthly returns are used.
[3] Risk is defined as annualized standard deviation.
[4] Correlation with the S&P GSCI Total Return.
Source: Author's research based on data from S&P Dow Jones Indices.

As can be seen in the table, over the past 40+ years, the S&P GSCI generated 6.8% in geometrically compounded annualized returns, with about three-quarters derived from interest rates (collateral return). The commodity price spot return component of the index (which has varied over time) contributed to approximately 45% of the total return (3.0% out of 6.8%), whereas the roll return subtracted from the overall return by –1.3% (or 130 bps) on an annualized basis. Investors can see the effect of commodities on inflation via the price return.

The volatility and correlations of the components of index returns are driven by the changes in the spot price return (effectively the same annualized standard deviation of 19.8% as the S&P GSCI with a 97% correlation). The roll return and collateral return do not drive, in general, the monthly returns historically. This link between commodity futures prices and commodity total return indexes helps to define commodities as a separate and investable asset class.

In summary, the total return on a fully collateralized commodity futures contract can be described as the spot price return plus the roll return plus collateral return (risk-free rate return). With an index, a return from rebalancing the index's component weights—a **rebalance return**—would also be added. Using historical data (at the risk of it becoming outdated over time), one can demonstratively use the total return deconstruction to analyze commodities.

EXAMPLE 17

Total Returns for Futures Contracts (1)

A commodity futures market with pricing in backwardation will exhibit which of the following characteristics?

A The roll return is usually negative.

B Rolling an expiring futures contract forward will require buying more contracts in order to maintain the same dollar position in the futures markets.

C Rolling an expiring futures contract forward will require buying fewer contracts in order to maintain the same dollar position in the futures markets.

Solution:

B is correct. Commodity futures markets in backwardation exhibit price curves in which longer-dated futures prices are priced lower than near-dated contracts and the nearest-dated contract is priced lower than the current spot price. With a lower futures price on the futures curve, rolling contracts forward in backwardation would require purchasing more contracts to maintain the same dollar position. A is incorrect because the roll return is usually positive, not negative, in markets in backwardation. C is incorrect because an investor would need to purchase more, not fewer, contracts in markets in backwardation to maintain his or her total dollar position.

EXAMPLE 18

Total Returns for Futures Contracts (2)

An investor has realized a 5% price return on a commodity futures contract position and a 2.5% roll return after all her contracts were rolled forward. She had held this position for one year with collateral equal to 100% of the position at a risk-free rate of 2% per year. Her total return on this position (annualized excluding leverage) was:

A 5.5%.

B 7.3%.

C 9.5%.

Solution:

C is correct. Total return on a commodity futures position is expressed as

Total return = Price return + Roll return + Collateral return.

In this case, she held the contracts for one year, so the price return of 5% is an annualized figure. In addition, the roll return is also an annual 2.5%. Her collateral return equals 2% per year × 100% initial collateral investment = 2%.

So, her total return (annualized) is

Total return = 5% + 2.5% + 2% = 9.5%.

EXAMPLE 19

Total Returns for Futures Contracts (3)

An investor has a $10,000 position in long futures contracts (for a hypothetical commodity) that he wants to roll forward. The current contracts, which are close to expiration, are valued at $4.00 per contract, whereas the longer-term contract he wants to roll into is valued at $2.50 per contract. What are the transactions—in terms of buying and selling new contracts—he needs to execute in order to maintain his current exposure?

A Close out (sell) 2,500 near-term contracts and initiate (buy) 4,000 of the longer-term contracts.

B Close out (buy) 2,500 near-term contracts and initiate (sell) 4,000 of the longer-term contracts.

C Let the 2,500 near-term contracts expire and use any proceeds to purchase an additional 2,500 of the longer-term contracts.

Solution:

A is correct. To roll over the same level of total exposure ($10,000), he will need to do the following:

Sell

$10,000/$4.00 per contract = 2,500 existing contracts.

And replace this position by purchasing

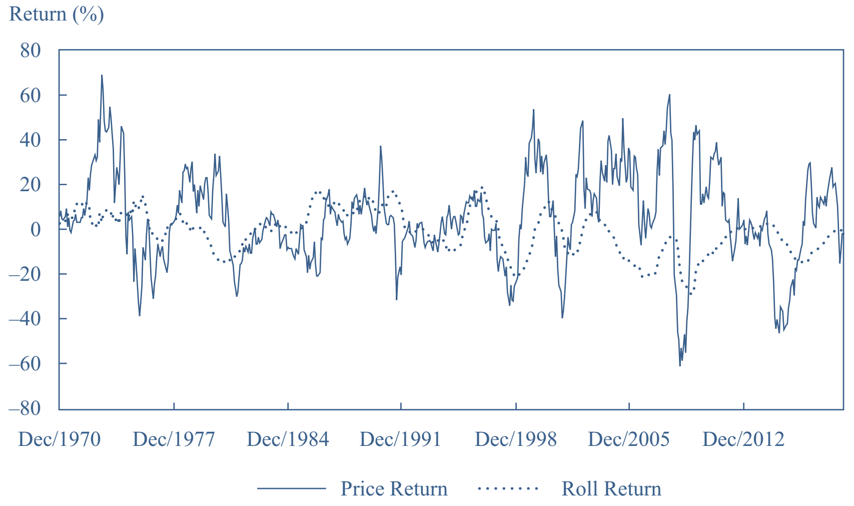
$10,000/$2.50 per contract = 4,000 existing contracts.

9 CONTANGO, BACKWARDATION, AND THE ROLL RETURN

h contrast roll return in markets in contango and markets in backwardation;

To reiterate, contango and backwardation—and the resulting roll return—fundamentally reflect underlying supply and demand expectations and are accounting mechanisms for the commodity term structure. We can gain a sense of these patterns by again examining the history of an index. Recall that from January 1970 to March 2019, the historical roll return of the S&P GSCI subtracted 1.3% from the average annual total return, with a standard deviation of 4.7%. That historical roll return varied over this time period, as depicted in Exhibit 13.

Exhibit 13 Historical One-Year S&P GSCI Price and Roll Return (Monthly Returns, January 1970–December 2019)



Note: The roll return is rolling monthly.

As the graph shows, periods of either backwardation or contango do not persist indefinitely. A simple review of the Exhibit 13 history demonstrates as much. Furthermore, with a correlation of 3%, roll return is not very indicative of price return, also contrary to popular belief. Positive price returns are associated with negative roll returns as well as positive roll returns. In some cases, certain sectors are indeed associated with contango, as can be seen in Exhibit 14.

Exhibit 14 Average Annual Sector Roll Return and Standard Deviation[a]

	S&P GSCI Total	Energy	Industrial Metals	Agriculture	Livestock	Precious Metals	Softs
Mean roll return (annual)[b]	−1.3%	−1.5%	−1.3%	−4.5%	−1.1%	−5.1%	−5.5%
Standard deviation of the mean (annual)[b]	0.4%	0.8%	0.5%	0.4%	0.5%	0.2%	0.6%
Maximum roll return (annual)[b]	18.9%	31.5%	45.9%	29.2%	35.5%	−0.4%	25.6%
Minimum roll return (annual)[b]	−29.6%	−39.5%	−16.6%	−18.6%	−31.2%	−15.4%	−24.9%

[a] The periods covered vary by sector:

- S&P GSCI total: December 1969–March 2019
- Energy: December 1982–March 2019
- Industrial metals: December 1976–March 2019
- Agriculture: December 1969–March 2019
- Livestock: December 1969–March 2019
- Precious metals: December 1972–March 2019
- Softs: December 1994–March 2019

[b] Calculated using rolling 12-month periods of monthly data.
Sources: Based on data from Bloomberg and Coloma Capital Futures.

Exhibit 14 highlights a few important factors. First, industrial metals, agriculture, livestock, precious metals, and softs have statistically strong negative mean roll returns. Only energy has a statistical possibility of a positive mean roll return, but that opportunity has diminished after 2010. Note from our comparison of the commodity sectors that industrial metals, agriculture, livestock, precious metals, and softs are stored for extended periods in warehouses, silos, and feedlots. In fact, precious metals historically have had negative roll returns because of gold's perpetual storage as an alternative currency. Historically, energy is consumed on a real-time basis apart from various strategic reserves, with the minimal storage buffer thus creating a lower or negative convenience yield. However, since 2010, the emergence of shale oil production in the United States has increased oil's convenience yield to the point that historical scarcity risk is much lower than before. Also, oil supply risk has shifted to China during this period as that country took over the United States' position as the lead oil importer. Finally, OPEC (with the inclusion of Russia and a few other non-OPEC members) regained some pricing power as the cartel achieved some success with supply restriction. Bringing it all together, one can conclude that indexes and long-only strategies that overweight agriculture, livestock, precious metals, and softs should expect to see negative roll returns (or roll yields). Energy commodities (apart from natural gas) have an opportunity for positive roll return, assuming producers successfully withhold supply from the market.

In conclusion, roll return can have an important impact on any single period return but overall has been relatively modest compared with price return. Furthermore, roll return is very sector dependent, which leads to a conclusion that sector diversification or concentration will have a profound impact on an investor's overall roll return based on a diversified portfolio of commodity futures.

EXAMPLE 20

Roll Return

When measuring its contribution to the total return of a commodity futures position, the roll return:

A typically has a significant contribution to total return over both single and multiple periods.

B typically has a modest contribution to total return in any single period but can be significant over multiple periods.

C is always close to zero.

Solution:

B is correct. Historically, the roll return has had a relatively modest impact on overall commodity futures return in the short term but can be meaningful over longer time periods. A is incorrect because the roll return is typically modest over shorter periods of time, as noted earlier. C is incorrect because futures contracts generate positive or negative roll returns, depending on the commodity and prevailing market conditions.

COMMODITY SWAPS

10

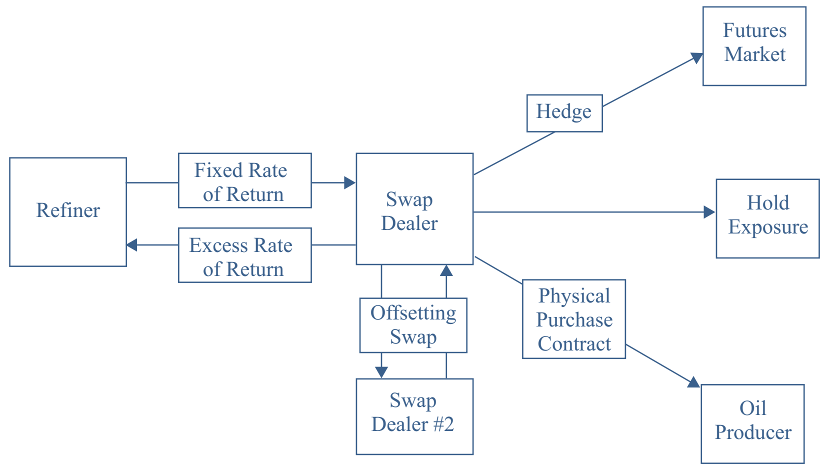
i. describe how commodity swaps are used to obtain or modify exposure to commodities;

Instead of futures, some investors can gain market exposure to or hedge risk of commodities via swaps. A **commodity swap** is a legal contract involving the exchange of payments over multiple dates as determined by specified reference prices or indexes relating to commodities. In the world of commodities, a series of futures contracts often forms the basis of the reference prices. For example, an independent oil refiner may want to hedge its oil purchases over an extended period. The refiner may not want to manage a large number of futures contracts but maintain flexibility with regard to its oil supply source. By entering into a swap contract—particularly one that is cash settled instead of physically settled—the refiner can be protected from a price spike and yet maintain flexibility of delivery.

Based on this example, one can see why commercial participants use swaps: The instrument provides both risk management and risk transfer while eliminating the need to set up and manage multiple futures contracts. Swaps also provide a degree of customization not possible with standardized futures contracts. The refiner in the example may negotiate a swap for a specific quality of crude oil (e.g., Heavy Louisiana Sweet instead of West Texas Intermediate, or WTI) as its reference price or a blend of crudes that shifts throughout the year depending on the season. Customization through the use of a swap may also have value by changing the quantity of crude oil hedged over time, such as lowering the exposure during the planned shutdown and maintenance periods at the refinery.

On the other side of the transaction from the refiner (or other hedging or speculating entity) would be a swap dealer, typically a financial intermediary, such as a bank or trading company. The dealer, in turn, may hedge its price risk exposure assumed in the swap through the futures market or, alternatively, negotiate its own swap with another party or arrange an oil purchase contract with a crude oil producer. The dealer may also choose to keep the price risk exposure, seeking to profit from its market information. A diagram demonstrating this swap transaction is shown in Exhibit 15.

Exhibit 15 Swap Market Participant Structure

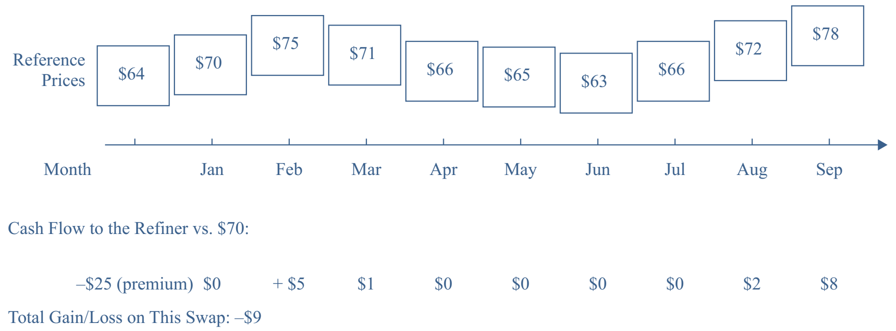


To further understand the diagram in Exhibit 15, assume we had the following scenario:

1 An oil refiner goes long a swap at the end of December that pays the amount exceeding $70 per barrel every month-end through September.
2 The oil refiner would pay a swap counterparty a premium (in this example, $25) for this privilege because it is effectively long a series of call options.

The flow of funds in the swap transaction would be as shown in Exhibit 16.

Exhibit 16 Flow of Funds for Swap Transaction Example



Total gain/loss on this swap to the refiner is –$9 (found by summing the cash flows and ignoring present value calculations or other considerations).

Although this example of a swap lost money and effectively increased the refiner's cost of a barrel of oil by $1 for this time period (given that the net loss on the swap was $9 over nine months), the swap protected the company against the risk of a cash squeeze during those months when an oil price spike could have impaired the liquidity of the company. The swap also defined the cost up front, giving a measure of cash flow predictability. Note that accounting standards and practices for swaps may also have an impact on the attractiveness of swaps. Given that oil prices are subject to many events beyond a company's control, a company looking to protect itself from financing risk may find that a swap can be a valuable tool.

There are many types of swaps available in the marketplace because they are not standardized, exchange-traded contracts like futures. The previous example of the refiner is an example of an "excess return swap." In an excess return swap, the payments to either party are driven primarily by the changes in price of each of the futures contracts that make up the index. The net change in the prices of the underlying futures contracts is defined as the "excess" return, and the excess return is multiplied by the contract's notional amount to determine the payments between buyer and seller.

10.1 Total Return Swap

Another common swap in commodities is a "total return swap." In a total return swap, the change in the level of the index will be equal to the returns generated by the change in price of each of the futures contracts that make up the index plus a return based on interest earned on any cash collateral posted on the purchase of the futures contracts that make up the index. If the level of the index increases, the swap buyer receives payment net of the fee paid to the seller; if the level of the index decreases between two valuation dates, the swap seller receives payment (plus the fee charged to the buyer). This type of swap is generally used by large institutional investors (e.g., pension plans) as opposed to commodity producers or buyers. With a total return swap, the investor seeks exposure to commodity returns, often because of the low return correlation of commodities with other asset classes (e.g., stocks or bonds) or as a reflection of the view that commodities provide a valuable inflation hedge for asset/liability matching (ALM). Therefore, such investors would engage in a total return swap that provides them with long exposure to the future returns from a commodity index that is used as the reference price. Again, accounting treatment with respect to futures often drives these decisions.

As an example of a total return swap, assume an investor who manages a defined benefit retirement plan desires commodity exposure for the reasons noted earlier. Given the size of the portfolio manager's plan assets (assume £2 billion), the manager is seeking approximately 5% exposure of plan assets to commodities. More specifically, the manager has decided that this £100 million exposure (5% of £2 billion) should be to the (hypothetical) China Futures Commodity Index (CFCI) and should remain for five years. Based on this decision, the manager issues a request for proposals (RFP) and, after evaluating the various bidders, contracts with a Swiss bank for a total return swap that will provide the desired exposure.

If on the first day of the swap agreement the CFCI increased by 1%, then the swap dealer would owe the manager £1 million (£100 million × 1%). If on the second day the CFCI declined by 5%, then the manager would owe £5 million to the dealer. Commonly, the dealer will hedge its short index exposure with futures or the physical commodity investments. Because the manager would be seeking the risk–return exposure offered by commodities, the manager would not generally hedge its exposure.

10.2 Basis Swap

Another common commodity swap is a basis swap, in which periodic payments are exchanged based on the values of two related commodity reference prices that are not perfectly correlated. These swaps are often used to adjust for the difference (called the basis) between a highly liquid futures contract in a commodity and an illiquid but related material. For example, a swap may pay the difference between the average daily prices of Brent crude oil (very liquid) and heavy crude oil available for delivery in the Gulf of Mexico (less liquid). This can be a very valuable arrangement for, in this example, refineries on the US Gulf Coast that have heavily invested in processing cheaper heavy crudes that come from such countries as Mexico or Venezuela. Because prices of these crudes do not always move in tandem with more common crudes, such as Brent, they derive a price basis between the two. It should be noted that "basis" has other meanings as well, depending on the commodity in question. For example, in grains, the basis may refer to the difference between the soybean contract and physical soybeans available for delivery at the Mississippi River.

10.3 Variance Swaps and Volatility Swaps

Two final types of relatively common commodity swaps are variance swaps and volatility swaps. Variance swaps of commodities are similar in concept to variance swaps of equities in that there is a variance buyer and a variance seller. Two parties agree to periodically exchange payments based on the proportional difference between an observed/actual variance in the price levels of a commodity (over consecutive time periods), and some fixed amount of variance established at the outset of the contract. If this difference is positive, the variance swap buyer receives a payment; if it is negative, the variance swap seller receives payment. Often the variance differences (observed versus fixed) are capped to limit upside and losses.

Volatility commodity swaps are very similar to variance swaps, with the exception that the direction and amount of payments are determined relative to the observed versus expected volatility for a reference price commodity. In this arrangement, the two sides are not speculating on the level or direction of prices but instead on how volatile prices will be versus expectations. A volatility seller will profit if realized volatility is lower than expectations, whereas the counterparty volatility buyer anticipates higher than expected volatility.

EXAMPLE 21

Commodity Swaps (1)

A portfolio manager enters into a \$100 million (notional) total return commodity swap to obtain a long position in commodity exposure. The position is reset monthly against a broad-based commodity index. At the end of the first month, the index is up 3%, and at the end of the second month, the index declines 2%. What are two payments that would occur between the portfolio manager and the swap dealer on the other side of the swap transaction?

A No payments are exchanged because a net cash flow only occurs when the swap agreement expires.

B \$3 million would be paid by the swap dealer to the portfolio manager (after Month 1), and \$2 million would be paid by the portfolio manager to the swap dealer (after Month 2).

C \$3 million would be paid by the portfolio manager to the swap dealer (after Month 1), and \$2 million would be paid by the swap dealer to the portfolio manager (after Month 2).

Solution:

B is correct. Because the portfolio manager has a long position in the total return commodity swap, he or she will receive payments when the commodity index rises and make payments when the commodity index declines. The payment calculations after the first two months are as follows:

Month 1: \$100 million × 3% = \$3 million.

Month 2: \$100 million × −2% = −\$2 million.

A is incorrect because swap payments are made periodically (in this case monthly) and not withheld to the end of the contract. C is incorrect because the payments would be in the opposite direction for each month.

EXAMPLE 22

Commodity Swaps (2)

In a commodity volatility swap, the direction and amount of payments are determined relative to the observed versus reference:

A direction in the price of a commodity.

B variance for the price of a commodity.

C volatility for the price of a commodity.

Solution:

C is correct. In a commodity volatility swap, the two sides of the transaction are speculating on expected volatility. A volatility seller will profit if realized volatility is lower than expectations, whereas the volatility buyer benefits from higher than expected volatility. A is incorrect because a volatility swap is based on price volatility, not direction. B is incorrect because a volatility swap is based on price volatility as opposed to price variance (price volatility squared).

COMMODITY INDEXES

11

j describe how the construction of commodity indexes affects index returns.

As in other parts of the investment universe, indexes have been created to portray the aggregate movement of commodity prices, investment vehicles, and investing approaches. In fact, one could say that an asset class does not exist without the presence of at least one representative index.

Commodity indexes play three primary roles in commodity sector investments. First, an index can be used as a benchmark to evaluate broader moves in commodity pricing. Second, as a broad indicator, an index can be used for macroeconomic or forecasting purposes by examining statistically significant relationships between movements in the commodity index and other macroeconomic variables. Finally, an index can act as the basis for an investment vehicle or contract providing the information needed to record, monitor, and evaluate price changes that affect contract value.

Although there are a number of commodity indexes, the following are used most frequently for the purposes just mentioned: (1) the S&P GSCI; (2) the Bloomberg Commodity Index (BCOM), formerly known as the Dow Jones–UBS Commodity Index (DJ–UBS); (3) the Deutsche Bank Liquid Commodity Index (DBLCI); (4) the

Thomson Reuters/CoreCommodity CRB Index (TR/CC CRB); and (5) the Rogers International Commodities Index (RICI). The following are key characteristics that differentiate each of these indexes:

- The *breadth* of coverage (number of commodities and sectors) included in each index, noting that some commodities have multiple reference contracts (e.g., for crude oil, the common contracts are for West Texas Intermediate in the United States and Brent crude for Europe).
- The relative *weightings* assigned to each component/commodity and the related methodology for how these weights are determined.
- The *rolling methodology* for determining how those contracts that are about to expire are rolled over into future months. This decision has a direct impact on the roll return (or yield) of the overall commodity. Recall that roll return is one of the three key components of overall commodity returns.
- The methodology and frequency for *rebalancing* the weights of the individual commodities, sectors, and contracts in the index to maintain the relative weightings assigned to each investment. As with stocks and bonds within a portfolio, the opportunity to earn positive rebalance returns for commodities depends on the correlation of the underlying components of the index and the propensity of underperforming components to revert back to the mean. For example, a drought may cause cotton prices to increase, but a strong crop the following year will cause prices to collapse. A rebalance sale of the overvalued cotton exposure into an undervalued exposure should "lock in" some of that gain. The rebalance return will likely vary depending on the methodology used by the index.
- The *governance* of indexes is important because it is the process by which all the aforementioned rules are implemented. For example, some indexes are rules-based, whereas others are selection-based. The rules-based indexes follow a quantitative methodology, whereas selection-based indexes are more qualitative in that an index committee picks the commodities. Also, governance oversees the independence of index providers so that, according to best practices of the Index Industry Association, the asset price should be independent from the index provider, which, in turn, should be independent from the product provider (e.g., the exchange-traded fund or swap provider).

For the index to be a viable and useful construct, it should be investable; that is, investors or their agents should be able to replicate the methodology outlined to translate the index concept into a representation of the asset class. For this reason, index providers and investors must be mindful of the venues (physical or electronic) for trading each commodity index, the liquidity and turnover of contracts based on each commodity index, and the term structure of each index (i.e., how far into the future the index extends and which months it covers). The weighting method for components in an index is key to diversification and—combined with rebalancing frequency—influences the opportunity to earn positive rebalance returns.

An index that requires investments in exchanges all over the world is more difficult and expensive for an investor to replicate. An emphasis on illiquid contracts has a negative impact on transaction costs. Contracts without a full yield curve may be a challenge to analyze and trade. In other words, seemingly small execution concerns are magnified when constructing a benchmark that represents an entire asset class, such as commodities. And indexes that choose (perhaps inadvertently) contracts that more commonly trade in backwardation may appear to improve forward-looking performance (because this generates a positive roll return), whereas those that more commonly trade in contango may hurt performance. Exhibit 17 summarizes the various elements of the main indexes discussed.

Exhibit 17 Overview of Major Commodity Indexes

	Index				
Element	S&P GSCI	BCOM	DBLCI	TR/CC CRB	RICI
Adoption date	1991	1998	2003	2005 (current version)	1998
Number of commodities	24	23	14	19	38
Weighting method	Production weighted	Production and liquidity weighted	Fixed weight	Fixed weight	Fixed weight
Rolling methodology	Nearby most liquid contract, monthly	Front month to next or second month	Optimized on roll return	Front month to next month	Front month to next month
Rebalancing frequency	Annually	Annually	Annually	Monthly	Monthly
Individual investor funds available?	Yes	Yes	Yes	Yes in some jurisdictions as well as an exchange-traded fund on a related index	Yes

Note: Information is as of 30 April 2019.
Sources: Information from respective sponsor websites, Bloomberg, and authors' research.

Exhibit 17 helps distinguish the key characteristics that differentiate these five commercially important commodity indexes. In terms of coverage (the number of commodities and sectors included in the index), all five of these indexes have broad sector coverage, including energy, grains, livestock, precious metals, industrial metals, and softs. The only exception is the DBLCI, which does not have any livestock exposure. At the other extreme, the RICI includes relatively exotic (and thus illiquid) commodities, such as lumber, oats, and rubber. As a further example of its unique nature, the RICI once included adzuki beans (the red beans found in many Asian cuisines) and palm oil.

11.1 S&P GSCI

The S&P GSCI is the second oldest of the selected commodity indexes. The index is based on 24 commodities and applies liquidity screens to include only those contracts with an established minimum level of trading volume and available historical pricing. It uses a world production value-weighting scheme that gives the largest weight to the most valuable commodity on the basis of physical trade value. It should be no surprise that crude oil has the highest single weight and energy has the highest sector weight (historically as high as 80%) in this index. This approach is most similar to a market-capitalization weighted index of nearly all major bond and stock market indexes. Like some market-capitalization indexes (particularly in emerging or frontier markets), the resulting weights of the S&P GSCI can be highly concentrated. The rolling methodology focuses on owning the front (i.e., near-term) contracts to address the highest liquidity and where supply and demand shocks are most likely to have an impact.

11.2 Bloomberg Commodity Index

The BCOM (formerly the DJ–UBS) is based on 23 commodities. It includes liquidity as both a weighting factor and a screening factor, although the index is selection-based, meaning a committee uses judgment to pick the included commodities. The rules of index construction also place caps on the size of the sectors (33% maximum) and floors on individual commodities (2% minimum). These differences mean that very different index composition and weights can occur. For example, the energy sector currently dominates the S&P GSCI (as high as 80% weight), whereas the BCOM's exposure is much lower (approximately 30%). However, exposure to natural gas as a single component of energy is higher in the BCOM (approximately 9%) than in the S&P GSCI (approximately 3%). Given that natural gas had an annualized roll cost of about 19% (often the highest roll cost of all the commodities), the higher weighting of natural gas in the BCOM implies that the index has to find other sources of return (e.g., price return and rebalance return) to overcome the drag that natural gas inventory storage creates through negative roll return. The rolling methodology focuses on owning the front (i.e., near-term) contracts.

11.3 Deutsche Bank Liquid Commodity Index

The DBLCI uses a fixed-weighting scheme to allocate exposure. The most notable/unique feature of this index is its rolling methodology. Instead of focusing on near-term contracts, it is optimized based on the time value of maximized backwardation/minimized contango for the contracts that fall within the next 12 calendar months. As an example, a June 2014 copper futures contract may be at 1% backwardation versus a May 2014 copper contract. But if the July 2014 copper contract is at a 3% backwardation (1.5% per month, or 3% divided by two months) versus the 1% backwardation per month on the June 2014 contract, then the DBLCI will roll to the July 2014 contract in preference to the June 2014 contract. Therefore, one could argue the DBLCI takes an active decision with regard to roll return positioning as compared with the other indexes.

11.4 Thomson Reuters/CoreCommodity CRB Index

The TR/CC CRB consists of 19 commodities and is a continuation of the first investable commodity index published by the Commodities Research Bureau in 1978 (although an earlier iteration started in 1957). It uses a fixed-weighting scheme to allocate exposure. An index management committee decides the weights based on a number of factors, including diversification, sector representation, liquidity, and economic importance. It also clusters the fixed weights into a number of tiers. As a result, constituents are moved from tier to tier. The rolling methodology focuses on owning the front (i.e., near-term) contracts that mechanically focus on the front month or second front month and do not require a particular calculation.

11.5 Rogers International Commodity Index

The RICI uses a fixed-weighting scheme to allocate exposure among 38 different commodities and was designed by investor Jim Rogers in the late 1990s. An index management committee decides the weights based on a number of factors, including diversification, sector representation, liquidity, and economic importance. Like the TR/CC CRB Index, it also clusters the fixed weights into a number of tiers. As a result, constituents are moved from tier to tier as they gain or lose relative importance as seen by the committee. Energy is the largest weight but is still a highly diversified basket.

Some energy constituents are denominated in non-US dollar terms—such as rubber (traded in Japan in Japanese yen) and cocoa (traded in London in British pounds)—which potentially adds a foreign exchange exposure element to the index returns.

11.6 Rebalancing Frequency

Rebalancing frequency plays a role in index returns, especially for those indexes that rebalance more frequently, such as the TR/CC CRB and RICI. Theoretically, from portfolio management theory, rebalancing is more important if a market is frequently mean reverting because there are more peaks to sell and valleys to buy. However, frequent rebalancing can lead to underperformance in a trending market because the outperforming assets are sold but continue up in price, whereas the underperforming assets are purchased but still drift lower.

The relative performance of the monthly rebalanced indexes (TR/CC CRB and RICI) versus the annual rebalance of the other indexes will depend on the length of time of price trends: More frequent mean reversions should favor the former two indexes, but a longer-term trend will more likely favor the annually rebalancing indexes. If an index uses a floating weighting scheme, such as production value (fully or partially), then the higher (lower) futures prices usually coincide with higher (lower) physical prices. Therefore, with this kind of approach, the magnitude of rebalancing weights is generally lower than for a fixed-weight scheme because the post-rebalance weights will generally drift in line with the current portfolio weights. As a result, the S&P GSCI and BCOM indexes typically have lower rebalancing costs and—in a trending market—have an opportunity to outperform their fixed-weight index counterparts, particularly those that have a relatively frequent rebalance period.

11.7 Commodity Index Summary

There is no dominant index based on a particular methodology. Relative performance will occur based on the circumstances of the markets and the time period examined. Evaluating which index is superior for a *long-term* investment generates modest if any value. Per the authors' research, these indexes all have been highly correlated (well above 70%) with each other and have had low (roughly 0%) correlations with traditional asset classes (e.g., US large-cap stocks, US bonds, international stocks). As with equities, for which there are many different index providers, commodity indexes act in parallel even when their returns (and Sharpe ratios) frequently differ dramatically over time.

EXAMPLE 23

Commodity Indexes (1)

All else being equal, compared with an equally weighted commodity index, a production value-weighted index (such as the S&P GSCI) will be:

A less sensitive to energy sector returns.

B more sensitive to energy sector returns.

C equally sensitive to energy sector returns.

Solution:

B is correct. The energy sector will make up a sizable portion of a production value-weighted index and thus will be a meaningful driver of returns for such an index. A is incorrect because a production value-weighted index will be more, not less, sensitive to the energy sector. C is incorrect because a production value-weighted index will be more, not equally, sensitive to the energy sector.

EXAMPLE 24

Commodity Indexes (2)

Which of the following statements is *not* correct regarding commodity futures indexes?

A Commodity sectors in backwardation typically improve index returns.

B An index that invests in several futures exchanges provides a high degree of diversification.

C Total returns of the major commodity indexes have low correlation with traditional asset classes, such as equities and bonds.

Solution:

B is correct. Commodity futures exchanges throughout the world are highly correlated and thus provide little diversification benefits. A is incorrect because markets in backwardation typically have positive roll yields and thus will likely improve index returns (although the price return may still not be positive and thus the total return may still be negative). C is incorrect because commodity index returns do indeed have historically low correlation with equities and bonds.

SUMMARY

- Commodities are a diverse asset class comprising various sectors: energy, grains, industrial (base) metals, livestock, precious metals, and softs (cash crops). Each of these sectors has a number of characteristics that are important in determining the supply and demand for each commodity, including ease of storage, geopolitics, and weather.
- Fundamental analysis of commodities relies on analyzing supply and demand for each of the products as well as estimating the reaction to the inevitable shocks to their equilibrium or underlying direction.
- The life cycle of commodities varies considerably depending on the economic, technical, and structural (i.e., industry, value chain) profile of each commodity as well as the sector. A short life cycle allows for relatively rapid adjustment to outside events, whereas a long life cycle generally limits the ability of the market to react.
- The valuation of commodities relative to that of equities and bonds can be summarized by noting that equities and bonds represent financial assets whereas commodities are physical assets. The valuation of commodities is not based on

the estimation of future profitability and cash flows but rather on a discounted forecast of future possible prices based on such factors as the supply and demand of the physical item.

- The commodity trading environment is similar to other asset classes, with three types of trading participants: (1) informed investors/hedgers, (2) speculators, and (3) arbitrageurs.
- Commodities have two general pricing forms: spot prices in the physical markets and futures prices for later delivery. The spot price is the current price to deliver or purchase a physical commodity at a specific location. A futures price is an exchange-based price agreed on to deliver or receive a defined quantity and often quality of a commodity at a future date.
- The difference between spot and futures prices is generally called the basis. When the spot price is higher than the futures price, it is called backwardation, and when it is lower, it is called contango. Backwardation and contango are also used to describe the relationship between two futures contracts of the same commodity.
- Commodity contracts can be settled by either cash or physical delivery.
- There are three primary theories of futures returns.
  - In insurance theory, commodity producers who are long the physical good are motived to sell the commodity for future delivery to hedge their production price risk exposure.
  - The hedging pressure hypothesis describes when producers along with consumers seek to protect themselves from commodity market price volatility by entering into price hedges to stabilize their projected profits and cash flow.
  - The theory of storage focuses on supply and demand dynamics of commodity inventories, including the concept of "convenience yield."
- The total return of a fully collateralized commodity futures contract can be quantified as the spot price return plus the roll return plus the collateral return (risk-free rate return).
- The roll return is effectively the weighted accounting difference (in percentage terms) between the near-term commodity futures contract price and the farther-term commodity futures contract price.
- A commodity swap is a legal contract between two parties calling for the exchange of payments over multiple dates as determined by several reference prices or indexes.
- The most relevant commodity swaps include excess return swaps, total return swaps, basis swaps, and variance/volatility swaps.
- The five primary commodity indexes based on assets are (1) the S&P GSCI; (2) the Bloomberg Commodity Index, formerly the Dow Jones–UBS Commodity Index; (3) the Deutsche Bank Liquid Commodity Index; (4) the Thomson Reuters/CoreCommodity CRB Index; and (5) the Rogers International Commodities Index.
- The key differentiating characteristics of commodity indexes are
  - the breadth and selection methodology of coverage (number of commodities and sectors) included in each index, noting that some commodities have multiple reference contracts,
  - the relative weightings assigned to each component/commodity and the related methodology for how these weights are determined,
  - the methodology and frequency for rolling the individual futures contracts,

- the methodology and frequency for rebalancing the weights of the individual commodities and sectors, and
- the governance that determines which commodities are selected.

REFERENCES

Cootner, Paul H. 1960. "Returns to Speculators: Telser versus Keynes." *Journal of Political Economy* 68 (4): 396–404.

De Roon, Frans A., Theo E. Nijman, and Chris Veld. 2000. "Hedging Pressure Effects in Futures Markets." *Journal of Finance* 55 (3): 1437–56.

Erb, Claude B., and Campbell R. Harvey. 2006. "The Strategic and Tactical Value of Commodity Futures." *Financial Analysts Journal* 62 (2): 69–97.

Fama, Eugene F., and Kenneth R. French. 1987. "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums and the Theory of Storage." *Journal of Business* 60 (1): 55–73.

Gorton, Gary, and K. Geert Rouwenhorst. 2006. "Facts and Fantasies about Commodity Futures." *Financial Analysts Journal* 62 (2): 47–68.

Hicks, John R. 1939. *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory*. London: Oxford University Press.

Ilmanen, Antti. 2011. *Expected Returns: An Investor's Guide to Harvesting Market Rewards*. Hoboken, NJ: John Wiley & Sons.

Kaldor, Nicholas. 1939. "Speculation and Economic Stability." *Review of Economic Studies* 7 (1): 1–27.

Keynes, John M. 1930. *The Applied Theory of Money*. vol. 2. A Treatise on Money. London: Macmillan.

Kolb, Robert W. 1992. "Is Normal Backwardation Normal?" *Journal of Futures Markets* 12 (1): 75–91.

PRACTICE PROBLEMS

The following information relates to Questions 1–8

Raffi Musicale is the portfolio manager for a defined benefit pension plan. He meets with Jenny Brown, market strategist with Menlo Bank, to discuss possible investment opportunities. The investment committee for the pension plan has recently approved expanding the plan's permitted asset mix to include alternative asset classes.

Brown proposes the Apex Commodity Fund (Apex Fund) offered by Menlo Bank as a potentially suitable investment for the pension plan. The Apex Fund attempts to produce trading profits by capitalizing on the mispricing between the spot and futures prices of commodities. The fund has access to storage facilities, allowing it to take delivery of commodities when necessary. The Apex Fund's current asset allocation is presented in Exhibit 1.

Exhibit 1 Apex Fund's Asset Allocation

Commodity Sector	Allocation (%)
Energy	31.9
Livestock	12.6
Softs	21.7
Precious metals	33.8

Brown explains that the Apex Fund has had historically low correlations with stocks and bonds, resulting in diversification benefits. Musicale asks Brown, "Can you identify a factor that affects the valuation of financial assets like stocks and bonds but does not affect the valuation of commodities?"

Brown shares selected futures contract data for three markets in which the Apex Fund invests. The futures data are presented in Exhibit 2.

Exhibit 2 Selected Commodity Futures Data*

Month	Gold Price	Coffee Price	Gasoline Price
July	1,301.2	0.9600	2.2701
September	1,301.2	0.9795	2.2076
December	1,301.2	1.0055	2.0307

* Gold: US$/troy ounce; coffee: US$/pound; gasoline: US$/gallon.

© 2019 CFA Institute. All rights reserved.

Menlo Bank recently released a report on the coffee market. Brown shares the key conclusion from the report with Musicale: "The coffee market had a global harvest that was greater than expected. Despite the large harvest, coffee futures trading activity is balanced between producers and consumers. This balanced condition is not expected to change over the next year."

Brown shows Musicale the total return of a recent trade executed by the Apex Fund. Brown explains that the Apex Fund took a fully collateralized long futures position in nearby soybean futures contracts at the quoted futures price of 865.0 (US cents/bushel). Three months later, the entire futures position was rolled when the near-term futures price was 877.0 and the farther-term futures price was 883.0. During the three-month period between the time that the initial long position was taken and the rolling of the contract, the collateral earned an annualized rate of 0.60%.

Brown tells Musicale that the pension fund could alternatively gain long exposure to commodities using the swap market. Brown and Musicale analyze the performance of a long position in an S&P GSCI total return swap having monthly resets and a notional amount of $25 million. Selected data on the S&P GSCI are presented in Exhibit 3.

Exhibit 3 Selected S&P GSCI Data

Reference Date	Index Level
April (swap initiation)	2,542.35
May	2,582.23
June	2,525.21

1 The Apex Fund is *most likely* to be characterized as:
   A a hedger.
   B a speculator.
   C an arbitrageur.

2 Which factor would *most likely* affect the supply or demand of all four sectors of the Apex Fund?
   A Weather
   B Spoilage
   C Government actions

3 The *most appropriate* response to Musicale's question regarding the valuation factor is:
   A storage costs.
   B transportation costs.
   C expected future cash flows.

4 Which futures market in Exhibit 2 is in backwardation?
   A Gold
   B Coffee
   C Gasoline

5 Based on the key conclusion from the Menlo Bank coffee market report, the shape of the coffee futures curve in Exhibit 2 is *most consistent* with the:
   A insurance theory.
   B theory of storage.

C hedging pressure hypothesis.

6 Based on Exhibit 2, which commodity's roll returns will *most likely* be positive?

A Gold

B Coffee

C Gasoline

7 The Apex Fund's three-month total return on the soybean futures trade is *closest* to:

A 0.85%.

B 1.30%.

C 2.22%.

8 Based on Exhibit 3, on the June settlement date, the party that is long the S&P GSCI total return swap will:

A owe a payment of $552,042.23.

B receive a payment of $1,502,621.33.

C receive a payment of $1,971,173.60.

The following information relates to Questions 9–15

Jamal Nabli is a portfolio manager at NextWave Commodities (NWC), a commodity-based hedge fund located in the United States. NWC's strategy uses a fixed-weighting scheme to allocate exposure among 12 commodities, and it is benchmarked against the Thomson Reuters/CoreCommodity CRB Index (TR/CC CRB). Nabli manages the energy and livestock sectors with the help of Sota Yamata, a junior analyst.

Nabli and Yamata meet to discuss a variety of factors that affect commodity values in the two sectors they manage. Yamata tells Nabli the following:

Statement 1 Storage costs are negatively related to futures prices.

Statement 2 In contrast to stocks and bonds, most commodity investments are made by using derivatives.

Statement 3 Commodities generate future cash flows beyond what can be realized through their purchase and sale.

Nabli and Yamata then discuss potential new investments in the energy sector. They review Brent crude oil futures data, which are presented in Exhibit 1.

Exhibit 1 Selected Data on Brent Crude Oil Futures

Spot Price	Near-Term Futures Price	Longer-Term Futures Price
77.56	73.64	73.59

Yamata presents his research related to the energy sector, which has the following conclusions:

- Consumers have been more concerned about prices than producers have.
- Energy is consumed on a real-time basis and requires minimal storage.

After concluding the discussion of the energy sector, Nabli reviews the performance of NWC's long position in lean hog futures contracts. Nabli notes that the portfolio earned a −12% price return on the lean hog futures position last year and a −24% roll return after the contracts were rolled forward. The position was held with collateral equal to 100% of the position at a risk-free rate of 1.2% per year.

Yamata asks Nabli to clarify how the state of the futures market affects roll returns. Nabli responds as follows:

Statement 4 Roll returns are generally negative when a futures market is in contango.

Statement 5 Roll returns are generally positive when a futures market is in backwardation.

As part of their expansion into new markets, NWC is considering changing its benchmark index. Nabli investigates two indexes as a possible replacement. These indexes both use similar weighting and rebalancing schemes. Index A includes contracts of commodities typically in contango, whereas Index B includes contracts of commodities typically in backwardation. Nabli asks Yamata how the two indexes perform relative to each other in a market that is trending upward.

Because of a substantial decline in drilling activity in the North Sea, Nabli believes the price of Brent crude oil will increase more than that of heavy crude oil. The actual price volatility of Brent crude oil has been lower than its expected volatility, and Nabli expects this trend to continue. Nabli also expects the level of the ICE Brent Index to increase from its current level. Nabli and Yamata discuss how to use swaps to take advantage of Nabli's expectations. The possible positions are (1) a basis swap long on Brent crude oil and short on heavy crude oil, (2) a long volatility swap on Brent crude oil, and (3) a short position in an excess return swap that is based on a fixed level (i.e., the current level) of the ICE Brent Index.

9 Which of Nabli's statements regarding the valuation and storage of commodities is correct?

A Statement 1

B Statement 2

C Statement 3

10 Based on Exhibit 1, Yamata should conclude that the:

A calendar spread for Brent crude oil is $3.97.

B Brent crude oil futures market is in backwardation.

C basis for the near-term Brent crude oil futures contract is $0.05 per barrel.

11 Based on Exhibit 1 and Yamata's research on the energy sector, the shape of the futures price curve for Brent crude oil is most consistent with the:

A insurance theory.

B theory of storage.

C hedging pressure hypothesis.

12 The total return (annualized excluding leverage) on the lean hog futures contract is:

A −37.2%.

B −36.0%.

C −34.8%.

13 Which of Nabli's statements about roll returns is correct?

A Only Statement 4

B Only Statement 5

C Both Statement 4 and Statement 5

14 The *best* response to Nabli's question about the relative performance of the two indexes is that Index B is *most likely* to exhibit returns that are:

A lower than those of Index A.

B the same as those of Index A.

C higher than those of index A.

15 Given Nabli's expectations for crude oil, the *most appropriate* swap position is the:

A basis swap.

B volatility swap.

C excess return swap.

The following information relates to Questions 16–22

Mary McNeil is the corporate treasurer at Farmhouse, which owns and operates several farms and ethanol production plants in the United States. McNeil's primary responsibility is risk management. Katrina Falk, a recently hired junior analyst at Farmhouse, works for McNeil in managing the risk of the firm's commodity price exposures. Farmhouse's risk management policy requires the use of futures to protect revenue from price volatility, regardless of forecasts of future prices, and prohibits risk managers from taking speculative positions.

McNeil meets with Falk to discuss recent developments in two of Farmhouse's commodity markets, grains and livestock. McNeil asks Falk about key characteristics of the two markets that affect revenues and costs. Falk tells McNeil the following:

Statement 1 The life cycle for livestock depends on the product and varies widely by product.

Statement 2 Grains have uniform, well-defined seasons and growth cycles specific to geographic regions.

A material portion of Farmhouse's revenue comes from livestock exports, and a major input cost is the cost of grains imported from outside the United States. Falk and McNeil next discuss three conclusions that Falk reached in an analysis of the grains and livestock markets:

Conclusion 1 Assuming demand for grains remains constant, extreme heat in the regions from which we import our grains will result in a benefit to us in the form of lower grain prices.

Conclusion 2 New tariffs on cattle introduced in our primary export markets will likely result in higher prices for our livestock products in our local market.

Conclusion 3 Major improvements in freezing technology allowing for longer storage will let us better manage the volatility in the prices of our livestock products.

McNeil asks Falk to gather spot and futures price data on live cattle, wheat, and soybeans, which are presented in Exhibit 1. Additionally, she observes that (1) the convenience yield of soybeans exceeds the costs of its direct storage and (2) commodity producers as a group are less interested in hedging in the forward market than commodity consumers are.

Exhibit 1 Selected Commodity Price Data*

Market	Live Cattle Price	Wheat Price	Soybeans Price
Spot	109	407	846
Futures	108	407	850

* Live cattle: US cents per pound; wheat and soybeans: US cents per bushel.

A key input cost for Farmhouse in producing ethanol is natural gas. McNeil uses positions in natural gas (NG) futures contracts to manage the risk of natural gas price volatility. Three months ago, she entered into a long position in natural gas futures at a futures price of $2.93 per million British thermal units (MMBtu). The current price of the same contract is $2.99. Exhibit 2 presents additional data about the three-month futures position.

Exhibit 2 Selected Information—Natural Gas Futures Three-Month Position*

			Prices	
Commodity	Total Current $ Exposure	Position	Near-Term Futures (Current Price)	Farther-Term Futures
Natural Gas (NG)	5,860,000	Long	2.99	3.03

* NG: $ per MMBtu; 1 contract = 10,000 MMBtu.

The futures position is fully collateralized earning a 3% rate. McNeil decides to roll forward her current exposure in the natural gas position.

Each month, McNeil reports the performance of the energy futures positions, including details on price returns, roll returns, and collateral returns, to the firm's executive committee. A new committee member is concerned about the negative roll returns on some of the positions. In a memo to McNeil, the committee member asks her to explain why she is not avoiding positions with negative roll returns.

16 With respect to its risk management policy, Farmhouse can be *best* described as:

A a trader.

B a hedger.

C an arbitrageur.

17 Which of Falk's statements regarding the characteristics of the grains and livestock markets is correct?

A Only Statement 1

B Only Statement 2

C Both Statement 1 and Statement 2

18 Which of Falk's conclusions regarding commodity markets is correct?

A Conclusion 1

B Conclusion 2

C Conclusion 3

19 Which commodity market in Exhibit 1 is currently in a state of contango?

A Wheat

B Soybeans

C Live cattle

20 Based on Exhibit 1 and McNeil's two observations, the futures price of soybeans is *most* consistent with the:

A insurance theory.

B theory of storage.

C hedging pressure hypothesis.

21 Based on Exhibit 2, the total return from the long position in natural gas futures is *closest* to:

A 1.46%.

B 3.71%.

C 4.14%.

22 The *most appropriate* response to the new committee member's question is that:

A roll returns are negatively correlated with price returns.

B such roll returns are the result of futures markets in backwardation.

C such positions may outperform other positions that have positive roll returns.

SOLUTIONS

1 C is correct. Commodity arbitrage involves an ability to inventory physical commodities and the attempt to capitalize on mispricing between the commodity (along with related storage and financing costs) and the futures price. The Apex Fund has access to storage facilities and uses these facilities in the attempt to capitalize on mispricing opportunities.

2 C is correct. Government actions can affect the supply or demand of all four sectors of the Apex Fund. With respect to energy, environmental mandates imposed by governments have tightened pollution standards, which have led to increasing processing costs that negatively affect demand. The supply of livestock, such as hogs and cattle, is affected by government-permitted use of drugs and growth hormones. Softs, or cash crops, can be affected by government actions, such as the attempt to maintain strategic stockpiles to control domestic prices. The level of demand and relative value of a precious metal, such as gold, is directly linked to government actions associated with managing to inflation targets.

3 C is correct. Expected future cash flows affect the valuation of financial assets, such as stocks and bonds, but do not affect the valuation of commodities. Financial assets (stocks and bonds) are valued based on expected future cash flows. In contrast, the valuation of a commodity is based on a discounted forecast of a future commodity price, which incorporates storage and transportation costs.

4 C is correct. When the near-term (i.e., closer to expiration) futures contract price is higher than the longer-term futures contract price, the futures market for the commodity is in backwardation. Because gasoline is the only one of the three futures markets in Exhibit 2 in which the near-term futures contract price ($2.2701) is higher than the longer-term contract price ($2.0307), the gasoline futures market is the only one in backwardation.

5 B is correct. The theory of storage focuses on the level of commodity inventories and the state of supply and demand. A commodity that is regularly stored should have a higher price in the future (contango) to account for those storage costs. Because coffee is a commodity that requires storage, its higher future price is consistent with the theory of storage.

6 C is correct. Roll returns are generally positive (negative) when the futures market is in backwardation (contango) and zero when the futures market is flat. Because the gasoline market is in backwardation, its roll returns will most likely be positive.

7 A is correct. The total return on the trade represents the sum of three components: price return, roll return, and collateral return.

Price return = (Current price − Previous price)/Previous price = (877.0 − 865.0)/865.0 = 1.387%.

Roll return = [(Near-term futures contract closing price − Farther-term futures contract closing price)/Near-term futures contract closing price] × Percentage of the position in the futures contract being rolled.

Because the entire position is being rolled, the percentage of the position in the futures contract being rolled is equal to 100%. So:

Roll return = [(877.0 − 883.0)/877.0] × 100% = −0.684%.

Collateral return = [3 months/12 months] × 0.60% = 0.15%.

Total return = 1.387% − 0.684% + 0.15% = 0.853%.

8 A is correct. The total return swap involves a monthly cash settlement (reset) based on the performance of the underlying reference asset (S&P GSCI) given a notional amount of $25 million. If the level of the index increases between the two valuation dates (in this case, May and June), the long position (the swap buyer) receives payment. If the level of the index decreases between the two valuation dates, the swap seller receives payment.

The return on the reference index for the month of June is [(2,525.21 − 2,582.23)/2,582.23], which is equivalent to −2.2082%. Therefore, the swap buyer (long position) must pay the swap seller a cash settlement for the month of June. The June payment calculation is equal to $25,000,000 × −2.2082%, or −$552,042.23.

9 B is correct. The most common way to invest in commodities is via derivatives, and commodities do not generate future cash flows beyond what can be realized through their purchase and sale. Also, storage costs are positively related to futures prices. Physical assets have to be stored, and storage incurs costs (rent, insurance, spoilage, etc.). Therefore, a commodity that is regularly stored should have a higher price in the future to account for those storage costs.

10 B is correct. The Brent crude oil futures market is in a state of backwardation. Commodity futures markets are in a state of backwardation when the spot price is greater than the price of near-term (i.e., nearest-to-expiration) futures contracts and, correspondingly, the price of near-term futures contracts is greater than that of longer-term contracts. The calendar spread is the difference between the near-term futures contract price and the longer-term futures contract price, which is $73.64 − $73.59 = $0.05. The basis for the near-term Brent crude oil futures contract is the difference between the spot price and the near-term futures price: $77.56 − $73.59 = $3.97.

11 B is correct. The Brent crude oil futures market is in a state of backwardation: The spot price is greater than the price of near-term (i.e., nearest-to-expiration) futures contracts. Commodities (in this case, Brent crude oil) are physical assets, not virtual assets, such as stocks and bonds. Physical assets have to be stored, and storage incurs costs (rent, insurance, inspections, spoilage, etc.). According to the theory of storage, a commodity that is consumed along a value chain that allows for just-in-time delivery and use (i.e., minimal inventories and storage) can avoid these costs. Yamata's research concluded that energy is consumed on a real-time basis and requires minimal storage. In this situation, demand dominates supply, and current prices are higher than futures prices (state of backwardation).

12 C is correct. The contract was held for one year, so the price return of −12% is an annualized figure. Additionally, the −24% roll return is also annualized. Nabli's collateral return equals 1.2% per year × 100% initial collateral investment = 1.2%. Therefore, the total return (annualized) is calculated as follows:

Total return = Price return + Roll return + Collateral return.

Total return = −12% + (−24%) + 1.2% = −34.8%.

13 C is correct. Roll returns are generally negative (positive) when the futures market is in contango (backwardation) and zero when the futures market is flat.

14 C is correct. Index B is likely to have higher performance than Index A in a market that is trending upward. Indexes that (perhaps inadvertently) contain contracts that more commonly trade in backwardation may improve forward-looking performance because this generates a positive roll return. Similarly, indexes that contain contracts that more commonly trade in contango may hurt performance for the same reason (i.e., negative roll return).

15 A is correct. Nabli expects the price of Brent crude oil to increase more than that of heavy crude oil, and Nabli can take advantage of this prediction by entering into a basis swap that is long Brent crude oil and short heavy crude oil. Nabli should take a short (not long) position in a volatility swap to take advantage of his prediction that Brent crude oil's price volatility will be lower than its expected volatility. Nabli should take a long (not short) position in an excess return swap to take advantage of his expectation that the level of the ICE Brent Index will increase faster than leading oil benchmarks.

16 B is correct. Hedgers trade in the futures markets to hedge their exposures related to the commodity, as stated in Farmhouse's risk management policy.

17 C is correct. The life cycle of livestock does vary widely by product. Grains have uniform, well-defined seasons and growth cycles specific to geographic regions. Therefore, both statements are correct.

18 C is correct. Commodity prices are affected by supply and demand, and improvements in freezing technology can improve the firm's ability to store its products for longer periods and manage the volatility of supply and demand. For example, during times of excess supply, a livestock producer, such as Farmhouse, can freeze its products and offer them during better market supply conditions.

19 B is correct. The futures market for soybeans is in a state of contango because the spot price is lower than the futures price.

20 C is correct. In Exhibit 1, the spot price of soybeans is less than the futures price. This observation can be explained only by the hedging pressure hypothesis. According to this hypothesis, hedging pressure occurs when both producers and consumers seek to protect themselves from commodity market price volatility by entering into price hedges to stabilize their projected profits and cash flows. If consumers are more interested in hedging than producers are, the futures price will exceed the spot price.

In contrast, the insurance theory predicts that the futures price has to be lower than the current spot price as a form of payment or remuneration to the speculator who takes on the price risk and provides price insurance to the commodity seller. Similarly, the theory of storage also predicts that when a commodity's convenience yield is greater than its direct storage costs, the futures price will be lower than the spot price.

21 A is correct. The total return for a fully collateralized position is the sum of the price return, the roll return, and the collateral return:

Price return = (Current price − Previous price)/Previous price

= (2.99 − 2.93)/2.93

= 2.05%.

Roll return = (Near-term futures closing price − Farther-term futures closing price)/Near-term futures closing price × Percentage of position in futures contract being rolled

= [(2.99 − 3.03)/2.99] × 100%

= −1.34%.

Collateral return = Annual rate × Period length as a fraction of the year

= 3% × 0.25

= 0.75%.

Therefore, the total return for three months = 2.05% − 1.34% + 0.75% = 1.46%.

22 C is correct. Investment positions are evaluated on the basis of total return, and the roll return is part of the total return. Even though negative roll return negatively affects the total return, this effect could be more than offset by positive price and collateral returns. Therefore, it is possible that positions with negative roll returns outperform positions with positive roll returns, depending on the price and collateral returns.

Portfolio Management

STUDY SESSIONS

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to explain and demonstrate the use of portfolio theory in risk and return estimation, security selection, and other practical applications. The candidate should also be able to explain the portfolio management process.

Portfolio management and risk management are key investment activities. Incorporating investor objectives, constraints, capital market expectations, and relevant risk considerations, together with portfolio construction, execution, and evaluation represent core activities in the investment process.

© 2021 CFA Institute. All rights reserved.

PORTFOLIO MANAGEMENT
STUDY SESSION 15

Portfolio Management (1)

This study session begins by examining exchange-traded funds (ETFs), including the creation and trading of ETFs, costs and risks of using ETFs, and how ETFs are used in strategic, tactical, and portfolio efficiency applications. Multifactor models including the arbitrage pricing theory (APT) and Carhart (4 factor) model are introduced as alternatives to the capital asset pricing model (CAPM). Considerations and applications of the three multifactor model types (macroeconomic, fundamental, statistical) are presented. Value at risk (VaR) and its use in measuring and managing market risk is discussed next. The three VaR approaches (parametric, historical simulation, Monte Carlo) along with the advantages and limitations of each are examined. The session ends with an introduction to backtesting and simulation.

READING ASSIGNMENTS

Reading 38	Exchange-Traded Funds: Mechanics and Applications by Joanne M. Hill, PhD, and Dave Nadig
Reading 39	Using Multifactor Models by Jerald E. Pinto, PhD, CFA, and Eugene L. Podkaminer, CFA
Reading 40	Measuring and Managing Market Risk by Don M. Chance, PhD, CFA, and Michelle McCarthy Beck
Reading 41	Backtesting and Simulation by Yin Luo, CPA, PStat, CFA and Sheng Wang

© 2021 CFA Institute. All rights reserved.

READING

38

Exchange-Traded Funds: Mechanics and Applications

by Joanne M. Hill, PhD, and Dave Nadig

Joanne M. Hill, PhD (USA). Dave Nadig is at ETFTrends.com and etfdb.com (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** explain the creation/redemption process of ETFs and the function of authorized participants;
☐	**b.** describe how ETFs are traded in secondary markets;
☐	**c.** describe sources of tracking error for ETFs;
☐	**d.** describe factors affecting ETF bid–ask spreads;
☐	**e.** describe sources of ETF premiums and discounts to NAV;
☐	**f.** describe costs of owning an ETF;
☐	**g.** describe types of ETF risk;
☐	**h.** identify and describe portfolio uses of ETFs.

INTRODUCTION

1

Exchange-traded funds (ETFs) have grown rapidly since their invention in the early 1990s, in large part because of their low associated cost, exchange access, holdings transparency, and range of asset classes available. Growth in ETFs has also been driven by the increased use of index-based investing. ETF investors need to understand how these products work and trade and how to choose from the numerous options available. Although many ETFs are organized under the same regulation as mutual fund products, there are important differences related to trading and tax efficiency. ETFs have features that can make them more tax efficient than traditional mutual funds, and not all ETFs are organized like mutual funds. ETFs can be based on derivative strategies, use leverage and shorting, and be offered in alternate structures, such as exchange-traded notes (ETNs), which have their own unique risks.

© 2021 CFA Institute . All rights reserved.

Understanding how ETF shares are created and redeemed is key to understanding how these products can add value in a portfolio. Because so many ETFs track indexes, understanding their index tracking or tracking error is also critical. Investors should also understand how to assess an ETF's trading costs, including differences between the ETF's market price and the fair value of its portfolio holdings.

We start with a discussion of the primary and secondary markets for ETFs, including the creation/redemption process, before moving on to important investor considerations, such as costs and risks. We then explain how ETFs are use in strategic, tactical, and portfolio efficiency applications.

2 ETF MECHANICS

a explain the creation/redemption process of ETFs and the function of authorized participants

b describe how ETFs are traded in secondary markets

Exchange-traded funds function differently from mutual funds because of their structure, with the key difference in an ETF's method of share creation and redemption. Mutual fund shares must be purchased or sold at the end of the day from the fund manager (or via a broker) at the closing net asset value (NAV) of the fund's holdings, in a cash-for-shares or shares-for-cash swap. In contrast, an ETF trades intraday, or during the trading day, just like a stock. ETF shares are created or redeemed in kind, in a shares-for-shares swap.

ETFs are intrinsically linked to the creation/redemption process. Creation/redemption enables ETFs to operate at lower cost and with greater tax efficiency than mutual funds and generally keeps ETF prices in line with their NAVs. Unlike stocks, which come to market via an initial public offering of fixed size, ETFs can be created or redeemed continuously. ETF transactions take place in two interrelated markets. Understanding how this mechanism works is key to understanding both the benefits and potential risks of ETFs.

The primary market for ETF trading is that which exists on an over-the-counter (OTC) basis between **authorized participants** (APs), a special group of institutional investors, and the ETF issuer, or sponsor. This process is referred to as **creation/redemption**. These primary market transactions are the only way that shares of the ETF can be created or redeemed. The "trade" in this market is in kind: A pre-specified basket of securities (which can include cash) is exchanged for a certain number of shares in the ETF.

ETF shares trade in the secondary market on exchanges. For investors, exchange trading is the only way to buy or sell ETFs. Like stocks, ETFs are bought and sold on exchanges through a brokerage account. This secondary market trading is perhaps the most novel feature of ETFs.

In-kind creation/redemption creates the unique benefits ETFs offer—as well as some of their risks. Here we explain ETFs' unique creation/redemption mechanism, the role of APs, and how the creation/redemption mechanism affects ETF design. ETF trading and settlement on primary and secondary markets is also covered.

2.1 The Creation/Redemption Process

The best way to understand the creation/redemption process is to step through the process from an investor's perspective.

Imagine you're an investor and you want to invest in an ETF. The process is simple: You place a buy order in your brokerage account the same way you would place an order to buy any publicly listed equity security, and your broker submits that order to the public market to find a willing seller: another investor or a market maker (i.e., a broker/dealer who stands ready to take the opposite side of the transaction). The order is executed, and you receive shares of the ETF in your brokerage account just as if you transacted in a stock.

At this point, the ETF manager (also referred to as the ETF issuer or sponsor) is not involved in the transaction. The ETF issuer does not know that you have bought these shares, nor does it receive an inflow of money to invest. Shares simply transfer in the open market, the secondary market for ETF shares, from one investor (the seller) to another (the buyer) and go through a settlement process based on the local exchange where the transaction took place. The process sounds simple, but if you can only buy ETF shares from another investor, where do the shares come from initially? How does money get invested into the fund?

The only investors who can create or redeem new shares of an ETF are a special group of institutional investors called *authorized participants*. APs are large broker/dealers, often market makers, who are authorized by the ETF issuer to participate in the creation/redemption process. The AP creates new ETF shares by transacting in kind with the ETF issuer. This in-kind swap happens off the exchange, in the primary market for the ETF, where APs transfer securities to (for creations) or receive securities from (for redemptions) the ETF issuer, in exchange for ETF shares. This is a prescribed, structured transaction with its own set of rules.

Each business day, the ETF manager publishes a list of required in-kind securities for each ETF. For instance, an S&P 500 Index ETF will typically list the index securities in quantities that reflect the index weighting. The list of securities specific to each ETF and disclosed publicly each day is called the **creation basket**. This basket also serves as the portfolio for determining the intrinsic net asset value of the ETF based on prices during the trading day.

To create new shares, an AP acquires the securities in the creation basket in the specified share amounts (generally by transacting in the public markets or using securities the AP happens to have in inventory). The AP then delivers this basket of securities to the ETF manager in exchange for an equal value in ETF shares. This exchange of shares happens after markets are closed through the settlement process. Importantly, the pricing of both the ETF and the basket is of minimal concern in this exchange: If the issuer receives 100 shares of a certain stock as part of the creation basket, the price the AP might have paid to acquire that stock or what its price happens to be at the end of the day is not relevant to the exchange taking place. Because it is an in-kind transaction, all that matters is that 100 shares of the required stock move from the AP's account to the ETF's account. Similarly, when the issuer delivers ETF shares to an AP, the ETF's closing NAV is not relevant.

These transactions between the AP and the ETF manager are done in large blocks called **creation units**, usually but not always equal to 50,000 shares of the ETF. This in-kind exchange involves the basket of underlying securities in exchange for a number of ETF shares of equal value.

The process also works in reverse: If the AP has a block of ETF shares it no longer wants (usually because it bought them from other market participants), the AP presents these shares for redemption to the ETF manager and receives in return the basket of underlying securities, which the AP can then sell in the market if it chooses. This basket often has the same security composition as the creation basket, but it may be different if the ETF portfolio manager is trying to sell particular securities for tax, compliance, or investment reasons. The basket of securities the AP receives when it redeems the ETF shares is called the **redemption basket**.

Although the actual process of exchanging baskets and blocks of ETF shares happens after the markets are closed, the AP is able to execute ETF trades throughout the trading day because the AP knows the security composition of the basket needed for ETF share creation or redemption, because of the fund's daily holdings disclosure to APs. If, during the course of the trading day, the AP wants to sell 50,000 shares of an ETF to investors in the secondary market, the AP can do so while simultaneously buying the securities in the creation basket. If the ETF and the securities in the creation basket are fairly priced, the AP faces no economic exposure in this transaction, because the value of the ETF shares sold and the value of the creation basket purchased are identical.

Why would APs engage in these transactions? Because there's a financial incentive to do so. The creation/redemption mechanism is key to keeping the price of an ETF in a tight range around the NAV of the portfolio of securities it holds, and it rewards the AP for this activity.

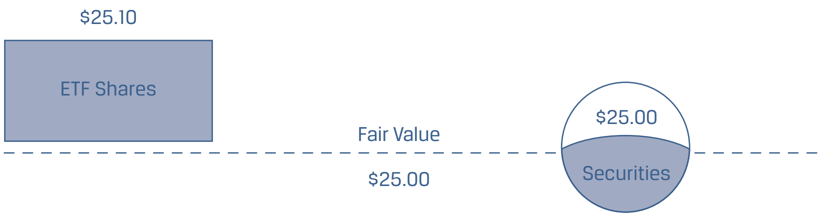
When the value of the security basket is different from the value of the corresponding ETF shares it represents, a potential arbitrage opportunity exists for APs to step in and transact in the ETF market. If the current per-share market value of the basket of underlying securities is greater than the quoted price of the ETF shares, the AP can simultaneously sell (or short) the basket of securities and buy ETF shares, to make a profit. In this situation, where the ETF share is undervalued, the ETF is said to be trading at a discount. If shares of the ETF are quoted at a higher price than the per-share market value of the basket of securities, the ETF is trading at a premium, and the AP can make a profit by simultaneously selling the ETF shares in the market and buying the basket of securities.

Because prices of the ETF and the basket securities are continuously changing on the basis of market conditions, APs monitor both for discrepancies, looking for opportunities to make arbitrage profits. The factors that drive the width of the ETF's bid–ask spread and trading range around intraday NAV include the cost of arbitrage (buying the securities and selling the ETF) and a risk premium to compensate for volatility and liquidity risk (ongoing volume in the securities and the ETF).

The *arbitrage gap*—the price(s) at which it makes sense for ETF market makers to step in and create or redeem shares—vary with the liquidity of the underlying securities and a variety of related costs; in some ETFs, the gap can be as small as the minimum tick size in the local market (e.g., –\$0.01 in the US markets), whereas for other ETFs with underlying securities that are hard to trade (e.g., high-yield bonds), the arbitrage gap can be more than 1% wide. For any ETF, however, the gap creates a band or range around its fair value inside which the ETF will trade. In other words, arbitrage keeps the ETF trading at or near its fair value.

ETF share creation. Let us examine how this works in practice. In the scenario shown in Exhibit 1, the ETF is trading in the market at \$25.10. The fair value of the ETF based on its underlying securities, however, is only \$25.00. So, an AP will step in to transact and buy the basket of securities (at ETF fair value of \$25.00) and simultaneously sell ETF shares on the open market for \$25.10, realizing the \$0.10 per share difference. (The AP may choose to create additional ETF shares by exchanging the basket securities for ETF shares with the fund's issuer).

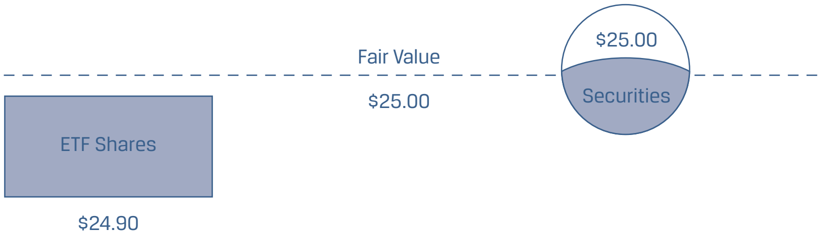
Exhibit 1 An ETF Share Price at a Premium to NAV



This action puts downward pressure on the ETF price because the AP is selling shares out into the market and puts upward pressure on the prices of the underlying securities because the AP went out into the market and bought the underlying shares. APs will repeat this process until no further arbitrage opportunity exists.

ETF share redemption As shown in Exhibit 2, the price of the ETF is $24.90. The fair value of the underlying stocks is $25.00. Here, the AP market maker steps in and purchases ETF shares on the open market while simultaneously selling the stocks on the exchange, realizing the $0.10 per share price difference. Once again, if the share price continues to be at a discount, the AP will continue this process until no further arbitrage opportunity exists. (The AP may choose to redeem ETF shares by exchanging them for the basket securities with the fund's issuer).

Exhibit 2 An ETF Share Price at a Discount to NAV



These profit-making scenarios do not include the costs that the APs incur related to ETF trading or any fees the issuer may charge for creating or redeeming shares. The AP generally pays all trading costs associated with buying or selling the securities in the baskets or the ETF shares and pays an additional fee to the ETF provider to cover processing fees associated with creation/redemption activities. APs may also have settlement costs, taxes, or other expenses based on their local markets and the markets for the underlying securities of the ETF.

The scenarios also do not account for risks in trading the basket of securities. If the underlying securities are difficult to access contemporaneously (for instance, if a US-listed ETF holds Japanese securities), then the AP will have to wait before completing one half of the transaction (e.g., selling the ETF shares but waiting until the Japanese market opens to buy the basket securities). These timing differences create uncertainty, which will generally cause the AP to wait for a wider arbitrage gap before stepping in. Similarly, if the basket securities are illiquid (such as high-yield bonds), the AP may need additional time to buy or sell the holdings. In both cases, the AP bears the market risk of the basket transaction.

A significant advantage of the ETF creation/redemption process is that the AP absorbs all costs of transacting the securities for the fund's portfolio. APs pass these costs to investors in the ETF's bid–ask spread, incurred by ETF buyers and sellers.

Thus, non-transacting shareholders of an ETF are shielded from the negative impact of transaction costs caused by other investors entering and exiting the fund. In contrast, when investors enter or exit a traditional mutual fund, the mutual fund manager incurs costs to buy or sell investments arising from this activity, which affect all fund shareholders. This makes the ETF structure inherently more fair: Frequent ETF traders bear the cost of their activity, whereas buy-and-hold ETF shareholders are shielded from those costs.

Additionally, because creation and redemption happen in kind, they allow the ETF's portfolio managers to manage the cost basis of their holdings by selecting low-basis holdings for redemptions, leading to greater tax efficiency. Put simply, when an issuer is presented with a redemption request from an AP, the issuer can select which tax lots of the underlying securities to deliver. In addition, issuers may choose to publish customized redemption baskets, which allows them to target specific low-basis securities for removal from the portfolio. By delivering out shares that were originally acquired at low costs, the issuer can continuously raise the average acquired cost (or cost basis) of each position, thereby minimizing the position's unrealized gains.

The ETF issuer has the ability to determine how the process works for a fund. If the issuer requires that a creation basket be 200,000 shares instead of 50,000 shares, the AP will have less incentive to step in to arbitrage when net new demand is lower than 200,000 shares per day. Basket sizes range from 10,000 shares to 600,000 shares. If the ETF holds highly illiquid securities, the issuer can alter the basket that APs must deliver, thereby lowering the costs of creation. In the most extreme case, the fund may allow for the creation of ETF shares in exchange for cash. Issuers can also charge minimal or large fees for creation and redemption, which affect an AP's profit consideration and transactions, to keep prices in line with fair value. Consider the fee of $50 for the Vanguard Short-Term Inflation-Protected Securities ETF (VTIP) versus the fee of $28,000 for the Vanguard FTSE All-World Ex-US Small-Cap ETF (VSS).

Creation/Redemption Asset Class Differences

The creation/redemption mechanism described is broadly representative of how most ETFs work, regardless of their particular legal structure. Depending on the asset class, however, some differences exist.

Fixed-income ETFs generally hold large amounts of bonds, which may be illiquid to trade (for example, a high-yield municipal bond ETF holds securities that might trade only every few days). Because of this, ETF issuers may choose not to do in-kind creations and redemptions but instead accept equivalent cash value. This makes the process easier for APs, encouraging greater ETF activity, but does result in trading costs and tax impact for the ETF. ETF issuers must balance those costs against the benefit of having the AP participate more actively in the market, keeping spreads tight and the price of the ETF close to fair value.

Similarly, many leveraged and inverse ETFs and commodity ETFs may use cash creation/redemption because it makes managing their underlying swap positions easier. Because swaps are generally negotiated OTC transactions, it would be difficult to have APs participate in increasing or decreasing those swap positions.

2.2 Trading and Settlement

There is much confusion in the investor community regarding the underlying mechanics of ETF trading and settlement. Whether this confusion relates to shorting, how shares are created/redeemed or settled, or how they trade, ETFs are potentially confusing to many investors. From the perspective of an investor buying on the open market, ETFs go through the same settlement and clearing process as other listed stocks. This section explains that process as it applies in two regions.

US settlement: National Security Clearing Corporation and Depository Trust Company. In the United States, all trades that have been entered into on a given business day are submitted at the end of the day to the National Security Clearing Corporation (NSCC). As long as both parties of a transaction agree that Party 1 sold to Party 2 *N* shares of XYZ stock, the NSCC becomes the guarantor of that transaction—the entity that ensures all parties are immunized against the financial impact of any operational problems—on the evening of the trade, and the trade is considered "cleared." After this point, the buyer is guaranteed beneficial ownership in the stock (or ETF) as of the time the trade was marked "executed," even if something (e.g., bankruptcy) happens to the seller before the trade is settled.

The Depository Trust Company (DTC), of which the NSCC is a subsidiary, holds the book of accounts—the actual list of security holders and ownership. This information is aggregated at the member firm level, rather than at the individual investor level. For instance, the DTC keeps track of how many shares of Microsoft are currently held by J.P. Morgan or Charles Schwab, but Charles Schwab is responsible for keeping track of which of its customers own how many shares.

After each trade is cleared, the DTC then adds up the total of all trades in a process of continuous net settlement. For example, suppose at the end of a trading day the following is true:

- E*TRADE owes Schwab 1,000 shares of SPY.
- Schwab owes Bank of America Merrill Lynch 1,000 shares of SPY.

Then, from the DTC's perspective, Schwab is "whole": It both is owed and owes 1,000 shares of SPY. To settle the day's transactions, E*TRADE's account will be debited the 1,000 shares of SPY and Bank of America Merrill Lynch will be credited 1,000 shares.

The NSCC has two days to complete this process and have each firm review its records and correct any discrepancies. We refer to this two-day period as T+2 (trade date + 2 days). This T+2 settlement process works for the vast majority of ETF transactions.

Market makers receive special treatment on settlement requirements. Because the role of market makers is to make a continuous market in a given security by standing ready to buy or sell the security on the basis of demand/supply imbalances, they are more likely to end up truly short at the end of a given day. Because of the time required to create or borrow ETF shares, market makers are given up to six days to settle their accounts.

European trading and settlement. In Europe, the majority of ETF owners are institutional investors. Additionally, the market is fragmented across multiple exchanges, jurisdictions, and clearinghouses. This fragmentation results in the use of many different trading strategies by investors in both the primary and secondary markets for ETFs. Fundamentally, trading works the same as in the United States: An investor purchases shares in the secondary market from a market maker or other counterparty. APs use the creation/redemption mechanism, which helps keep the ETF share price in line with its fair value.

The majority of trading happens in negotiated OTC trades between large institutions, and although those trades are reported, they do not appear as "live" or published bids and asks on the public markets prior to their execution. Most ETFs in Europe are also cross-listed on multiple exchanges and may have different share classes available that vary in their treatment of distributions or currency hedging. The fragmented European settlement process means that trades are cleared to one of 29 central securities depositories (or CSDs). This has no direct impact on investors other than the inherent complexity of such a system, which may result in wider spreads and higher local market trading costs.

3 UNDERSTANDING ETFS

c describe sources of tracking error for ETFs

Among the most important questions an investor can ask about an ETF is, Does the fund deliver on its promise? The best-managed ETFs charge low and predictable investment costs, closely track the indexes on which they are based, and provide investors with the lowest possible tax exposure for the investment objective. Additionally, these funds provide complete, accurate information in their prospectuses and marketing materials and explain the fund's structure, composition, performance, and risks. To best understand an ETF's ability to meet expectations, its expense ratio, index tracking, tax treatment, and potential costs and risks should be considered.

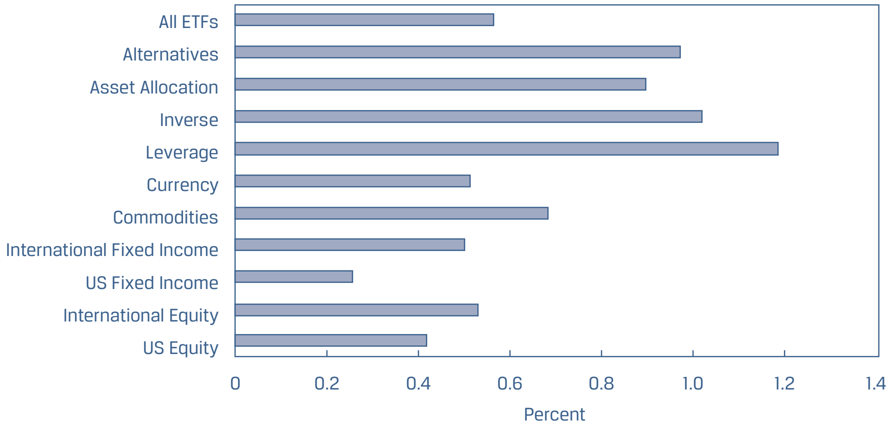
3.1 Expense Ratios

Fund expense ratios are often one of the first factors investors look at when evaluating ETFs. ETFs generally charge lower fees than mutual funds, in part because ETF providers do not have to keep track of individual investor accounts, since ETF shares are held by and transacted through brokerage firms. Nor do ETF issuers bear the costs of communicating directly with individual investors. In addition, index-based portfolio management, used by most ETFs, does not require the security and macroeconomic research carried out by active managers, which increases fund operating costs.

The actual costs to manage an ETF vary, depending on portfolio complexity (number of securities held, frequency of rebalancing or strategy implementation, difficulty in maintaining portfolio exposures), issuer size (economies of scale apply), and the competitive landscape.

ETF expense ratios have been one of the most visible areas of competitive differentiation for issuers, which has led to an overall decline in fees. Exhibit 3 shows average US-domiciled ETF expense ratios by asset class at the end of 2018.

Exhibit 3 Average US-Domiciled ETF Expense Ratios by Asset Class at the End of 2018



Sources: ETF.com and FactSet, as of 31 December 2018.

Because the average numbers include complex and expensive funds, they dramatically overstate the cost of accessing the most common ETF investment strategies and indexes.

As of the end of 2018, expense ratios for broad-based, capitalization-weighted indexes were as low as 0.03% for US equities, 0.11% for emerging market equities, and 0.04% for US bonds.

3.2 Index Tracking/Tracking Error

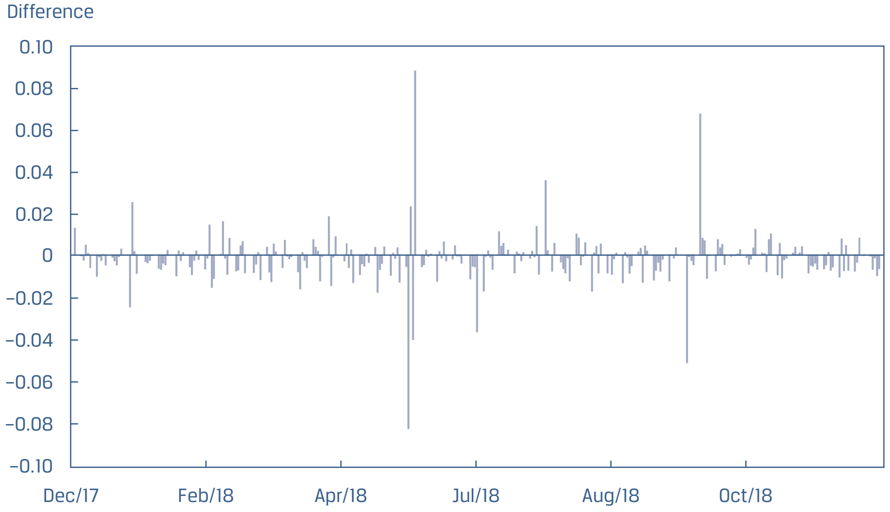
Even though an ETF's expense ratio is useful, it does not fully reflect the cost of holding an ETF. To understand how well an ETF delivers on its mandate, it is critical to assess the ETF's ability to track its underlying index.

For index-tracking ETFs, which represented 98% of the US ETF market as measured by assets under management (AUM) as of December 2018, ETF managers attempt to deliver performance that tracks the fund's benchmark as closely as possible (after subtracting fees). This can be measured by comparing ETF performance with index returns. The comparison can be done using daily or periodic returns but should always include both a central tendency, such as mean or median, and an expression of variability, such as standard deviation or range.

Daily differences.

Index tracking is often evaluated using the one-day difference in returns between the fund, as measured by its NAV, and its index. Exhibit 4 shows the daily tracking difference between the iShares MSCI Emerging Markets ETF (EEM) and its underlying index, the MSCI Emerging Markets Index (EMI), for a one-year period. EMI is a multicurrency international index containing hundreds of illiquid securities in more than 20 emerging markets. The index represents large- and mid-cap stocks in each of these markets. At the end of November 2018, EEM held approximately 900 of the 1,150 constituents in EMI.

Exhibit 4 EEM Daily Tracking Difference Relative to EMI, One-Year Period Ending 30 November 2018



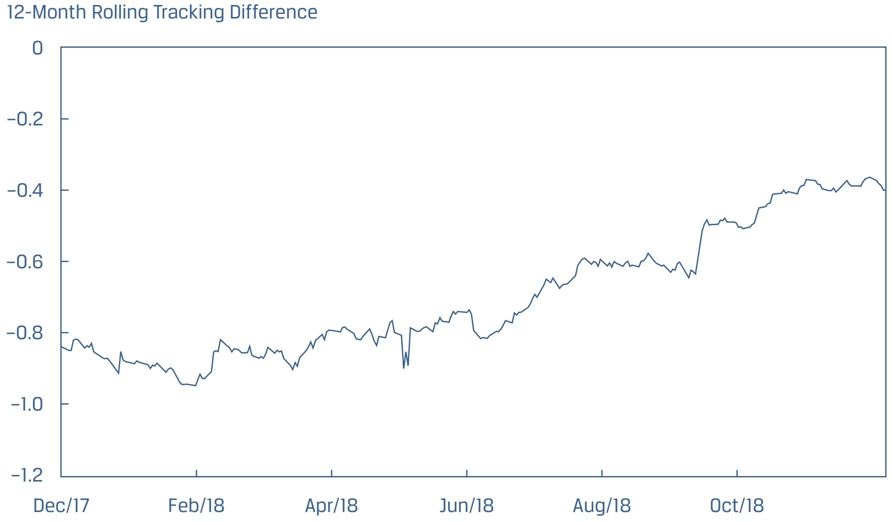
Source: FactSet.

Periodic tracking. Tracking error is defined as the standard deviation of differences in daily performance between the index and the fund tracking the index, and a reported tracking error number is typically for a 12-month period. Over the period shown, EEM's standard deviation of daily performance differences to its index was 0.012%.

But importantly, tracking error does not reveal the extent to which the fund is under- or overperforming its index or anything about the distribution of errors. Daily tracking error could be concentrated over a few days or more consistently experienced. Therefore, tracking error should be assessed with the mean or median values.

An alternative approach is to look at tracking differences calculated over a longer holding period. A series of rolling holding periods can be used to represent both central tendencies and variability. This approach allows investors to see the cumulative effect of portfolio management and expenses over an extended period. Exhibit 5 shows the 12-month rolling return (or cumulative annual) tracking difference between EEM and its index.

Exhibit 5 EEM 12-Month Rolling Tracking Difference Relative to EMI, One-Year Periods Ending 30 November 2018



Source: FactSet.

One benefit of the rolling annual analysis is that it allows for comparison with other annual metrics, such as the fund's expense ratio. All else equal, one would normally expect an index fund to underperform its benchmark on an annual basis by the amount of its expense ratio. In Exhibit 5, EEM's median tracking difference of 0.79% exceeded its 0.69% expense ratio by 0.10%. Notably, the range of EEM's annual tracking difference showed some variability, with underperformance as low as 0.38% and as high as 0.95%.

Sources of tracking error. Numerous factors can account for differences between an ETF's expected and actual performance and the range of results with respect to its index. Because of this, funds tracking the same underlying index can have very different index tracking results. Sources of benchmark tracking error include the following:

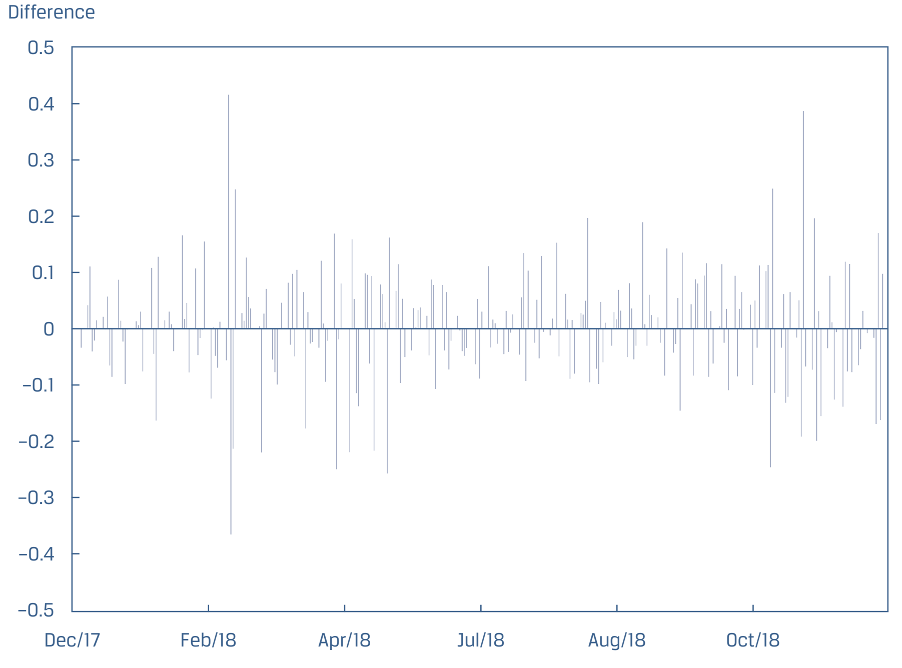
- Fees and expenses—Index calculation generally assumes that trading is frictionless and occurs at the closing price. A fund's operating fees and expenses reduce the fund's return relative to the index.
- Representative sampling/optimization—Rather than fully replicate the index, funds may hold only a subset of index securities to track the benchmark index.
- Depositary receipts and other ETFs—Funds may hold securities that are different from those in the index, such as American depositary receipts (ADRs), global depositary receipts (GDRs), and other ETFs.
- Index changes—Funds may trade index changes at times and prices that are different from those of the benchmark tracked.
- Fund accounting practices—Fund accounting practices may differ from the index calculation methodology—for example, valuation practices for foreign exchange and fixed income.

- Regulatory and tax requirements—Funds may be subject to regulatory and tax requirements that are different from those assumed in index methodology, such as with foreign dividend withholding.
- Asset manager operations—ETF issuers may attempt to offset costs through security lending and foreign dividend recapture. These act as "negative" costs, which enhance fund performance relative to the index.

Fees and expenses. As outlined in the prior sections, fund operating expenses vary by ETF, but all else equal, one would normally expect an index fund to underperform its benchmark on an annual basis by the amount of its expense ratio.

Representative sampling/optimization. For funds tracking index exposure to small or illiquid markets, owning every index constituent can be difficult and costly. Therefore, fund managers may choose to optimize their portfolios by holding only a portion, or representative sample, of index securities. A striking example is the SPDR S&P Emerging Asia Pacific ETF (GMF). As of 7 December 2018, GMF held only 763 of the 2,342 securities in the S&P Asia Pacific Emerging BMI Index. As shown in Exhibit 6, sampling has caused some sizable discrepancies between the fund's daily return and the index.

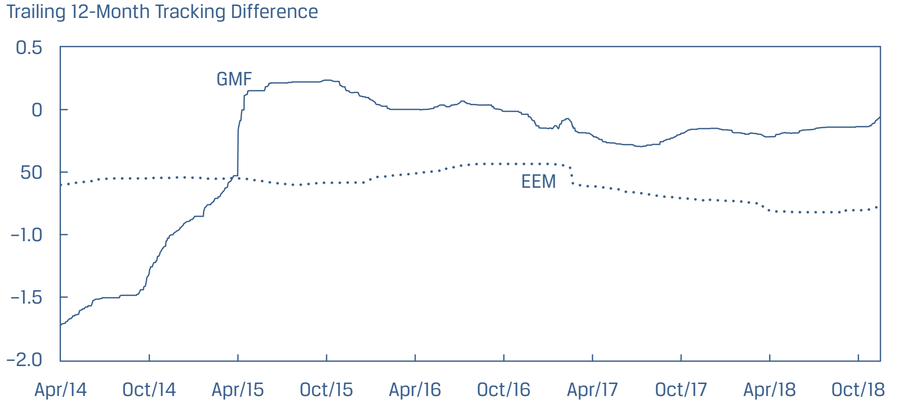
Exhibit 6 GMF Daily Tracking Difference Relative to the S&P Asia Pacific Emerging BMI Index, One-Year Period Ending 30 November 2018



Source: FactSet.

Sampling, or optimization, can affect long-term tracking in two ways. First, it can make the median value unpredictive of future median values, especially if market regimes shift. Second, it dramatically expands the range of results. Exhibit 7 and the table below illustrate these effects, using trailing 12-month (TTM) rolling comparisons. Exhibit 7 contrasts EEM's median trailing 12-month tracking difference with GMF's more variable results.

Exhibit 7 Trailing 12-Month Tracking Difference: EEM and GMF



Source: FactSet.

A high level of optimization causes GMF's portfolio to underperform in certain market regimes and outperform in others. Looking at the differences between GMF and its underlying index explains why. As of 30 November 2018, GMF's median constituent market cap was $2.8 billion, whereas the S&P Asia Pacific Emerging BMI's was $0.695 billion, indicating that by holding approximately one-third of index constituents, GMF's portfolio omits many of the index's mid-caps and small-caps. Therefore, GMF will likely underperform the index during times when emerging market mid-caps and small-caps outperform emerging market large-caps, and vice versa.

As illustrated in the following table, GMF's tracking range—the spread between its maximum and minimum trailing 12-month tracking difference—is nearly 4 times that of EEM. A higher level of optimization within GMF causes it to have a wider range of tracking difference relative to its index.

EEM and GMF Tracking Range, One Year Ending 30 November 2018

	EEM	GMF
Maximum TTM	−0.38%	1.14%
Minimum TTM	−0.95%	−0.81%
Range	0.57%	1.95%

Source: FactSet.

Representative sampling/optimization, therefore, enhances or detracts from fund returns relative to the index depending on whether ETF portfolio holdings outperform or underperform those in the index. Compared with a full replication approach, representative sampling/optimization introduces greater potential for tracking error.

Depositary receipts and ETFs. When local market shares are illiquid, ETF portfolio managers may choose to hold depositary receipts instead of local constituent shares. Although the economic exposure is equivalent, exchange trading hours for these securities differ. Differences in trading hours and security prices create discrepancies between portfolio and index values. Similarly, ETF issuers may choose to hold ETFs as underlying holdings. This also creates discrepancies between fund NAV and index value, because the ETFs' holdings are valued at their closing market price and not their NAV.

Index changes. An index provider will periodically change index constituents or weights to comply with its index methodology. In the real world, portfolio managers may transact these changes before or after the effective date or time of the index change/closing prices, at different prices. The more volatile the market, the wider the bid–offer spreads and range of traded prices. ETF portfolio managers can use the creation/redemption process to manage rebalance trades, by cooperating with APs to ensure market-on-close pricing on the rebalance date, thus minimizing this source of tracking error.

Fund accounting practices. Differences in valuation practices between the fund and its index can create discrepancies that magnify daily tracking differences. Some ETF issuers follow the index industry's convention of establishing (striking) currency valuations using WM/Reuters rates, which are set at 4:00 p.m. GMT (11:00 a.m. EST), whereas others conform to established mutual fund industry practices of striking currency valuations at the close of ETF trading. In the United States, equity markets close at 4:00 p.m. ET. Many fixed-income ETF portfolios value bond positions at the time of the equity market close, in keeping with ETF industry custom. However, fixed-income indexes often follow the bond market's practice of valuing bonds at an earlier time. These practices may create valuation discrepancies between the ETF's NAV and the index value, particularly in volatile segments of the bond market, such as long-dated maturities. Valuation discrepancies can also occur for ETFs holding futures, foreign securities, physical metals, and currencies held in specie.

Regulatory and tax requirements. Regulatory and tax requirements may cause a fund to mis-track its index. For example, non-domestic holders of a nation's securities owe tax on distributions received from securities of companies domiciled in that nation. The tax withholding rate charged is determined by treaty and investor domicile. Index providers who offer a "net" return series adjust the dividends received to account for the tax charged, usually from the point of view of US-domiciled investors. Index providers may use rates different from those experienced by the ETF, however, which can create return differences between the ETF and its index. For many years, Brazil imposed a tax on foreign investments coming into the country. Although this tax did not affect the closing prices of the local stocks and, therefore, was not reflected in index calculation methodology, non-local ETFs domiciled outside Brazil paid this tax whenever they acquired Brazilian stocks. This caused fund underperformance relative to the index.

Asset manager operations. ETF issuers may engage in security lending or foreign dividend recapture to generate additional income to offset fund expenses. These can be considered "negative" costs. Many ETFs (and mutual funds) lend a portion of their portfolio holdings to short sellers. In exchange, the ETF receives a fee and earns interest on the collateral posted by the borrower (generally, overnight fixed-income securities), which creates income for the portfolio. Because the index calculation does not account for securities-lending income, it is a source of tracking error. Asset managers may work with foreign governments to minimize tax paid on distributions received.

3.3 Tax Issues

Two kinds of tax-based evaluations must be made for all ETFs: First, the investor must consider the likelihood of an ETF distributing capital gains to shareholders. Second, the investor must consider what happens when the investor sells the ETF. These two actions are distinct; the tax efficiency of a fund regarding its capital gains distributions has no relation to its tax efficiency at the time of investor sale.

3.3.1 *Capital Gains Distributions*

The issue of capital gains distributions affects all investors in taxable accounts. In general, funds must distribute any capital gains realized during the year. Funds typically make these distributions at year-end, although they may make them quarterly or on another periodic schedule.

ETFs are said to be "tax fair" and "tax efficient" because they have certain advantages over traditional mutual funds regarding capital gains distributions. On average, they distribute less in capital gains than competing mutual funds for two primary reasons.

Tax fairness. In a traditional mutual fund, when an investor sells, the fund must (with a few exceptions) sell portfolio securities to raise cash to pay the investor. Any securities sold at a profit incur a capital gains charge, which is distributed to remaining shareholders. Put another way, in a traditional mutual fund, shareholders may have to pay tax liabilities triggered by other shareholders redeeming out of the fund.

In contrast, an investor sells ETF shares to another investor in the secondary market. The ETF manager typically does not know that the sale is occurring and does not need to alter the portfolio to accommodate this transaction. Thus, the selling activities of individual investors in the secondary market do not require the fund to trade out of its underlying positions. If an AP redeems ETF shares, this redemption occurs in kind. In markets where redemptions in kind are allowed, this is not a taxable event. Thus, redemptions do not trigger capital gain realizations. This aspect is why ETFs are considered "tax fair": The actions of investors selling shares of the fund do not influence the tax liabilities for remaining fund shareholders.

Tax efficiency. The redemption process allows portfolio managers to manage the fund's tax liability. When an authorized participant submits shares of an ETF for redemption, the ETF manager can choose which underlying share lots to deliver in the redemption basket. By choosing shares with the largest unrealized capital gains—that is, those acquired at the lowest cost basis—ETF managers can use the in-kind redemption process to reduce potential capital gains in the fund. Tax lot management allows portfolio managers to limit the unrealized gains in a portfolio.

3.3.2 *Other Distributions*

Other events, such as security dividend distributions, can trigger tax liabilities for investors but the treatment varies by region, so investors must ensure they understand the tax treatment specific to each fund's domicile, legal structure, and portfolio type.

3.3.3 *Taxes on Sale*

In most jurisdictions, ETFs are taxed according to their underlying holdings. For example, in the United States, an ETF holding equities or bonds will itself be subject to the same capital gain, dividend, and return-of-capital tax rules that apply to its underlying stock or bond holdings. There can be nuances in individual tax jurisdictions, however, that require investor analysis. For example, in the United States, exchange-traded notes tracking commodity indexes are treated differently from exchange-traded funds holding commodity futures contracts, creating a preferential tax treatment. A thorough analysis of ETF efficiency should take into account the ETF structure, the local market's taxation regime, and the individual tax situation of the end investor.

3.4 ETF Trading Costs

d describe factors affecting ETF bid–ask spreads

e describe sources of ETF premiums and discounts to NAV

In comparing ETF and mutual fund costs, the usual starting point is management fees, which are often lower for an ETF because most are index based and traded in a highly competitive market. Other important costs should be considered, however.

An ETF has the advantage that it can be purchased whenever exchanges are open—as well as at closing NAV of the fund (similar to mutual fund purchases and sales) when a transaction is large enough to qualify for a creation or redemption. ETF investors usually pay a commission and incur a trading cost related to the liquidity factors associated with the ETF. The trading, or market impact, costs are influenced by the bid–ask spread of the ETF, the size of the trade relative to the normal trading activity of the ETF, and the ease of hedging the ETF by the market-making community. The closing price of the ETF on the exchange may include a premium or discount to the NAV, driven by supply and demand factors on the exchange and the market impact costs of executing an exchange transaction.

3.4.1 *ETF Bid–Ask Spreads*

One of the most important drivers of ETF bid–ask spreads and liquidity is the market structure and liquidity of the underlying securities held. Fixed-income securities, which trade in a dealer market, tend to have much wider bid–ask spreads than large-capitalization stocks. The bid–ask spread of an ETF holding stocks traded in other markets and time zones is influenced by whether the markets for the underlying stocks are open during the hours in which the ETF trades. For specialized ETFs—such as those tracking commodities, volatility futures, or even small-cap stocks—bid–ask spreads can be wide simply because the risk of holding a position even for a short period of time can be high. For some ETFs, even though the underlying securities are liquid, bid–ask spreads may be wide simply because the ETF trades so infrequently the market maker or liquidity provider may need to carry ETF positions for some time before they accumulate sufficient size to create or redeem. Generally, as long as the liquidity in the underlying securities is adequate or hedging instruments can be easily sourced, an ETF trade can usually be executed in a cost-effective manner.

The primary factors that determine the width of the quoted bid–ask spread for a particular transaction size are the amount of ongoing order flow in the ETF, as measured by daily share volume ; the amount of competition among market makers for that ETF ; and the actual costs and risks for the liquidity provider. The bid–ask spread represents the market maker's price for taking the other side of the ETF transaction, which includes the costs and risks to carry the position on his books or to hedge the position using underlying securities or closely related ETFs or derivatives.

More specifically, ETF bid–ask spreads are generally less than or equal to the combination of the following:

- ± Creation/redemption fees and other direct trading costs, such as brokerage and exchange fees
- + Bid–ask spreads of the underlying securities held in the ETF
- + Compensation (to market maker or liquidity provider) for the risk of hedging or carrying positions for the remainder of the trading day

- \+ Market maker's desired profit spread, subject to competitive forces
- – Discount related to the likelihood of receiving an offsetting ETF order in a short time frame

Large, actively traded ETFs have narrow bid–offer spreads and the capacity (or liquidity) for large transaction sizes. For very liquid US-listed ETFs, such as SPY (the SPDR S&P 500 ETF), or EEM (the iShares MSCI Emerging Markets ETF), buyers and sellers are active throughout the trading day and market makers have a high likelihood of finding the other side or hedging larger orders. Therefore, because most of these ETF trades are matched quickly and never involve the creation/redemption process, the first three factors do not contribute heavily in their spreads. For liquid ETFs, the bid–ask spread can be significantly tighter than the spreads on the underlying securities.

The quoted ETF bid–ask spread, however, is generally for a specific, usually small, trade size and does not always reflect ETF liquidity for larger transactions (more than 10% of average daily volume). Larger trades may best be handled by negotiation, involving work with capital market specialists at ETF managers and broker/dealer ETF desks to understand the various ETF execution options and associated trading costs.

Exhibit 8 shows the asset-weighted average and median bid–ask spreads for various ETF categories traded in the United States.

Exhibit 8 Average and Median Bid–Ask Spreads for US-Traded ETFs

US-Traded ETF Category	AUM ($ millions)	Average Spread ($ asset-weighted)	Median Spread
US Equity	1,871,942	0.03%	0.16%
International Equity	731,251	0.05%	0.24%
US Fixed Income	589,851	0.02%	0.14%
International Fixed Income	65,159	0.06%	0.24%
Commodities	62,620	0.05%	0.24%
Leveraged	29,633	0.29%	0.32%
Inverse	11,315	0.10%	0.21%
Asset Allocation	9,318	0.21%	0.29%
Alternatives	4,388	0.18%	0.38%
All US-Traded ETFs*	3,377,276	0.04%	0.20%

* Includes currency ETFs in addition to ETFs listed. Total currency ETF assets are $1,799 million.
Source: FactSet, as of the end of December 2018, based on 60-day averages.

US equity and fixed-income ETFs have the tightest asset-weighted spreads. International equity and international fixed-income spreads are wider, because the underlying securities trade in different market structures, making it difficult to price simultaneously, and because the underlying security exchanges may be closed during a portion of the US trading day. ETF categories representing longer-term strategies, such as asset allocation and alternatives, are less actively traded and have lower asset levels and wider spreads, in part because they have less ongoing two-way order flow and, therefore, depend more on market makers to source liquidity through the underlying securities. Bid–ask spreads are dynamic, vary by trade, and tend to widen when market volatility increases or when significant information relating to the underlying index securities is expected.

Understanding spreads for non-equity ETFs is more complex. Although the fixed-income ETFs give investors access to a portfolio of debt securities trading with transparent bid–ask spreads in the stock market (via the ETF), the actual market for the underlying bonds is far less transparent with OTC trading, in which traders at banks and large bond desks offering quotes on demand without posting bids or offers on an exchange.

Unlike actively traded US Treasury securities, both corporate debt and high-yield bonds, as well as some municipals and international bonds, trade actively only around the time of issuance, after which they may be held until maturity. Therefore, bond ETFs that track indexes containing corporate and high-yield debt often invest only in a subset of the most liquid high-yield securities. Their bid–ask spreads tend to be wider than those of ETFs based on stocks or US Treasuries because of the risk to dealers in hedging inventory and the default risk of the securities, especially in periods of weak economic conditions.

3.4.2 *Premiums and Discounts*

In addition to commissions and bid–ask spreads, ETF premiums and discounts are also important components of ETF trading costs.

At the end of the trading day, each ETF has an end-of-day NAV at which shares can be created or redeemed and with which the ETF's closing price can be compared. Most investors rely on return calculations based on this closing NAV. NAV is intended to be an accurate assessment of the ETF's fair value. This is the case when the underlying securities trade on the same exchange as the one where the ETF is listed (or trades), because these securities trade in the same market structure and have the same closing price time as the ETF.

During the trading day, exchanges disseminate ETF **iNAVs**, or "indicated" NAVs; iNAVs are intraday "fair value" estimates of an ETF share based on its creation basket composition for that day. An ETF is said to be trading at a premium when its share price is higher than iNAV and at a discount if its price is lower than iNAV.

The calculation for end-of-day and intraday premiums/discounts is as follows:

End-of-day ETF premium or discount (%) = (ETF price − NAV per share)/NAV per share.

Intraday ETF premium or discount (%) = (ETF price − iNAV per share)/iNAV per share.

Like tracking error, premiums/discounts are driven by a number of factors, including timing differences and stale pricing.

Timing differences. NAV is often a poor fair value indicator for ETFs that hold foreign securities because of differences in exchange closing times between the underlying (e.g., foreign stocks, bonds, or commodities) and the exchange where the ETF trades. For example, if a commodity held in the fund stops trading in the futures market at 3:00 p.m., the issuer may elect to retain that price for a 4:00 p.m. valuation. If a fund holds securities in a different currency, it may choose to "strike" or value the currency at 4:00 p.m. ET—or occasionally, at 4:00 p.m. London time. In the case where international stocks are held in US-traded ETFs, the NAV may be based on a market closing price in Asia or Europe that occurred hours ahead of when the ETF stops trading on the US exchange.

Because bonds do not trade on an exchange, no true "closing prices" are available for valuing the bonds in a portfolio. Instead, ETF issuers rely on bids from bond desks or pricing services for proxy prices. In the case of bonds that have not traded near the close of the dealer market, index providers and bondholders typically use pricing services for bond valuation. These pricing services often use more liquid bonds that have similar features to estimate where the non-traded bond would have closed.

Sometimes, bond pricing model inputs reflect the price at which a dealer is willing to buy the bonds and the risk and cost to a dealer in carrying the bonds in inventory. In such cases, the ETF's closing price is often higher than the bid prices of the underlying bond holdings used to calculate NAV, making it appear that the ETF is at a premium. During times of market stress, few bonds may trade, leaving pricing services without updated inputs for their models. Like ETFs holding foreign securities, this causes NAVs to be "stale" and, in this case, with possibly too high a valuation given market conditions. In this case, fixed-income ETFs with sufficient trading volume may appear to be trading at discounts to NAV. In these cases, by reflecting the market's most current assessment of value, liquid ETFs become "price discovery" vehicles.

ETFs also provide price discovery for after-hours markets. For example, US-listed ETFs holding European stocks trade until 4:00 p.m. ET, hours after European markets have closed. In these cases, premiums or discounts resulting from closed underlying markets are not mispricing; rather, they are the market's best estimate as to where the fund holdings would trade if the underlying markets were open.

Stale pricing. ETFs that trade infrequently may also have large premiums or discounts to NAV. If the ETF has not traded in the hours leading up to the market close, NAV may have significantly risen or fallen during that time owing to market movement. In this case, comparing the last ETF trade price—for example, at 1:00 p.m.—with the end-of-day 4:00 p.m. NAV would result in a premium (or discount) if the market and corresponding NAV fell (or rose) sharply between 1:00 and 4:00 p.m.

This situation can be compounded if days or weeks elapse between the ETF's trades. Some premium/discount calculations use a strict last price input, whereas others use a closing midpoint. The strict pricing will quote the last trade price, no matter how distant the ETF trade date, which can lead to severe premiums or discounts because NAVs are updated on the basis of the latest market closing prices while the ETF price remains unchanged at last trade.

Comparison of US ETF Trading Costs

A good way to assess the liquidity and potential trading costs of ETFs is to compare various measures of trading activity among similar funds. Exhibit 9 shows trading measures for some of the most liquid ETFs—the SPDR S&P 500 ETF (SPY), the iShares Core S&P 500 ETF (IVV), and the Vanguard S&P 500 ETF (VOO) benchmarked to the S&P 500 Index; another large-cap ETF, the iShares MSCI USA Equal Weighted ETF (EUSA), benchmarked to the MSCI USA Equal Weighted Index; and a liquid small-cap ETF, the iShares Russell 2000 ETF (IWM), benchmarked to the Russell 2000 Index.

Exhibit 9 Selected US Equity Index ETF Trading Measure Comparison

ETF Ticker	SPY	IVV	VOO	EUSA	IWM
Benchmark Index	**S&P 500**	**S&P 500**	**S&P 500**	**MSCI USA Equal Weighted**	**Russell 2000**
Volume in US dollars					
Daily average volume	24.47 billion	1.22 billion	819.28 million	1.32 million	3.90 billion
Median volume	20.23 billion	1.08 billion	739.69 million	0.94 million	3.81 billion
Other trading characteristics					

(continued)

Exhibit 9 (Continued)

ETF Ticker	SPY	IVV	VOO	EUSA	IWM
Benchmark Index	S&P 500	S&P 500	S&P 500	MSCI USA Equal Weighted	Russell 2000
Average spread (%)	0.00%	0.01%	0.01%	0.12%	0.01%
Average spread ($)	$0.01	$0.03	$0.03	$0.07	$0.01
Median premium/discount (%)[a]	0.00%	0.00%	0.00%	0.04%	0.01%
Maximum premium (%)[a]	0.12%	0.13%	0.18%	0.96%	0.12%
Maximum discount (%)[a]	–0.19%	–0.11%	–0.08%	–0.38%	–0.13%

[a] Over previous 12 months.
Source: FactSet, as of 7 November 2018.

SPY, the largest ETF by AUM and the first ETF traded in the United States, is one of the most liquid securities in the world. IVV and VOO, with the same benchmark, are used more by intermediate- and longer-horizon investors but also have very tight spreads because of liquidity in the underlying securities and ease of hedging for market makers. SPY trades a median of $20 billion a day, compared with a median of $1 billion for IVV. The average bid–ask spread shows that both are highly liquid. In addition, both have tight premiums and discounts to NAV.

In contrast, EUSA has a larger spread, 0.12%. The lower liquidity and higher trading cost for EUSA can be attributed to the fact that the benchmark index does not have futures and other index products available for hedging use by market makers. The MSCI USA Equal Weighted Index also includes close to 600 stocks—100 more than the S&P 500 Index has.

IWM, benchmarked to the Russell 2000 Index of US small-cap stocks, holds far more securities than any of the previously mentioned ETFs, and many are small-cap stocks that have wide spreads. IWM, however, trades with spreads and premiums/discounts close to those of SPY.

How is that possible? First, trading activity in IWM is high (median daily dollar volume of $4 billion) and continuous throughout the trading day. Second, the Russell 2000 Index has an active futures market, making it easy for market makers and APs to quickly hedge the risk of large trades.

Exhibit 10 shows three US fixed-income ETFs—one US-Treasury based and two benchmarked to US high-Yield indexes. All three are among the most liquid fixed-income ETFs and have tight average bid–ask spreads. The iShares iBoxx $ High Yield Corporate Bond ETF (HYG) is the most liquid, with median daily volume of $1.4 billion and a higher median premium (0.20%) than the iShares 20+ Year Treasury Bond ETF (TLT). These positive median premiums indicate that the SPDR Bloomberg Barclays High Yield Bond ETF (JNK) and HYG have been in a net demand position over most of the 12-month period covered in Exhibit 10 and investors have typically paid above fair value for ETF access to a high-yield portfolio.

The maximum premium and discount have generally been much larger for bond ETFs compared with the equity ETFs shown in Exhibit 9. This is because the underlying fixed-income securities trade in a dealer market and are not continuously priced. In this case, the fixed-income ETFs, which trade on an exchange with more continuous pricing, may be a better reflection of true supply and demand for the portfolio because the underlying bonds may not trade as frequently, particularly in extreme market conditions.

Exhibit 10 Selected US Fixed-Income ETF Trading Measure Comparison

ETF Ticker	TLT	JNK	HYG
Benchmark Index	ICE US Treasury 20+ Year Bond Index	Bloomberg Barclays High Yield Very Liquid Index	Markit iBoxx USD Liquid High Yield Index
Volume in US dollars			
Daily average volume	1.04 billion	0.46 billion	1.50 billion
Median volume	0.97 billion	0.41 billion	1.44 billion
Other trading characteristics			
Average spread (%)	0.01%	0.03%	0.01%
Average spread ($)	$0.01	$0.01	$0.01
Median premium/discount (%)[a]	0.03%	0.10%	0.20%
Maximum premium (%)[a]	0.68%	0.41%	0.59%
Maximum discount (%)[a]	–0.52%	–0.67%	–0.75%

[a] Over the previous 12 months.
Source: FactSet, as of 7 November 2018.

3.5 Total Costs of ETF Ownership

f describe costs of owning an ETF

Exhibit 11 provides a summary of cost factors when considering ETFs and mutual funds. Some of these costs are explicit, whereas others are implicit and reflected in net investment returns. Both ETFs and mutual funds typically pay lower institutional commission rates for trades because of their asset size. ETF transaction costs are incurred at purchase and sale regardless of holding period, whereas other costs, such as management fees, increase as the holding period lengthens. Ongoing costs, such as management fees, portfolio turnover, and security lending proceeds, have a consistent impact on investment returns based on holding period. ETF trading costs, such as commissions and bid–ask spreads, are incurred only at purchase and sale, and their return impact diminishes over longer holding periods, whereas management fees and other ongoing costs become a more significant proportion of total costs. Tracking error can be considered a positive or negative implicit cost.

For active short-term ETF investors who trade frequently, the cost of entering and exiting their ETF positions (commissions, bid–ask spreads, premiums/discounts) is a far more significant consideration than management fees, tracking error, and other costs that accumulate over longer holding periods.

ETFs may trade at market prices higher (premiums) or lower (discounts) than NAV, which is based on closing prices for the fund's underlying securities. Premiums and discounts may reflect a lag in the timing of the underlying security valuations relative to current market conditions and can be considered positive costs (in the case of premiums) or negative costs (in the case of discounts).

There are additional implicit trading costs of fund management, such as portfolio turnover costs that are reflected in fund returns. These are incurred within the fund as the portfolio manager buys and sells securities to execute the investment strategy and manage fund cash flows. Portfolio turnover costs reduce returns and affect performance for all investors in the fund. Many ETFs are based on indexes that have lower portfolio turnover than actively managed funds. Taxable gains incurred upon sale can be considered positive costs for the investor, whereas taxable losses represent negative costs. Security lending income for the fund represents negative costs.

Exhibit 11 Cost Factor Comparison—ETFs and Mutual Funds

Fund Cost Factor	Function of Holding Period?	Explicit/ Implicit	ETFs	Mutual Funds
Management fee	Y	E	X (often less)	X
Tracking error	Y	I	X (often less than comparable index mutual funds)	(index funds only)
Commissions	N	E	X (some free)	
Bid–ask spread	N	I	X	
Premium/discount to NAV	N	I	X	
Portfolio turnover (from investor flows and fund management)	Y	I	X (often less)	X
Taxable gains/losses to investors	Y	E	X (often less)	X
Security lending	Y	I	X (often more)	X

Trading costs vs. management fees. To illustrate the effect of management fees versus trading costs, consider an investor who pays a commission of \$10 on a \$20,000 trade (0.05% each way) combined with a 0.15% bid–ask spread on purchase and sale. The round-trip trading cost is, therefore, 0.25% and is calculated as follows:

Round-trip trading cost (%) = (One-way commission % × 2) + (½ Bid–ask spread % × 2)

= (0.05% × 2) + (½ × 0.15% × 2)

= 0.10% + 0.15%

= 0.25%.

For a round-trip trade that happens over a year, 0.25% can be larger than the annual expense ratios of many ETFs. If held for less than a year, the trading costs may be far larger than the expense ratio paid on the ETF.

To see the impact of holding period, consider the 3-month versus 12-month versus 3-year holding period costs for an ETF with a 0.15% annual fee, one-way commissions of 0.05%, and a bid–ask spread of 0.15%. Holding period costs can be calculated as follows:

Holding period cost (%) = Round-trip trade cost (%) + Management fee for period (%).

Specific holding period costs can be calculated as follows:

3-month holding period cost (%) = 0.25% + 3/12 × 0.15%

= 0.29%.

12-month holding period cost (%) = 0.25% + 12/12 × 0.15%

= 0.40%.

3-year holding period cost (%) = 0.25% + 36/12 × 0.15%

= 0.70%.

Exhibit 12 illustrates that for holding periods of 3 and 12 months, trading costs represent the largest proportion of annual holding costs (0.86% and 0.625%, respectively). Excluding the compounding effect, for a three-year holding period, management fees represent a much larger proportion of holding costs (0.64%).

Exhibit 12 ETF Management Fee and Trading Cost Comparison

Holding Period:	3 Months	12 Months	3 Years
Commission	0.10%	0.10%	0.10%
Bid–ask spread	0.15%	0.15%	0.15%
Management fee	0.0375%	0.15%	0.45%
Total	**0.29%**	**0.40%**	**0.70%**
Trading costs (% of total)	0.86%	0.625%	0.36%
Management fees (% of total)	0.14%	0.375%	0.64%

For broad-based, capitalization-weighted equity index ETFs that have the lowest fees, trading costs represent the largest cost in using an ETF. The longer an ETF is held, the greater the proportion of total costs represented by the management fee component.

Tactical traders will generally choose an ETF on the basis of its liquidity and trading costs (e.g., commissions, bid–ask spreads). In many cases, shorter-term tactical traders may use an ETF with a higher management fee but a tighter bid–ask spread and more active or continuous two-way trading flow to avoid incurring the capital commitment cost of a market maker or the cost of arbitrage for the ETF versus the underlying securities. The size of the management fee is typically a more significant consideration for longer-term buy-and-hold investors.

4 ETF RISKS

g describe types of ETF risk

ETFs introduce several unique risks because of their structure, fund holdings, and underlying exposure.

4.1 Counterparty Risk

Some ETP (exchange-traded product) legal structures involve dependence on a counterparty. A counterparty failure can put the investor's principal at risk of default or affect a portion of the assets via settlement risk. Likewise, counterparty activity can affect a fund's economic exposure. Therefore, investors should carefully assess counterparty risk.

Although exchange-traded notes (ETNs) trade on exchanges and have a creation/redemption mechanism, they are not truly funds because they do not hold underlying securities. ETNs are unsecured debt obligations of the institution that issues them and are structured as a promise to pay a pattern of returns based on the return of the stated index minus fund expenses. The issuer of the note takes responsibility for setting up the counterbalancing hedges it believes necessary to meet the obligations.

In the United States, ETNs are registered under the Securities Act of 1933 because they are general obligation debt securities of a bank and are not managed by an investment firm for a fee. Similar ETN structures exist in most markets where ETFs are listed.

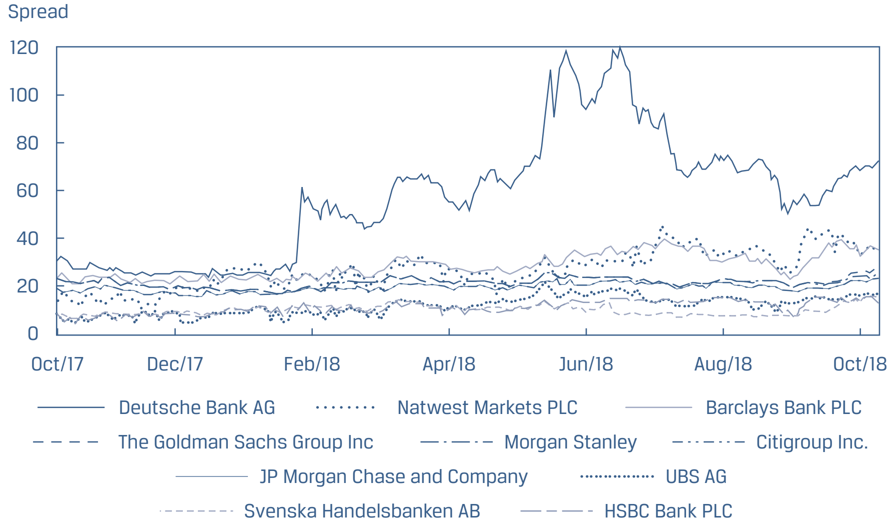
ETNs have the largest potential counterparty risk of all exchange-traded products because they are unsecured, unsubordinated debt notes and, therefore, are subject to default by the ETN issuer. Theoretically, an ETN's counterparty risk is 100% in the event of an instantaneous default by the underwriting bank, and should an issuing bank declare bankruptcy, any ETNs issued by the bank would effectively be worthless. Because baskets of notes may be redeemed back to the issuer at NAV, however, it is likely that only an extremely rapid and catastrophic failure would take investors by surprise. This happened once, in 2008, with three Lehman Brothers–backed ETNs, but it has not happened since.

In the United States, some funds offering exposure to non-US-dollar currencies achieve this via offshore bank deposits. These funds bear default risk at the deposit-holding bank.

Because ETNs and deposit-based ETFs are backed by banks, their default risk can be monitored via the issuing bank's credit default swap (CDS) pricing.

The credit spreads for one-year CDSs by issuer at the end of October 2018 are shown in Exhibit 13.

Exhibit 13 One-Year CDS Spreads for ETN Issuers, 24 October 2018



Source: Bloomberg

The quoted CDS rates represent the cost to insure debt, in basis points per year; so, for example, investors could "insure" $1 million in Goldman Sachs bonds for just under $30,000 per year. Although the insurance rate should never be considered an estimate of actual default risk for a 12-month period, it does provide a reasonable gauge of the relative risk of the various issuers. In general, a one-year CDS rate above 5% should raise significant concerns among investors because it foretells a significant default risk in the year to come.

Settlement risk. A fund that uses OTC derivatives, such as swaps, to gain market exposure has settlement risk; that is, mark-to-market (unrealized) gains are subject to counterparty default. Such ETFs include many European swap-based funds (or synthetic ETFs), funds using leverage (or geared funds), some currency funds, and some actively managed portfolios. To minimize settlement risk, OTC contracts are typically settled frequently—usually on a daily or weekly basis. This frequent settlement reduces the exposure the swap partners face if a company goes bankrupt, but there is a theoretical risk of counterparty default between settlement periods. In addition, the majority of the contract collateral is held in low-risk instruments, such as US T-bills, at a custodian bank.

Swap exposures are not unique to ETFs. Many mutual funds also use swaps and other derivatives to gain exposure. With ETFs, swap exposures are somewhat transparent because these holdings are disclosed daily by the ETF provider, although full information on counterparties and terms may not be disclosed.

Security lending. ETF issuers (in addition to traditional mutual fund managers and institutions) lend their underlying securities to short sellers, earning additional income for the fund's investors. Securities lent are generally overcollateralized, to 102% (domestic) or 105% (international), so that the risk from counterparty default is low. Cash collateral is usually reinvested into extremely short-term fixed-income securities with minimal associated risk. At the time of writing, there has been no instance of shareholder loss resulting from security lending in an ETF since ETF product inception in the early 1990s. A well-run security lending program can generate significant income for the ETF issuer, sometimes entirely offsetting the fund's operating expenses. Most ETF issuers credit all profits from this activity back to shareholders, although information about issuer lending programs is sometimes not well disclosed.

4.2 Fund Closures

Similar to mutual fund closures, ETF issuers may decide to close an ETF. In such a case, the fund generally sells its underlying positions and returns cash to investors. This activity can trigger capital gain events for investors and the need to find a replacement investment. Primary reasons for a fund to close include regulation, competition, and corporate activity. "Soft" closures—which do not involve an actual fund closing—include creation halts and changes in investment strategy.

Regulations. Security regulators can change the regulations governing certain types of funds, resulting in forced closure of those funds. For example, commodity futures are under constant regulatory scrutiny, and position limits can make it impossible for some funds to function. In 2018, the Israeli security regulator banned the ETN structure, forcing over 700 products to close and reopen as traditional ETFs.

Competition. Investors have benefited from a growing number of ETFs and increased competition. As ETFs proliferate, some funds fail to attract sufficient assets and are shut down by the ETF issuer. A fund's assets under management, in addition to those of any competitor, and the ETF's average daily liquidity are indications of market support. Low AUM and trading volumes over a significant period could indicate potential fund closure.

Corporate actions. Mergers and acquisitions between ETF providers can prompt fund closures. When ETF families merge or are sold to other ETF providers, new ETF owners may close underperforming ETFs (from an asset-gathering perspective) and invest in new, higher-growth opportunities.

Creation and redemption halts. ETN issuers may halt creations and redemptions. An example of this scenario is when an ETN issuer no longer wants to add debt to its balance sheet related to the index on which the ETN is based. This situation occurred in September 2018, when ETN issuer UBS issued a "sales halt" for its ETRACS Monthly Pay 2xLeveraged Mortgage REIT ETN (MORL), effectively suspending further sales from its outstanding inventory of the ETN and preventing new shares from being created. When creations are halted, the ETN can trade at a substantial premium over fair value, as the arbitrage mechanism breaks down. In this case, MORL traded at a premium of more than 5%. Although all ETFs can theoretically close creations in extraordinary situations, in practice, it happens more commonly with ETNs.

Change in investment strategy. Some ETF issuers find it easier to repurpose a low-asset ETF from their existing lineup than to close one fund and open another. Issuers simply announce a change in the fund's underlying index—a common occurrence in the ETF industry. Although most index changes result in small adjustments to an ETF's portfolio and economic exposure, these "soft closures" can sometimes result in a complete overhaul, changing exposures to countries, industries, or even asset classes.

4.3 Investor-Related Risk

ETFs provide access to sometimes complex asset classes and strategies. For all ETFs, it is important that investors understand the underlying exposure provided by the ETF; otherwise, ETFs may introduce risks to investors who do not fully understand them. For many investors, leveraged and inverse ETFs fall into this category by failing to meet investor expectations. Index methodology (e.g., constituent universe, weighting approach) and the fund's portfolio construction approach are central to understanding an ETF's underlying exposure and related performance.

Leveraged and inverse funds generally offer levered (or geared), inverse, or levered and inverse exposure to a given index and have a daily performance objective that is a multiple of index returns. These products must reset or adjust their exposure daily to deliver the target return multiple each day.

For example, consider a fund offering 300% exposure (3 times, or 3×) to the FTSE 100 Index with a net asset value of £100. It uses swaps to obtain a notional exposure of £300. If the one-day FTSE 100 Index return is 5%, the £300 in exposure becomes £315 (a 5% increase), and the ETF's end-of-day NAV is £115: $100 \times (1 + 3 \times 5\%)$.

In order to deliver 300% of the index's daily performance for the following day, the ETF, now valued at £115, requires notional exposure of £345 for 3 times exposure. Because at the end of the day the ETF has only £315 in exposure, it must reset its exposure—in this case, increasing notional swap exposure by £30.

Exhibit 14 outlines this example.

Exhibit 14 Example of Levered 3× ETF Exposure

	Index Level	One-Day Index Return (%)	3× ETF NAV (£)	Notional Swap Exposure (£)	3× Swap Exposure (£)	Swap Exposure Adjustment (£)
Day 1	100	—	100	300	300	0
Day 2	105	5%	115	315	345	30

If these ETFs are held for longer than a one-day period, the math of compounding and resetting exposure is such that an investor will not see the return multiple—for example, a 200% or −100% return in the case of a 2× ETF or inverse ETF, respectively—over her holding period.

Exhibit 15 presents a levered, inverse fund offering 2 times (−200%) exposure to the S&P 500 Index. The fund (−2× ETF) has a starting net asset value of $100 and uses swaps to obtain notional exposure.

Exhibit 15 Example of Levered and Inverse 2× ETF Daily Return vs. Holding Period Return

	Index Level	One-Day Index Return (%)	Index Period Return (%)	−2× ETF NAV	One-Day ETF Return (%)	2× ETF Holding Period Return (%)
Day 1	100	—		100	—	—
Day 2	110	10%	10%	80	−20%	−20%
Day 3	99	−10%	−1%	96	20%	−4%

Day 1: Both the index and the −2× fund are at a starting level of 100.
Day 2: The index increases to 110, a one-day return of 10%.

The −2× ETF daily return is calculated as follows:

= −2 × [(110 − 100)/100]
= −2 × (10%)
= −20%.

The −2× ETF NAV is calculated as follows:

= 100 × (1 + −0.2)
= 80.

Day 3: The index falls to 99, a one-day return of −10%.

The 2× ETF daily return is calculated as follows:

= −2 × [(99 − 110)/110]
= −2 × (−10%)
= 20%.

The −2× ETF NAV is calculated as follows:

= 80 × (1 + 0.2).
= 96.

This example shows the fund delivering its promised performance, −2× the daily index return, but it also shows how the return may not be what is naively expected over periods longer than a day.

Over the three days, the index return is −1%: (99 − 100)/100. A naive expectation might assume that over the same period, the −2× ETF would return 2% (= −2 × −1%). Over the three days, the fund's actual return was −4%: (96 − 100)/100.

Because of these compounding effects in leveraged ETFs, the funds are generally not intended to be buy-and-hold products for more than a one-month horizon. If investors are planning to hold them long term, they must rebalance the funds periodically to maintain the desired net exposure.

5 ETFS IN PORTFOLIO MANAGEMENT

h identify and describe portfolio uses of ETFs

ETFs have become valuable tools for both institutional and retail investors. Available on a wide range of passive, systematic (rules-based) active, and traditional active strategies and segments of the stock, bond, and commodity markets, ETFs are used for both top-down (based on macro views) and bottom-up (focused on security selection) investment approaches. In addition to their use in implementing long-term strategic exposure to asset classes and risk factors, ETFs are used for tactical tilts, portfolio rebalancing, and risk management.

5.1 ETF Strategies

Most institutional asset managers and hedge fund managers, Registered Investment Advisers (RIAs), and financial advisers use ETFs for a wide range of strategies. These strategies serve many different investment objectives—some strategic, some tactical, and some dynamic, where the timing of changes is based on market conditions. Other ETF applications help in managing portfolios more efficiently and are used primarily for operational purposes. As we discuss the diverse set of strategies that can be found in an ETF structure, it is apparent that they are not easily classified as either active or passive. Except for core asset class and portfolio efficiency investment applications that use ETFs based on market-capitalization weighted benchmarks, almost all ETF-related strategies have some component of active investing, either within the ETF strategy or in the way the ETF is used.

Not all strategies are suitable in an ETF structure. The disclosure of holdings may be undesirable for an active manager who invests in less liquid securities or pursues either a concentrated investment strategy or one that relies on an approach that cannot be easily described (such as a "black box" methodology) or disclosed without compromising the strategy. The liquidity of the underlying investments must also be high enough to accommodate daily creations and redemptions. Such factors as tax efficiency, low fees, and available product make ETFs competitive alternatives to traditional mutual funds and active managers. The primary applications in which ETFs are used include the following:

Portfolio efficiency: The use of ETFs to better manage a portfolio for efficiency or operational purposes. Applications include cash or liquidity management, rebalancing, portfolio completion, and active manager transition management.

Asset class exposure management: The use of ETFs to achieve or maintain core exposure to key asset classes, market segments, or investment themes on a strategic, tactical, or dynamic basis.

Active and factor investing: The use of ETFs to target specific active or factor exposures on the basis of an investment view or risk management need.

5.2 Efficient Portfolio Management

ETFs are useful tools for managing portfolio activity necessitated by cash flows and changes in external managers. In addition, ETFs can be used to easily accommodate portfolio rebalancing needs and unwanted gaps in portfolio exposure.

Portfolio liquidity management. One of the primary institutional applications of ETFs is cash flow management. ETFs can be used to invest excess cash balances quickly (known as cash equitization), enabling investors to remain fully invested in target benchmark exposure, thereby minimizing potential cash drag. Cash drag refers to a fund's mis-tracking relative to its index that results from holding uninvested cash. Managers may also use ETFs to transact small cash flows originating from dividends, income, or shareholder activity. Some portfolio managers hold small portions of their funds in ETFs in anticipation of future cash outflows. Transacting the ETF may incur lower trading costs and be easier operationally than liquidating underlying securities or requesting funds from an external manager.

Portfolio rebalancing. Many investors rebalance portfolios on the basis of a specified time interval, usually at least quarterly, and some may adjust whenever the market value of a portfolio segment, or allocation, deviates from its target weight by a threshold, such as 2%. For tighter rebalancing thresholds and more frequent rebalancing time intervals, using liquid ETFs with tight bid–ask spreads allows the portfolio manager to execute the rebalance in a single ETF trade and ensures the portfolio remains fully invested according to its target weights. For investors who have the ability to sell short, reducing exposure associated with a rebalance can be done quickly using an ETF, and as the underlying securities are sold off, the short position can be covered.

Portfolio completion strategies. ETFs can also be used for completion strategies to fill a temporary gap in exposure to an asset class, sector, or investment theme or factor. Gaps may arise with changes in external managers or when an existing manager takes an active view that moves the portfolio out of a market segment to which the investor wishes to have continued exposure. The investor may want to retain the manager but use a tactical ETF strategy to maintain exposure to the desired market segment. If external managers are collectively underweighting or overweighting an industry or segment, such as technology, international small-cap stocks, or high-yield bonds, ETFs can be used to adjust exposure up or down to the desired level without making changes to underlying external manager allocations.

Transition management. Transition management refers to the process of hiring and firing managers—or making changes to allocations with existing managers—while trying to keep target allocations in place. Because ETFs exist on most domestic, international, and global equity benchmarks, a newly appointed transition manager can invest in an ETF to maintain market exposure as she undergoes the process of selling the unwanted positions of the manager she is replacing (the terminated manager). The new transition manager can then take her time to invest in positions for her strategy and gradually reduce the ETF holding.

Asset owners can use ETFs to maintain desired market or asset class exposure in the absence of having an external manager in place. For example, if a fixed-income manager benchmarked to the Bloomberg Barclays US Aggregate Bond Index is terminated, the asset owner may wish to invest in the iShares Core US Aggregate Bond ETF (AGG) to maintain benchmark exposure until a replacement manager can be hired. In some cases, asset owners will "fund" new managers with ETF positions. The new manager will then sell off his ETF positions in the benchmark index as he invests in the underlying securities that meet his desired investment objectives and valuation criteria.

For very large asset owners, there are three potential drawbacks to using ETFs for portfolio management: (1) Given the asset owner size, they may be able to negotiate lower fees for a dedicated separately managed account (SMA) or find lower-cost commingled trust accounts that offer lower fees for large investors, (2) an SMA can be customized to the investment goals and needs of the investor, and (3) many regulators require large ETF holdings (as a percentage of ETF assets) to be disclosed to the public. This can detract from the flexibility in managing the ETF position and increase the cost of shifting investment holdings.

Exhibit 16 provides a summary of ETF portfolio efficiency applications, covering their roles in the portfolio, and examples by benchmark type. Applications include (1) transacting cash flows for benchmark exposure, (2) rebalancing to target asset class or risk factor weights, (3) filling exposure gaps in portfolio holdings of other strategies and funds, and (4) temporarily holding during transitions of strategies or managers.

Exhibit 16 ETF Portfolio Applications—Portfolio Efficiency

Portfolio Application	Role in Portfolio	Examples of ETFs by Benchmark Type
Cash Equitization/ Liquidity Management	Minimize cash drag by staying fully invested to benchmark exposure, transact small cash flows	Liquid ETFs benchmarked to asset category
Portfolio Rebalancing	Maintain exposure to target weights (asset classes, sub-asset classes)	Domestic equity, international equity, domestic fixed income
Portfolio Completion	Fill gaps in strategic exposure (countries, sectors, industries, themes, factors)	International small cap, Canada, bank loans, real assets, health care, technology, quality, ESG
Manager Transition Activity	Maintain interim benchmark exposure during manager transitions	ETFs benchmarked to new manager's target benchmark

5.3 Asset Class Exposure Management

Investors have used index exposure in core asset classes for decades, but one of the fastest-growing areas of ETF usage, especially by institutional investors, is fixed income. Since the financial crisis of 2008, the reduced capital available for banks (to participate in dealer bond markets) has contributed to greater use of fixed-income ETFs for core exposure. Except for the largest institutional investors, trading portfolios of bonds is much more difficult and expensive than similar portfolio trades in stocks. Fixed-income ETFs, especially those benchmarked to indexes containing corporates and high-yield securities, provide bond investors with a more efficient (lower cost, more continuous pricing, agency market) and liquid means of obtaining core fixed-income exposure.

Core exposure to an asset class or sub-asset class. The primary strategic use of ETFs is to gain core index exposure to various asset classes and sub-asset classes. ETFs make doing so easy—across global equities, bonds, commodities, and currencies—and investors regularly use ETFs for broad portfolio diversification. Investors also use ETFs for more targeted strategic exposure to such segments as high-yield debt, bank loans, and commodities (including crude oil, gold and other metals, and agricultural products).

A financial adviser can use ETFs to build a diversified portfolio on the basis of ETF recommendations from his firm's wealth management research team. Benchmarked to broad asset classes, portfolio choices for equity ETF exposure might include domestic large- and small-cap equities, sectors, such risk factors as dividend growth or momentum, industries, and international regions or countries with or without currency exposure. Choices for fixed-income ETF exposure might include government and corporate debt of various maturities, emerging market debt, bank loans, and possibly floating interest rate strategies. Commodity ETF exposure could include gold and other metals, broad commodity indexes, agriculture products, and oil. Similarly, brokerage firms and robo-advisers may offer more-automated solutions that select an ETF allocation based on the investor's risk and return profile. These firms offer a range of ETF investment choices from a preapproved product list to fit different asset class and risk factor categories.

Tactical strategies. ETFs can also be used to implement market views and adjust portfolio risk on a more short-term, tactical basis. Some financial advisers and institutional investors allocate a portion of their portfolios for opportunistic trading based on their firm's (or strategist's) research or short-term outlook. Others make tactical adjustments in a range around target weights for asset classes or categories within an asset class. ETFs based on risk factors, country exposure, credit or duration exposure, currencies, or even volatility, crude oil, or metals can be used to express tactical views. To profit from an expected price decline, investors can sell ETFs short in a margin account.

Thematic ETFs are also used to implement investment views. Thematic ETFs hold stocks passively but allow investors to take an active view on a market segment they believe will deliver strong returns. These ETFs typically cover a narrow or niche area of the market not well represented by an industry. Examples include focused areas of technology, such as cybersecurity and robotics. Other themes accessed via ETFs are global infrastructure, regional banks, semiconductors, and gold mining. Generally, thematic ETFs are tactical tools that serve as substitutes for buying individual stocks or an industry ETF that is too broad to adequately represent the investor's investment view. Holdings may overlap with those of other ETFs or other portfolio positions but play a role when the investor wants to overweight this segment in the portfolio. Thematic ETFs should be evaluated similarly to stocks because they tend to have comparable levels of volatility and represent specialized active views.

ETFs that have the highest trading volumes in their asset class category are generally preferred for tactical trading applications, and the liquidity in many of the largest ETFs offered in each region makes them well suited for this purpose. Trading costs and liquidity, rather than management fees, are the important criteria in selecting an ETF for tactical adjustments. To identify the most commonly used ETFs for tactical strategies, one can look at the ratio of average dollar volume to average assets for the ETF.

Exhibit 17 provides a summary of ETF asset class exposure applications, covering their roles in the portfolio, categories of use, and examples by benchmark type. These applications relate to using ETFs for strategic, tactical, and dynamic asset class exposure.

Exhibit 17 ETF Portfolio Applications—Asset Class Exposure Management

Portfolio Application	Category	Role in Portfolio	Examples of ETFs by Benchmark Type
Core asset class or market	Strategic or tactical	Core long-term, strategic weighting Tactical tilt to enhance returns or modify risk Ease of access vs. buying underlying securities	Domestic equity, international equity, fixed income, commodities
Equity style, country, or sector; fixed income or commodity segment	Strategic or tactical	Tactical tilt to enhance returns or modify risk depending on short-term views Hedge index exposure of active stocks or bond strategy Ease of access vs. buying underlying securities	Value, growth, Japanese, Chinese, UK, Canadian, or Mexican equities; corporate or high-yield debt; gold; oil; agriculture
Equity sector, industry, investment theme	Dynamic or tactical	Tactical or dynamic active tilt to enhance returns or modify risk Efficient implementation of a thematic/industry vs. single-stock view Capture performance on an emerging theme or innovation not reflected in industry categories	Technology, financials, oil and gas, biotech, infrastructure, robotics, gold mining, buybacks, internet innovation, cybersecurity

5.4 Active and Factor Investing

In the mid-2000s, quantitative or rules-based strategies became available in ETFs. These strategies had “active” weights different from market capitalization and were able to disclose holdings because the stock selection and weighting was not chosen by a discretionary portfolio manager but, rather, by a set of quantitative rules, disclosed in the index methodology.

The first smart beta ETFs were indexes weighted by company fundamentals, such as dividends, or quantitatively screened on stock features. Although adoption was initially slow, institutional investors and RIAs now use smart beta ETF strategies to gain systematic active exposure to persistent common return drivers or factors. Global assets in smart beta equity funds, including both single-factor and multi-factor strategies, now represent approximately 20% of ETF assets.

Active ETFs, where the investment strategy is benchmarked but managed with discretion, have also gained assets, especially in fixed income, but they still represent a relatively small percentage of global ETF assets, at 2%–3%.

Factor (smart beta) ETFs. Factor ETFs are usually benchmarked to an index created with predefined rules for screening and/or weighting constituent holdings. The strategy index rules are structured around return drivers or factors, such as value, dividend yield, earnings or dividend growth, quality, stock volatility, or momentum. Some of these factors, such as size, value, and momentum, have academic support as equity risk premiums that may be rewarded over the long term. Within each single factor category, a range of offerings from competing ETF providers exists, differentiated by the criteria used to represent the factor and the weights applied to constituent holdings (equal, factor, or cap weighted). Their application is typically in providing longer-term, buy-and-hold exposure to a desired factor based on an investment view. Factor ETFs can be used to add risk factor allocations that might not be present in a benchmark or portfolio—for example, adding an equity index ETF with stocks screened for quality to add desired exposure to a quality factor.

Multi-factor ETFs that combine several factors also exist. They may adjust their weights dynamically as market opportunities and risk change. In a multi-factor ETF, strategy design involves factor selection, factor strategy construction, and a weighting scheme across factors that is managed over time. A multi-factor approach typically has lower return volatility than a single-factor approach over time but may also have less return potential for investors who want to capitalize on factor timing.

The success of active strategy ETFs is related to (1) whether the factor, as represented by a target benchmark factor index, performs well relative to expectations and (2) how effective the selected ETF is at delivering the benchmark factor return. Just as with traditional active investing, the success of active investing with ETFs depends on the skill of the ETF portfolio manager as well as the end investor’s decision to undertake the investment strategy.

Risk management. Some smart beta ETFs are constructed to deliver lower or higher risk than that of their asset class benchmark. For example, low-volatility factor ETFs select stocks on the basis of their relative return volatility and seek to represent a portfolio that offers a lower or target volatility return profile. These low-volatility rules-based factor ETFs have gained assets within each segment of the global equity market (domestic, developed international, and emerging markets) as investors have moved to lower volatility in portfolios. Other ETFs based on the beta characteristics of the constituent stocks can be used to adjust the portfolio’s beta profile to desired levels.

ETFs are also used to manage other portfolio risks, such as currency and duration risk. ETFs that provide international exposure with a hedge on all or part of the associated currency risk are available. With respect to interest rate risk management, several smart beta fixed-income ETFs hold long positions in corporate or high-yield bonds and hedge out the duration risk of these bonds with futures or short positions in government bonds. These ETFs enable investors to add a position to their portfolio that seeks returns from taking credit risk with minimal sensitivity to movements in interest rates. Active investors with a negative macro view can use inverse asset class or factor ETF exposure to temporarily reduce benchmark holding risk. Doing so allows them to implement a macro view on a short-term basis and minimize turnover in underlying portfolio holdings.

Alternatively weighted ETFs. ETFs that weight their constituents by means other than market capitalization, such as equal weighting or weightings based on fundamentals, can also be used to implement investment views—for example, ETFs that weight constituent stocks on the basis of their dividend yields. These ETFs select or overweight stocks with higher dividend yields, subject to other fundamental criteria or constraints, and are used by investors seeking income-generating strategies.

Discretionary active ETFs. The largest active ETFs are in fixed income, where passive management is much less prominent than in equities. The PIMCO Active Bond ETF (BOND) launched in 2012 with an investment objective similar to that of the world's largest mutual fund at that time, the PIMCO Total Return Fund. Shorter-maturity, actively managed ETFs are also available in fixed income. Other active ETFs include exposure to senior bank loans, floating rate debt, and mortgage securities. Active equity ETFs have also been launched in areas of the technology industry.

"Liquid alternative" ETFs are based on strategy indexes that attempt to deliver absolute return performance and/or risk diversification of stock and bond holdings. Some of the first liquid alternative ETFs used rules-based strategies to replicate broad hedge fund indexes. Other strategy indexes offer transparent, rules-based, "hedge fund–like" strategies in specific types of alternatives. Such strategies include long–short, managed futures, private equity, and merger arbitrage.

Dynamic asset allocation and multi-asset strategies. ETF availability across a wide range of equity and bond risk exposures has fostered greater use of dynamic, top-down investment strategies based on return and risk forecasts. Asset managers, hedge funds, and asset owners have increasingly used ETFs for discretionary asset allocation or global macro strategies. Dynamic asset allocation ETF strategies are also available in commodities. Although some strategies allocate holdings on the basis of their relative risk contribution and others are return focused, all involve adjustments back to target weights, as defined by a dynamic investment process. Some pension and sovereign wealth funds implement these strategies in house, whereas other investors hire asset managers that offer multi-asset strategies. Implementation is done using ETFs, along with futures and swaps where available and when they are more efficient to trade.

Proper use of an active or factor strategy ETF requires investors to research and assess the index construction methodology and performance history and to ensure consistency with their investment view.

Exhibit 18 provides a summary of active and factor ETF portfolio applications, covering their roles in the portfolio, categories of use, and examples by benchmark type. These applications relate to ETFs as alternatives to other fund products, such as active mutual funds. In these cases, ETF evaluation is based on features of the investment approach, holdings, cost, risk, and return potential, as well as the impact to the portfolio's overall risk and return.

Exhibit 18 ETF Portfolio Applications—Active and Factor Investing

Portfolio Application	Category	Role in Portfolio	Examples of ETFs by Benchmark Type
Factor exposure	Strategic, dynamic, or tactical	Capture risk premium for one or more factors driving returns or risk Overweight or underweight depending on factor return or risk outlook Seek to capture alpha from rules-based screening and rebalancing (systematic active)	Quality, dividend growth, value, momentum, low volatility, liquidity screen, multi-factor
Risk management	Dynamic or tactical	Adjust equity beta, duration, credit, or currency risk	Currency-hedged, low-volatility, or downside-risk-managed ETFs
Leveraged and inverse exposure	Tactical	Access leveraged or short exposure for short-term tilts or risk management Limit losses on shorting to invested funds	ETFs representing asset classes, countries, or industries with leveraged or inverse daily return targets
Alternative weighting	Strategic, dynamic, or tactical	Seek outperformance from weighting based on one or more fundamental factors Balance or manage risk of security holdings	ETFs weighted by fundamentals, dividends, or risk; equal-weighted ETFs
Active strategies within an asset class	Strategic	Access discretionary active management in an ETF structure	ETFs from reputable fixed income or equity managers with active approach or theme
Dynamic asset allocation and multi-asset strategies	Dynamic or tactical	Seek returns from active allocation across asset classes or factors based on return or risk outlook Invest in a multi-asset-class strategy in single product	ETFs that allocate across asset categories or investment themes based on quantitative or fundamental factors

SUMMARY

We have examined important considerations for ETF investors, including how ETFs work and trade, tax efficient attributes, and key portfolio uses. The following is a summary of key points:

- ETFs rely on a creation/redemption mechanism that allows for the continuous creation and redemption of ETF shares.
- The only investors who can create or redeem new ETF shares are a special group of institutional investors called authorized participants.
- ETFs trade on both the primary market (directly between APs and issuers) and on the secondary markets (exchange-based or OTC trades, such as listed equity).
- End investors trade ETFs on the secondary markets, like stocks.
- Holding period performance deviations (tracking differences) are more useful than the standard deviation of daily return differences (tracking error).
- ETF tracking differences from the index occur for the following reasons:
  - fees and expenses,
  - representative sampling/optimization,

- use of depositary receipts and other ETFs,
- index changes,
- fund accounting practices,
- regulatory and tax requirements, and
- asset manager operations.

- ETFs are generally taxed in the same manner as the securities they hold, with some nuances:
  - ETFs are more tax fair than traditional mutual funds, because portfolio trading is generally not required when money enters or exits an ETF.
  - Owing to the creation/redemption process, ETFs can be more tax efficient than mutual funds.
  - ETF issuers can redeem out low-cost-basis securities to minimize future taxable gains.
  - Local markets have unique ETF taxation issues that should be considered.
- ETF bid–ask spreads vary by trade size and are usually published for smaller trade sizes. They are tightest for ETFs that are very liquid and have continuous two-way order flow. For less liquid ETFs, the following factors can determine the quoted bid–ask spread of an ETF trade:
  - Creation/redemption costs, brokerage and exchange fees
  - Bid–ask spread of underlying securities held by the ETF
  - Risk of hedging or carry positions by liquidity provider
  - Market makers' target profit spread
- ETF bid–ask spreads on fixed income relative to equity tend to be wider because the underlying bonds trade in dealer markets and hedging is more difficult. Spreads on ETFs holding international stocks are tightest when the underlying security markets are open for trading.
- ETF premiums and discounts refer to the difference between the exchange price of the ETF and the fund's calculated NAV, based on the prices of the underlying securities and weighted by the portfolio positions at the start of each trading day. Premiums and discounts can occur because NAVs are based on the last traded prices, which may be observed at a time lag to the ETF price, or because the ETF is more liquid and more reflective of current information and supply and demand than the underlying securities in rapidly changing markets.
- Costs of ETF ownership may be positive or negative and include both explicit and implicit costs. The main components of ETF cost are
  - the fund management fee;
  - tracking error;
  - portfolio turnover;
  - trading costs, such as commissions, bid–ask spreads, and premiums/discounts;
  - taxable gains/losses; and
  - security lending.
- Trading costs are incurred when the position is entered and exited. These one-time costs decrease as a portion of total holding costs over longer holding periods and are a more significant consideration for shorter-term tactical ETF traders.

- Other costs, such as management fees and portfolio turnover, increase as a proportion of overall cost as the investor holding period lengthens. These costs are a more significant consideration for longer-term buy-and-hold investors.
- ETFs are different from exchange-traded notes, although both use the creation/redemption process.
  - Exchange-traded notes carry unique counterparty risks of default.
  - Swap-based ETFs may carry counterparty risk.
  - ETFs, like mutual funds, may lend their securities, creating risk of counterparty default.
  - ETF closures can create unexpected tax liabilities.
- ETFs are used for core asset class exposure, multi-asset, dynamic, and tactical strategies based on investment views or changing market conditions; for factor or smart beta strategies with a goal to improve return or modify portfolio risk; and for portfolio efficiency applications, such as rebalancing, liquidity management, completion strategies, and transitions.
- ETFs are useful for investing cash inflows, as well as for raising proceeds to provide for client withdrawals. ETFs are used for rebalancing to target asset class weights and for "completion strategies" to fill a temporary gap in an asset class category, sector, or investment theme or when external managers are underweight. When positions are in transition from one external manager to another, ETFs are often used as the temporary holding and may be used to fund the new manager.
- All types of investors use ETFs to establish low-cost core exposure to asset classes, equity style benchmarks, fixed-income categories, and commodities.
- For more tactical investing, thematic ETFs are used in active portfolio management and represent narrow or niche areas of the equity market not well represented by industry or sector ETFs.
- Systematic, active strategies that use rules-based benchmarks for exposure to such factors as size, value, momentum, quality, or dividend tilts or combinations of these factors are frequently implemented with ETFs.
- Multi-asset and global asset allocation or macro strategies that manage positions dynamically as market conditions change are also areas where ETFs are frequently used.
- Proper utilization requires investors to carefully research and assess the ETF's index construction methodology, costs, risks, and performance history.

PRACTICE PROBLEMS

1 Which of the following statements regarding exchange-traded funds (ETFs) is correct? ETFs:

A disclose their holdings on a quarterly basis.

B trade in both primary and secondary markets.

C offer a creation/redemption mechanism that allows any investor to create or redeem shares.

2 The list of securities that a particular ETF wants to own, which is disclosed daily by all ETFs, is referred to as the:

A creation unit.

B creation basket.

C redemption basket.

3 When an authorized participant transacts to create or redeem ETF shares, the related costs are ultimately borne:

A solely by the ETF sponsor.

B solely by the AP.

C proportionally by all existing ETF shareholders.

4 Assuming arbitrage costs are minimal, which of the following is *most likely* to occur when the share price of an ETF is trading at a premium to its intraday NAV?

A New ETF shares will be created by the ETF sponsor.

B Redemption baskets will be received by APs from the ETF sponsor.

C Retail investors will exchange baskets of securities that the ETF tracks for creation units.

5 An ETF's reported tracking error is typically measured as the:

A standard deviation of the difference in daily returns between an ETF and its benchmark.

B difference in annual return between an ETF and its benchmark over the past 12 months.

C annualized standard deviation of the difference in daily returns between an ETF and its benchmark.

6 To best assess an ETF's performance, which reflects the impact of portfolio rebalancing expenses and other fees, an investor should:

A review daily return differences between the ETF and its benchmark.

B perform a rolling return assessment between the ETF and its benchmark.

C compare the ETF's annual expense ratio with that of other ETFs in its asset class category.

7 An ETF's tracking error, as traditionally reported, indicates to investors:

A whether the ETF is underperforming or outperforming its underlying index.

B the magnitude by which an ETF's returns deviate from its benchmark over time.

C the distribution of differences in daily returns between the ETF and its benchmark.

© 2020 CFA Institute. All rights reserved.

8 For a typical ETF, which of the following sources of tracking error is *most likely* to be the smallest contributor to tracking error?

- **A** Representative sampling
- **B** Fees and expenses incurred by the ETF
- **C** Changes to the underlying index securities

9 Which of the following statements relating to capital gains in ETFs and mutual funds is correct?

- **A** ETFs tend to distribute less in capital gains than mutual funds do.
- **B** Mutual funds may elect not to distribute all realized capital gains in a given year.
- **C** The selling of ETF shares by some investors may create capital gains that affect the remaining ETF investors in terms of taxes.

10 Which of the following statements regarding distributions made by ETFs is correct?

- **A** Return-of-capital (ROC) distributions are generally not taxable.
- **B** ETFs generally reinvest any dividends received back into the ETF's holdings.
- **C** A dividend distribution is a distribution paid to investors in excess of an ETF's earnings.

11 Such factors as regulations, competition, and corporate actions relate to:

- **A** fund-closure risk.
- **B** counterparty risk.
- **C** expectation-related risk.

12 John Smith has invested in an inverse ETF. Smith is a novice investor who is not familiar with inverse ETFs, and therefore, he is unsure how the ETF will perform because of a lack of understanding of the ETF's risk and return characteristics. This risk is *best* described as:

- **A** counterparty risk.
- **B** holdings-based risk.
- **C** expectation-related risk.

13 Investors buying ETFs:

- **A** incur management fees that decrease with the length of the holding period.
- **B** are assured of paying a price equal to the NAV if they purchase shares at the market close.
- **C** incur trading costs in the form of commissions and bid–ask spreads at the time of purchase.

14 Consider an ETF with the following trading costs and management fees:

- Annual management fee of 0.40%
- Round-trip trading commissions of 0.55%
- Bid–offer spread of 0.20% on purchase and sale

Excluding compound effects, the expected total holding-period cost for investing in the ETF over a nine-month holding period is *closest* to:

- **A** 1.05%.
- **B** 1.15%.
- **C** 1.25%.

15 The bid–ask spread for very liquid, high-volume ETFs will be *least* influenced by the:

A market maker's desired profit spread.

B creation/redemption fees and other direct costs.

C likelihood of receiving an offsetting ETF order in a short time frame.

16 Factor (smart beta) strategy ETFs are *least likely* to be used by investors:

A to modify portfolio risk.

B for tactical trading purposes.

C to seek outperformance versus a benchmark.

17 Which of the following statements regarding applications of ETFs in portfolio management is correct?

A Equity ETFs tend to be more active than fixed-income ETFs.

B The range of risk exposures available in the futures market is more diverse than that available in the ETF space.

C ETFs that have the highest trading volumes in their asset class category are generally preferred for tactical trading applications.

The following information relates to questions 18–23

Howie Rutledge is a senior portfolio strategist for an endowment fund. Rutledge meets with recently hired junior analyst Larry Stosur to review the fund's holdings.

Rutledge asks Stosur about the mechanics of exchange-traded funds (ETFs). Stosur responds by making the following statements:

Statement 1 Unlike mutual fund shares that can be shorted, ETF shares cannot be shorted.

Statement 2 In the ETF creation/redemption process, the authorized participants (APs) absorb the costs of transacting securities for the ETF's portfolio.

Statement 3 If ETF shares are trading at a discount to NAV and arbitrage costs are sufficiently low, APs will buy the securities in the creation basket and exchange them for ETF shares from the ETF sponsor.

Rutledge notes that one holding, ETF 1, is trading at a premium to its intraday NAV. He reviews the ETF's pricing and notes that the premium to the intraday NAV is greater than the expected arbitrage costs.

Stosur is evaluating three ETFs for potential investment. He notes that the ETFs have different portfolio characteristics that are likely to affect each ETF's tracking error. A summary of the characteristics for the ETFs is presented in Exhibit 1.

Exhibit 1 ETF Characteristics Affecting Tracking Error

	ETF 2	ETF 3	ETF 4
Portfolio Construction Approach	Full Replication	Representative Sampling	Full Replication
Type of Foreign Holdings	Local shares	ADRs*	ADRs*
Engagement in Securities Lending	Yes	Yes	No

*ADRs are American Depositary Receipts.

Rutledge and Stosur discuss the factors that influence ETF bid–ask spreads. Stosur tells Rutledge that quoted bid–ask spreads for a particular transaction size are (1) negatively related to the amount of the ongoing order flow in the ETF, (2) positively related to the costs and risks for the ETF liquidity provider, and (3) positively related to the amount of competition among market makers for the ETF.

As ETF shares may trade at prices that are different from the NAV, Rutledge examines selected data in Exhibit 2 for three ETFs that might have this problem.

Exhibit 2 Selected Data on ETFs

	ETF 5	ETF 6	ETF 7
Percentage of Foreign Holdings	10%	50%	90%
Trading Frequency	High	Low	Low

Rutledge considers a new ETF investment for the fund. He plans to own the ETF for nine months. The ETF has the following trading costs and management fees:

- Annual management fee of 0.32%
- Round-trip trading commissions of 0.20%
- Bid–offer spread of 0.10% on purchase and sale

Rutledge asks Stosur to compute the expected total holding period cost for investing in the ETF.

18 Which of Stosur's statements regarding ETF mechanics is correct?

A Statement 1

B Statement 2

C Statement 3

19 Given the current pricing of ETF 1, the *most likely* transaction to occur is that:

A new ETF shares will be created by the APs.

B redemption baskets will be received by APs from the ETF sponsor.

C retail investors will exchange baskets of securities that the ETF tracks for creation units.

20 Which ETF in Exhibit 1 is *most likely* to have the lowest tracking error?

A ETF 2

B ETF 3

C ETF 4

21 Stosur's statement about quoted bid–ask spreads is *incorrect* with respect to the:

A amount of the ongoing order flow in the ETF.

B costs and risks for the ETF liquidity providers.

C amount of competition among market makers for the ETF.

22 Which ETF in Exhibit 2 is *most likely* to trade at the largest premium or discount relative to NAV?

A ETF 5

B ETF 6

C ETF 7

23 Excluding the compounding effect, the expected total holding period cost for investing in the ETF over a nine-month holding period is *closest* to:

A 0.54%.

B 0.62%.

C 0.64%.

SOLUTIONS

1 B is correct. ETFs trade in both primary and secondary markets. The primary market for ETF trading is that which exists on an over-the-counter basis between authorized participants (APs), a special group of institutional investors, and the ETF issuer or sponsor. This process is referred to as creation/redemption, and it is only through these primary market transactions that shares of the ETF can be created or destroyed. ETFs also trade in the secondary market on exchanges. Secondary market trading happens between any pair of market participants—individual or institutional investors, market makers, and so on.

2 B is correct. Each day, ETF managers publicly disclose a list of securities that they want to own, which is referred to as the creation basket. This basket also serves as the portfolio for determining the intrinsic net asset value (NAV) of the ETF on the basis of prices during the trading day.

3 B is correct. The AP generally absorbs all the costs associated with buying or selling the securities in the baskets or the ETF shares and pays an additional fee to the ETF provider to cover processing fees associated with creation/redemption activities. APs pass these costs to investors in the ETF's bid–ask spread, which is incurred by investors entering (ETF share buyers) and exiting (ETF share sellers) the fund.

4 A is correct. When the share price of an ETF is trading at a premium to its intraday NAV and assuming arbitrage costs are minimal, APs will step in and take advantage of the arbitrage. Specifically, APs will step in and buy the basket of securities that the ETF tracks (the creation basket) and exchange it with the ETF provider for new ETF shares (a creation unit). These new shares received by APs can then be sold on the open market to realize arbitrage profits.

5 C is correct. An ETF's tracking error is typically reported as the annualized standard deviation of the daily differential returns of the ETF and its benchmark.

6 B is correct. A rolling return assessment, referred to in the ETF industry as the "tracking difference," provides a more informative picture of the investment outcome for an investor in an ETF. Such an analysis allows investors to see the cumulative effect of portfolio management and expenses over an extended period. It also allows for comparison with other annual metrics such as a fund's expense ratio. Tracking error, as a statistic, reveals only ETF tracking variability; it does not reveal to investors whether the fund is over- or underperforming its index or whether that tracking error is concentrated over a few days or is more consistently experienced. An ETF's expense ratio does not fully reflect the investor experience. That is, the expense ratio does not reflect the cost of portfolio rebalancing or other fees, making it an inferior assessment measure relative to a rolling return assessment.

7 B is correct. An ETF's tracking error is typically reported as the annualized standard deviation of the daily differential returns of the ETF and its benchmark. Therefore, an ETF's reported tracking error indicates to investors the magnitude by which an ETF's returns deviate from those of its benchmark over time.

8 C is correct. Although additions and deletions of securities from the underlying benchmark index may occur and result in tracking error, such index changes generally occur infrequently (often quarterly). In addition, ETF portfolio managers may work with APs for index rebalance trades to ensure market-on-close

pricing to minimize this source of tracking error. Therefore, the resulting tracking error caused by index changes will not likely be as large as the tracking error caused by representative sampling or by fees and expenses incurred by the ETF.

9 A is correct. ETFs tend to distribute far less in capital gains relative to mutual funds. This is mostly due to the fact that ETFs have historically had significantly lower turnover than mutual funds have had.

10 A is correct. Return-of-capital distributions are amounts paid out in excess of an ETF's earnings and serve to reduce an investor's cost basis by the amount of the distribution. These distributions are generally not taxable.

11 A is correct. Fund-closure risk is the risk that an ETF may shut down. The reasons that lead to an ETF closing down often have to do with changes in regulations, increased competition, and corporate activity (merger and acquisition activity within the ETF industry).

12 C is correct. Expectation-related risk is the risk that some ETF investors may not fully understand how more complex ETFs will perform because of a lack of understanding of sophisticated assets classes and strategies.

13 C is correct. ETF trading costs in the form of commissions and bid–ask spreads are paid by investors buying or selling ETF shares on an exchange. These trading costs are influenced by the bid–ask spread of the ETF, the size of the trade relative to the normal trading activity of the ETF, and the ease of hedging the ETF by the market-making community. Even the closing price of the ETF on the exchange includes a premium or discount to the NAV, driven by supply and demand factors on the exchange and the market impact costs of executing an exchange transaction. The purchase and sale trading costs of an ETF are paid regardless of holding period, whereas other costs, such as management fees, increase as the holding period lengthens.

14 A is correct. The expected total holding-period cost for investing in the ETF over a nine-month holding period is calculated as follows:

Total holding-period cost = Annual management fee + Round-trip trading commissions + Bid–offer spread on purchase/sale.

Total holding-period cost = (9/12) × (0.40%) + 0.55% + 0.20% = 1.05%.

15 B is correct. ETF bid–ask spreads are generally less than or equal to the combination of the following:

- ± Creation/redemption fees and other direct costs, such as brokerage and exchange fees
- + Bid–ask spread of the underlying securities held by the ETF
- + Compensation for the risk of hedging or carrying positions by liquidity providers (market makers) for the remainder of the trading day
- + Market maker's desired profit spread
- – Discount related to the likelihood of receiving an offsetting ETF order in a short time frame

For very liquid and high-volume ETFs, buyers and sellers are active throughout the trading day. Therefore, because most of these ETF trades are matched extremely quickly and never involve the creation/redemption process, the first three factors listed do not contribute heavily to their bid–ask spreads. So, creation/redemption fees and other direct costs are not likely to have much influence on these ETFs' bid–ask spreads.

16 B is correct. Factor strategy ETFs are usually benchmarked to an index created with predefined rules for screening and/or weighting stock holdings and are considered longer-term, buy-and-hold investment options rather than tactical trading instruments. The strategy index rules are structured around return drivers or factors, such as value, dividend yield, earnings or dividend growth, quality, stock volatility, or momentum. Investors using factor-based investing seek outperformance versus a benchmark or portfolio risk modification.

17 C is correct. ETFs that have the highest trading volumes in their asset class category are generally preferred for tactical trading applications.

18 B is correct. Statement 2 is correct. A significant advantage of the ETF creation/redemption process is that the AP absorbs all costs of transacting the securities for the fund's portfolio. APs pass these costs to investors in the ETF's bid–ask spread, incurred by ETF buyers and sellers. Thus, non-transacting shareholders of an ETF are shielded from the negative impact of transaction costs caused by other investors entering and exiting the fund. In contrast, when investors enter or exit a traditional mutual fund, the mutual fund manager incurs costs to buy or sell investments arising from this activity, which affects all fund shareholders. This makes the ETF structure inherently fairer: Frequent ETF traders bear the cost of their activity, while buy-and-hold ETF shareholders are shielded from those costs. Investors cannot short mutual fund shares, but they can short ETF shares. Also, if ETF shares are trading at a discount to NAV and arbitrage costs are sufficiently low, APs will buy ETF shares and exchange them for the securities in the redemption basket. Statement 3 describes the scenario that would occur if the ETF shares are trading at a premium to NAV.

A is incorrect because Statement 1 is incorrect. Investors cannot short mutual fund shares, but they can short ETF shares.

C is incorrect because Statement 3 is incorrect. If ETF shares are trading at a discount to NAV and arbitrage costs are sufficiently low, APs will buy ETF shares and exchange them for the securities in the redemption basket. Statement 3 describes the scenario that would occur if ETF shares are trading at a premium to NAV.

19 A is correct. When the share price of an ETF is trading at a premium to its intraday NAV and arbitrage costs are minimal, APs will step in and take advantage of the arbitrage. Specifically, APs will buy the basket of securities that the ETF tracks (the creation basket) and exchange it with the ETF sponsor for new ETF shares (a creation unit). These new ETF shares received by APs can then be sold on the open market to realize arbitrage profits.

B is incorrect because in the case of an ETF trading at a premium to NAV, the APs will not receive redemption baskets of securities. Instead, the APs will deliver creation baskets to the ETF sponsor and receive new ETF shares.

C is incorrect because only APs can deliver creation baskets or receive redemption baskets from the ETF sponsors. Retail investors can buy and sell ETF shares on the open market.

20 A is correct. Compared with a full replication approach, ETF portfolios managed using a representative sampling/optimization approach are likely to have greater tracking error. Also, differences in trading hours for depositary receipts and local constituent shares create discrepancies between the portfolio and index values. These discrepancies can lead to greater tracking error for portfolios holding ADRs in lieu of the underlying local shares. Further, ETF sponsors that engage in securities lending can generate additional portfolio income to help offset fund expenses, thereby lowering tracking error. ETF 2 uses a full replication approach, holds only local foreign shares, and engages in securities lending. Therefore, ETF 2 will likely have the lowest tracking error out of the ETFs in Exhibit 1. ETF 3 will likely have greater tracking error than ETF 2 because it is managed using a representative sampling approach and is invested in depositary receipts in lieu of local shares. ETF 4 will likely have greater tracking error than ETF 2 because it is invested in depositary receipts in lieu of local shares and does not engage in securities lending.

21 C is correct. Several factors determine the width of an ETF's quoted bid–ask spread. First, the amount of ongoing order flow in the ETF is negatively related to the bid–ask spread (more flow means lower spreads). Second, the actual costs and risks for the liquidity provider are positively related to spreads (more costs and risks mean higher spreads); the spread is compensation to the liquidity provider for incurring these costs and risks. Finally, the amount of competition among market makers for that ETF is negatively related to the bid–ask spread (more competition means lower spreads).

A is incorrect because Stosur is correct in stating that the quoted bid–ask spread for a particular transaction size is negatively related to the amount of the ongoing order flow in the ETF (more flow means lower spreads).

B is incorrect because Stosur is correct in stating that the quoted bid–ask spread for a particular transaction size is positively related to the costs and risks for the ETF liquidity provider (more costs and risks mean higher spreads). The bid–ask spread represents the market maker's price for taking the other side of the ETF transaction, which includes the costs and risks to carry the position on its books and/or to hedge the position using underlying securities or closely related ETFs or derivatives.

22 C is correct. ETFs that trade infrequently may have large premiums or discounts to NAV, because the ETF may not have traded in the hours leading up to the market close and NAV may have significantly risen or fallen during that time because of market movement. Furthermore, NAV is often a poor fair value indicator for ETFs holding foreign securities because of differences in exchange closing times between the underlying (e.g., foreign stocks, bonds, or commodities) and the exchange where the ETF trades. Therefore, ETF 7 is most likely to have the largest discount or premium because it has a low trading frequency and has the highest percentage of foreign holdings among the three ETFs.

A is incorrect because ETF 5 has the lowest percentage of foreign holdings among the three ETFs and is the one ETF with a high trading frequency. Therefore, relative to ETF 7, with its low trading frequency and high foreign holdings, ETF 5 is likely to trade at smaller premiums or discounts.

B is incorrect because ETF 6 has a lower percentage of foreign holdings than ETF 7. Even though both ETF 6 and ETF 7 have the same low trading frequency, the lower percentage of foreign holdings for ETF 6 is likely to result in it trading at smaller premiums or discounts.

23 A is correct. The expected total holding period cost for investing in the ETF over the nine-month holding period is calculated as follows:

Total expected holding period cost = Annual management fee + Round-trip trading commissions + Bid–offer spread on purchase/sale.

Total expected holding period cost = (9/12) × (0.32%) + 0.20% + 0.10% = 0.54%.

READING

39

Using Multifactor Models

by Jerald E. Pinto, PhD, CFA, and Eugene L. Podkaminer, CFA

Jerald E. Pinto, PhD, CFA, is at CFA Institute (USA). Eugene L. Podkaminer, CFA, is at Franklin Templeton Investments (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** describe arbitrage pricing theory (APT), including its underlying assumptions and its relation to multifactor models;
☐	**b.** define arbitrage opportunity and determine whether an arbitrage opportunity exists;
☐	**c.** calculate the expected return on an asset given an asset's factor sensitivities and the factor risk premiums;
☐	**d.** describe and compare macroeconomic factor models, fundamental factor models, and statistical factor models;
☐	**e.** explain sources of active risk and interpret tracking risk and the information ratio;
☐	**f.** describe uses of multifactor models and interpret the output of analyses based on multifactor models;
☐	**g.** describe the potential benefits for investors in considering multiple risk dimensions when modeling asset returns.

1 BACKGROUND AND USES OF MULTIFACTOR MODELS

As used in investments, a **factor** is a variable or a characteristic with which individual asset returns are correlated. Models using multiple factors are used by asset owners, asset managers, investment consultants, and risk managers for a variety of portfolio construction, portfolio management, risk management, and general analytical

CFA Institute would like to thank Yin Luo, CPA, PStat, CFA, and Sheng Wang, both of Wolfe Research LLC (USA), for contributing the case "Constructing Multifactor Portfolios" in Section 8.

© 2021 CFA Institute. All rights reserved.

purposes. In comparison to single-factor models (typically based on a market risk factor), multifactor models offer increased explanatory power and flexibility. These comparative strengths of multifactor models allow practitioners to

- build portfolios that replicate or modify in a desired way the characteristics of a particular index;
- establish desired exposures to one or more risk factors, including those that express specific macro expectations (such as views on inflation or economic growth), in portfolios;
- perform granular risk and return attribution on actively managed portfolios;
- understand the comparative risk exposures of equity, fixed-income, and other asset class returns;
- identify active decisions relative to a benchmark and measure the sizing of those decisions; and
- ensure that an investor's aggregate portfolio is meeting active risk and return objectives commensurate with active fees.

Multifactor models have come to dominate investment practice, having demonstrated their value in helping asset managers and asset owners address practical tasks in measuring and controlling risk. We explain and illustrate the various practical uses of multifactor models.

We first describe the modern portfolio theory background of multifactor models. We then describe arbitrage pricing theory and provide a general expression for multifactor models. We subsequently explore the types of multifactor models and certain applications. Lastly, we summarize major points.

1.1 Multifactor Models and Modern Portfolio Theory

In 1952, Markowitz introduced a framework for constructing portfolios of securities by quantitatively considering each investment in the context of a portfolio rather than in isolation; that framework is widely known today as modern portfolio theory (MPT). Markowitz simplified modeling asset returns using a multivariate normal distribution, which completely defines the distribution of returns in terms of mean returns, return variances, and return correlations. One of the key insights of MPT is that any value of correlation among asset returns of less than one offers the potential for risk reduction by means of diversification.

In 1964, Sharpe introduced the capital asset pricing model (CAPM), a model for the expected return of assets in equilibrium based on a mean–variance foundation. The CAPM and the literature that developed around it has provided investors with useful and influential concepts—such as alpha, beta, and systematic risk—for thinking about investing. The concept of systematic risk, for example, is critical to understanding multifactor models: An investment may be subject to many different types of risks, but they are generally not equally important so far as investment valuation is concerned. Risk that can be avoided by holding an asset in a portfolio, where the risk might be offset by the various risks of other assets, should not be compensated by higher expected return, according to theory. By contrast, investors would expect compensation for bearing an asset's non-diversifiable risk: **systematic risk**. Theory indicates that only systematic risk should be **priced risk**. In the CAPM, an asset's systematic risk is a positive function of its beta, which measures the sensitivity of an asset's return to the market's return. According to the CAPM, differences in mean return are explained by a single factor: market portfolio return. Greater risk with respect to the market factor, represented by higher beta, is expected to be associated with higher return.

The accumulation of evidence from the equity markets during the decades following the CAPM's development have provided clear indications that the CAPM provides an incomplete description of risk and that models incorporating multiple sources of systematic risk more effectively model asset returns. Bodie, Kane, and Marcus (2017) provide an introduction to the empirical evidence. There are, however, various perspectives in practice on how to model risk in the context of multifactor models. We will examine some of these—focusing on macroeconomic factor models and fundamental factor models—in subsequent sections.

ARBITRAGE PRICING THEORY AND MULTIFACTOR MODELS 2

a describe arbitrage pricing theory (APT), including its underlying assumptions and its relation to multifactor models

b define arbitrage opportunity and determine whether an arbitrage opportunity exists

c calculate the expected return on an asset given an asset's factor sensitivities and the factor risk premiums

In the 1970s, Ross (1976) developed the arbitrage pricing theory (APT) as an alternative to the CAPM. APT introduced a framework that explains the expected return of an asset (or portfolio) in equilibrium as a linear function of the risk of the asset (or portfolio) with respect to a set of factors capturing systematic risk. Unlike the CAPM, the APT does not indicate the identity or even the number of risk factors. Rather, for any multifactor model assumed to generate returns ("return-generating process"), the theory gives the associated expression for the asset's expected return.

Suppose that K factors are assumed to generate returns. Then the simplest expression for a multifactor model for the return of asset i is given by

$$R_i = a_i + b_{i1}I_1 + b_{i2}I_2 + \ldots + b_{iK}I_K + \varepsilon_i, \quad (1)$$

where

R_i = the return to asset i

a_i = an intercept term

I_k = the return to factor k, $k = 1, 2, \ldots, K$

b_{ik} = the sensitivity of the return on asset i to the return to factor k, $k = 1, 2, \ldots, K$

ε_i = an error term with a zero mean that represents the portion of the return to asset i not explained by the factor model

The intercept term a_i is the expected return of asset i given that all the factors take on a value of zero. Equation 1 presents a multifactor return-generating process (a time-series model for returns). In any given period, the model may not account fully for the asset's return, as indicated by the error term. But error is assumed to average to zero. Another common formulation subtracts the risk-free rate from both sides of Equation 1 so that the dependent variable is the return in excess of the risk-free rate and one of the explanatory variables is a factor return in excess of the risk-free rate. (The Carhart model described next is an example.)

Based on Equation 1, the APT provides an expression for the expected return of asset i assuming that financial markets are in equilibrium. The APT is similar to the CAPM, but the APT makes less strong assumptions than the CAPM. The APT makes just three key assumptions:

1 A factor model describes asset returns.

2 With many assets to choose from, investors can form well-diversified portfolios that eliminate asset-specific risk.

3 No arbitrage opportunities exist among well-diversified portfolios.

Arbitrage is a risk-free operation that requires no net investment of money but earns an expected positive net profit. (Note that “arbitrage,” or the phrase “risk arbitrage,” is also sometimes used in practice to describe investment operations in which significant risk is present). An **arbitrage opportunity** is an opportunity to conduct an arbitrage—an opportunity to earn an expected positive net profit without risk and with no net investment of money.

In the first assumption, the number of factors is not specified. The second assumption allows investors to form portfolios with factor risk but without asset-specific risk. The third assumption is the condition of financial market equilibrium.

Empirical evidence indicates that Assumption 2 is reasonable (Fabozzi, 2008). When a portfolio contains many stocks, the asset-specific or non-systematic risk of individual stocks makes almost no contribution to the variance of portfolio returns.

According to the APT, if these three assumptions hold, the following equation holds:

$$E(R_p) = R_F + \lambda_1\beta_{p,1} + \ldots + \lambda_K\beta_{p,K}, \tag{2}$$

where

$E(R_p)$ = the expected return to portfolio p

R_F = the risk-free rate

λ_j = the expected reward for bearing the risk of factor j

$\beta_{p,j}$ = the sensitivity of the portfolio to factor j

K = the number of factors

The APT equation, Equation 2, says that the expected return on any well-diversified portfolio is linearly related to the factor sensitivities of that portfolio. The equation assumes that a risk-free rate exists. If no risk-free asset exists, in place of R_F we write λ_0 to represent the expected return on a risky portfolio with zero sensitivity to all the factors. The number of factors is not specified but must be much lower than the number of assets, a condition fulfilled in practice.

The **factor risk premium** (or **factor price**), λ_j, represents the expected reward for bearing the risk of a portfolio with a sensitivity of 1 to factor j and a sensitivity of 0 to all other factors. The exact interpretation of “expected reward” depends on the multifactor model that is the basis for Equation 2. For example, in the Carhart four-factor model, shown later in Equation 3a and 3b, the risk premium for the market factor is the expected return of the market in excess of the risk-free rate. Then, the factor risk premiums for the other three factors are the mean returns of the specific portfolios held long (e.g., the portfolio of small-cap stocks for the “small minus big” factor) minus the mean return for a related but opposite portfolio (e.g., a portfolio of large-cap stocks, in the case of that factor). A portfolio with a sensitivity of 1 to factor j and a sensitivity of 0 to all other factors is called a **pure factor portfolio** for factor j (or simply the **factor portfolio** for factor j).

For example, suppose we have a portfolio with a sensitivity of 1 with respect to Factor 1 and a sensitivity of 0 to all other factors. Using Equation 2, the expected return on this portfolio is $E_1 = R_F + \lambda_1 \times 1$. If $E_1 = 0.12$ and $R_F = 0.04$, then the risk premium for Factor 1 is

$$0.12 = 0.04 + \lambda_1 \times 1.$$

$$\lambda_1 = 0.12 - 0.04 = 0.08, \text{ or } 8\%.$$

EXAMPLE 1

Determining the Parameters in a One-Factor APT Model

Suppose we have three well-diversified portfolios that are each sensitive to the same single factor. Exhibit 1 shows the expected returns and factor sensitivities of these portfolios. Assume that the expected returns reflect a one-year investment horizon. To keep the analysis simple, all investors are assumed to agree upon the expected returns of the three portfolios as shown in the exhibit.

Exhibit 1 Sample Portfolios for a One-Factor Model

Portfolio	Expected Return	Factor Sensitivity
A	0.075	0.5
B	0.150	2.0
C	0.070	0.4

We can use these data to determine the parameters of the APT equation. According to Equation 2, for any well-diversified portfolio and assuming a single factor explains returns, we have $E(R_p) = R_F + \lambda_1\beta_{p,1}$. The factor sensitivities and expected returns are known; thus there are two unknowns, the parameters R_F and λ_1. Because two points define a straight line, we need to set up only two equations. Selecting Portfolios A and B, we have

$$E(R_A) = 0.075 = R_F + 0.5\lambda_1$$

and

$$E(R_B) = 0.150 = R_F + 2\lambda_1.$$

From the equation for Portfolio A, we have $R_F = 0.075 - 0.5\lambda_1$. Substituting this expression for the risk-free rate into the equation for Portfolio B gives

$$0.15 = 0.075 - 0.5\lambda_1 + 2\lambda_1.$$

$$0.15 = 0.075 + 1.5\lambda_1.$$

So, we have $\lambda_1 = (0.15 - 0.075)/1.5 = 0.05$. Substituting this value for λ_1 back into the equation for the expected return to Portfolio A yields

$$0.075 = R_F + 0.05 \times 0.5.$$

$$R_F = 0.05.$$

So, the risk-free rate is 0.05 or 5%, and the factor premium for the common factor is also 0.05 or 5%. The APT equation is

$$E(R_p) = 0.05 + 0.05\beta_{p,1}.$$

From Exhibit 1, Portfolio C has a factor sensitivity of 0.4. Therefore, according to the APT, the expected return of Portfolio C should be

$$E(R_B) = 0.05 + (0.05 \times 0.4) = 0.07,$$

which is consistent with the expected return for Portfolio C given in Exhibit 1.

EXAMPLE 2

Checking Whether Portfolio Returns Are Consistent with No Arbitrage

In this example, we examine how to tell whether expected returns and factor sensitivities for a set of well-diversified portfolios may indicate the presence of an arbitrage opportunity. Exhibit 2 provides data on four hypothetical portfolios. The data for Portfolios A, B, and C are repeated from Exhibit 1. Portfolio D is a new portfolio. The factor sensitivities given relate to the one-factor APT model $E(R_p) = 0.05 + 0.05\beta_{p,1}$ derived in Example 1. As in Example 1, all investors are assumed to agree upon the expected returns of the portfolios. The question raised by the addition of this new Portfolio D is whether the addition of this portfolio created an arbitrage opportunity. If a portfolio can be formed from Portfolios A, B, and C that has the same factor sensitivity as Portfolio D but a different expected return, then an arbitrage opportunity exists: Portfolio D would be either undervalued (if it offers a relatively high expected return) or overvalued (if it offers a relatively low expected return).

Exhibit 2 Sample Portfolios for a One-Factor Model

Portfolio	Expected Return	Factor Sensitivity
A	0.0750	0.50
B	0.1500	2.00
C	0.0700	0.40
D	0.0800	0.45
0.5A + 0.5C	0.0725	0.45

Exhibit 2 gives data for an equally weighted portfolio of A and C. The expected return and factor sensitivity of this new portfolio are calculated as weighted averages of the expected returns and factor sensitivities of A and C. Expected return is thus (0.50)(0.0750) + (0.50)(0.07) = 0.0725, or 7.25%. The factor sensitivity is (0.50)(0.50) + (0.50)(0.40) = 0.45. Note that the factor sensitivity of 0.45 matches the factor sensitivity of Portfolio D. In this case, the configuration of expected returns in relation to factor risk presents an arbitrage opportunity involving Portfolios A, C, and D. Portfolio D offers, at 8%, an expected return that is too high given its factor sensitivity. According to the assumed APT model, the expected return on Portfolio D should be $E(R_D) = 0.05 + 0.05\beta_{D,1} = 0.05 + (0.05 \times 0.45) = 0.0725$, or 7.25%. Portfolio D is undervalued relative to its factor risk. We will buy D (hold it long) in the portfolio that exploits the arbitrage opportunity (the **arbitrage portfolio**). We purchase D using the proceeds from selling short an equally weighted portfolio of A and C with exactly the same 0.45 factor sensitivity as D.

The arbitrage thus involves the following strategy: Invest $10,000 in Portfolio D and fund that investment by selling short an equally weighted portfolio of Portfolios A and C; then close out the investment position at the end of one year (the investment horizon for expected returns). Exhibit 3 demonstrates the arbitrage profits to the arbitrage strategy. The final row of the exhibit shows the net cash flow to the arbitrage portfolio.

Exhibit 3 Arbitrage Opportunity within Sample Portfolios

	Initial Cash Flow	Final Cash Flow	Factor Sensitivity
Portfolio D	–$10,000.00	$10,800.00	0.45
Portfolios A and C	$10,000.00	–$10,725.00	–0.45
Sum	$0.00	$75.00	0.00

As Exhibit 3 shows, if we buy $10,000 of Portfolio D and sell $10,000 of an equally weighted portfolio of Portfolios A and C, we have an initial net cash flow of $0. The expected value of our investment in Portfolio D at the end of one year is $10,000(1 + 0.08) = $10,800. The expected value of our short position in Portfolios A and C at the end of one year is –$10,000(1.0725) = –$10,725. So, the combined expected cash flow from our investment position in one year is $75.

What about the risk? Exhibit 3 shows that the factor risk has been eliminated: Purchasing D and selling short an equally weighted portfolio of A and C creates a portfolio with a factor sensitivity of 0.45 – 0.45 = 0. The portfolios are well diversified, and we assume any asset-specific risk is negligible.

Because an arbitrage is possible, Portfolios A, C, and D cannot all be consistent with the same equilibrium. If Portfolio D actually had an expected return of 8%, investors would bid up its price until the expected return fell and the arbitrage opportunity vanished. Thus, arbitrage restores equilibrium relationships among expected returns.

The Carhart four-factor model, also known as the four-factor model or simply the Carhart model, is a frequently referenced multifactor model in current equity portfolio management practice. Presented in Carhart (1997), it is an extension of the three-factor model developed by Fama and French (1992) to include a momentum factor. According to the model, three groups of stocks tend to have higher returns than those predicted solely by their sensitivity to the market return:

- Small-capitalization stocks
- Low price-to-book stocks, commonly referred to as "value" stocks
- Stocks whose prices have been rising, commonly referred to as "momentum" stocks

On the basis of that evidence, the Carhart model posits the existence of three systematic risk factors beyond the market risk factor. They are named, in the same order as above, the following:

- Small minus big (SMB)
- High minus low (HML)
- Winners minus losers (WML)

Equation 3a is the Carhart model, in which the excess return on the portfolio is explained as a function of the portfolio's sensitivity to a market index (RMRF), a market capitalization factor (SMB), a book-to-market factor (HML), which is essentially the reciprocal of the aforementioned price-to-book ratio, and a momentum factor (WML).

$$R_p - R_F = a_p + b_{p1}\text{RMRF} + b_{p2}\text{SMB} + b_{p3}\text{HML} + b_{p4}\text{WML} + \varepsilon_p, \tag{3a}$$

where

R_p and R_F = the return on the portfolio and the risk-free rate of return, respectively

a_p = "alpha" or return in excess of that expected given the portfolio's level of systematic risk (assuming the four factors capture all systematic risk)

b_p = the sensitivity of the portfolio to the given factor

RMRF = the return on a value-weighted equity index in excess of the one-month T-bill rate

SMB = small minus big, a size (market capitalization) factor; SMB is the average return on three small-cap portfolios minus the average return on three large-cap portfolios

HML = high minus low, the average return on two high book-to-market portfolios minus the average return on two low book-to-market portfolios

WML = winners minus losers, a momentum factor; WML is the return on a portfolio of the past year's winners minus the return on a portfolio of the past year's losers. (Note that WML is an equally weighted average of the stocks with the highest 30% 11-month returns lagged 1 month minus the equally weighted average of the stocks with the lowest 30% 11-month returns lagged 1 month.)

ε_p = an error term that represents the portion of the return to the portfolio, *p*, not explained by the model

Following Equation 2, the Carhart model can be stated as giving equilibrium expected return as

$$E(R_p) = R_F + \beta_{p,1}\text{RMRF} + \beta_{p,2}\text{SMB} + \beta_{p,3}\text{HML} + \beta_{p,4}\text{WML} \tag{3b}$$

because the expected value of alpha is zero.

The Carhart model can be viewed as a multifactor extension of the CAPM that explicitly incorporates drivers of differences in expected returns among assets variables that are viewed as anomalies from a pure CAPM perspective. (The term "anomaly" in this context refers to an observed capital market regularity that is not explained by, or contradicts, a theory of asset pricing.) From the perspective of the CAPM, there are size, value, and momentum anomalies. From the perspective of the Carhart model, however, size, value, and momentum represent systematic risk factors; exposure to them is expected to be compensated in the marketplace in the form of differences in mean return.

Size, value, and momentum are common themes in equity portfolio construction, and all three factors continue to have robust uses in active management risk decomposition and return attribution.

TYPES OF MULTIFACTOR MODELS

3

d describe and compare macroeconomic factor models, fundamental factor models, and statistical factor models

f describe uses of multifactor models and interpret the output of analyses based on multifactor models

Having introduced the APT, it is appropriate to examine the diversity of multifactor models in current use.

In the following sections, we explain the basic principles of multifactor models and discuss various types of models and their application. We also expand on the APT, which relates the expected return of investments to their risk with respect to a set of factors.

3.1 Factors and Types of Multifactor Models

Many varieties of multifactor models have been proposed and researched. We can categorize most of them into three main groups according to the type of factor used:

- In a **macroeconomic factor model**, the factors are surprises in macroeconomic variables that significantly explain returns. In the example of equities, the factors can be understood as affecting either the expected future cash flows of companies or the interest rate used to discount these cash flows back to the present. Among macroeconomic factors that have been used are interest rates, inflation risk, business cycle risk, and credit spreads.
- In a **fundamental factor model**, the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices. Among the fundamental factors that have been used are the book-value-to-price ratio, market capitalization, the price-to-earnings ratio, and financial leverage.
- In a **statistical factor model**, statistical methods are applied to historical returns of a group of securities to extract factors that can explain the observed returns of securities in the group. In statistical factor models, the factors are actually portfolios of the securities in the group under study and are therefore defined by portfolio weights. Two major types of factor models are factor analysis models and principal components models. In factor analysis models, the factors are the portfolios of securities that best explain (reproduce) historical *return covariances*. In principal components models, the factors are portfolios of securities that best explain (reproduce) the historical *return variances*.

A potential advantage of statistical factor models is that they make minimal assumptions. But the interpretation of statistical factors is generally difficult in contrast to macroeconomic and fundamental factors. A statistical factor that is a portfolio with weights that are similar to market index weights might be interpreted as “the market factor,” for example. But in general, associating a statistical factor with economic meaning may not be possible. Because understanding statistical factor models requires substantial preparation in quantitative methods, a detailed discussion of statistical factor models is outside the scope of our coverage.

Our discussion concentrates on macroeconomic factor models and fundamental factor models. Industry use has generally favored fundamental and macroeconomic models, perhaps because such models are much more easily interpreted and rely less on data-mining approaches. Nevertheless, statistical factor models have proponents and are also used in practical applications.

3.2 The Structure of Fundamental Factor Models

We earlier gave the equation of a macroeconomic factor model as

$$R_i = a_i + b_{i1}F_1 + b_{i2}F_2 + \ldots + b_{iK}F_K + \varepsilon_i.$$

We can also represent the structure of fundamental factor models with this equation, but we need to interpret the terms differently.

In fundamental factor models, the factors are stated as *returns* rather than return *surprises* in relation to predicted values, so they do not generally have expected values of zero. This approach changes the meaning of the intercept, which is no longer interpreted as the expected return. Note that if the coefficients were not standardized, as described in the following paragraph, the intercept could be interpreted as the risk-free rate because it would be the return to an asset with no factor risk (zero factor betas) and no asset-specific risk (with standardized coefficients, the intercept is not interpreted beyond being an intercept in a regression included so that the expected asset-specific risk equals zero).

Factor sensitivities are also interpreted differently in most fundamental factor models. In fundamental factor models, the factor sensitivities are attributes of the security. An asset's sensitivity to a factor is expressed using a **standardized beta**: the value of the attribute for the asset minus the average value of the attribute across all stocks divided by the standard deviation of the attribute's values across all stocks.

$$b_{ik} = \frac{\text{Value of attribute } k \text{ for asset } i - \text{Average value of attribute } k}{\sigma(\text{Values of attribute } k)} \tag{4}$$

Consider a fundamental model for equities that uses a dividend yield factor. After standardization, a stock with an average dividend yield will have a factor sensitivity of 0; a stock with a dividend yield one standard deviation above the average will have a factor sensitivity of 1; and a stock with a dividend yield one standard deviation below the average will have a factor sensitivity of −1. Suppose, for example, that an investment has a dividend yield of 3.5% and that the average dividend yield across all stocks being considered is 2.5%. Further, suppose that the standard deviation of dividend yields across all stocks is 2%. The investment's sensitivity to dividend yield is (3.5% − 2.5%)/2% = 0.50, or one-half standard deviation above average. The scaling permits all factor sensitivities to be interpreted similarly, despite differences in units of measure and scale in the variables. The exception to this interpretation is factors for binary variables, such as industry membership. A company either participates in an industry or does not. The industry factor is represented by dummy variables: The value of the variable is 1 if the stock belongs to the industry and 0 if it does not.

A second distinction between macroeconomic multifactor models and fundamental factor models is that with the former, we develop the factor (surprise) series first and then estimate the factor sensitivities through regressions. With the latter, we generally specify the factor sensitivities (attributes) first and then estimate the factor returns through regressions.

Financial analysts use fundamental factor models for a variety of purposes, including portfolio performance attribution and risk analysis. (*Performance attribution* consists of return attribution and risk attribution. *Return attribution* is a set of techniques used to identify the sources of the excess return of a portfolio against its benchmark. *Risk attribution* addresses the sources of risk, identifying the sources of portfolio volatility for absolute mandates and the sources of tracking risk for relative mandates.) Fundamental factor models focus on explaining the returns to individual stocks using observable fundamental factors that describe either attributes of the securities themselves or attributes of the securities' issuers. Industry membership, price-to-earnings ratio, book-value-to-price ratio, size, and financial leverage are examples of fundamental factors.

Example 4 discusses a study that examined macroeconomic, fundamental, and statistical factor models.

We encounter a range of distinct representations of risk in the fundamental models that are currently used in practical applications. Diversity exists in both the identity and exact definition of factors as well as in the underlying functional form and estimation procedures. Despite the diversity, we can place the factors of most fundamental factor models for equities into three broad groups:

- **Company fundamental factors**. These are factors related to the company's internal performance. Examples are factors relating to earnings growth, earnings variability, earnings momentum, and financial leverage.
- **Company share-related factors**. These factors include valuation measures and other factors related to share price or the trading characteristics of the shares. In contrast to the previous category, these factors directly incorporate investors' expectations concerning the company. Examples include price multiples, such as earnings yield, dividend yield, and book to market. Market capitalization falls under this heading. Various models incorporate variables relating to share price momentum, share price volatility, and trading activity that fall in this category.
- **Macroeconomic factors**. Sector or industry membership factors fall under this heading. Various models include such factors as CAPM beta, other similar measures of systematic risk, and yield curve level sensitivity—all of which can be placed in this category.

For global factor models, in particular, a classification of country, industry, and style factors is often used. In that classification, country and industry factors are dummy variables for country and industry membership, respectively. Style factors include those related to earnings, risk, and valuation that define types of securities typical of various styles of investing.

3.3 Fixed-Income Multifactor Models

While the previous discussion focuses on equity applications, similar approaches are equally suited to fixed income. In addition, some of the same broad factor groupings are relevant for bonds.

3.3.1 *Macroeconomic Multifactor Models*

Macroeconomic models, as discussed earlier, are easily translatable to fixed-income investing. For instance, surprises to economic growth, interest rates, and inflation will impact bond pricing, often mechanically.

Consider a bond factor model in which the returns are correlated with two factors. Following our earlier discussion, returns for bonds are assumed to be correlated with surprises in inflation rates and surprises in GDP growth. The return to *bond i*, R_i, can be modeled as

$$R_i = a_i + b_{i1}F_{INFL} + b_{i2}F_{GDP} + \varepsilon_i,$$

where

R_i = the return to bond i

a_i = the expected return to bond i

b_{i1} = the sensitivity of the return on bond i to inflation rate surprises

F_{INFL} = the surprise in inflation rates

b_{i2} = the sensitivity of the return on bond i to GDP growth surprises

F_{GDP} = the surprise in GDP growth (assumed to be uncorrelated with F_{INFL})

ε_i = an error term with a zero mean that represents the portion of the return to bond i not explained by the factor model

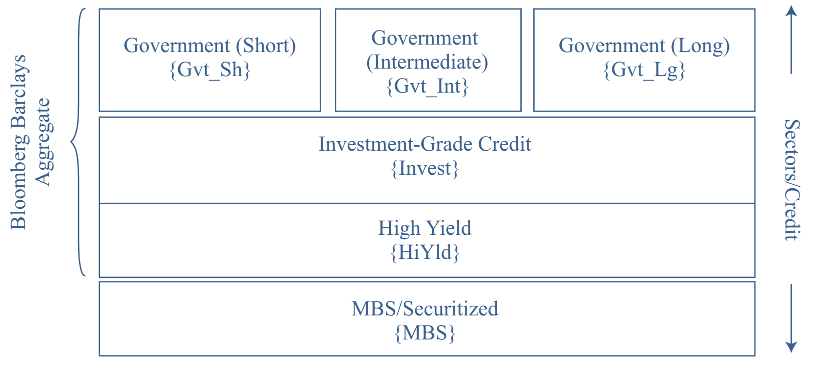
3.3.2 *Fundamental Multifactor Models*

Fundamental factor approaches have been developed to address the unique aspects of fixed income by using, for example, the following categories:

- Duration (ranging from cash to long-dated bonds)
- Credit (ranging from government securities to high yield)
- Currency (ranging from home currency to foreign developed and emerging market currencies)
- Geography (specific developed and emerging markets)

A simplified structure, shown in Exhibit 4, divides the US Barclays Bloomberg Aggregate index, a standard bond benchmark, into sectors, where each has such unique factor exposures as spread or duration. This factor model was developed by Dopfel (2004), and the factors have been chosen to cover three macro sectors plus high yield. The government sector is further broken down into three maturity buckets to help explain duration exposures.

Exhibit 4 A Simple Fixed-Income Fundamental Framework



Source: Dopfel (2004).

These components can be thought of as both macroeconomic and fundamental. They are macroeconomically oriented because spread, or expected return above similar duration government bonds, is closely related to the growth factor and is sometimes expressed as simply credit spread. Fundamentally, duration can also be thought of as a factor. This simplistic approach can be extended to encompass global fixed-income markets or adapted to a specific country's market:

$$R_i = a_i + b_{i1}F_{Gvt\_Sh} + b_{i2}F_{Gvt\_Int} + b_{i3}F_{Gvt\_Lg} + b_{i4}F_{Invest} + b_{i5}F_{HiYld} + b_{i6}F_{MBS} + \varepsilon_i, \text{where}$$

R_i = the return to bond i

a_i = the expected return to bond i

b_{ik} = the sensitivity of the return on bond i to factor k

F_k = factor k, where k represents "Gov't (Short)," "Gov't (Long)," and so on

ε_i = an error term with a zero mean that represents the portion of the return to bond i not explained by the factor model

The historic style factor weights, b_{ik}, are determined by a constrained regression (the constraint being that the total "weights" add up to 100%) of the portfolio returns against the listed style factors.

This framework lends itself readily to performance and risk attribution, along with portfolio construction. When evaluating a fixed-income manager, such characteristics as spread, duration, yield, and quality can be incorporated. This type of framework can also be extended to ESG (environmental, social, and governance) considerations as these should be generally unrelated to the basic duration and spread foundation presented. For instance, each box in Exhibit 4 could also contain E, S, and G scores, which after the initial disaggregation of a fixed-income return stream into duration and spread components could be used to model the overall portfolio's aggregate scores. For forward-looking portfolio construction purposes, a desired loading on duration, spread, and ESG scores could be handled with a quantitative objective function.

3.3.3 *Risk and Style Multifactor Models*

Another category of multifactor approach incorporates risk, or style, factors, several of which can thematically apply across asset classes. Examples of such factors include momentum, value, carry, and volatility. Many of these are similar in construction to those commonly used in equity portfolios. Examples include defining value as real (inflation-adjusted) yield, momentum as the previous 12-month excess return, and carry as the term spread. An illustrative example of risk factor approaches, in this case across asset classes, can be found in Exhibit 5.

Exhibit 5 An Illustration of Factor Approaches across Asset Classes

	Factor/Asset Class	Equity	Credit	Treasury	Commodities	Currency
Macro	Economic Growth	xx	x			
	Rates		x	xx		
	Inflation			x	xx	x

(continued)

Exhibit 5 (Continued)

	Factor/Asset Class	Equity	Credit	Treasury	Commodities	Currency
Style	Value	xx	x		x	x
	Size	xx				
	Momentum	xx	xx	xx	xx	xx
	Carry	x	xx	xx	xx	xx
	Low-Volume	xx	x			

Note: Double check marks denote strong alignment between risk factor and asset class; single check marks denote moderate alignment.
Source: Podkaminer (2017).

Of the three types of multifactor models (macroeconomic, fundamental, and statistical), statistical models can be most easily applied to various asset classes, including fixed income, as no asset-class-specific tuning is required given the minimal required assumption set. This is in contrast to macroeconomic and fundamental models, which both require adjustments and repurposing to ensure the frameworks are fit for the specifics of bond investing. Example 3 shows how expected return could be expressed.

EXAMPLE 3

Calculating Factor-Based Expected Returns at the Portfolio Level

A fixed-income portfolio has the following estimated exposures: 35% intermediate government bonds, 40% investment-grade credit, 5% securitized, and 20% high yield. The expected component returns are

Short government bonds: 0.25%

Intermediate government bonds: 1.50%

Long government bonds: 3.00%

Investment-grade credit: 4.25%

MBS/Securitized: 1.75%

High yield: 5.75%

Express the expected return of the portfolio.

Solution

Expected return could be expressed as

$$E(R) = 3.46\% = (0.35)(1.50\%) + (0.40)(4.25\%) + (0.05)(1.75\%) + (0.20)(5.75\%).$$

EXAMPLE 4

Reconciling Bond Portfolio Characteristics Using Style Factors

Talia Ayalon is evaluating intermediate duration (between 5 and 7 years) investment-grade fixed-income strategies using the framework presented in Exhibit 4. One of the strategies has the following sector attribution (totaling to 100%):

Gov't (Short) 2%	Gov't (Intermediate) 4%	Gov't (Long) 14%
Investment-Grade Credit 56%		
MBS/Securitized 6%		
High Yield 18%		

Are these sector exposures consistent with an intermediate duration investment-grade approach? Why or why not?

Suggested answer:

No, the sector exposures are inconsistent with the stated approach for two reasons: 1) The 18% exposure to high yield constitutes a significant amount of below investment-grade exposure. A true investment-grade portfolio would, for example, not have exposure to high yield. 2) The loading to longer duration sectors implies a longer-than-intermediate duration for the portfolio.

4 MACROECONOMIC FACTOR MODELS

c calculate the expected return on an asset given an asset's factor sensitivities and the factor risk premiums

d describe and compare macroeconomic factor models, fundamental factor models, and statistical factor models

The representation of returns in macroeconomic factor models assumes that the returns to each asset are correlated with only the surprises in some factors related to the aggregate economy, such as inflation or real output. We can define *surprise* in general as the actual value minus predicted (or expected) value. A factor's surprise is the component of the factor's return that was unexpected, and the factor surprises constitute the model's independent variables. This idea contrasts with the representation of independent variables as returns in Equation 2, reflecting the fact that how the independent variables are represented varies across different types of models.

Suppose that K macro factors explain asset returns. Then in a macroeconomic factor model, Equation 5 expresses the return of asset i:

$$R_i = a_i + b_{i1}F_1 + b_{i2}F_2 + \ldots + b_{iK}F_K + \varepsilon_i, \tag{5}$$

where

R_i = the return to asset i

a_i = the expected return to asset i

b_{ik} = the sensitivity of the return on asset i to a surprise in factor k, k = 1, 2, . . ., K

F_k = the surprise in the factor k, k = 1, 2, . . ., K

ε_i = an error term with a zero mean that represents the portion of the return to asset i not explained by the factor model

Surprise in a macroeconomic factor can be illustrated as follows: Suppose we are analyzing monthly returns for stocks. At the beginning of each month, we have a prediction of inflation for the month. The prediction may come from an econometric model or a professional economic forecaster, for example. Suppose our forecast at the beginning of the month is that inflation will be 0.4% during the month. At the end of the month, we find that inflation was actually 0.5% during the month. During any month,

Actual inflation = Predicted inflation + Surprise inflation.

In this case, actual inflation was 0.5% and predicted inflation was 0.4%. Therefore, the surprise in inflation was 0.5% − 0.4% = 0.1%.

What is the effect of defining the factors in terms of surprises? Suppose we believe that inflation and gross domestic product (GDP) growth are two factors that carry risk premiums; that is, inflation and GDP represent priced risk. (GDP is a money measure of the goods and services produced within a country's borders.) We do not use the predicted values of these variables because the predicted values should already be reflected in stock prices and thus in their expected returns. The intercept a_i, the expected return to asset i, reflects the effect of the predicted values of the macroeconomic variables on expected stock returns. The surprise in the macroeconomic variables during the month, however, contains new information about the variable. As a result, this model structure analyzes the return to an asset in three components: the asset's expected return, its unexpected return resulting from new information about the factors, and an error term.

Consider a factor model in which the returns to each asset are correlated with two factors. For example, we might assume that the returns for a particular stock are correlated with surprises in inflation rates and surprises in GDP growth. For stock i, the return to the stock can be modeled as

$$R_i = a_i + b_{i1}F_{INFL} + b_{i2}F_{GDP} + \varepsilon_i,$$

where

R_i = the return to stock i

a_i = the expected return to stock i

b_{i1} = the sensitivity of the return on stock i to inflation rate surprises

F_{INFL} = the surprise in inflation rates

b_{i2} = the sensitivity of the return on stock i to GDP growth surprises

F_{GDP} = the surprise in GDP growth (assumed to be uncorrelated with F_{INFL})

ε_i = an error term with a zero mean that represents the portion of the return to asset i not explained by the factor model

Consider first how to interpret b_{i1}. The factor model predicts that a 1 percentage point surprise in inflation rates will contribute b_{i1} percentage points to the return to stock i. The slope coefficient b_{i2} has a similar interpretation relative to the GDP growth factor. Thus, slope coefficients are naturally interpreted as the factor sensitivities of the asset. A *factor sensitivity* is a measure of the response of return to each unit of increase in a factor, holding all other factors constant. (Factor sensitivities are sometimes called *factor betas* or *factor loadings*.)

Now consider how to interpret the intercept a_i. Recall that the error term has a mean or average value of zero. If the surprises in both inflation rates and GDP growth are zero, the factor model predicts that the return to asset i will be a_i. Thus, a_i is the expected value of the return to stock i.

Finally, consider the error term, ε_i. The intercept a_i represents the asset's expected return. The term $(b_{i1}F_{INFL} + b_{i2}F_{GDP})$ represents the return resulting from factor surprises, and we have interpreted these as the sources of risk shared with other assets. The term ε_i is the part of return that is unexplained by expected return or the factor surprises. If we have adequately represented the sources of common risk (the factors), then ε_i must represent an asset-specific risk. For a stock, it might represent the return from an unanticipated company-specific event.

The risk premium for the GDP growth factor is typically positive. The risk premium for the inflation factor, however, is typically negative. Thus, an asset with a positive sensitivity to the inflation factor—an asset with returns that tend to be positive in response to unexpectedly high inflation—would have a lower required return than if its inflation sensitivity were negative; an asset with positive sensitivity to inflation would be in demand for its inflation-hedging ability.

This discussion has broader applications. It can be used for various asset classes, including fixed income and commodities. It can also be used in asset allocation, where asset classes can be examined in relation to inflation and GDP growth, as illustrated in the following exhibit. In Exhibit 6, each quadrant reflects a unique mix of inflation and economic growth expectations. Certain asset classes or securities can be expected to perform differently in various inflation and GDP growth regimes and can be plotted in the appropriate quadrant, thus forming a concrete illustration of a two-factor model.

Exhibit 6 Growth and Inflation Factor Matrix

	Inflation	
Growth	**Low Inflation/Low Growth** Cash Government bonds	**High Inflation/Low Growth** ■ Inflation-linked bonds ■ Commodities ■ Infrastructure
	Low Inflation/High Growth ■ Equity ■ Corporate debt	**High Inflation/High Growth** ■ Real assets (real estate, timberland, farmland, energy)

Note: Entries are assets likely to benefit from the specified combination of growth and inflation.

In macroeconomic factor models, the time series of factor surprises are constructed first. Regression analysis is then used to estimate assets' sensitivities to the factors. In practice, estimated sensitivities and intercepts are often acquired from one of the many consulting companies that specialize in factor models. When we have the parameters for the individual assets in a portfolio, we can calculate the portfolio's parameters as a weighted average of the parameters of individual assets. An individual asset's weight in that calculation is the proportion of the total market value of the portfolio that the individual asset represents.

EXAMPLE 5

Estimating Returns for a Two-Stock Portfolio Given Factor Sensitivities

Suppose that stock returns are affected by two common factors: surprises in inflation and surprises in GDP growth. A portfolio manager is analyzing the returns on a portfolio of two stocks, Manumatic (MANM) and Nextech (NXT). The following equations describe the returns for those stocks, where the factors F_{INFL} and F_{GDP} represent the surprise in inflation and GDP growth, respectively:

$$R_{MANM} = 0.09 - 1F_{INFL} + 1F_{GDP} + \varepsilon_{MANM}.$$

$$R_{NXT} = 0.12 + 2F_{INFL} + 4F_{GDP} + \varepsilon_{NXT}.$$

One-third of the portfolio is invested in Manumatic stock, and two-thirds is invested in Nextech stock.

1 Formulate an expression for the return on the portfolio.

2 State the expected return on the portfolio.

3 Calculate the return on the portfolio given that the surprises in inflation and GDP growth are 1% and 0%, respectively, assuming that the error terms for MANM and NXT both equal 0.5%.

In evaluating the equations for surprises in inflation and GDP, convert amounts stated in percentage terms to decimal form.

Solution to 1:

The portfolio's return is the following weighted average of the returns to the two stocks:

$$\begin{aligned} R_P &= (1/3)(0.09) + (2/3)(0.12) + [(1/3)(-1) + (2/3)(2)]F_{INFL} + [(1/3)(1) + \\ &\quad (2/3)(4)]F_{GDP} + (1/3)\varepsilon_{MANM} + (2/3)\varepsilon_{NXT} \\ &= 0.11 + 1F_{INFL} + 3F_{GDP} + (1/3)\varepsilon_{MANM} + (2/3)\varepsilon_{NXT}. \end{aligned}$$

Solution to 2:

The expected return on the portfolio is 11%, the value of the intercept in the expression obtained in the solution to 1.

Solution to 3:

$$\begin{aligned} R_P &= 0.11 + 1F_{INFL} + 3F_{GDP} + (1/3)\varepsilon_{MANM} + (2/3)\varepsilon_{NXT} \\ &= 0.11 + 1(0.01) + 3(0) + (1/3)(0.005) + (2/3)(0.005) \\ &= 0.125\text{, or } 12.5\%. \end{aligned}$$

5 FUNDAMENTAL FACTOR MODELS

c calculate the expected return on an asset given an asset's factor sensitivities and the factor risk premiums

d describe and compare macroeconomic factor models, fundamental factor models, and statistical factor models

EXAMPLE 6

Comparing Types of Factor Models

Connor (1995) contrasted a macroeconomic factor model with a fundamental factor model to compare how well the models explain stock returns.

Connor reported the results of applying a macroeconomic factor model to the returns for 779 large-cap US stocks based on monthly data from January 1985 through December 1993. Using five macroeconomic factors, Connor was able to explain approximately 11% of the variance of return on these stocks. Exhibit 7 shows his results.

Exhibit 7 The Explanatory Power of the Macroeconomic Factors

Factor	Explanatory Power from Using Each Factor Alone	Increase in Explanatory Power from Adding Each Factor to All the Others
Inflation	1.3%	0.0%
Term structure	1.1%	7.7%
Industrial production	0.5%	0.3%
Default premium	2.4%	8.1%
Unemployment	−0.3%	0.1%
All factors (total explanatory power)		10.9%

Notes: The explanatory power of a given model was computed as 1 − [(Average asset −Specific variance of return across stocks)/(Average total variance of return across stocks)]. The variance estimates were corrected for degrees of freedom, so the marginal contribution of a factor to explanatory power can be zero or negative. Explanatory power captures the proportion of the total variance of return that a given model explains for the average stock.
Source: Connor (1995).

Connor also reported a fundamental factor analysis of the same companies. The factor model employed was the BARRA US-E2 model (as of 2019, the current version is E4). Exhibit 8 shows these results. In the exhibit, "variability in markets" represents the stock's volatility, "success" is a price momentum variable, "trade activity" distinguishes stocks by how often their shares trade, and "growth" distinguishes stocks by past and anticipated earnings growth (explanations of variables are from Grinold and Kahn 1994).

Exhibit 8 The Explanatory Power of the Fundamental Factors

Factor	Explanatory Power from Using Each Factor Alone	Increase in Explanatory Power from Adding Each Factor to All the Others
Industries	16.3%	18.0%
Variability in markets	4.3%	0.9%
Success	2.8%	0.8%

(continued)

Exhibit 8 (Continued)

Factor	Explanatory Power from Using Each Factor Alone	Increase in Explanatory Power from Adding Each Factor to All the Others
Size	1.4%	0.6%
Trade activity	1.4%	0.5%
Growth	3.0%	0.4%
Earnings to price	2.2%	0.6%
Book to price	1.5%	0.6%
Earnings variability	2.5%	0.4%
Financial leverage	0.9%	0.5%
Foreign investment	0.7%	0.4%
Labor intensity	2.2%	0.5%
Dividend yield	2.9%	0.4%
All factors (total explanatory power)		42.6%

Source: Connor (1995).

As Exhibit 8 shows, the most important fundamental factor is "industries," represented by 55 industry dummy variables. The fundamental factor model explained approximately 43% of the variation in stock returns, compared with approximately 11% for the macroeconomic factor model. Because "industries" must sum to the market and the market portfolio is not incorporated in the macroeconomic factor model, some advantage to the explanatory power of the fundamental factor may be built into the specific models being compared. Connor's article also does not provide tests of the statistical significance of the various factors in either model; however, Connor's research is strong evidence for the usefulness of fundamental factor models. Moreover, this evidence is mirrored by the wide use of those models in the investment community. For example, fundamental factor models are frequently used in portfolio performance attribution. Typically, fundamental factor models employ many more factors than macroeconomic factor models, giving a more detailed picture of the sources of an investment manager's returns.

We cannot conclude from this study, however, that fundamental factor models are inherently superior to macroeconomic factor models. Each major type of model has its uses. The factors in various macroeconomic factor models are individually backed by statistical evidence that they represent systematic risk (i.e., risk that cannot be diversified away). The same may not be true of each factor in a fundamental factor model. For example, a portfolio manager can easily construct a portfolio that excludes a particular industry, so exposure to a particular industry is not systematic risk.

The two types of factors, macroeconomic and fundamental, have different implications for measuring and managing risk, in general. The macroeconomic factor set is parsimonious (five variables in the model studied) and allows a portfolio manager to incorporate economic views into portfolio construction by adjustments to portfolio exposures to macro factors. The fundamental factor set examined by Connor is large (67 variables, including the 55 industry dummy variables); at the expense of greater complexity, it can give a more detailed picture of risk in terms that are easily related to company and security characteristics. Connor found that the macroeconomic factor model had no marginal explanatory power when added to the fundamental factor model, implying that the fundamental risk attributes capture all the risk characteristics represented by the macroeconomic factor betas. Because the fundamental factors supply such a detailed description of the characteristics of a stock and its issuer, however, this finding is not necessarily surprising.

FACTOR MODELS IN RETURN ATTRIBUTION

6

- **c** calculate the expected return on an asset given an asset's factor sensitivities and the factor risk premiums
- **f** describe uses of multifactor models and interpret the output of analyses based on multifactor models
- **g** describe the potential benefits for investors in considering multiple risk dimensions when modeling asset returns

The following sections present selected applications of multifactor models in investment practice. The applications discussed are return attribution, risk attribution, portfolio construction, and strategic portfolio decisions. We begin by discussing portfolio return attribution and risk attribution, focusing on the analysis of benchmark-relative returns. After discussing performance attribution and risk analysis, we explain the use of multifactor models in creating a portfolio with a desired set of risk exposures.

Additionally, multifactor models can be used for asset allocation purposes. Some large, sophisticated asset owners have chosen to define their asset allocation opportunity sets in terms of macroeconomic or thematic factors and aggregate factor exposures (represented by pure factor portfolios as defined earlier). Many others are examining their traditionally derived asset allocation policies using factor models to map asset class exposure to factor sensitivities. The trend toward factor-based asset allocation has two chief causes: First is the increasing availability of sophisticated factor models (like the BARRA models used in the following examples); second is the more intense focus by asset owners on the many dimensions of risk.

6.1 Factor Models in Return Attribution

Multifactor models can help us understand in detail the sources of a manager's returns relative to a benchmark. For simplicity, in this section we analyze the sources of the returns of a portfolio fully invested in the equities of a single national equity market, which allows us to ignore the roles of country selection, asset allocation, market timing, and currency hedging. The same methodology can, however, be applied across asset classes and geographies.

Analysts often favor fundamental multifactor models in decomposing (separating into basic elements) the sources of returns. In contrast to statistical factor models, fundamental factor models allow the sources of portfolio performance to be described using commonly understood terms. Fundamental factors are also thematically understandable and can be incorporated into simple narratives for clients concerning return or risk attribution.

Also, in contrast to macroeconomic factor models, fundamental models express investment style choices and security characteristics more directly and often in greater detail.

We first need to understand the objectives of active managers. As mentioned previously, managers are commonly evaluated relative to a specified benchmark. Active portfolio managers hold securities in different-from-benchmark weights in an attempt to add value to their portfolios relative to a passive investment approach. Securities held in different-from-benchmark weights reflect portfolio manager expectations that differ from consensus expectations. For an equity manager, those expectations may relate to common factors driving equity returns or to considerations unique to a company. Thus, when we evaluate an active manager, we want to ask such questions as, Did the manager have insights that were effectively translated into returns in excess of those that were available from a passive alternative? Analyzing the sources of returns using multifactor models can help answer these questions.

The return on a portfolio, R_p, can be viewed as the sum of the benchmark's return, R_B, and the **active return** (portfolio return minus benchmark return):

$$\text{Active return} = R_p - R_B. \tag{6}$$

With the help of a factor model, we can analyze a portfolio manager's active return as the sum of two components. The first component is the product of the portfolio manager's factor tilts (over- or underweights relative to the benchmark factor sensitivities) and the factor returns; we call this component the return from factor tilts. The second component of active return reflects the manager's skill in individual asset selection (ability to overweight securities that outperform the benchmark or underweight securities that underperform the benchmark); we call this component security selection. Equation 7 shows the decomposition of active return into those two components, where k represents the factor or factors represented in the benchmark portfolio:

$$\begin{aligned}\text{Active return} = &\sum_{k=1}^{K}\left[\left(\text{Portfolio sensitivity}\right)_k - \left(\text{Benchmark sensitivity}\right)_k\right] \\ &\times \left(\text{Factor return}\right)_k + \text{Security selection}\end{aligned} \tag{7}$$

In Equation 7, the portfolio's and benchmark's sensitivities to each factor are calculated as of the beginning of the evaluation period.

EXAMPLE 7

Four-Factor Model Active Return Decomposition

As an equity analyst at a pension fund sponsor, Ronald Service uses the Carhart four-factor multifactor model of Equation 3a to evaluate US equity portfolios:

$$R_p - R_F = a_p + b_{p1}\text{RMRF} + b_{p2}\text{SMB} + b_{p3}\text{HML} + b_{p4}\text{WML} + \varepsilon_p.$$

Service's current task is to evaluate the performance of the most recently hired US equity manager. That manager's benchmark is an index representing the performance of the 1,000 largest US stocks by market value. The manager describes himself as a "stock picker" and points to his performance in beating the benchmark as evidence that he is successful. Exhibit 9 presents an analysis based on the Carhart model of the sources of that manager's active return during the year, given an assumed set of factor returns. In Exhibit 9, the entry "A. Return from Factor Tilts = 2.1241%" is the sum of the four numbers above it. The entry "B. Security Selection" gives security selection as equal to –0.05%. "C. Active Return" is found as the sum of these two components: 2.1241% + (–0.05%) = 2.0741%.

Exhibit 9 Active Return Decomposition

	Factor Sensitivity				Contribution to Active Return	
Factor	Portfolio (1)	Benchmark (2)	Difference (3) = (1) − (2)	Factor Return (4)	Absolute (3) × (4)	Proportion of Total Active
RMRF	0.95	1.00	−0.05	5.52%	−0.2760%	−13.3%
SMB	−1.05	−1.00	−0.05	−3.35%	0.1675%	8.1%
HML	0.40	0.00	0.40	5.10%	2.0400%	98.4%
WML	0.05	0.03	0.02	9.63%	0.1926%	9.3%
				A. Return from Factor Tilts =	2.1241%	102.4%
				B. Security Selection =	−0.0500%	−2.4%
				C. Active Return (A + B) =	2.0741%	100.0%

From his previous work, Service knows that the returns to growth-style portfolios often have a positive sensitivity to the momentum factor (WML). By contrast, the returns to certain value-style portfolios, in particular those following a contrarian strategy, often have a negative sensitivity to the momentum factor. Using the information given, address the following questions (assume the benchmark chosen for the manager is appropriate):

1 Determine the manager's investment mandate and his actual investment style.

2 Evaluate the sources of the manager's active return for the year.

3 What concerns might Service discuss with the manager as a result of the return decomposition?

Solution to 1:

The benchmarks chosen for the manager should reflect the baseline risk characteristics of the manager's investment opportunity set and his mandate. We can ascertain whether the manager's actual style follows the mandate by examining the portfolio's actual factor exposures:

- The sensitivities of the benchmark are consistent with the description in the text. The sensitivity to RMRF of 1 indicates that the assigned benchmark has average market risk, consistent with it being a broad-based index; the negative sensitivity to SMB indicates a large-cap orientation. The mandate might be described as large-cap without a value/growth bias (HML is zero) or a momentum bias (WML is close to zero).
- Stocks with high book-to-market ratios are generally viewed as value stocks. Because the equity manager has a positive sensitivity to HML (0.40), it appears that the manager has a value orientation. The manager is approximately neutral to the momentum factor, so the equity manager is not a momentum investor and probably not a contrarian value investor. In summary, these considerations suggest that the manager has a large-cap value orientation.

Solution to 2:

The dominant source of the manager's positive active return was his positive active exposure to the HML factor. The bet contributed approximately 98% of the realized active return of about 2.07%. The manager's active exposure to the overall market (RMRF) was unprofitable, but his active exposures to small stocks (SMB) and to momentum (WML) were profitable. The magnitudes of the manager's active exposures to RMRF, SMB, and WML were relatively small, however, so the effects of those bets on active return were minor compared with his large and successful bet on HML.

Solution to 3:

Although the manager is a self-described "stock picker," his active return from security selection in this period was actually negative. His positive active return resulted from the concurrence of a large active bet on HML and a high return to that factor during the period. If the market had favored growth rather than value without the manager doing better in individual security selection, the manager's performance would have been unsatisfactory. Service's conversations with the manager should focus on evidence that he can predict changes in returns to the HML factor and on the manager's stock selection discipline.

7 FACTOR MODELS IN RISK ATTRIBUTION

e explain sources of active risk and interpret tracking risk and the information ratio

f describe uses of multifactor models and interpret the output of analyses based on multifactor models

g describe the potential benefits for investors in considering multiple risk dimensions when modeling asset returns

Building on the discussion of active returns, this section explores the analysis of active risk. A few key terms are important to the understanding of how factor models are used to build an understanding of a portfolio manager's risk exposures. We will describe them briefly before moving on to the detailed discussion of risk attribution.

Active risk can be represented by the standard deviation of active returns. A traditional term for that standard deviation is **tracking error** (TE). **Tracking risk** is a synonym for tracking error that is often used in the CFA Program curriculum. We will use the abbreviation TE for the concept of active risk and refer to it usually as tracking error:

$$\text{TE} = s(R_p - R_B). \quad (8)$$

In Equation 8, $s(R_p - R_B)$ indicates that we take the sample standard deviation (indicated by s) of the time series of differences between the portfolio return, R_p, and the benchmark return, R_B. We should be careful that active return and tracking error are stated on the same time basis. As an approximation assuming returns are serially uncorrelated, to annualize a daily TE based on daily returns, we multiply daily TE by $(250)^{1/2}$ based on 250 trading days in a year. To annualize a monthly TE based on monthly returns, we multiply monthly TE by $(12)^{1/2}$.

As a broad indication of the range for tracking error, in US equity markets a well-executed passive investment strategy can often achieve a tracking error on the order of 0.10% or less per year. A low-risk active or enhanced index investment strategy, which makes tightly controlled use of managers' expectations, often has a tracking error goal of 2% per year. A diversified active large-cap equity strategy that might be benchmarked to the S&P 500 Index would commonly have a tracking error in the range of 2%–6% per year. An aggressive active equity manager might have a tracking error in the range of 6%–10% or more.

Somewhat analogous to the use of the traditional Sharpe measure in evaluating absolute returns, the **information ratio** (IR) is a tool for evaluating mean active returns per unit of active risk. The historical or *ex post* IR is expressed as follows:

$$\mathrm{IR} = \frac{\bar{R}_p - \bar{R}_B}{s\left(R_p - R_B\right)} \tag{9}$$

In the numerator of Equation 9, $\bar{R}_p$ and $\bar{R}_B$ stand for the sample mean return on the portfolio and the sample mean return on the benchmark, respectively. The equation assumes that the portfolio being evaluated has the same systematic risk as its benchmark. To illustrate the calculation, if a portfolio achieved a mean return of 9% during the same period that its benchmark earned a mean return of 7.5% and the portfolio's tracking error (the denominator) was 6%, we would calculate an information ratio of (9% − 7.5%)/6% = 0.25. Setting guidelines for acceptable active risk or tracking error is one of the methods that some investors use to ensure that the overall risk and style characteristics of their investments are in line with their chosen benchmark.

Note that in addition to focusing exclusively on *active* risk, multifactor models can also be used to decompose and attribute sources of *total* risk. For instance, a multi-asset class multi-strategy long/short fund can be evaluated with an appropriate multifactor model to reveal insights on sources of total risk.

EXAMPLE 8

Creating Active Manager Guidelines

The framework of active return and active risk is appealing to investors who want to manage the risk of investments. The benchmark serves as a known and continuously observable reference standard in relation to which quantitative risk and return objectives may be stated and communicated. For example, a US public employee retirement system invited investment managers to submit proposals to manage a "low-active-risk US large-cap equity fund" that would be subject to the following constraints:

- Shares must be components of the S&P 500.
- The portfolio should have a minimum of 200 issues. At time of purchase, the maximum amount that may be invested in any one issuer is 5% of the portfolio at market value or 150% of the issuers' weight within the S&P 500, whichever is greater.
- The portfolio must have a minimum information ratio of 0.30 either since inception or over the last seven years.
- The portfolio must also have tracking risk of less than 3% with respect to the S&P 500 either since inception or over the last seven years.

Once a suitable active manager is found and hired, these requirements can be written into the manager's guidelines. The retirement system's individual mandates would be set such that the sum of mandates across managers would equal the desired risk exposures.

Analysts use multifactor models to understand a portfolio manager's risk exposures in detail. By decomposing active risk, the analyst's objective is to measure the portfolio's active exposure along each dimension of risk—in other words, to understand the sources of tracking error. This can even be done at the level of individual holdings. Among the questions analysts will want to answer are the following:

- What active exposures contributed most to the manager's tracking error?
- Was the portfolio manager aware of the nature of his active exposures, and if so, can he articulate a rationale for assuming them?
- Are the portfolio's active risk exposures consistent with the manager's stated investment philosophy?
- Which active bets earned adequate returns for the level of active risk taken?

In addressing these questions, analysts often choose fundamental factor models because they can be used to relate active risk exposures to a manager's portfolio decisions in a fairly direct and intuitive way. In this section, we explain how to decompose or explain a portfolio's active risk using a multifactor model.

We previously addressed the decomposition of active return; now we address the decomposition of active risk. In analyzing risk, it is more convenient to use variances rather than standard deviations because the variances of uncorrelated variables are additive. We refer to the variance of active return as **active risk squared**:

$$\text{Active risk squared} = s^2(R_p - R_B). \tag{10}$$

We can separate a portfolio's active risk squared into two components:

- **Active factor risk** is the contribution to active risk squared resulting from the portfolio's different-from-benchmark exposures relative to factors specified in the risk model.
- **Active specific risk** or **security selection risk** measures the active non-factor or residual risk assumed by the manager. Portfolio managers attempt to provide a positive average return from security selection as compensation for assuming active specific risk.

As we use the terms, "active specific risk" and "active factor risk" refer to variances rather than standard deviations. When applied to an investment in a single asset class, active risk squared has two components:

$$\text{Active risk squared} = \text{Active factor risk} + \text{Active specific risk}. \tag{11}$$

Active factor risk represents the part of active risk squared explained by the portfolio's active factor exposures. Active factor risk can be found indirectly as the risk remaining after active specific risk is deducted from active risk squared. Active specific risk can be expressed as

$$\text{Active specific risk} = \sum_{i=1}^{n}\left(w_i^a\right)^2 \sigma_{\varepsilon_i}^2$$

where w_i^a is the ith asset's active weight in the portfolio (that is, the difference between the asset's weight in the portfolio and its weight in the benchmark) and $\sigma_{\varepsilon_i}^2$ is the residual risk of the ith asset (the variance of the ith asset's returns left unexplained by the factors).

The direct procedure for calculating active factor risk is as follows. A portfolio's active factor exposure to a given factor *j*, b_j^a, is found by weighting each asset's sensitivity to factor *j* by its active weight and summing the terms: $b_j^a = \sum_{i=1}^{n} w_i^a b_{ji}$. Then active factor risk equals $\sum_{i=1}^{K}\sum_{j=1}^{K} b_i^a b_j^a \operatorname{cov}\left(F_i, F_j\right)$.

EXAMPLE 9

A Comparison of Active Risk

Richard Gray is comparing the risk of four US equity managers who share the same benchmark. He uses a fundamental factor model, the BARRA US-E4 model, which incorporates 12 style factors and a set of 60 industry factors. The style factors measure various fundamental aspects of companies and their shares, such as size, liquidity, leverage, and dividend yield. In the model, companies have non-zero exposures to all industries in which the company operates. Exhibit 10 presents Gray's analysis of the active risk squared of the four managers, based on Equation 11 (note that there is a covariance term in active factor risk, reflecting the correlation of industry membership and the risk indexes, which we assume is negligible in this example). In Exhibit 10, the column labeled "Industry" gives the portfolio's active factor risk associated with the industry exposures of its holdings; the "Style Factor" column gives the portfolio's active factor risk associated with the exposures of its holdings to the 12 style factors.

Exhibit 10 Active Risk Squared Decomposition

	Active Factor				
Portfolio	Industry	Style Factor	Total Factor	Active Specific	Active Risk Squared
A	12.25	17.15	29.40	19.60	49
B	1.25	13.75	15.00	10.00	25
C	1.25	17.50	18.75	6.25	25
D	0.03	0.47	0.50	0.50	1

Note: Entries are in % squared.

Using the information in Exhibit 10, address the following:

1. Contrast the active risk decomposition of Portfolios A and B.
2. Contrast the active risk decomposition of Portfolios B and C.
3. Characterize the investment approach of Portfolio D.

Solution to 1:

Exhibit 11 restates the information in Exhibit 10 to show the proportional contributions of the various sources of active risk. (e.g., Portfolio A's active risk related to industry exposures is 25% of active risk squared, calculated as 12.25/49 = 0.25, or 25%).

The last column of Exhibit 11 now shows the square root of active risk squared—that is, active risk or tracking error.

Exhibit 11 Active Risk Decomposition (restated)

	Active Factor (% of total active)				
Portfolio	**Industry**	**Style Factor**	**Total Factor**	**Active Specific (% of total active)**	**Active Risk**
A	25%	35%	60%	40%	7%
B	5%	55%	60%	40%	5%
C	5%	70%	75%	25%	5%
D	3%	47%	50%	50%	1%

Portfolio A has assumed a higher level of active risk than B (7% versus 5%). Portfolios A and B assumed the same proportions of active factor and active specific risk, but a sharp contrast exists between the two in the types of active factor risk exposure. Portfolio A assumed substantial active industry risk, whereas Portfolio B was approximately industry neutral relative to the benchmark. By contrast, Portfolio B had higher active bets on the style factors representing company and share characteristics.

Solution to 2:

Portfolios B and C were similar in their absolute amounts of active risk. Furthermore, both Portfolios B and C were both approximately industry neutral relative to the benchmark. Portfolio C assumed more active factor risk related to the style factors, but B assumed more active specific risk. It is also possible to infer from the greater level of B's active specific risk that B is somewhat less diversified than C.

Solution to 3:

Portfolio D appears to be a passively managed portfolio, judging by its negligible level of active risk. Referring to Exhibit 11, Portfolio D's active factor risk of 0.50, equal to 0.707% expressed as a standard deviation, indicates that the portfolio's risk exposures very closely match the benchmark.

The discussion of performance attribution and risk analysis has used examples related to common stock portfolios. Multifactor models have also been effectively used in similar roles for portfolios of bonds and other asset classes. For example, such factors as duration and spread can be used to decompose the risk and return of a fixed-income manager.

FACTOR MODELS IN PORTFOLIO CONSTRUCTION

8

f describe uses of multifactor models and interpret the output of analyses based on multifactor models

g describe the potential benefits for investors in considering multiple risk dimensions when modeling asset returns

Equally as important to the use of multifactor models in analyzing a portfolio's active returns and active risk is the use of such multifactor models in portfolio construction. At this stage of the portfolio management process, multifactor models permit the portfolio manager to make focused bets or to control portfolio risk relative to the benchmark's risk. This greater level of detail in modeling risk that multifactor models afford is useful in both passive and active management.

- *Passive management*. In managing a fund that seeks to track an index with many component securities, portfolio managers may need to select a sample of securities from the index. Analysts can use multifactor models to replicate an index fund's factor exposures, mirroring those of the index tracked.
- *Active management*. Many quantitative investment managers rely on multifactor models in predicting alpha (excess risk-adjusted returns) or relative return (the return on one asset or asset class relative to that of another) as part of a variety of active investment strategies. In constructing portfolios, analysts use multifactor models to establish desired risk profiles.
- *Rules-based active management (alternative indexes)*. These strategies routinely tilt toward such factors as size, value, quality, or momentum when constructing portfolios. As such, alternative index approaches aim to capture some systematic exposure traditionally attributed to manager skill, or "alpha," in a transparent, mechanical, rules-based manner at low cost. Alternative index strategies rely heavily on factor models to introduce intentional factor and style biases versus capitalization-weighted indexes.

In the following, we explore some of these uses in more detail. As indicated, an important use of multifactor models is to establish a specific desired risk profile for a portfolio. In the simplest instance, the portfolio manager may want to create a portfolio with sensitivity to a single factor. This particular (pure) factor portfolio would have a sensitivity of 1 for that factor and a sensitivity (or weight) of 0 for all other factors. It is thus a portfolio with exposure to only one risk factor and exactly represents the risk of that factor. As a pure bet on a source of risk, factor portfolios are of interest to a portfolio manager who wants to hedge that risk (offset it) or speculate on it. This simple case can be expanded to multiple factors where a factor replication portfolio can be built based either on an existing target portfolio or on a set of desired exposures. Example 10 illustrates the use of factor portfolios.

EXAMPLE 10

Factor Portfolios

Analyst Wanda Smithfield has constructed six portfolios for possible use by portfolio managers in her firm. The portfolios are labeled A, B, C, D, E, and F in Exhibit 12. Smithfield adapts a macroeconomic factor model based on research presented in Burmeister, Roll, and Ross (1994). The model includes five factors:

- Confidence risk, based on the yield spread between corporate bonds and government bonds. A positive surprise in the spread suggests that investors are willing to accept a smaller reward for bearing default risk and so that confidence is high.
- Time horizon risk, based on the yield spread between 20-year government bonds and 30-day Treasury bills. A positive surprise indicates increased investor willingness to invest for the long term.
- Inflation risk, measured by the unanticipated change in the inflation rate.
- Business cycle risk, measured by the unexpected change in the level of real business activity.
- Market timing risk, measured as the portion of the return on a broad-based equity index that is unexplained by the first four risk factors.

Exhibit 12 Factor Portfolios

	Portfolios					
Risk Factor	**A**	**B**	**C**	**D**	**E**	**F**
Confidence risk	0.50	0.00	1.00	0.00	0.00	0.80
Time horizon risk	1.92	0.00	1.00	1.00	1.00	1.00
Inflation risk	0.00	0.00	1.00	0.00	0.00	–1.05
Business cycle risk	1.00	1.00	0.00	0.00	1.00	0.30
Market timing risk	0.90	0.00	1.00	0.00	0.00	0.75

Note: Entries are factor sensitivities.

- A portfolio manager wants to place a bet that real business activity will increase.
  - **A** Determine and justify the portfolio among the six given that would be most useful to the manager.
  - **B** Would the manager take a long or short position in the portfolio chosen in Part A?
- A portfolio manager wants to hedge an existing positive (long) exposure to time horizon risk.
  - **A** Determine and justify the portfolio among the six given that would be most useful to the manager.
  - **B** What type of position would the manager take in the portfolio chosen in Part A?

Solution to 1A:

Portfolio B is the most appropriate choice. Portfolio B is the factor portfolio for business cycle risk because it has a sensitivity of 1 to business cycle risk and a sensitivity of 0 to all other risk factors. Portfolio B is thus efficient for placing a pure bet on an increase in real business activity.

Solution to 1B:

The manager would take a long position in Portfolio B to place a bet on an increase in real business activity.

Solution to 2A:

Portfolio D is the appropriate choice. Portfolio D is the factor portfolio for time horizon risk because it has a sensitivity of 1 to time horizon risk and a sensitivity of 0 to all other risk factors. Portfolio D is thus efficient for hedging an existing positive exposure to time horizon risk.

Solution to 2B:

The manager would take a short position in Portfolio D to hedge the positive exposure to time horizon risk.

The following case was written by Yin Luo, CPA, PStat, CFA, and Sheng Wang, both of Wolfe Research LLC (USA).

Constructing Multifactor Portfolios

In practice, most stock selection models use some common multifactor structure. Here, we describe constructing two types of multifactor portfolios—a benchmark portfolio and a risk parity portfolio—that target desired risk exposures to eight fundamental factors. The benchmark portfolio equally weights the pure factors, whereas the risk parity portfolio weights the pure factors based on equal risk contribution. We focus on the benchmark and risk parity portfolios because their factor weighting schemes are clear and objective.

Setting the Scene: Pure Factor Portfolios

For demonstration purposes, we use fundamental factor models and choose common company- and company share–related factors from each main investment style (i.e., value, growth, price momentum, analyst sentiment, and quality):

1. *Defensive value*: Trailing earnings yield—companies with high earnings yield are preferred.
2. *Cyclical value*: Book-to-market ratio—companies with high book-to-market ratios (i.e., cheap stock valuations) are bought.
3. *Growth*: Consensus FY1/FY0 EPS growth—companies with high expected earnings growth are preferred.
4. *Price momentum*: 12M total return excluding the most recent month—companies with positive price momentum are preferred.
5. *Analyst sentiment*: 3M EPS revision—companies with positive earnings revisions are bought.
6. *Profitability*: Return on equity (ROE)—companies with high ROEs are bought.
7. *Leverage*: Debt/equity ratio—companies with low financial leverage are preferred.
8. *Earnings quality*: Non-cash earnings—companies with low accruals are bought. Research suggests that net income with low levels of non-cash items (i.e., accruals) is less likely to be manipulated.

The stock universe for this demonstration consists of the Russell 3000 Index (US), the S&P/TSX Composite Index (Canada), the MSCI China A Index (China), and the S&P Global Broad Market Index (all other countries). A pure factor portfolio is formed for each of the eight factors by buying the top 20% of stocks and shorting the bottom 20% of stocks ranked by the factor. Stocks held long and short are equally weighted, and the eight factor portfolios are each rebalanced monthly. Note that this demonstration does not account for transaction costs or other portfolio constraints. Other methods for forming pure factor portfolios include ranking stocks by Pearson IC (correlation between prior period factor scores and current period stock returns) or by Spearman Rank IC (correlation between prior period ranked factor scores and current period ranked stock returns), as well as ranking by other univariate regression methods. However, for simplicity, we follow the long–short portfolio approach.

A straightforward way to combine these pure factor portfolios into a multifactor portfolio is equal weighting. We call the equally weighted multifactor portfolio the "benchmark (BM) portfolio." The experience in practice is that portfolios constructed using this simple weighting scheme typically perform at least as well as those using more sophisticated optimization techniques.

Risk parity is a common alternative portfolio construction technique used in the asset allocation space. Risk parity accounts for the volatility of each factor and the correlations of returns among all factors to be combined into the multifactor portfolio. The objective is for each factor to contribute equally to the overall (or targeted) risk of the portfolio. Thus, a risk parity (RP) multifactor portfolio can be created by equally weighting the risk contribution of each of the eight pure factors mentioned.

Constructing and Backtesting Benchmark and Risk Parity Multifactor Portfolios

To create a successful multifactor portfolio strategy, the investment manager needs to perform backtesting to assess factor performance and effectiveness. In a typical backtest, a manager first forms her investment hypothesis, determines her investment rules and processes, collects the required data, and creates the portfolio, and then she periodically rebalances and evaluates the portfolio.

In the rolling window backtesting methodology, analysts use a rolling window framework, fit factors based on the rolling window, rebalance the portfolio periodically, and then track performance. Thus, backtesting is a proxy for actual investing. As new information arrives, investment managers readjust their models and rebalance their stock positions, typically monthly. Thus, they repeat the same in-sample training/out-of-sample testing process. If the investment strategy's performance in out-of-sample periods is desirable and the strategy makes intuitive sense, then it is deemed successful.

The following exhibit illustrates rolling window backtesting of the defensive value factor from November 2011 to April 2012. On 30 November 2011, we compute each stock's trailing 12-month earnings yield, then buy the 20% of stocks with the highest earnings yield and short the bottom quintile of stocks, and assess performance using returns in the next month, December 2011, the out-of-sample (OOS) period. The process is repeated on 31 December 2011, and so on, and finally, we compute the average monthly return, volatility, Sharpe ratio, and drawdown from the test results of the six OOS periods.

An Example of Rolling Window Backtesting of the Defensive Value Factor

	2010:12	2011:01	2011:02	2011:03	2011:04	2011:05	2011:06	2011:07	2011:08	2011:09	2011:10	2011:11	2011:12	2012:01	2012:02	2012:03	2012:04	2012:05
11/30/2011									In-Sample (Last 12M EPS/Price)				OOS					
12/31/2011										In-Sample (Last 12M EPS/Price)				OOS				
1/31/2012											In-Sample (Last 12M EPS/Price)				OOS			
2/29/2012												In-Sample (Last 12M EPS/Price)				OOS		
3/31/2012													In-Sample (Last 12M EPS/Price)				OOS	
4/30/2012														In-Sample (Last 12M EPS/Price)				OOS

Source: Wolfe Research Luo's QES.

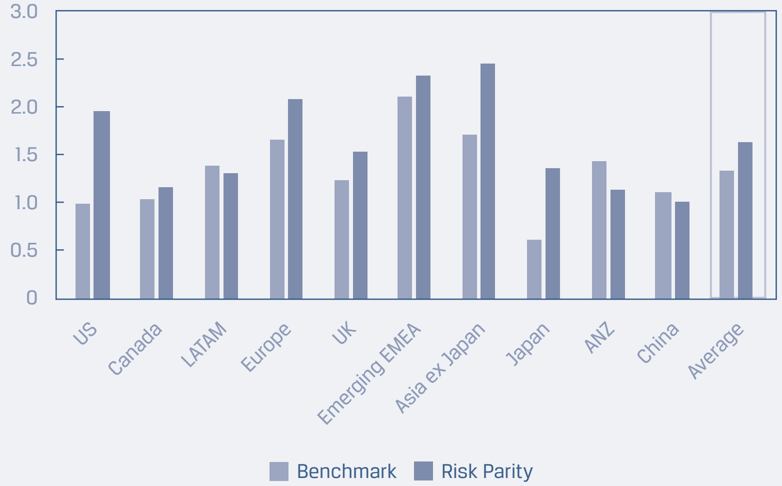
Constructing and backtesting multifactor portfolios is similar to the method just described, except that the rolling window procedure is implemented twice. First, we form the eight pure factor portfolios for each month from 1988 until May 2019 by implementing the rolling window procedure. Then, we combine the underlying factor portfolios into the multifactor portfolios using the two approaches—equally weighting all factors (i.e., benchmark, or BM, allocation) and equally risk weighting all factors (i.e., risk parity, or RP, allocation).

Importantly, the process for creating the multifactor portfolios requires a second implementation of the rolling window procedure to avoid look-ahead bias; note this second rolling window covers the same time span as the first one (i.e., 1988 until May 2019). At each month-end, the previous five years of monthly data are used to estimate the variance–covariance matrix for the eight factor portfolios. Once the covariance matrix is estimated, we optimize and compute the weights for each of the eight pure factor portfolios and then form the RP portfolio. Finally, we compute the returns of the two multifactor portfolios (BM and RP) during this out-of-sample period using the weights at the end of the previous month and the returns of the eight underlying factor portfolios for the current month. This process is repeated every month over the entire horizon of 1988 until May 2019.

We created and backtested the multifactor portfolios using both the equal weighting (BM) scheme and risk parity (RP) scheme for each of 10 markets, including the United States. Both multifactor portfolios are rebalanced monthly to maintain equal factor weights or equal factor risk contributions. As noted previously, the key input to the RP allocation is the monthly variance–covariance matrix for the eight underlying factor portfolios derived from the rolling (five-year) window procedure. To be clear, each of the eight factor portfolios is a long–short portfolio. However, our factor allocation strategies to form the BM and RP multifactor portfolios are long only, meaning the weights allocated to each of the eight factor portfolios are restricted to be non-negative. Therefore, factor weights for the BM and RP portfolios are positive and add to 100%.

In the United States over the period 1993–2019, the weights of the eight factor portfolios in the RP allocation are relatively stable. Interestingly, book-to-market and earnings quality factor portfolios receive the largest allocations, whereas ROE and price momentum factor portfolios have the lowest weights. The RP multifactor portfolio provides a lower cumulative return than does the BM multifactor portfolio; however, the RP portfolio's volatility is substantially lower than that of the BM portfolio. Consequently, in the United States, the RP portfolio's Sharpe ratio is nearly double that of the BM portfolio, as shown in the following exhibit. Outperformance of the RP portfolio in terms of Sharpe ratio is also apparent across most markets examined.

Average Sharpe Ratios for Multifactor Portfolios: Equally Weighted vs. Risk Parity Weighted (1993–2019)



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

9 FACTOR MODELS IN STRATEGIC PORTFOLIO DECISIONS

f describe uses of multifactor models and interpret the output of analyses based on multifactor models

g describe the potential benefits for investors in considering multiple risk dimensions when modeling asset returns

Multifactor models can help investors recognize considerations that are relevant in making various strategic decisions. For example, given a sound model of the systematic risk factors that affect assets' mean returns, the investor can ask, relative to other investors,

- What types of risk do I have a comparative advantage in bearing?
- What types of risk am I at a comparative disadvantage in bearing?

For example, university endowments, because they typically have very long investment horizons, may have a comparative advantage in bearing business cycle risk of traded equities or the liquidity risk associated with many private equity investments. They may tilt their strategic asset allocation or investments within an asset class to capture the associated risk premiums for risks that do not much affect them. However, such investors may be at a comparative disadvantage in bearing inflation risk to the extent that the activities they support have historically been subject to cost increases running above the average rate of inflation.

This is a richer framework than that afforded by the CAPM, according to which all investors optimally should invest in two funds: the market portfolio and a risk-free asset. Practically speaking, a CAPM-oriented investor might hold a money market fund and a portfolio of capitalization-weighted broad market indexes across many asset classes, varying the weights in these two in accordance with risk tolerance. These types of considerations are also relevant to individual investors. An individual investor who depends on income from salary or self-employment is sensitive to business cycle risk, in particular to the effects of recessions. If this investor compared two stocks with the same CAPM beta, given his concern about recessions, he might be very sensitive to receiving an adequate premium for investing in procyclical assets. In contrast, an investor with independent wealth and no job-loss concerns would have a comparative advantage in bearing business cycle risk; his optimal risky asset portfolio might be quite different from that of the investor with job-loss concerns in tilting toward greater-than-average exposure to the business cycle factor, all else being equal. Investors should be aware of which priced risks they face and analyze the extent of their exposure.

A multifactor approach can help investors achieve better-diversified and possibly more-efficient portfolios. For example, the characteristics of a portfolio can be better explained by a combination of SMB, HML, and WML factors in addition to the market factor than by using the market factor alone.

Thus, compared with single-factor models, multifactor models offer a richer context for investors to search for ways to improve portfolio selection.

SUMMARY

In our coverage of multifactor models, we have presented concepts, models, and tools that are key ingredients to quantitative portfolio management and are used to both construct portfolios and to attribute sources of risk and return.

- Multifactor models permit a nuanced view of risk that is more granular than the single-factor approach allows.
- Multifactor models describe the return on an asset in terms of the risk of the asset with respect to a set of factors. Such models generally include systematic factors, which explain the average returns of a large number of risky assets. Such factors represent priced risk—risk for which investors require an additional return for bearing.
- The arbitrage pricing theory (APT) describes the expected return on an asset (or portfolio) as a linear function of the risk of the asset with respect to a set of factors. Like the CAPM, the APT describes a financial market equilibrium; however, the APT makes less strong assumptions.
- The major assumptions of the APT are as follows:
  - Asset returns are described by a factor model.
  - With many assets to choose from, asset-specific risk can be eliminated.
  - Assets are priced such that there are no arbitrage opportunities.
- Multifactor models are broadly categorized according to the type of factor used:
  - Macroeconomic factor models
  - Fundamental factor models
  - Statistical factor models
- In *macroeconomic* factor models, the factors are surprises in macroeconomic variables that significantly explain asset class (equity in our examples) returns. Surprise is defined as actual minus forecasted value and has an expected value of zero. The factors can be understood as affecting either the expected future cash flows of companies or the interest rate used to discount these cash flows back to the present and are meant to be uncorrelated.
- In *fundamental* factor models, the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices. Among the fundamental factors are book-value-to-price ratio, market capitalization, price-to-earnings ratio, and financial leverage.
- In contrast to macroeconomic factor models, in fundamental models the factors are calculated as returns rather than surprises. In fundamental factor models, we generally specify the factor sensitivities (attributes) first and then estimate the factor returns through regressions. In macroeconomic factor models, however, we first develop the factor (surprise) series and then estimate the factor sensitivities through regressions. The factors of most fundamental factor models may be classified as company fundamental factors, company share-related factors, or macroeconomic factors.
- In *statistical* factor models, statistical methods are applied to a set of historical returns to determine portfolios that explain historical returns in one of two senses. In factor analysis models, the factors are the portfolios that best explain (reproduce) historical return covariances. In principal-components models, the factors are portfolios that best explain (reproduce) the historical return variances.

- Multifactor models have applications to return attribution, risk attribution, portfolio construction, and strategic investment decisions.
- A factor portfolio is a portfolio with unit sensitivity to a factor and zero sensitivity to other factors.
- Active return is the return in excess of the return on the benchmark.
- Active risk is the standard deviation of active returns. Active risk is also called tracking error or tracking risk. Active risk squared can be decomposed as the sum of active factor risk and active specific risk.
- The information ratio (IR) is mean active return divided by active risk (tracking error). The IR measures the increment in mean active return per unit of active risk.
- Factor models have uses in constructing portfolios that track market indexes and in alternative index construction.
- Traditionally, the CAPM approach would allocate assets between the risk-free asset and a broadly diversified index fund. Considering multiple sources of systematic risk may allow investors to improve on that result by tilting away from the market portfolio. Generally, investors would gain from accepting above average (below average) exposures to risks that they have a comparative advantage (comparative disadvantage) in bearing.

REFERENCES

Bodie, Zvi, Alex Kane, and Alan J. Marcus. 2017. *Investments*. 11th ed. New York: McGraw-Hill Education.

Burmeister, Edwin, Richard Roll, and Stephen A. Ross. 1994. "A Practitioner's Guide to Arbitrage Pricing Theory." In *A Practitioner's Guide to Factor Models*. Charlottesville, VA: Research Foundation of the Institute of Chartered Financial Analysts.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52 (1): 57–82.

Connor, Gregory. 1995. "The Three Types of Factor Models: A Comparison of Their Explanatory Power." *Financial Analysts Journal* 51 (3): 42–46.

Dopfel, Frederick E. 2004. "Fixed-Income Style Analysis and Optimal Manager Structure." *Journal of Fixed Income* 14 (2): 32–43.

Fabozzi, Frank J. 2008. *Handbook of Finance, Financial Markets and Instruments*. Wiley.

Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47 (2): 427–65.

Grinold, Richard, and Ronald N. Kahn. 1994. "Multi-Factor Models for Portfolio Risk." In *A Practitioner's Guide to Factor Models*. Charlottesville, VA: Research Foundation of the Institute of Chartered Financial Analysts.

Podkaminer, Eugene L. 2017. "Smart Beta Is the Gateway Drug to Risk Factor Investing." *Journal of Portfolio Management* 43 (5): 130–34.

Ross, S. A. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (3): 341–60.

PRACTICE PROBLEMS

1 Compare the assumptions of the arbitrage pricing theory (APT) with those of the capital asset pricing model (CAPM).

2 Last year the return on Harry Company stock was 5 percent. The portion of the return on the stock not explained by a two-factor macroeconomic factor model was 3 percent. Using the data given below, calculate Harry Company stock's expected return.

Macroeconomic Factor Model for Harry Company Stock

Variable	Actual Value (%)	Expected Value (%)	Stock's Factor Sensitivity
Change in interest rate	2.0	0.0	–1.5
Growth in GDP	1.0	4.0	2.0

3 Assume that the following one-factor model describes the expected return for portfolios:

$$E(R_p) = 0.10 + 0.12\beta_{p,1}$$

Also assume that all investors agree on the expected returns and factor sensitivity of the three highly diversified Portfolios A, B, and C given in the following table:

Portfolio	Expected Return	Factor Sensitivity
A	0.196	0.80
B	0.156	1.00
C	0.244	1.20

Assuming the one-factor model is correct and based on the data provided for Portfolios A, B, and C, determine if an arbitrage opportunity exists and explain how it might be exploited.

4 Which type of factor model is most directly applicable to an analysis of the style orientation (for example, growth vs. value) of an active equity investment manager? Justify your answer.

5 Suppose an active equity manager has earned an active return of 110 basis points, of which 80 basis points is the result of security selection ability. Explain the likely source of the remaining 30 basis points of active return.

6 Address the following questions about the information ratio.

A What is the information ratio of an index fund that effectively meets its investment objective?

© 2017 CFA Institute. All rights reserved.

B What are the two types of risk an active investment manager can assume in seeking to increase his information ratio?

7 A wealthy investor has no other source of income beyond her investments and that income is expected to reliably meet all her needs. Her investment advisor recommends that she tilt her portfolio to cyclical stocks and high-yield bonds. Explain the advisor's advice in terms of comparative advantage in bearing risk.

The following information relates to Questions 8–13

Carlos Altuve is a manager-of-managers at an investment company that uses quantitative models extensively. Altuve seeks to construct a multi-manager portfolio using some of the funds managed by portfolio managers within the firm. Maya Zapata is assisting him.

Altuve uses arbitrage pricing theory (APT) as a basis for evaluating strategies and managing risks. From his earlier analysis, Zapata knows that Funds A and B in Exhibit 1 are well diversified. He has not previously worked with Fund C and is puzzled by the data because it is inconsistent with APT. He asks Zapata gather additional information on Fund C's holdings and to determine if an arbitrage opportunity exists among these three investment alternatives. Her analysis, using the data in Exhibit 1, confirms that an arbitrage opportunity does exist.

Exhibit 1 Expected Returns and Factor Sensitivities (One-Factor Model)

Fund	Expected Return	Factor Sensitivity
A	0.02	0.5
B	0.04	1.5
C	0.03	0.9

Using a two-factor model, Zapata now estimates the three funds' sensitivity to inflation and GDP growth. That information is presented in Exhibit 2. Zapata assumes a zero value for the error terms when working with the selected two-factor model.

Exhibit 2 Expected Returns and Factor Sensitivities (Two-Factor Model)

		Factor Sensitivity	
Fund	Expected Return	Inflation	GDP Growth
A	0.02	0.5	1.0
B	0.04	1.6	0.0
C	0.03	1.0	1.1

Altuve asks Zapata to calculate the return for Portfolio AC, composed of a 60% allocation to Fund A and 40% allocation to Fund C, using the surprises in inflation and GDP growth in Exhibit 3.

Exhibit 3 Selected Data on Factors

Factor	Research Staff Forecast	Actual Value
Inflation	2.0%	2.2%
GDP Growth	1.5%	1.0%

Finally, Altuve asks Zapata about the return sensitivities of Portfolios A, B, and C given the information provided in Exhibit 3.

8 Which of the following is *not* a key assumption of APT, which is used by Altuve to evaluate strategies and manage risks?

A A factor model describes asset returns.

B Asset-specific risk can be eliminated through diversification.

C Arbitrage opportunities exist among well-diversified portfolios.

9 The arbitrage opportunity identified by Zapata can be exploited with:

A Strategy 1: Buy \$50,000 Fund A and \$50,000 Fund B; sell short \$100,000 Fund C.

B Strategy 2: Buy \$60,000 Fund A and \$40,000 Fund B; sell short \$100,000 Fund C.

C Strategy 3: Sell short \$60,000 of Fund A and \$40,000 of Fund B; buy \$100,000 Fund C.

10 The two-factor model Zapata uses is a:

A statistical factor model.

B fundamental factor model.

C macroeconomic factor model.

11 Based on the data in Exhibits 2 and 3, the return for Portfolio AC, given the surprises in inflation and GDP growth, is *closest* to:

A 2.02%.

B 2.40%.

C 4.98%.

12 The surprise in which of the following had the greatest effect on fund returns?

A Inflation on Fund B

B GDP growth on Fund A

C GDP growth on Fund C

13 Based on the data in Exhibit 2, which fund is most sensitive to the combined surprises in inflation and GDP growth in Exhibit 3?

A Fund A

B Fund B

C Fund C

The following information relates to Questions 14–19

Hui Cheung, a portfolio manager, asks her assistant, Ronald Lam, to review the macroeconomic factor model currently in use and to consider a fundamental factor model as an alternative.

The current macroeconomic factor model has four factors:

$$R_i = a_i + b_{i1}F_{\text{GDP}} + b_{i2}F_{\text{CAP}} + b_{i3}F_{\text{CON}} + b_{i4}F_{\text{UNEM}} + \varepsilon_i,$$

Where F_{GDP}, F_{CAP}, F_{CON}, and F_{UNEM} represent unanticipated changes in four factors: gross domestic product, manufacturing capacity utilization, consumer spending, and the rate of unemployment, respectively. Lam assumes the error term is equal to zero when using this model.

Lam estimates the current model using historical monthly returns for three portfolios for the most recent five years. The inputs used in and estimates derived from the macroeconomic factor model are presented in Exhibit 1. The US Treasury bond rate of 2.5% is used as a proxy for the risk-free rate of interest.

Exhibit 1 Inputs for and Estimates from the Current Macroeconomic Model

	Factor Sensitivities and Intercept Coefficients				Factor
Factor	**Portfolio 1**	**Portfolio 2**	**Portfolio 3**	**Benchmark**	**Surprise (%)**
Intercept (%)	2.58	3.20	4.33		
F_{GDP}	0.75	1.00	0.24	0.50	0.8
F_{CAP}	−0.23	0.00	−1.45	−1.00	0.5
F_{CON}	1.23	0.00	0.50	1.10	2.5
F_{UNEM}	−0.14	0.00	−0.05	−0.10	1.0
	Annual Returns, Most Recent Year				
Return (%)	6.00	4.00	5.00	4.50	

Lam uses the macroeconomic model to calculate the tracking error and the mean active return for each portfolio. He presents these statistics in Exhibit 2.

Exhibit 2 Macroeconomic Factor Model Tracking Error and Mean Active Return

Portfolio	Tracking Error	Mean Active Return
Portfolio 1	1.50%	1.50%
Portfolio 2	1.30%	−0.50%
Portfolio 3	1.00%	0.50%

Lam considers a fundamental factor model with four factors:

$$R_i = a_j + b_{j1}F_{\text{LIQ}} + b_{j2}F_{\text{LEV}} + b_{j3}F_{\text{EGR}} + b_{j4}F_{\text{VAR}} + \varepsilon_j,$$

where F_{LIQ}, F_{LEV}, F_{EGR}, and F_{VAR} represent liquidity, financial leverage, earnings growth, and the variability of revenues, respectively.

Lam and Cheung discuss similarities and differences between macroeconomic factor models and fundamental factor models, and Lam offers a comparison of those models to statistical factor models. Lam makes the following statements.

Statement 1 The factors in fundamental factor models are based on attributes of stocks or companies, whereas the factors in macroeconomic factor models are based on surprises in economic variables.

Statement 2 The factor sensitivities are generally determined first in fundamental factor models, whereas the factor sensitivities are estimated last in macroeconomic factor models.

Lam also tells Cheung:

> An advantage of statistical factor models is that they make minimal assumptions, and therefore, statistical factor model estimation lends itself to easier interpretation than macroeconomic and fundamental factor models.

Lam tells Cheung that multifactor models can be useful in active portfolio management, but not in passive management. Cheung disagrees; she tells Lam that multifactor models can be useful in both active and passive management.

14 Based on the information in Exhibit 1, the expected return for Portfolio 1 is *closest* to:

- **A** 2.58%.
- **B** 3.42%.
- **C** 6.00%.

15 Based on Exhibit 1, the active risk for Portfolio 2 is explained by surprises in:

- **A** GDP.
- **B** consumer spending.
- **C** all four model factors.

16 Based on Exhibit 2, which portfolio has the best information ratio?

- **A** Portfolio 1
- **B** Portfolio 2
- **C** Portfolio 3

17 Which of Lam's statements regarding macroeconomic factor models and fundamental factor models is correct?

- **A** Only Statement 1
- **B** Only Statement 2
- **C** Both Statements 1 and 2

18 Is Lam's comment regarding statistical factor models correct?

- **A** Yes
- **B** No, because he is incorrect with respect to interpretation of the models' results
- **C** No, because he is incorrect with respect to the models' assumptions

19 Whose statement regarding the use of multifactor models in active and passive portfolio management is correct?

- **A** Lam only
- **B** Cheung only
- **C** Both Lam and Cheung

SOLUTIONS

1 APT and the CAPM are both models that describe what the expected return on a risky asset should be in equilibrium given its risk. The CAPM is based on a set of assumptions including the assumption that investors' portfolio decisions can be made considering just returns' means, variances, and correlations. The APT makes three assumptions:

   1 A factor model describes asset returns.

   2 There are many assets, so investors can form well-diversified portfolios that eliminate asset-specific risk.

   3 No arbitrage opportunities exist among well-diversified portfolios.

2 In a macroeconomic factor model, the surprise in a factor equals actual value minus expected value. For the interest rate factor, the surprise was 2 percent; for the GDP factor, the surprise was −3 percent. The intercept represents expected return in this type of model. The portion of the stock's return not explained by the factor model is the model's error term.

$$\begin{aligned} 5\% &= \text{Expected return} - 1.5(\text{Interest rate surprise}) + 2(\text{GDP surprise}) + \\ &\quad \text{Error term} \\ &= \text{Expected return} - 1.5(2\%) + 2(-3\%) + 3\% \\ &= \text{Expected return} - 6\% \end{aligned}$$

Rearranging terms, the expected return for Harry Company stock equals 5% + 6% = 11%.

3 According to the one-factor model for expected returns, the portfolio should have these expected returns if they are correctly priced in terms of their risk:

   Portfolio A $E(R_A) = 0.10 + 0.12\beta_{A,1} = 0.10 + (0.12)(0.80) = 0.10 + 0.10 = 0.20$

   Portfolio B $E(R_B) = 0.10 + 0.12\beta_{B,1} = 0.10 + (0.12)(1.00) = 0.10 + 0.12 = 0.22$

   Portfolio C $E(R_C) = 0.10 + 0.12\beta_{C,1} = 0.10 + (0.12)(1.20) = 0.10 + 0.14 = 0.24$

In the table below, the column for expected return shows that Portfolios A and C are correctly priced but Portfolio B offers too little expected return for its risk, 0.15 or 15%. By shorting Portfolio B (selling an overvalued portfolio) and using the proceeds to buy a portfolio 50% invested in A and 50% invested in C with a sensitivity of 1 that matches the sensitivity of B, for each monetary unit shorted (say each euro), an arbitrage profit of €0.22 − €0.15 = €0.07 is earned.

Portfolio	Expected Return	Factor Sensitivity
A	0.196	0.80
B	0.156	1.00
C	0.244	1.20
0.5A + 0.5C	0.22	1.00

4 A fundamental factor model. Such models typically include many factors related to the company (e.g., earnings) and to valuation that are commonly used indicators of a growth orientation. A macroeconomic factor model may provide relevant information as well, but typically indirectly and in less detail.

5 This remainder of 30 basis points would be attributable to the return from factor tilts. A portfolio manager's active return is the sum of two components, factor tilts and security selection. Factor tilt is the product of the portfolio manager's higher or lower factor sensitivities relative to the benchmark's factor sensitivities and the factor returns. Security selection reflects the manager's ability to overweight securities that outperform or underweight securities that underperform.

6 **A** An index fund that effectively meets its investment objective is expected to have an information ratio of zero, because its active return should be zero.

B The active manager may assume active factor risk and active specific risk (security selection risk) in seeking a higher information ratio.

7 This wealthy investor has a comparative advantage in bearing business cycle risk compared with the average investor who depends on income from employment. Because the average investor is sensitive to the business cycle and in particular the risk of recession, we would expect there to be a risk premium to hold recession-sensitive securities. Cyclical stocks and high-yield bonds are both very sensitive to the risk of recessions. Because the welfare of the wealthy investor is not affected by recessions, she can tilt her portfolio to include cyclical stocks and high yield bonds to attempt to capture the associated risk premiums.

8 C is correct. Arbitrage pricing theory (APT) is a framework that explains the expected return of a portfolio in equilibrium as a linear function of the risk of the portfolio with respect to a set of factors capturing systematic risk. A key assumption of APT is that, in equilibrium, there are no arbitrage opportunities.

9 C is correct. The expected return and factor sensitivities of a portfolio with a 60% weight in Fund A and a 40% weight in Fund B are calculated as weighted averages of the expected returns and factor sensitivities of Funds A and B:

$$\begin{aligned}\text{Expected return of Portfolio 60/40} &= (0.60)(0.02) + (0.40)(0.04)\\ &= 0.028\text{, or }2.8\%\\ \text{Factor sensitivity of Portfolio 60/40} &= (0.60)(0.5) + (0.40)(1.5)\\ &= 0.9\end{aligned}$$

Fund	Expected Return	Factor Sensitivity
A	0.02	0.5
B	0.04	1.5
C	0.03	0.9
Portfolio 60/40		
60%A + 40%B	0.028	0.900
Portfolio 50/50		
50%A + 50%B	0.030	1.000

The factor sensitivity of Portfolio 60/40 is identical to that of Fund C; therefore, this strategy results in no factor risk relative to Portfolio C. However, Fund C's expected return of 3.0% is higher than Portfolio 60/40's expected return of 2.8%. This difference supports Strategy 3: buying Fund C and selling short Portfolio 60/40 to exploit the arbitrage opportunity.

10 C is correct. In a macroeconomic factor model, the factors are surprises in macroeconomic variables, such as inflation risk and GDP growth, that significantly explain returns.

11 A is correct. The macroeconomic two-factor model takes the following form:

$$R_i = a_i + b_{i1}F_{\text{INF}} + b_{i2}F_{\text{GDP}} + \varepsilon_i,$$

where F_{INF} and F_{GDP} represent surprises in inflation and surprises in GDP growth, respectively, and a_i represents the expected return to asset i. Using this model and the data in Exhibit 2, the returns for Fund A and Fund C are represented by the following:

$$R_A = 0.02 + 0.5F_{\text{INF}} + 1.0F_{\text{GDP}} + \varepsilon_A$$

$$R_C = 0.03 + 1.0F_{\text{INF}} + 1.1F_{\text{GDP}} + \varepsilon_C$$

Surprise in a macroeconomic model is defined as actual factor minus predicted factor. The surprise in inflation is 0.2% (= 2.2% – 2.0%). The surprise in GDP growth is –0.5% (= 1.0% – 1.5%). The return for Portfolio AC, composed of a 60% allocation to Fund A and 40% allocation to Fund C, is calculated as the following:

$$\begin{aligned} R_{AC} &= (0.6)(0.02) + (0.4)(0.03) + [(0.6)(0.5) + (0.4)(1.0)](0.002) + [(0.6)(1.0) \\ &\quad + (0.4)(1.1)](-0.005) + 0.6(0) + 0.4(0) \\ &= 2.02\% \end{aligned}$$

12 C is correct. Surprise in a macroeconomic model is defined as actual factor minus predicted factor. For inflation, the surprise factor is 2.2% – 2.0% = 0.2%; for GDP growth, the surprise factor is 1.0% – 1.5% = –0.5%. The effect on returns is the product of the surprise and the factor sensitivity.

	Change in Portfolio Return due to Surprise in	
Fund	**Inflation**	**GDP Growth**
A	0.5 × 0.2% = 0.10%	1.0 × –0.5% = –0.50%
B	1.6 × 0.2% = 0.32%	0.0 × –0.5% = 0.00%
C	1.0 × 0.2% = 0.20%	1.1 × –0.5% = –0.55%

The effect of the GDP growth surprise on Fund C was the largest single-factor effect on Fund returns (–0.55%).

13 A is correct. The effect of the surprises in inflation and GDP growth on the returns of the three funds is calculated as the following.

	Change in Portfolio Return Because of Surprise in	
Fund	**Inflation**	**GDP Growth**
A	0.5 × 0.2% = 0.10%	1.0 × –0.5% = –0.50%
B	1.6 × 0.2% = 0.32%	0.0 × –0.5% = 0.00%
C	1.0 × 0.2% = 0.20%	1.1 × –0.5% = –0.55%

The combined effects for the three funds are the following.

Fund A: 0.10% + (–0.50%) = –0.40%

Fund B: 0.32% + (0.00%) = 0.32%

Fund C: 0.20% + (–0.55%) = –0.35%

Therefore, Fund A is the most sensitive to the surprises in inflation and GDP growth in Exhibit 3.

14 A is correct. When using a macroeconomic factor, the expected return is the intercept (when all model factors take on a value of zero). The intercept coefficient for Portfolio 1 in Exhibit 1 is 2.58.

15 C is correct. Active risk, also referred to as tracking risk or tracking error, is the sample standard deviation of the time series of active returns, where the active returns consist of the differences between the portfolio return and the benchmark return. Whereas GDP is the only portfolio non-zero sensitivity for Portfolio 2, the contribution to the portfolio's active return is the sum of the differences between the portfolio's and the benchmark's sensitivities multiplied by the factor return. Because all four of the factor sensitivities of Portfolio 2 are different from the factor sensitivities of the benchmark, all four factors contribute to the portfolio's active return and, therefore, to its active risk.

16 A is correct. Portfolio 1 has the highest information ratio, 1.0, and thus has the best mean active return per unit of active risk:

$$\mathrm{IR} = \frac{\bar{R}_P - \bar{R}_B}{s\left(R_P - R_B\right)}$$

$$= \frac{1.50\%}{1.50\%}$$

$$= 1.00$$

This information ratio exceeds that of Portfolio 2 (−0.38) or Portfolio 3 (0.50).

17 C is correct. In a macroeconomic factor model, the factors are surprises in macroeconomic variables that significantly explain returns. Factor sensitivities are generally specified first in fundamental factor models, whereas factor sensitivities are estimated last in macroeconomic factor models.

18 B is correct. An advantage of statistical factor models is that they make minimal assumptions. However, the interpretation of statistical factors is generally more difficult than the interpretation of macroeconomic and fundamental factor models.

19 B is correct. Analysts can use multifactor models in passively managed portfolios to replicate an index fund's factor exposures.

READING

40

Measuring and Managing Market Risk

by Don M. Chance, PhD, CFA, and Michelle McCarthy Beck

Don M. Chance, PhD, CFA, is at Louisiana State University (USA). Michelle McCarthy Beck is at TIAA (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** explain the use of value at risk (VaR) in measuring portfolio risk;
☐	**b.** compare the parametric (variance–covariance), historical simulation, and Monte Carlo simulation methods for estimating VaR;
☐	**c.** estimate and interpret VaR under the parametric, historical simulation, and Monte Carlo simulation methods;
☐	**d.** describe advantages and limitations of VaR;
☐	**e.** describe extensions of VaR;
☐	**f.** describe sensitivity risk measures and scenario risk measures and compare these measures to VaR;
☐	**g.** demonstrate how equity, fixed-income, and options exposure measures may be used in measuring and managing market risk and volatility risk;
☐	**h.** describe the use of sensitivity risk measures and scenario risk measures;
☐	**i.** describe advantages and limitations of sensitivity risk measures and scenario risk measures;
☐	**j.** explain constraints used in managing market risks, including risk budgeting, position limits, scenario limits, and stop-loss limits;
☐	**k.** explain how risk measures may be used in capital allocation decisions;
☐	**l.** describe risk measures used by banks, asset managers, pension funds, and insurers.

© 2020 CFA Institute. All rights reserved.

1 INTRODUCTION, UNDERSTANDING VALUE AT RISK VALUE AT RISK: FORMAL DEFINITION

a explain the use of value at risk (VaR) in measuring portfolio risk;

This reading is an introduction to the process of measuring and managing market risk. Market risk is the risk that arises from movements in stock prices, interest rates, exchange rates, and commodity prices. Market risk is distinguished from credit risk, which is the risk of loss from the failure of a counterparty to make a promised payment, and also from a number of other risks that organizations face, such as breakdowns in their operational procedures. In essence, market risk is the risk arising from changes in the markets to which an organization has exposure.

Risk management is the process of identifying and measuring risk and ensuring that the risks being taken are consistent with the desired risks. The process of managing market risk relies heavily on the use of models. A model is a simplified representation of a real world phenomenon. Financial models attempt to capture the important elements that determine prices and sensitivities in financial markets. In doing so, they provide critical information necessary to manage investment risk. For example, investment risk models help a portfolio manager understand how much the value of the portfolio is likely to change given a change in a certain risk factor. They also provide insight into the gains and losses the portfolio might reasonably be expected to experience and the frequency with which large losses might occur.

Effective risk management, though, is much more than just applying financial models. It requires the application of judgment and experience not only to know how to use the models appropriately but also to appreciate the strengths and limitations of the models and to know when to supplement or substitute one model with another model or approach.

Financial markets operate more or less continuously, and new prices are constantly being generated. As a result, there is a large amount of data on market risk and a lot of collective experience dealing with this risk, making market risk one of the easier financial risks to analyze. Still, market risk is not an easy risk to capture. Although a portfolio's exposures can be identified with some certainty, the potential losses that could arise from those exposures are unknown. The data used to estimate potential losses are generated from past prices and rates, not the ones to come. Risk management models allow the experienced risk manager to blend that historical data with their own forward-looking judgment, providing a framework within which to test that judgment.

We first lay a foundation for understanding value at risk, discuss three primary approaches to estimating value at risk, and cover the primary advantages and limitations as well as extensions of value at risk. We then address the sensitivity measures used for equities, fixed-income securities, and options and also cover historical and hypothetical scenario risk measures. Next, we discuss the use of constraints in risk management, such as risk budgeting, position limits, scenario limits, stop-loss limits, and capital allocation as risk management tools. Lastly, we describe various applications and limitations of risk measures as used by different types of market participants and summarize our discussion.

1.1 Understanding Value at Risk

Value at risk (VaR) was developed in the late 1980s, and over the next decade, it emerged as one of the most important risk measures in global financial markets.

1.1.1 *Value at Risk: Formal Definition*

Value at risk is the minimum loss that would be expected a certain percentage of the time over a certain period of time given the assumed market conditions. It can be expressed in either currency units or as a percentage of portfolio value. Although this statement is an accurate definition of VaR, it does not provide sufficient clarity to fully comprehend the concept. To better understand what VaR means, let us work with an example. Consider the statement:

The 5% VaR of a portfolio is €2.2 million over a one-day period.

The following three points are important in understanding the concept of VaR:

- VaR can be measured in either currency units (in this example, the euro) or in percentage terms. In this example, if the portfolio value is €400 million, the VaR expressed in percentage terms would be 0.55% (€2.2 million/€400 million = 0.0055).
- VaR is a *minimum* loss. This point cannot be emphasized enough. VaR is often mistakenly assumed to represent *how much one can lose*. If the question is, "how much can one lose?" there is only one answer: *the entire portfolio*. In a €400 million portfolio, assuming no leverage, the most one can lose is €400 million.
- A VaR statement references a time horizon: losses that would be expected to occur over a given period of time. In this example, that period of time is one day. (If VaR is measured on a daily basis, and a typical month has 20–22 business days, then 5% of the days equates to about one day per month.)

These are the explicit elements of a VaR statement: the *frequency* of losses of a given *minimum magnitude* expressed either in *currency* or *percentage* terms. Thus, the VaR statement can be rephrased as follows: A loss of at least €2.2 million would be expected to occur about once every month.

A 5% VaR is often expressed as its complement—a 95% level of confidence. In this reading, we will typically refer to the notion as a 5% VaR, but we should be mindful that it does imply a 95% level of confidence.

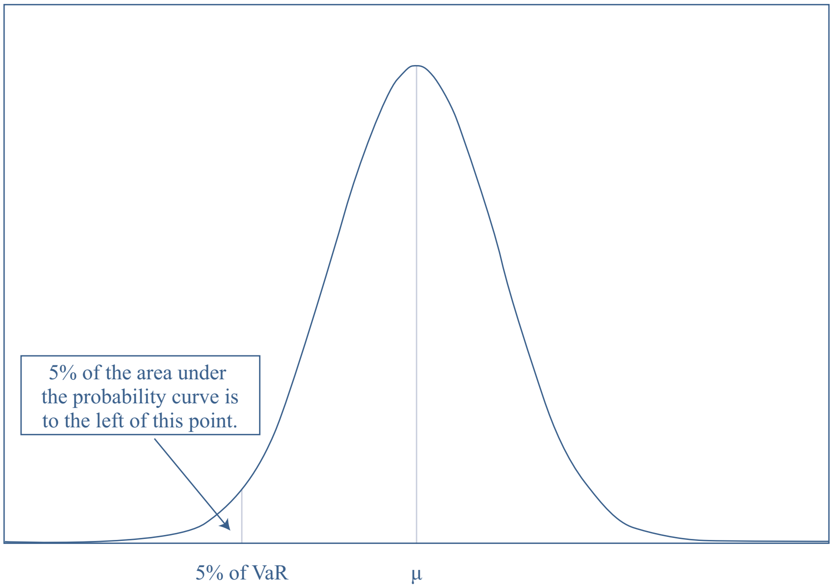
Using the example given, it is correct to say any of the following:

- €2.2 million is the minimum loss we would expect 5% of the time.
- 5% of the time, losses would be at least €2.2 million.
- We would expect a loss of no more than €2.2 million 95% of the time.

The last sentence is sometimes mistakenly phrased as "95% of the time we would expect to lose less than €2.2 million," but this statement could be taken to mean that 95% of the time we would incur losses, although those losses would be less than €2.2 million. In fact, a large percentage of the time we will make money.

Exhibit 1 illustrates the concept of VaR using the 5% case. It depicts a probability distribution of returns from a hypothetical portfolio. The distribution chosen is the familiar normal distribution, known sometimes as the bell curve, but that distribution is only one curve that might be used. In fact, there are compelling arguments that the normal distribution is not the right one to use for financial market returns. We discuss these arguments later.

Exhibit 1 Illustration of 5% VaR in the Context of a Probability Distribution



Note that the distribution in Exhibit 1 is centered on the value μ. [The symbol μ (Greek: *mu*) is a common symbol used to represent an expected value.] Near the left tail of the distribution is the notation "5% VaR," indicating that 5% of the area under the curve is to the left of the point of the VaR (i.e., the probability of observing a value less than the VaR is 5%).

Thus, it is apparent that VaR is simply a point on the probability distribution of profits or returns from a portfolio. Given the characteristics of the normal distribution, a 5% VaR is equivalent to the point on the distribution that is 1.65 standard deviations below the expected value. Although the concept of VaR can be easily visualized in this manner, actually measuring the VaR is a challenge.

Before we take on that challenge, however, note that there is no formal requirement that VaR be measured at a 5% threshold. It is also common to use a 1% threshold (2.33 standard deviations from the expected value), and some investment managers use a one standard deviation movement (equal to a 16% VaR)—both assuming a normal distribution. There is no definitive rule for what VaR cutoff should be used. A specification with a higher confidence level will produce a higher VaR. It is up to the decision maker to choose an appropriate level.

VaR and Standard Deviations

The 16% VaR relates to a one standard deviation move as follows: In a normal distribution, 50% of the outcomes are to the right of the expected value and 50% are to the left. A one standard deviation interval implies that 68% of the outcomes lie within one standard deviation of the expected value; thus, 34% of the outcomes lie one standard deviation to the left of the expected value and 34% of the outcomes one standard deviation to the right. Adding the 50% of the outcomes that lie to the right of the expected value to the 34% of the outcomes that lie one standard deviation below the expected value means that 84% of all outcomes lie to the right of the point that is one standard deviation to the left of the expected value. Therefore, 16% of all outcomes lie below this point. Thus, a one standard deviation movement is equivalent to a 16% VaR (or an 84% level of confidence).

Just as there is no formal requirement that VaR be measured at a 5% cutoff, there is also no formal requirement that VaR be measured using a daily loss estimate. One could reasonably measure VaR on a weekly, bi-weekly, monthly, quarterly, semiannually, or annual basis. Choosing the VaR threshold and the time horizon are examples of why VaR is not a precise measure but in fact entails considerable judgment.

We should also reiterate that VaR can be expressed as a rate of return or in monetary terms. It is typically easier to process the data necessary to estimate VaR in terms of returns, but VaR is most frequently expressed in terms of profits or losses. This point will become clearer as we work through examples.

EXAMPLE 1

Definition of VaR

1 Given a VaR of $12.5 million at 5% for one month, which of the following statements is correct?
   - A There is a 5% chance of losing $12.5 million over one month.
   - B There is a 95% chance that the expected loss over the next month is less than $12.5 million.
   - C The minimum loss that would be expected to occur over one month 5% of the time is $12.5 million.

2 Which of the following statements is **not** correct?
   - A A 1% VaR implies a downward move of 1%.
   - B A one standard deviation downward move is equivalent to a 16% VaR.
   - C A 5% VaR implies a move of 1.65 standard deviations less than the expected value.

Solution to 1:

C is correct because it is the only statement that accurately expresses the VaR. A is incorrect because VaR does not give the likelihood of losing a specific amount. B is incorrect because VaR is not an expected loss; rather, it is a minimum loss.

Solution to 2:

A is correct. A 1% VaR (99% confidence) is the point on the distribution 2.33 standard deviations below the expected value. Answers B and C correctly describe a 16% and 5% VaR, respectively.

To this point, we have given only the conceptual definition of VaR. Defining something is one thing; measuring it can be quite challenging. Such is the case for VaR.

2 ESTIMATING VAR

Three methods are typically used to estimate VaR: the parametric (variance–covariance) method, the historical simulation method, and the Monte Carlo simulation method. Each of these will be discussed in turn.

The first step of every VaR calculation, regardless of the VaR method used, is to convert the set of holdings in the portfolio into a set of exposures to **risk factors**, a process called **risk decomposition**. In some instances, this process can be very simple:

An equity security can be the risk factor itself. In other instances, the process can be highly complex. For example, a convertible bond issued by a foreign entity has both currency and equity risk factors as well as exposures to multiple points on a yield curve of a given credit quality. Fixed-income instruments and derivatives products often contain distinct risk exposures that require decomposition in order to accurately capture their loss potential.

The second step of VaR estimation requires gathering a data history for each of the risk factors in the VaR model. The three methods use different approaches to specifying these inputs, which will be discussed in the following sections. We will see that the parametric and Monte Carlo methods do not formally require a data history. They require only that the user enter estimates of certain parameters into the computational procedure (expected return, standard deviation, and for some models, skewness and kurtosis). One of the most common sources for estimating parameter inputs for any financial model is historical data, but the user could substitute estimates based on judgement or alternative forecasting models. Indeed, shortly we will override some historical estimates with our own judgement. Nonetheless, the collection of a data history is typically used at least as a starting point in the parametric and Monte Carlo methods, and it is absolutely required for the historical simulation method.

The third step of each method is where the differences between the three VaR methods are most apparent: how each method uses the data to make an estimate of the VaR.

Although most portfolios contain a large number of individual securities and other assets, we will use a two-asset portfolio to illustrate the three VaR methods. Using a limited number of assets permits us to closely observe the essential elements of the VaR estimation procedure without getting mired in the complex mathematics required to accommodate a large number of assets. The objective is to understand the concept of VaR, be aware of how it is estimated, know how it is used, appreciate the benefits of VaR, and be attentive to its limitations. We can achieve these objectives by keeping the portfolio fairly simple.

Our example portfolio has a market value of $150 million and consists of two ETFs—SPDR S&P 500 ETF (SPY), representing the US equity exposure, and SPDR Portfolio Long-Term Corporate Bond ETF (SPLB), representing a corporate bond exposure. We will allocate 80% of the portfolio to SPY and 20% of the portfolio to SPLB. For the sake of simplicity, the two securities will represent the risk factors and the return history of each ETF will serve as the risk factor history used in the VaR model. We have collected a set of two years of daily total return data, reflecting both capital appreciation and dividends on each ETF. The period used for this historical data set is called the **lookback period**. The question of exactly how much data are required to be a representative data set is a complex question that is common to all estimation problems in economics and finance. We will discuss some of the issues on this matter later in this reading.

Exhibit 2 provides statistical summary information based on the two years of daily data in the lookback period, covering the period of 1 July 2015 through 28 June 2019.

Exhibit 2 Statistical Estimates from Daily Return Data, 1 July 2015–28 June 2019

	Daily		Annualized	
	Average Return	Standard Deviation	Average Return	Standard Deviation
SPY	0.047%	0.86%	12.51%	13.64%
SPLB	0.031%	0.49%	8.03%	7.73%

Note: The correlation of SPLB and SPY = −0.0607.

SPY produced an annualized average return of about 12.5% with a standard deviation of 13.6%, significantly different from the long-term historical performance of the S&P 500 Index of approximately 10.5% average return and 20% standard deviation. SPLB produced an annualized average return of 8% with a standard deviation of about 7.7%. These numbers compare with an average annual return for long-term corporate bonds of slightly more than 6% and a standard deviation of about 8.5% (historical data are drawn from Malkiel 2007). Although the average return of SPLB in the last four years was higher than that of the overall long-term corporate bond sector, the standard deviations were similar.

The risk and return parameters for each risk factor in Exhibit 2 illustrate how one might collect historical data. It is necessary, however, to critically assess the data and apply judgment to modify the inputs if the lookback period is not representative of the expected performance of the securities (or risk factors) going forward. Exercising our judgment, and believing that we have no information to suggest that future performance will deviate from the long-run historical performance, we adjust our inputs and use returns of 10.5% for SPY and 6% for SPLB, with standard deviations of 20% for SPY and 8.5% for SPLB. These adjustments align the inputs more closely with the long-run historical performance of each sector. In practice, users will want to use estimates they believe are reflective of current expectations, though clearly one user's estimates could differ widely from another's.

Although the returns and standard deviations experienced over the lookback period have been adjusted to more closely align with long-run historical experience, we will use a correlation estimate approximately equal to the observed correlation over our lookback period. We are assuming that the recent historical relationship of equity and fixed-income returns is a reasonable assumption moving forward. To keep the numbers simple, we round the observed correlation of −0.0607 to −0.06.

Exhibit 3 illustrates our input assumptions for the VaR estimations.

Exhibit 3 Input Assumptions, 1 July 2015–28 June 2019

		Annualized	
	Allocation	Return	Standard Deviation
SPY	80%	10.5%	20.0%
SPLB	20%	6.0%	8.5%

Note: The correlation of SPLB and SPY = −0.06.

3 THE PARAMETRIC METHOD OF VAR ESTIMATION

b compare the parametric (variance–covariance), historical simulation, and Monte Carlo simulation methods for estimating VaR;

c estimate and interpret VaR under the parametric, historical simulation, and Monte Carlo simulation methods;

The **parametric method** of estimating VaR is sometimes referred to as the analytical method and sometimes the variance–covariance method. The parametric method begins, as does each method, with a risk decomposition of the portfolio holdings. It typically assumes that the return distributions for the risk factors in the portfolio are normal. It then uses the expected return and standard deviation of return for each risk factor to estimate the VaR.

Note that we said that this method *typically* uses the normal distribution. Indeed, that is the common case in practice, but there is no formal requirement that the normal distribution be used. The normal distribution conveniently requires only two parameters—the expected value and standard deviation—to encompass everything there is to know about it. If other distributions are used, additional parameters of the distribution, such as skewness and kurtosis, would be required. We will limit the presentation here to the normal distribution, but be aware that other, more accurately representative distributions could be used but would add complexity to the VaR estimation process.

Recall that in defining VaR, we identified a VaR threshold—a point in the left tail of the distribution, typically either the 5% left tail, the 1% left tail, or a one standard deviation move (16%). If the portfolio is characterized by normally distributed returns and the expected value and standard deviation are known, it is a simple matter to identify any point on the distribution. A normal distribution with expected value μ and standard deviation σ can be converted to a standard normal distribution, which is a special case of the normal distribution in which the expected value is zero and the standard deviation is one. A standard normal distribution is also known as a z-distribution. If we have observed a return R from a normal distribution, we can convert to its equivalent z-distribution value by the transformation:

$$z = \frac{R - \mu}{\sigma}$$

In a standard normal (z) distribution, a 5% VaR is 1.65 standard deviations below the expected value of zero. A 1% VaR is 2.33 standard deviations below the expected value of zero. A 16% VaR is one standard deviation below the expected value of zero. Thus, in our example, for a 5% VaR, we wish to know the return that is 1.65 standard deviations to the left of the expected return.

To estimate this VaR, we need the expected return and volatility of the portfolio. The expected return is estimated from the following equation:

$$E(R_p) = w_{SPY}E(R_{SPY}) + w_{SPLB}E(R_{SPLB}), \tag{1}$$

where the expected return of the portfolio, $E(R_p)$, is equal to the portfolio weights of SPY (w_{SPY}) and SPLB (w_{SPLB}) multiplied by the expected return of each asset, $E(R_{SPY})$ and $E(R_{SPLB})$.

The volatility of the portfolio, σ_p, is estimated from the following equation:

$$\sigma_p = \sqrt{w_{SPY}^2\sigma_{SPY}^2 + w_{SPLB}^2\sigma_{SPLB}^2 + 2w_{SPY}w_{SPLB}\rho_{SPY,SPLB}\sigma_{SPY}\sigma_{SPLB}} \tag{2}$$

where σ_{SPY} and σ_{SPLB} are the standard deviations (volatilities) of SPY and SPLB, respectively; $\rho_{SPY,SPLB}$ is the correlation between the returns on SPY and SPLB, respectively; and $\rho_{SPY,SPLB}\sigma_{SPY}\sigma_{SPLB}$ is the covariance between SPY and SPLB.

Recall that we estimated these parameters from the historical data, with some modifications to make them more consistent with long-run values. The formal calculations for our portfolio based on these adjusted estimates are as follows:

$$E\left(R_p\right) = 0.8(0.105) + 0.2(0.06) = 0.096000$$

$$\sigma_p = \sqrt{(0.8)^2(0.2)^2 + (0.2)^2(0.085)^2 + 2(0.8)(0.2)(-0.06)(0.2)(0.085)}$$

$$=0.159883.$$

Thus, our portfolio, consisting of an 80% position in SPY and a 20% position in SPLB, is estimated to have an expected return of 9.6% and a volatility of approximately 15.99%.

But these inputs are based on annual returns. If we want a one-day VaR, we should adjust the expected returns and volatilities to their daily counterparts. Assuming 250 trading days in a year, the expected return is adjusted by dividing by 250 and the standard deviation is adjusted by dividing by the square root of 250. (Note that the variance is converted by dividing by time, 250 days; thus, the standard deviation must be adjusted by using the square root of time, 250 days.) Thus, the daily expected return and volatility are

$$E\left(R_p\right) = \frac{0.096}{250} = 0.000384 \tag{3}$$

and

$$\sigma_p = \frac{0.159883}{\sqrt{250}} = 0.010112. \tag{4}$$

It is important to note that we have assumed that the statistical properties of the return distribution are constant across the year. Earlier, we annualized the daily data in Exhibit 2 in order to see how our estimates compared with long-term estimates. We made some modest adjustments to the annualized data and then, in Equations 3 and 4, returned to using daily data. To estimate an annual VaR, we would need to use annual data, but we would need a longer lookback period in order to have sufficient data points.

> It is important to note that we cannot estimate a daily VaR and annualize it to arrive at an annual VaR estimate. First, to assume that a daily distribution of returns can be extrapolated to an annual distribution is a bold assumption. Second, annualizing the daily VaR is not the same as adjusting the expected return and the standard deviation to annual numbers and then calculating the annual VaR. The expected return is annualized by multiplying the daily return by 250, and the standard deviation is annualized by multiplying the daily standard deviation by the square root of 250. Thus, we can annualize the data and estimate an annual VaR, but we cannot estimate a daily VaR and annualize it without assuming a zero expected return.

Having calculated the daily expected return and volatility, the parametric VaR is now easily obtained. With the distribution centered at the expected return of 0.0384% and a one standard deviation move equal to 0.996%, a 5% VaR is obtained by identifying the point on the distribution that lies 1.65 standard deviations to the left of the mean. It is now easy to see why parametric VaR is so named: The expected values, standard deviations, and covariances are the *parameters* of the distributions.

The following step-by-step procedure shows how the VaR is derived:

$$\{[E(Rp) - 1.65\sigma_p](-1)\}(\$150{,}000{,}000)$$

Step 1 Multiply the portfolio standard deviation by 1.65.

$0.010112 \times 1.65 = 0.016685$

Step 2 Subtract the answer obtained in Step 1 from the expected return.

$0.000384 - 0.016685 = -0.016301$

Step 3 Because VaR is expressed as an absolute number (despite representing an expected loss), change the sign of the value obtained in Step 2.

Change −0.016301 to 0.016301

Step 4 Multiply the result in Step 3 by the value of the portfolio.

$\$150,000,000 \times 0.016301 = \$2,445,150$

Thus, using the parametric method, our estimate of VaR is $2,445,150, meaning that on 5% of trading days the portfolio would be expected to incur a loss of at least $2,445,150. Note that asset managers may stop at Step 3 because at that point the measure is expressed as a percentage of the value of the portfolio, which is the unit this group more commonly uses.

EXAMPLE 2

Parametric VaR

1 The parameters of normal distribution required to estimate parametric VaR are:
   - A expected value and standard deviation.
   - B skewness and kurtosis.
   - C standard deviation and skewness.

2 Assuming a daily expected return of 0.0384% and daily standard deviation of 1.0112% (as in the example in the text), which of the following is *closest* to the 1% VaR for a $150 million portfolio? Express your answer in dollars.
   - A $3.5 million
   - B $2.4 million
   - C $1.4 million

3 Assuming a daily expected return of 0.0384% and daily standard deviation of 1.0112% (as in the example in the text), the daily 5% parametric VaR is $2,445,150. Rounding the VaR to $2.4 million, which of the following values is *closest* to the annual 5% parametric VaR? Express your answer in dollars.
   - A $38 million
   - B $25 million
   - C $600 million

Solution to 1:

A is correct. The parameters of a normal distribution are the expected value and standard deviation. Skewness, as mentioned in B and C, and kurtosis, as mentioned in B, are characteristics used to describe a *non*-normal distribution.

Solution to 2:

A is correct and is obtained as follows:

Step 1 $2.33 \times 0.010112 = 0.023561$

Step 2 $0.000384 - 0.023561 = -0.023177$

Step 3 Convert -0.023177 to 0.023177

Step 4 $0.023177 \times \$150$ million = $3,476,550

B is the estimated VaR at a 5% threshold, and C is the estimated VaR using a one standard deviation threshold.

Solution to 3:

B is correct. It is found by annualizing the daily return and standard deviation and using these figures in the calculation. The annual return and standard deviation are, respectively, 0.096000 (0.000384×250) and 0.159885 ($0.010112 \times \sqrt{250}$).

Step 1 $0.159885 \times 1.65 = 0.263810$

Step 2 $0.096000 - 0.263810 = -0.167810$

Step 3 Convert -0.167810 to 0.167810

Step 4 $0.167810 \times \$150$ million = $25,171,500

A incorrectly multiplies the daily VaR by the square root of the number of trading days in a year ($\sqrt{250}$), and C incorrectly multiplies the daily VaR by the approximate number of trading days in a year (250). Neither A nor C make the appropriate adjustment to annualize the standard deviation.

To recap, we see that the parametric VaR method generally makes the assumption that the distribution of returns on the risk factors is normal. Under that assumption, all of the information about a normal distribution is contained in the expected value and standard deviation. Therefore, finding the 5% VaR requires only that we locate the point in the distribution beyond which 5% of the outcomes occur. Although normality is the general assumption of the parametric method, it is not an absolute requirement. Other distributions could be accommodated by incorporating skewness and kurtosis, the third and fourth parameters of the distribution, but that added complexity is not needed to demonstrate the general approach to parametric VaR and is rarely done in practice.

The major advantage of the parametric method is its simplicity and straightforwardness. The assumption of the normal distribution allows us to easily estimate the parameters using historical data, although judgment is required to adjust the parameters when the historical data may be misleading. The parametric method is best used in situations in which one is confident that the normal distribution can be applied as a reasonable approximation of the true distribution and the parameter estimates are reliable or can be turned into reliable estimates by suitable adjustments. It is important to understand that VaR under the parametric method is very sensitive to the parameter estimates, especially the covariances.

One of the major weaknesses of the parametric method is that it can be difficult to use when the investment portfolio contains options. When options are exercised, they pay off linearly with the underlying; however, if never exercised, an option loses 100% of its value. This characteristic leads to a truncated, non-normal distribution that does not lend itself well to the parametric method. But some adjustments can render options more responsive to the parametric method. These adjustments are helpful but not perfect, limiting the usefulness of the parametric method when options are in the portfolio. Additionally, although the expected return and volatility of the underlying fixed income or equity security may be stable over the life of the option, the distribution of the option changes continuously as the value of the underlying, the volatility of the underlying, and the time to expiration all change.

4 THE HISTORICAL SIMULATION METHOD OF VAR ESTIMATION

b compare the parametric (variance–covariance), historical simulation, and Monte Carlo simulation methods for estimating VaR;

c estimate and interpret VaR under the parametric, historical simulation, and Monte Carlo simulation methods;

The **historical simulation method** of VaR uses the *current* portfolio and reprices it using the actual *historical* changes in the key factors experienced during the lookback period. We begin, as with the parametric method, by decomposing the portfolio into risk factors and gathering the historical returns of each risk factor from the chosen lookback period. Unlike the parametric method, however, we do not characterize the distribution using estimates of the mean return, the standard deviation, or the correlations among the risk factors in the portfolio. Instead, we reprice the current portfolio given the returns that occurred on each day of the historical lookback period and sort the results from largest loss to greatest gain. To estimate a one-day VaR at a 5% confidence interval, we choose the point on the resulting distribution beyond which 5% of the outcomes result in larger losses.

Illustrating this point using a full four years of daily observations would be tedious and consume a great deal of space, so we will condense the process quite a bit and then extrapolate the methodology. Exhibit 4 shows the daily returns on the SPY, the SPLB, and our 80% SPY/20% SPLB portfolio over the first five days of our historical data set. Please note that fixed weights are assumed for all days. Neither historical simulation nor Monte Carlo simulation is intended to be a replication of sequences of prices. They are intended to create a sample of one-day returns for a portfolio of given weights.

Exhibit 4 First Five Days of Historical Returns on the SPY/SPLB Portfolio Using the 1 July 2015–28 June 2019 Data

Day	SPY Return	SPLB Return	Portfolio Return
1	0.80%	–0.53%	0.53%
2	–0.09%	0.45%	0.02%
3	–0.28%	1.47%	0.07%
4	–0.63%	0.28%	0.56%
5	–1.68%	–0.23%	–1.39%

Notes: The Day 1 portfolio return is obtained by multiplying each holding (SPY, SPLB) by its respective weight in the portfolio (80%/20%) and adding the two results together: 0.80(0.008) + 0.20(–0.0053). Although Exhibit 4 shows only five days of returns, we would, of course, use all of the data at our disposal that is reasonably representative of possible future outcomes.

The historical simulation VaR extracts the portfolio return that lies at the appropriate confidence interval along the distribution. Using Excel's "=percentile(x,y)" function, we calculated the following historical simulation VaRs for our sample portfolio:

- 1% VaR (99% confidence) $2,643,196
- 5% VaR (95% confidence) $1,622,272
- 16% VaR (84% confidence) $880,221

Now, it will be interesting to compare this result with the parametric VaR estimates. Exhibit 5 shows the results side-by-side with the parameters used. The historical simulation method does not directly use these parameters but uses the data itself, and these numbers are the parameters implied by the data itself.

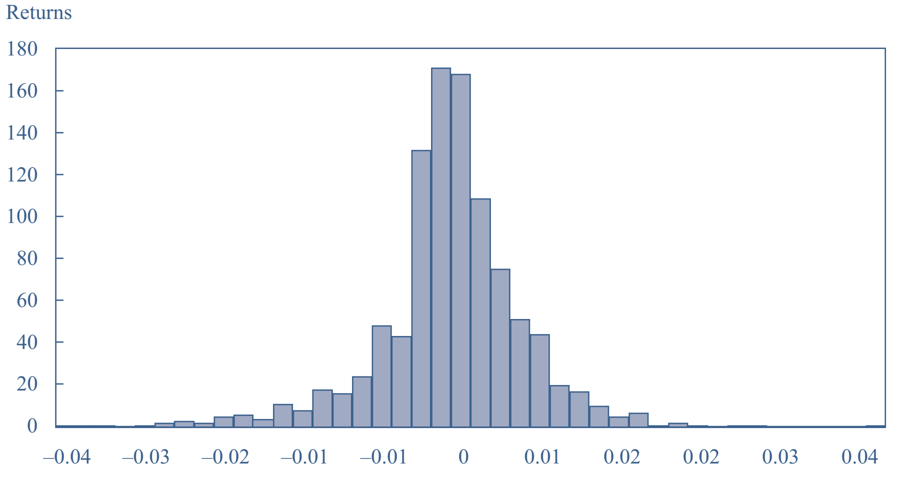
Exhibit 5 Comparison of Historical and Parametric VaR Estimates Using 1 July 2015–28 June 2019 Data

	Historical Simulation Method		Parametric Method	
1% VaR	$2,643,196		$3,476,550	
5% VaR	$1,622,272		$2,445,150	
16% VaR	$880,221		$1,459,200	
	Average Return	**Standard Deviation**	**Average Return**	**Standard Deviation**
SPY	12.51%	13.64%	10.50%	20.00%
SPLB	8.03%	7.73%	6.00%	8.50%
Correlation of SPY and SPLB	–0.061		–0.06	

The historical simulation VaRs are much smaller, and the differences stem primarily from the adjustments we made to the historical parameters. We adjusted the volatility and the average return estimates of SPY to more closely reflect the historical norms and slightly raised the volatility of SPLB. Recall, in particular, that our factor history for the S&P 500 exhibited abnormally low volatility relative to the long-run experience.

Additionally, our calculations using the historical simulation method were not constrained by the assumption of a normal distribution as was the case with the parametric method. Exhibit 6 is a histogram of the portfolio returns used in the historical simulation results, overlaid with a normal distribution.

Exhibit 6 Histogram of Historical Portfolio Returns (80% SPY and 20% SPLB) Using 1 July 2015–28 June 2019 Data



As can be seen, the resulting distribution under the historical simulation method is a departure from a normal distribution. This point again highlights the importance of understanding the underlying assumptions of any VaR model.

There is *no single right way* of estimating VaR. Each method provides an estimate of VaR and is highly sensitive to the input parameters, and similar to many estimation models, they will disagree.

Both the parametric and historical simulation methods in their most basic forms have the limitation that, as with most samples, all observations are weighted equally. The historical simulation method can adjust for this problem, however, by using a weighting methodology that gives more weight to more recent observations and less weight to more distant observations.

The primary advantage of the historical simulation method compared with the parametric method is that the historical simulation method estimates VaR based on what actually happened, so it cannot be dismissed as introducing impossible outcomes. Yet, therein also lies the primary weakness of the historical simulation method: There can be no certainty that a historical event will re-occur or that it would occur in the same manner or with the same likelihood as represented by the historical data. If one uses a relatively short historical data set, such as from January 1987 through December 1988 (a period encompassing the "Black Monday" of 19 October 1987, when stock markets around the world collapsed in a very short time), an occurrence of this magnitude might be projected to occur once every two years, surely an overstatement of its probability. Thus, the historical simulation method is best used when the distribution of returns during the lookback period are expected to be representative of the future.

The historical method is capable of handling the adjustment of one time horizon to another; that is, the information derived from daily data can be extrapolated to estimate an annual VaR, provided the distribution can be assumed to be stationary. In other words, one can convert each daily return to an annual return and then estimate the annual VaR. Although using annual data to estimate an annual VaR is always preferred, that would require a much longer lookback period.

We noted earlier that the parametric method is not well suited for options. Because the historical simulation method captures the returns that actually occurred regardless of the type of financial instrument used, it can accommodate options.

EXAMPLE 3

Historical Simulation VaR

1 Which of the following statements about the historical simulation method of estimating VaR is *most* correct?

  A A 5% historical simulation VaR is the value that is 5% to the left of the expected value.

  B A 5% historical simulation VaR is the value that is 1.65 standard deviations to the left of the expected value.

  C A 5% historical simulation VaR is the fifth percentile, meaning the point on the distribution beyond which 5% of the outcomes result in larger losses.

2 Which of the following is a limitation of the historical simulation method?

  A The past may not repeat itself.

  B There is a reliance on the normal distribution.

  C Estimates of the mean and variance could be biased.

Solution to 1:

C is correct. In the historical method, the portfolio returns are arrayed lowest to highest and the observation at the fifth percentile (95% of the outcomes are better than this outcome) is the VaR. A is not correct because it draws a point on the distribution relative to the expected value rather than using the 5% of the outcomes that are in the left-most of the distribution. B confuses the parametric and historical methods. In the parametric method, the 5% VaR lies 1.65 standard deviations below the mean.

Solution to 2:

A is correct. The historical simulation method estimates VaR based on the historical distribution of the risk factors. B is not correct; the historical simulation method does not rely on any particular distribution because it simply uses whatever distribution applied in the past. C is not correct because the historical distribution does not formally estimate the mean and variance.

5 THE MONTE CARLO SIMULATION METHOD OF VAR ESTIMATION

b compare the parametric (variance–covariance), historical simulation, and Monte Carlo simulation methods for estimating VaR;

c estimate and interpret VaR under the parametric, historical simulation, and Monte Carlo simulation methods;

Monte Carlo simulation is a method of estimating VaR in which the user develops his own assumptions about the statistical characteristics of the distribution and uses those characteristics to generate random outcomes that represent hypothetical returns to a portfolio with the specified characteristics. This method is widely used in the sciences to estimate the statistical distribution of scientific phenomena and has many applications in business and finance. For example, a corporation considering the investment of a large amount of capital in a new project with many uncertain variables could simulate the possible values of these variables and thus gain an understanding of the distribution of the possible returns from this investment. Or, complex options can often be priced by simulating outcomes of the underlying, determining the payoffs of the option, and then averaging the option payoffs and discounting that value back to the present. The reference to the famous Mediterranean casino city allegedly came from an observation made by a scientist that the method is similar to tossing dice at a casino.

Monte Carlo simulation avoids the complexity inherent in the parametric method when the portfolio has a large number of assets. (A large number of assets makes the parameters of the distribution difficult to extract.) There can be many risk factors, and the interactions among these risk factors can be too complex to specify. Moreover, Monte Carlo simulation does not need to be constrained by the assumption of normal distributions. Rather than attempt to determine the expected return and volatility of a combination of multiple statistical processes, one would simply simulate these processes, tabulate the statistical results of the simulations, and thereby gain a measure of the combined effects of these complex component processes on the overall risk.

Monte Carlo simulation requires the generation of random values of the underlying unknowns. In our example, the unknowns are the returns on the two risk factors, represented by the SPY and SPLB ETFs. We can, of course, assume that the statistical properties of the historical returns—their averages, volatilities, and correlation—are appropriate for use in a simulation, or we can modify those values to conform to what we expect to be relevant for the future. For illustrative purposes here, we will simply use the inputs we used in the parametric method.

Recall that we previously assumed for the sake of simplicity that the two securities represent the risk factors. We now decompose the portfolio holdings into these risk factors. First we simulate the returns of these two risk factors, and then we re-price our exposures to the risk factors under the range of simulated returns, recording the results much as we do in the historical simulation method. We then sort the results in order from worst to best. A 5% Monte Carlo VaR would simply be the fifth percentile of the simulated values instead of the historical values.

Yet, it is not quite that simple. We must first decide how many random values to generate. There is no industry standard. The more values we use, the more reliable our answers are but the more time-consuming the procedure becomes. In addition, we cannot just simulate values of two random variables without accounting for the correlation between the two. For example, if you spin two roulette wheels, you can assume they are independent of each other in much the same manner as are two uncorrelated assets. But most assets have at least a small degree of correlation. In our example, we used the historical correlation of about −0.06. Monte Carlo simulation must take that relationship into account.

For simplicity, this reading will not go into detail on either the mathematical techniques that can account for the correlations among risk factor returns or the specific method used to simulate outcomes given average values and volatilities for each risk factor. Both are beyond the scope of this reading.

For this example, we will use 10,000 simulated returns on SPY and SPLB drawn from a normal distribution. Of course, non-normal distributions can be used—and they commonly are in practice—but we want to keep the illustration simple to facilitate comparisons between methods. Each set of simulated returns combines to produce a sample with the expected returns and volatilities as we specified. In addition, the returns will have the pre-specified correlation of −0.06. Each pair of returns is weighted 80/20 as desired. We generate the 10,000 outcomes, sort them from worst to best, and either select the outcome at the 5th percentile for a 5% VaR, the outcome at the 1st percentile for a 1% VaR, or the outcome at the 16th percentile if we want to evaluate the impact of a one standard deviation move. Using the parameters specified in our example, the simulation returns a distribution from which we can draw the following VaR numbers:

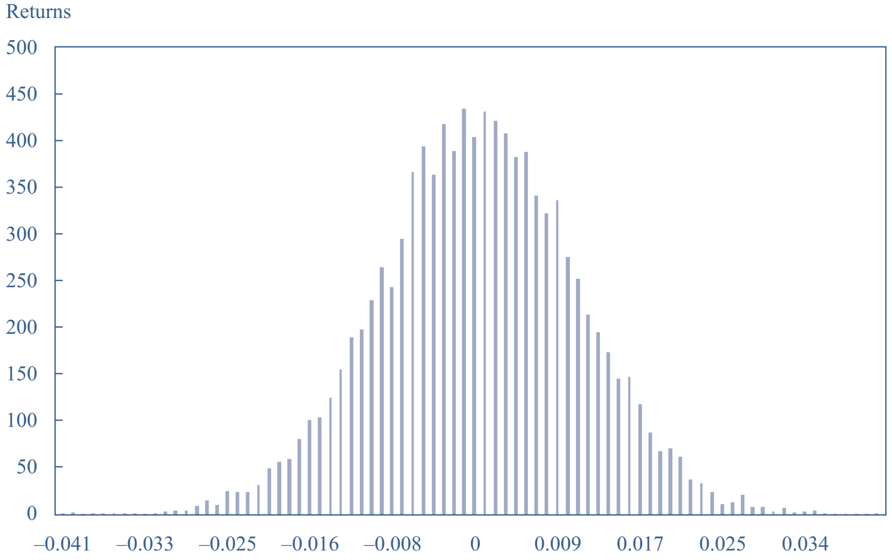
1% VaR = $3,541,035

5% VaR = $2,517,702

16% VaR = $1,524,735

Note that these results are fairly close to VaR under the parametric VaR method, where the 5% VaR was $2,445,150. The slight difference arises from the fact that Monte Carlo simulation only *samples* from a population with certain parameters while the parametric method *assumes* those parameters. A sample of a distribution will not produce statistics that match the parameters precisely except in extremely large sample sizes, much larger than the 10,000 used here. Exhibit 7 displays a histogram of the simulated returns overlaid with a bell curve representing a normal distribution. Note how the simulated returns appear more normally distributed than do the historical values, as illustrated in Exhibit 6. This is because we explicitly assumed a normal distribution when running the simulation to generate the values in our example.

Exhibit 7 Monte Carlo Simulated Returns 80/20 Portfolio of SPY and SPLB



Although we conveniently assumed a normal distribution, one of the advantages of the Monte Carlo method is that it can accommodate virtually *any* distribution. In fact, the flexibility of the Monte Carlo method to handle more complex distributions is its primary attraction. The Monte Carlo and historical simulation methods are much more capable than the parametric method of accurately incorporating the effects of option positions or bond positions with embedded options.

Similar to the historical simulation method, you can scale daily returns to annual returns and extrapolate an estimate of the annual VaR by running a Monte Carlo simulation on these annual returns.

At one time, calculating VaR using the Monte Carlo simulation method was slow, but with the speed of today's computers, it is relatively easy and fast to simulate extremely complex processes for portfolios with thousands of exposures.

EXAMPLE 4

Monte Carlo Simulation VaR

1 When will the Monte Carlo method of estimating VaR produce virtually the same results as the parametric method?

   A When the Monte Carlo method assumes a non-normal distribution.

   B When the Monte Carlo method uses the historical return and distribution parameters.

   C When the parameters and the distribution used in the parametric method are the same as those used in the Monte Carlo method and the Monte Carlo method uses a sufficiently large sample.

2 Which of the following is an advantage of the Monte Carlo method?

   A The VaR is easy to calculate with a simple formula.

   B It is flexible enough to accommodate many types of distributions.

   C The number of necessary simulations is determined by the parameters.

Solution to 1:

C is correct. The Monte Carlo method simulates outcomes using whatever distribution is specified by the user. *If* a normal distribution is used *and* a sufficiently large number of simulations are run, the parameters of the Monte Carlo sample will converge with those used in the parametric method and the overall VaR should be very close to that of the parametric method. A is incorrect because the parametric method is not well-adapted to a non-normal distribution. B is incorrect because neither the Monte Carlo method nor the parametric method focuses on historical outcomes.

Solution to 2:

B is correct. The method can handle any distribution. A is incorrect because Monte Carlo simulation is not a simple formula. C is incorrect; there is no industry-wide agreement as to the necessary number of simulations.

6 ADVANTAGES AND LIMITATIONS OF VAR AND EXTENSIONS OF VAR

d describe advantages and limitations of VaR;

e describe extensions of VaR;

The concept of VaR is solidly grounded in modern portfolio analysis. Nonetheless, the implementation of VaR, both in the estimation procedure and in the application of the concept, presents a number of advantages and limitations.

6.1 Advantages of VaR

The use of VaR as a risk measure has the following advantages:

- *Simple concept.* VaR is relatively easy to understand. Although the methodology is fairly technical, the concept itself is not very difficult. So, decision makers without technical backgrounds should be able to grasp the likelihood of possible losses that might endanger the organization. Reporting that a daily 5% VaR is, for example, €2.2 million allows the user to assess the risk in the context of the capital deployed. If a portfolio is expected to incur losses of a minimum of €2.2 million on 5% of the trading days, about once a month, this information is valuable in the context of the size of the portfolio.
- *Easily communicated concept.* VaR captures a considerable amount of information into a single number. If the recipient of the information fully understands the meaning and limitations of VaR, it can be a very significant and practical piece of information.
- *Provides a basis for risk comparison.* VaR can be useful in comparing risks across asset classes, portfolios, and trading units—giving the risk manager a better picture of which constituents are contributing the least and the most to the overall risk. As such, the risk manager can be better informed as he looks for potential hot spots in the organization. This point will be discussed further in a later section.
- *Facilitates capital allocation decisions.* The ability to compare VaR across trading units or portfolio positions provides management with a benchmark that can be used in capital allocation decisions. A proprietary trading firm,

for example, can find that its VaR in equity trading is $20 million and its VaR in fixed-income trading is $10 million. If its equity trading portfolio is not expected to take more risk than its fixed-income trading portfolio, then the equity trading activities are taking too much risk or there is too much capital allocated to equity trading. The firm should either make adjustments to realign its VaR or allocate capital in proportion to the relative risks. If a firm is looking to add a position to a portfolio or change the weights of existing portfolio positions, certain extensions of VaR allow the manager to assess the risk of these changes. This topic will be covered in more detail later.

- *Can be used for performance evaluation.* Risk-adjusted performance measurement requires that return or profit be adjusted by the level of risk taken. VaR can serve as the basis for risk adjustment. Without this adjustment, more profitable units could be perceived as more successful; however, when adjusted by VaR, a less profitable unit that poses less risk of loss may be judged more desirable.
- *Reliability can be verified.* VaR is easily capable of being verified, a process known as backtesting. For example, if the daily VaR is $5 million at 5%, we would expect that on 5% of trading days a loss of at least $5 million would be incurred. To determine whether a VaR estimate is reliable, one can determine over a historical period of time whether losses of at least $5 million were incurred on 5% of trading days, subject to reasonable statistical variation.
- *Widely accepted by regulators.* In the United States, the SEC requires that the risk of derivatives positions be disclosed either in the form of a summary table, by sensitivity analysis (a topic we cover later), or by VaR. Thus, VaRs are frequently found in annual reports of financial firms. Global banking regulators also encourage banks to use VaR. These regulations require or encourage the use of VaR, but they do not prescribe how it should be implemented, which estimation method to use, or the maximum acceptable VaR.

6.2 Limitations of VaR

Despite its many advantages, users of VaR must also understand its limitations. The primary limitations of VaR are the following:

- *Subjectivity.* In spite of the apparent scientific objectivity on which it is based, VaR is actually a rather subjective method. As we saw in the descriptions of the three methods of estimating VaR, there are many decisions to make. At the fundamental level, decisions must be made as to the desired VaR cutoff (5%, 1%, or some other cutoff); over what time horizon the VaR will be measured; and finally, which estimation method will be used. As we have seen here, for each estimation method, there are numerous other discretionary choices to make about inputs, source of data, and so on.
- *Underestimating the frequency of extreme events.* In particular, use of the normal distribution in the parametric method and sometimes in the Monte Carlo method commonly underestimates the likelihood of extreme events that occur in the left tail of the distribution. In other words, there are often more extreme adverse events, called "left-tail events," than would be expected under a normal distribution. As mentioned previously, there is no particular requirement that one use the normal distribution. The historical simulation method uses whatever distribution the data produce. We chose to illustrate the Monte Carlo method with a normal distribution, and it is virtually always used in the

parametric method. Nonetheless, the tendency to favor the normal distribution and other simple and symmetrical distributions often leads to an understatement of the frequency of left-tail events.

- *Failure to take into account liquidity*. If some assets in a portfolio are relatively illiquid, VaR could be understated, even under normal market conditions. Additionally, liquidity squeezes are frequently associated with tail events and major market downturns, thereby exacerbating the risk. Although illiquidity in times of stress is a general problem that affects virtually all of a firm's financial decisions, reliance on VaR in non-normal market conditions will lead the user to underestimate the magnitude of potential losses.
- *Sensitivity to correlation risk*. Correlation risk is the risk that during times of extreme market stress, correlations among all assets tend to rise significantly. Thus, markets that provide a reasonable degree of diversification under normal conditions tend to decline together under stressed market conditions, thereby no longer providing diversification.
- *Vulnerability to trending or volatility regimes*. A portfolio might remain under its VaR limit every day but lose an amount approaching this limit each day. Under such circumstances, the portfolio could accumulate substantial losses without technically breaching the VaR constraint. Also, during periods of low volatility, VaR will appear quite low, underestimating the losses that could occur when the environment returns to a normal level of volatility.
- *Misunderstanding the meaning of VaR*. VaR is not a worst-case scenario. Losses can and will exceed VaR.
- *Oversimplification*. Although we noted that VaR is an easily communicated concept, it can also oversimplify the picture. And although VaR does indeed consolidate a considerable amount of information into a single number, that number should be interpreted with caution and an awareness of the other limitations as well as supported by additional risk measures.
- *Disregard of right-tail events*. VaR focuses so heavily on the left tail (the losses) that the right tail (potential gains) are often ignored. By examining both tails of the distribution, the user can get a better appreciation of the overall risk–reward trade-off, which is often missed by concentrating only on VaR.

These limitations are not unique to VaR; they apply equally to any technique or measure used to quantify the expected rewards and risks of investing.

EXAMPLE 5

Advantages and Limitations of VaR

1 Which of the following is **not** an advantage of VaR?

  A It is a simple concept to communicate.

  B There is widespread agreement on how to calculate it.

  C It can be used to compare risk across portfolios or trading units.

2 Which of the following is a limitation of VaR?

  A It requires the use of the normal distribution.

  B The maximum VaR is prescribed by federal securities regulators.

  C It focuses exclusively on potential losses, without considering potential gains.

Solution to 1:

B is correct. There is no consensus on how to calculate VaR. A and C are both advantages of VaR, as we noted that VaR is fairly simple to communicate and it can show the contribution of each unit to the overall VaR.

Solution to 2:

C is correct. VaR deals exclusively with left-tail or adverse events. A is wrong because although parametric VaR does generally use the normal distribution, the historical simulation method uses whatever distribution occurred in the past and Monte Carlo simulation uses whatever distribution the user chooses. B is incorrect because regulators do not specify maximum VaRs, although they may encourage and require that the measure be used.

6.3 Extensions of VaR

Clearly no single risk model can answer all of the relevant questions a risk manager may have. As a result, VaR has laid a foundation for a number of variations, each of which provides additional information.

As discussed previously, VaR is a minimum loss and is typically expressed as the minimum loss that can be expected to occur 5% of the time. An important and related measure can determine the average loss that would be incurred if the VaR cutoff is exceeded. This measure is sometimes referred to as the **conditional VaR (CVaR)**, although it is not technically a VaR measure. It is the average loss conditional on exceeding the VaR cutoff. So, VaR answers the question, “What is the minimum loss I can expect at a certain confidence?” And CVaR answers the question, “How much can I expect to lose if VaR is exceeded?” CVaR is also sometimes referred to as the **expected tail loss** or **expected shortfall**. CVaR is best derived using the historical simulation and Monte Carlo methods, in which one can observe all of the returns throughout the distribution and calculate the average of the losses beyond the VaR cutoff. The parametric method uses a continuous distribution, so obtaining the average loss beyond the VaR cutoff would require a level of mathematics beyond the scope of this reading.

Using our earlier example, in the historical simulation method, our sample of 500 historical returns was sorted from lowest to highest and the 5% VaR was $1,622,272. With 1,006 returns in the sample, 50 observations (5% of 1,006) lie below the VaR estimate. The average of these losses is $2,668,389. Thus, when the VaR is exceeded, we would expect an average loss of about $2.7 million.

For the Monte Carlo method, we generated 10,000 random values and obtained a 5% VaR of $2,517,705. Given 10,000 random values, 500 observations are in the lowest 5% of the VaR distribution. The CVaR using the Monte Carlo method would be the average of the 500 lowest values, which is $4,397,756.

Note that once again, the CVaR derived using the historical simulation method is lower than the CVaR derived using the Monte Carlo method. As explained earlier, this result can largely be attributed to the lower volatility of the S&P 500 component in the historical data series.

Beyond assessing tail loss, a risk manager often wants to know how the portfolio VaR will change if a position size is changed relative to the remaining positions. This effect can be captured by a concept called **incremental VaR (IVaR)**. Using our example, suppose the portfolio manager is contemplating increasing the risk by increasing the investment in SPY to 90% of the portfolio. We recalculate the VaR under the proposed allocation, and the incremental VaR is the difference between the “before” and “after” VaR. As an example, using the parametric method, the VaR would be expected to increase from $2,445,150 to $2,752,500; thus, the IVaR for the 5% case would be

$307,350. Or, the portfolio manager might wish to add a new asset, thereby reducing the exposure to the existing assets. The risk manager would calculate the VaR under the assumption that the change is made, and then the difference between the new VaR and the old VaR is the IVaR. This measure is useful because it reflects the effect of an anticipated change on the VaR. The risk manager could find that the new VaR will be unacceptably high or that it has possibly even decreased.

A related concept is called **marginal VaR (MVaR)**. It is conceptually similar to incremental VaR in that it reflects the effect of an anticipated change in the portfolio, but it uses formulas derived from calculus to reflect the effect of a very small change in the position. Some people interpret MVaR as a change in the VaR for a $1 or 1% change in the position, although that is not strictly correct. Nonetheless, this interpretation is a reasonable approximation of the concept behind marginal VaR, which is to reflect the impact of a small change. In a diversified portfolio, marginal VaR may be used to determine the contribution of each asset to the overall VaR; the marginal VaRs for all positions may be proportionately weighted to sum to the total VaR.

Both incremental and marginal VaR address the question of what impact a change in the portfolio holdings might have on the total VaR of the portfolio. Both take into account the potential diversifying effects of various positions or subportfolios, and thus they both can be useful in evaluating the potential effect of a trade before the trade is done.

Another related measure is ***ex ante* tracking error**, also known as **relative VaR**, which is a measure of the degree to which the performance of a given investment portfolio might deviate from its benchmark. It is computed using any of the standard VaR models, described earlier, but the portfolio to which VaR is applied contains the portfolio's holdings *minus* the holdings in the specified benchmark. In other words, the benchmark's holdings, weighted in proportion to the value of the subject portfolio, are entered into the VaR modeling process as short positions. VaR for this measure is typically expressed as a one standard deviation annualized measure. If the portfolio is a perfect match to the benchmark, *ex ante* tracking error will be at or near zero. The more the portfolio differs from the benchmark, the larger the *ex ante* tracking error will be.

EXAMPLE 6

Extensions of VaR

1 Conditional VaR measures the:

A VaR over all possible losses.

B VaR under normal market conditions.

C average loss, given that VaR is exceeded.

2 Which of the following correctly identifies incremental VaR?

A The change in VaR from increasing a position in an asset.

B The increase in VaR that might occur during extremely volatile markets.

C The difference between the asset with the highest VaR and the asset with the second highest VaR.

3 Which of the following statements is correct about marginal VaR?

A The marginal VaR is the same as the incremental VaR.

B The marginal VaR is the VaR required to meet margin calls.

C Marginal VaR estimates the change in VaR for a small change in a given portfolio holding.

Solution to 1:

C is correct. Conditional VaR is the average loss conditional on exceeding the VaR. A is not correct because CVaR is not concerned with losses that do not exceed the VaR threshold, and B is incorrect because VaR does not distinguish between normal and non-normal markets.

Solution to 2:

A correctly defines incremental VaR. Incremental VaR is the change in VaR from increasing a position in an asset, not a change in VaR from an increase in volatility. B is not correct because incremental volatility reflects the results of intentional changes in exposure, not uncontrollable market volatility. C is not correct because incremental VaR is not the difference in the VaRs of the assets with the greatest and second greatest VaRs.

Solution to 3:

C is correct. In A, marginal VaR is a similar concept to incremental VaR in that they both deal with the effect of changes in VaR, but they are not the same concept. B is incorrect because marginal VaR has nothing to do with margin calls.

7 OTHER KEY RISK MEASURES - SENSITIVITY RISK MEASURES; SENSITIVITY RISK MEASURES

f describe sensitivity risk measures and scenario risk measures and compare these measures to VaR;

g demonstrate how equity, fixed-income, and options exposure measures may be used in measuring and managing market risk and volatility risk;

h describe the use of sensitivity risk measures and scenario risk measures;

Just as no single measure of a person's health gives a complete picture of that person's physical condition, no single risk measure gives a full picture of a portfolio's risk profile. As we saw, although VaR has many advantages, it also has many limitations. Therefore, good risk managers will use a comprehensive set of risk tools. In this section, we will look at two additional classes of risk measures: those based on sensitivity analysis and those based on the use of hypothetical or historical scenarios. The former enable us to estimate how our estimated gains and losses change with changes in the underlying risk factors, whereas the latter are based on situations involving considerable market stress from which we estimate how our portfolio will perform.

7.1 Sensitivity Risk Measures

Equity, fixed-income, and options positions can be characterized by a number of exposure measures that reflect the sensitivities of these positions to movements in underlying risk factors. Sensitivity measures examine how performance responds to a single change in an underlying risk factor. Understanding and measuring how portfolio positions respond to the underlying sources of risk are primary objectives in managing risk.

7.1.1 *Equity Exposure Measures*

The primary equity exposure measure is the beta. In a simple world, a single market factor drives equity returns. The return on a stock is given by the familiar capital asset pricing model (CAPM):

$$E(R_i) = R_F + \beta_i[E(R_M) - R_F],$$

where $E(R_i)$ is the expected return on the asset or portfolio i, R_F is the risk-free rate, $E(R_m)$ is the expected return on the market portfolio, and β_i is the beta, which is the risk measure. The expression $E(R_m) - R_F$ is the equity risk premium, which is the return investors demand for investing in equities rather than risk-free instruments. It should be apparent from this often-used equation that beta measures the sensitivity of the security's expected return to the equity risk premium. The beta is defined as the covariance of the asset return with the market return divided by the variance of the market return. The broad market beta, which is an average of all individual betas, is 1.0. Assets with betas more (less) than 1 are considered more (less) volatile than the market as a whole. The CAPM has a number of extensions, including multifactor models, and risk measures derived from those models can also provide more nuanced information on equity risk exposures.

7.1.2 *Fixed-Income Exposure Measures*

The primary sensitivity exposure measures for fixed-income investments are duration and convexity. (Note that credit, a major factor driving non-government fixed-income markets, is covered elsewhere.) **Duration** is sometimes described as the weighted-average time to maturity of a bond, in which the bond is treated as partially maturing on each coupon payment date. Duration is a sensitivity measure. Under the assumption that all interest rates that affect a bond change by the same percentage, the duration is a measure of the sensitivity of the bond price to the interest rate change that characterizes all rates. This single rate can be viewed as the bond's yield, y. Given a bond priced at B and yield change of Δy, the rate of return or percentage price change for the bond is approximately given as follows:

$$\frac{\Delta B}{B} \approx -D\frac{\Delta y}{1+y}$$

where D is the duration.(The ≈ sign stands for the phrase "approximately equal" and reflects the fact that the relationship is not exact.) In this expression, it is easy to see that duration does reflect the sensitivity of a bond's price to its yield, although under the restrictive assumption of a single change to all rates. The assumption of a single change to all rates may seem fairly restrictive, but ultimately the assumption is encapsulated by assuming that a single discount rate, the yield, drives the bond price. Duration is considered to be a fairly good sensitivity measure. As previously mentioned, duration is a time measure, the weighted-average maturity of a bond, in which the bond is viewed as maturing progressively as it makes its coupon payments.

The relationship shown here is approximate. The formula is derived under the assumption that the yield change is infinitesimally small, and duration fails to accurately capture bond price movements when yield changes are relatively large. Thus, in the above expression, Δy is for small yield changes. It is not possible, however, to say how small a yield change must be before it is small enough for the expression to hold true. In addition, the expression holds only at any instant in time and only for that instant. Over longer periods, the relationship will be less accurate because of the passage of time and because Δy is likely to be larger. To accommodate longer periods of time and larger yield changes, we can incorporate a second factor called **convexity**, which is denoted C. Convexity describes the sensitivity of a bond's duration to changes in interest rates. Adding convexity to the expression, we obtain the following formula:

$$\frac{\Delta B}{B} \approx -D\frac{\Delta y}{1+y} + \frac{1}{2}C\frac{\Delta y^2}{(1+y)^2}$$

Convexity can play an important role as a risk measure for large yield changes and long holding periods.

Duration and convexity are essential tools in fixed-income risk management. They allow the risk manager to assess the potential losses to a fixed-income portfolio or position under a given change in interest rates.

7.1.3 *Options Risk Measures*

Derivatives have their own unique exposure measures. Because forwards, futures, and swaps have payoffs that are linear in relation to their underlying, they can often be evaluated using the same exposure measures as their underlying. Options, however, have non-linear payoffs, which result in them having their own family of exposure measures that incorporate this non-linear behavior.

Although options can be very risky instruments in and of themselves, they are a critical tool for effective risk management and are often used to create an exposure to offset an existing risk in the portfolio. The relative riskiness of an option arises from the high degree of leverage embedded in most options. An additional and very important risk can also arise from the sensitivity of an option to the volatility of the underlying security. We will expand on these points in the next few paragraphs.

The most fundamental risk of an option is its sensitivity to the price of the underlying. This sensitivity is called the option's **delta**. Although delta is derived by using mathematics beyond the scope of this reading, we can provide a simple and reasonably effective definition as follows:

$$\Delta\ (\text{delta}) \approx \frac{\text{Change in value of option}}{\text{Change in value of underlying}}$$

Call option deltas range from a value of 0 to a value of 1, whereas put option deltas range from a value of 0 to a value of –1. A value of 0 means that the option value does not change when the value of the underlying changes, a condition that is never absolutely true but can be roughly true for a very deep out-of-the-money option. A call delta of 1 means that the price of the call option changes in unison with the underlying, a condition that is also never absolutely true but is *approximately* true for very deep in-the-money calls. A put delta of –1 means that the price of the put option changes in unison with the underlying but in the opposite direction, a condition that is also never absolutely true but is *approximately* true for very deep in-the-money puts. As expiration approaches, an in-the-money call (put) delta approaches 1 (–1) and an out-of-the-money call (put) delta approaches 0.

Delta can be used to approximate the new price of an option as the underlying changes. For a call option, we can use the following formula:

$$c + \Delta c \approx c + \Delta_c \Delta S.$$

Here, c is the original price of the option and Δc is the change in the price. We approximate the change in the price as the product of the call's delta, Δ_c, and the change in the value of the underlying, ΔS. The same relationship would hold for puts, simply changing the c's to p's.

The delta of an option is somewhat analogous to the duration of a fixed-income security. It is a first-order effect, reflecting the direct change in the value of the option or fixed-income security when the underlying price or yield, respectively, changes. Just as duration captures the effect of only small changes in the yield over a short period of time, delta captures the effect of only small changes in the value of the underlying security over a short period of time. Similar to duration, which has the second-order effect of convexity, we can add a second-order effect for options called **gamma**. Gamma is a measure of how sensitive an option's delta is to a change in the underlying. It is a second-order effect in that it is measuring the sensitivity of the first-order effect, delta. Gamma can be interpreted in several ways. The delta reflects the direct change in the value of the underlying position, whereas gamma reflects the indirect change (i.e., the change in the change). Technically, it reflects the change in the delta, as indicated by the following:

$$\Gamma\ (\text{gamma}) \approx \frac{\text{Change in delta}}{\text{Change in value of underlying}}$$

As with convexity, gamma itself is not simple to interpret. For example, a call option might have a delta of 0.6 and a gamma of 0.02. It is not easy to determine whether the gamma is large or small. Using the equation just given, if the value of the underlying increases by 0.10 and the gamma is 0.02, then the delta would increase by 0.002 (0.10 × 0.02), from 0.6 to 0.602. Gammas get larger as the option approaches at-the-money, and they are large when options approach expiration, unless the option is deeply in or out of the money. Gamma reflects the uncertainty of whether the option will expire in or out of the money. When an option is close to expiration and roughly at the money, a small change in the price of the underlying will determine whether the option expires worthless or in the money. The uncertainty associated with this win-or-lose situation over a very short time frame leads to a large gamma.

Using delta and gamma, the new call price is

$$c + \Delta c \approx c + \Delta_c \Delta S + \frac{1}{2}\Gamma_c (\Delta S)^2$$

where Γ_c is the gamma of the call. This equation is similar to the corresponding expression that relates yield changes to bond price changes through duration and convexity. Indeed, as we said, gamma is a second-order effect, like convexity.

A third important sensitivity measure for options is **vega**, and it reflects the effect of volatility. Vega is a first-order effect reflecting the relationship between the option price and the volatility of the underlying. Vega is expressed by the following relationship:

$$\text{Vega} \approx \frac{\text{Change in value of option}}{\text{Change in volatility of underlying}}$$

Most options are very sensitive to the volatility of the underlying security. The effect of changing volatility can have a material impact on the value of the option, even when the value of the underlying is not changing.

Using delta, gamma, and vega, the new value of an option given an old value, a change in the value of the underlying, and a change in the volatility can be estimated as follows:

$$c + \Delta c \approx c + \Delta_c \Delta S + \frac{1}{2}\Gamma_c (\Delta S)^2 + \text{vega}(\Delta\sigma)$$

where $\Delta\sigma$ is the change in volatility.

The expression represents a composite sensitivity relationship for options. It reflects the expected response of an option value to changes in the value and volatility of the underlying, the two primary factors that change in an unpredictable manner and influence the option value. For portfolios that contain options, understanding these relationships and using them to assess the portfolio's response to market movements are essential elements of effective risk management.

These option measures are applicable not only to options but also to portfolios that contain options. For example, the delta of a portfolio consisting of a long position in an S&P 500 ETF and a short position in a call option on the ETF has a delta that is determined by both the ETF and the option. The ETF has a delta of 1; it changes one-for-one with the S&P 500. The option delta, as noted, has a delta between 0 and 1, though technically 0 and –1 because the option position is short. The ETF has no gamma or vega, so the portfolio gamma and vega are determined by the option. The overall deltas, gammas, and vegas are sums of the deltas, gammas, and vegas of the component positions, taking into account the relative amounts of money invested in each position. Risk managers need to know the overall deltas, gammas, vegas, durations, convexities, and betas to get a comprehensive picture of the sensitivity of the entire portfolio to the prices and volatilities of the underlying.

EXAMPLE 7

Sensitivity Risk Measures

1 Which of the following *most* accurately characterizes duration and convexity?

- A Sensitivity of bond prices to interest rates
- B First- and second-order effects of yield changes on bond prices
- C Weighted-average time to maturity based on the coupon payments and principal

2 Which of the following statements about the delta of a call option is **not** correct?

- A It ranges between 0 and 1.
- B It precisely captures the change in the call value when the underlying changes.
- C It approaches 1 for an in-the-money option and 0 for an out-of-the-money option.

3 Which of the following statements about gamma and vega are correct?

- A Gamma is a second-order effect, and vega is a first-order effect.
- B Gamma is the effect of volatility, and vega is the effect of changes in volatility.
- C Gamma is a second-order effect arising from changes in the sensitivity of volatility to the underlying price.

Solution to 1:

B is correct. Duration is the first-order effect and convexity the second-order effect of a change in interest rates on the value of a bond. A and C are correct with respect to duration, but not for convexity.

Solution to 2:

B is correct. A and C correctly characterize delta, whereas B states that delta is precise, which is incorrect because it gives an approximate relationship.

Solution to 3:

A is correct. B is not correct because gamma does not capture the effect of volatility. Vega is the effect of volatility, but it relates to the level and not the change in volatility. C is incorrect because although gamma is a second-order effect on the option value, it is not related to the sensitivity of volatility to the underlying price.

8 SCENARIO RISK MEASURES

f describe sensitivity risk measures and scenario risk measures and compare these measures to VaR;

h describe the use of sensitivity risk measures and scenario risk measures;

A scenario risk measure estimates the portfolio return that would result from a hypothetical change in markets (a hypothetical scenario) or a repeat of a historical event (a historical scenario). As an example, the risk manager might want to understand how her current portfolio would perform if an event, such as the Black Monday of October 1987, were to reoccur. The factor movements that characterized the historical event would be applied to the factor exposures of the current portfolio. Alternatively, the risk manager may develop a hypothetical scenario to describe a market event that has not occurred in the past but which he or she believes has some probability of occurring in the future. The two elements of scenario risk measures that set them apart from sensitivity risk measures are (1) the use of multiple factor movements used in the scenario measures versus the single factor movements typically used in risk sensitivity measures and (2) the typically larger size of the factor movement used in the scenario measures. Scenario risk measures are related to VaR in that they focus on extreme outcomes, but they are not bound by either recent historical events or assumptions about parameters or probability distributions. **Stress tests**, which apply extreme negative stress to a particular portfolio exposure, are closely related to scenario risk measures. Scenario analysis is an open-ended exercise that could look at positive or negative events, although its most common application is to assess the negative outcomes. Stress tests intentionally focus on extreme negative events to assess the impact of such an event on the portfolio.

The two types of scenario risk measures—historical scenarios and hypothetical scenarios—are discussed in the following sections.

8.1 Historical Scenarios

Historical scenarios are scenarios that measure the portfolio return that would result from a repeat of a particular period of financial market history. Historical scenarios used in risk management include such events as the currency crisis of 1997–1998, the market dislocation surrounding the failure of Long-Term Capital Management, the market rout of October 1987, the bursting of the technology bubble in 2001, and the financial crisis of 2008–2009. In order to create a historical scenario, the current set of portfolio holdings is placed into the appropriate valuation models.

Equity positions can often be modeled using their price histories as proxies for their expected behavior, although some practitioners model equities using factor analysis. Valuation models are needed for fixed-income and derivatives products because they have a maturity or an expiration feature that must be accommodated when modeling the portfolio. Historical prices for the fixed-income and derivatives positions currently held in the portfolio may not exist, as in the case of a bond that was issued after the historical period being modeled. Even when historical prices for specific instruments do exist, they may not be relevant to the current characteristics of the instrument. Take the case of a 5-year historical price series for a 10-year bond with 1 year remaining to maturity; the historical price series reflects the price volatility of what used to be a longer bond (e.g., five years ago, the bond had six years remaining to maturity; three years ago, the bond had four years remaining to maturity). The volatility of the bond when it had six years remaining to maturity would be higher than it is today, with only one year remaining to maturity. Using its historical price history would mischaracterize the risk of the current portfolio holding. For this reason, the historical yields, spreads, implied volatilities, prices of the underlying assets in derivatives contracts, and the other input parameters that drive the pricing of these instruments are more important in explaining the risks of these instruments than the price history of the instrument itself.

Some examples may help to show how fixed-income or derivatives valuation models are used in a historical scenario. In the case of a convertible bond, the bond's terms and conditions (e.g., coupon, conversion ratio, maturity) are entered into a convertible bond pricing model. In the case of standard bonds, the terms and conditions of these instruments (e.g., coupon, call features, put features, any amortization or sinking fund features, maturity) are entered into fixed-income pricing models. These modeled fixed-income or derivatives holdings, together with the equity holdings, are then re-priced under the conditions that prevailed during the "scenario period"—a given set of dates in the past. Changes in interest rates, credit spreads, implied volatility levels, and any asset underlying a derivatives product, as well as the historical price changes in the equity portfolio, would all be reflected in the re-priced portfolio. The value of each position is recorded before and after these changes in order to arrive at the gain or loss that would occur under the chosen scenario. Historical scenario events are specifically chosen to represent extreme market dislocations and often exhibit abnormally high correlations among asset classes. It is most common to run the scenario or stress test as if the total price action movement across the period occurs instantaneously, before any rebalancing or management action is possible. The output of the scenario can include

- the total return of the portfolio;
- for long-only asset managers, the total return of the portfolio relative to its benchmark;
- for pensions, insurers, and others whose liabilities are not already incorporated into the portfolio, the total return of the portfolio relative to the change in liabilities under the scenario; and
- any collateral requirements and other cash needs that will be driven by the changes specified in the scenario.

One variation of the historical scenario approach includes running the scenario over multiple days and incorporating actions that the manager might be expected to take during the period. Instead of assuming the shock is a single instant event, this approach assumes it takes place over a number of days and that on each day the portfolio manager can take such actions as selling assets or rebalancing hedges.

Many risk managers are skeptical of this approach because it produces smaller potential loss measures (by design) and does not answer important questions that have been relevant in real crises, such as, "What if the severe price action happens so quickly that the portfolio manager cannot take remedial actions?" Generally, risk managers prefer that a stress testing exercise be tailored to the *initial outcome of a large shock*, to ensure that the event is survivable by a portfolio that uses leverage, and that there will be no unacceptable counterparty exposures or portfolio concentrations before action can be taken to improve the situation. This method also helps to simulate the possibility that liquidity may be unavailable.

Risk managers seeking to measure the impact of a historical scenario need to ensure all relevant risk factors are included. For instance, foreign equities will need to be decomposed into foreign exchange exposure and equity exposure in the analysis. Stress tests typically take the explicit currency approach, which measures the currency exposure of each foreign equity. Alternatively, the risk manager may use an approach that incorporates implicit currency risks, such as companies that may be registered in one country but have earnings flowing in from other countries, and may hedge some of those revenues back to their base currency.

When the historical simulation fully revalues securities under rate and price changes that occurred during the scenario period, the results should be highly accurate. Sometimes, however, scenarios are applied to risk sensitivities rather than the securities themselves. This approach is a simpler form of analysis, but it should not be used for options or option-embedded securities. Although it may be tempting to use delta and gamma or duration and convexity to estimate the impact of a scenario on options or option-embedded securities, these measures are not suited for handling the kinds of extreme movements analyzed in scenario analysis. Although gamma and convexity are second-order adjustments that work with delta and duration to estimate extreme movements, they are inadequate for scenario analysis.

Even in simpler fixed-income cases in which no options are present, care needs to be taken to ensure the analysis does not oversimplify. Duration sensitivities can be used as the inputs to a scenario analysis for straightforward fixed-income instruments, but these sensitivities need to be mapped to the most relevant sectors, credit curves, and yield curve segments before beginning the analysis. If assets are mapped too broadly, the analysis will miss the important differences that could drive the most meaningful outcomes in a given scenario.

It is also important to pay careful attention to how securities or markets that did not yet exist at the time of the scenario are modeled. If, for instance, an analyst is measuring a current portfolio's sensitivity to a recurrence of the 1987 US stock market crash, the analyst needs to determine how to treat stocks in the portfolio that had an initial public offering after 1987. They may need to be mapped to a relevant index or to a similar company or be decomposed into the relevant statistical factors (such as growth, value, volatility, or momentum) by using a factor model before beginning the analysis. Similarly, because credit default swaps did not come into widespread use until 2002, historical scenarios for dates preceding this time would need to be adapted to appropriately reflect the impact of a repeat of that scenario on these new securities.

8.2 Hypothetical Scenarios

Scenarios have a number of benefits. They can reflect the impact of extreme market movements, and they make no specific assumptions regarding normality or correlation. Historical scenarios have the extra benefit of being uncontroversial; no one can claim it is impossible for such events to occur, because they did. One problem with scenario analysis, however, lies in ascribing the probability of a given scenario. Most would agree that it is improbable to assume that the exact historical scenario specified will actually occur in precisely the same way in the future. Another potential problem is that, because it has happened (particularly when it has happened recently), risk managers or portfolio managers are inclined to take precautions that make their portfolios safer for a replay of that historical crisis—and, in the process, make their portfolios more vulnerable to a crisis that has not yet happened.

For that reason, risk managers also use hypothetical scenarios—extreme movements and co-movements in different markets that have not necessarily previously occurred. The scenarios used are somewhat difficult to believe, and it is difficult to assess their probability. Still, they represent the only real method to assess portfolio outcomes under market movements that might be imagined but that have not yet been experienced.

To design an effective hypothetical scenario, it is necessary to identify the portfolio's most significant exposures. Targeting these material exposures and assessing their behavior in various environments is a process called **reverse stress testing**. The risk manager is seeking answers to such questions as the following: What are the top 10 exposures or risk drivers in my portfolio? What would make them risky? What are the top 10 benchmark-relative exposures? Under what scenario would hedges not hedge? Under what scenario would my securities lending activity, ordinarily thought to be riskless, be risky? The ideal use of hypothetical scenarios is, then, not to model every possible future state of every market variable, but rather to target those that are highly significant to the portfolio in order to assess, and potentially address, vulnerabilities.

Reverse stress testing is particularly helpful in estimating potential losses if more than one important exposure is affected in a market crisis, as often happens when participants "crowd" into the same exposures. Sometimes, apparently unrelated markets experience stress at the same time.

The risk manager might also choose to design a hypothetical geopolitical event, estimating its potential effect on markets and the resulting impact on the portfolio. To develop these scenarios, individuals with varying areas of expertise posit an event—such as an earthquake in Country Y, or Country X invades Country Z, or the banking system implodes in Region A. The group conducting the analysis identifies which markets are most likely to be affected as well as any identifiable secondary effects. The next step is to establish a potential range of movement for the affected markets. The final scenario is intended to meet the standard of "rare, but not impossible." The exercise is unlikely to be truly accurate in the face of the real event, but it will often help to identify unexpected portfolio vulnerabilities and outcomes and to think through counterparty credit and operational considerations that could exacerbate or accelerate the scenario.

Hypothetical scenarios are particularly beneficial in being able to stress correlation parameters. The scenario is not constrained to assume that assets will co-move as they have done in the past, which can help identify dangers that other forms of risk analysis may miss. Scenarios can be designed to highlight that correlations often increase in times of stress. This is often achieved by subjecting markets that typically have little or no correlation with one another to the same or similar movements, thereby simulating a temporarily higher correlation. Scenarios can also be devised to pinpoint times when hedging might work poorly—when assets, such as a bond and the credit default swap used to hedge it, that normally have a high correlation might temporarily decouple and move by different percentages or even in different directions. This often occurs when markets experience a "flight to quality"; the swap rate may move down as a result of their relative credit strength, whereas the bond yield might increase given its perceived credit risk.

Once a risk manager has completed a scenario analysis, common questions may be, "What do you do with a scenario analysis? What are the action steps?" If the portfolios are within all other rules and guidelines—their exposures have been kept within desired limits and their VaR or *ex ante* tracking error is within the desired range—scenario analysis provides one final opportunity to assess the potential for negative surprises during a given stress event. The action steps might be to trim back positions that are otherwise within all limits and that appear to present comfortable risk exposures under the current environment but would perform unacceptably during a plausible stress environment. In the case of asset management, where clients have elected to be in a given asset class and the asset manager is constrained by that investment mandate, action steps may include adjusting benchmark-relative risk, disclosing to clients the manager's concerns regarding the risks in the portfolio, or changing counterparty or operational procedures to avoid an unwanted event.

But a caution is in order: A portfolio that has no sensitivity to any stress event is unlikely to earn more than the risk-free rate, or in the case of long-only asset managers, outperform the benchmark index. Stress tests and scenarios analyses are best used in the effort to *understand* a portfolio's risk exposures, not to eliminate them. Effective risk management sets a tolerance range for a stress test or scenario that reflects a higher loss possibility than the investment manager would normally find acceptable. Scenarios should be periodically run again, and action should be taken only if the portfolio exceeds this relatively high tolerance level. It is also important to continually evaluate new threats and new market developments and to periodically refresh the set of scenarios, removing scenarios that are no longer meaningful for the portfolio.

Note also that scenario risk measures and stress tests are best used as the final screen in a series of position constraints that include position size limits, exposure limits, and VaR or *ex ante* tracking error limits. They do not serve well as the initial or primary screen, for reasons that will be discussed shortly.

Parties that use leverage, such as banks and hedge funds, are more likely to use single-factor stress tests rather than multifactor scenario analyses. The focus on a single factor helps in assessing whether a given exposure is likely to impair their capital under a given stress movement; these are pass/fail tests. If capital falls below an acceptable level, it could set off a chain reaction of margin calls, withdrawal of financing, and other actions that threaten the viability of the business.

EXAMPLE 8

Scenario Analysis

1 Which of the following is an example of a reverse stress test?
   - A Identify the top 10 exposures in the portfolio, and then generate a hypothetical stress that could adversely affect all 10 simultaneously.
   - B Find the worst single day's performance that could have occurred for the current portfolio had it been held throughout the past five years.
   - C Find the returns that occurred in all risk factors in the 2008 global financial crisis, reverse the sign on these, and apply them to today's portfolio.

2 Which kind of market participant is *least likely* to use scenario analysis as a pass/fail stress test?
   - A Bank
   - B Long-only asset manager
   - C Hedge fund using leverage

3 What is the *most* accurate approach to scenario analysis for a portfolio that uses options?
   - A Apply the scenario to option delta.
   - B Apply the scenario to option delta + gamma.
   - C Fully reprice the options using the market returns specified under the scenario.

Solution to 1:

A is correct. B is not a reverse stress test because reverse stress tests focus more narrowly on trouble spots for a specific portfolio. C would illustrate how the portfolio would have performed in an extremely strong market, quite unlike what occurred in 2008.

Solution to 2:

B is correct. Long-only asset managers do not typically use leverage and are thus less likely to become insolvent, making a pass/fail test for solvency less relevant to them. A and C are not correct because parties that use leverage, such as hedge funds and banks, are likely to use stress tests to determine what market movements could impair their capital and lead to insolvency.

Solution to 3:

C is correct. Both A and B risk misestimating the actual results of the scenario because both delta and gamma estimate how an option's value might change for a small move in the underlying asset, not the large movements typically used in a scenario analysis.

SENSITIVITY AND SCENARIO RISK MEASURES AND VAR

9

f describe sensitivity risk measures and scenario risk measures and compare these measures to VaR;

i describe advantages and limitations of sensitivity risk measures and scenario risk measures;

Although both VaR and sensitivity risk measures deal with related concepts, they have their own distinctions. VaR is a measure of losses and the probability of large losses. Sensitivity risk measures capture changes in the value of an asset in response to a change in something else, such as a market index, an interest rate, or an exchange rate; they do not, however, tell us anything about the probability of a given change in value occurring. For example, we could use duration to measure the change in a bond price for an instantaneous 1 bp change in the yield, but duration does not tell us anything about the likelihood of such a change occurring. Similar statements could be made about equities and the various option measures: Betas and deltas do not tell us how likely a change might be in the underlying risk factors, but given a change, they tell us how responsive the asset or derivative would be.

VaR gives us a broader picture of the risk in the sense that it accounts for the probability of losses of certain amounts. In this sense, it incorporates what we know about the probability of movements in the risk factors. Nonetheless, these sensitivity measures are still very useful in that they allow us to take a much more detailed look at the relationships driving the risk. It is one thing to say that a VaR is $2 million for one day at 5%. We know what that means. But it is equally important to understand what is driving the risk. Is it coming from high beta stocks, high duration bonds, or high delta options? If we find our VaR unacceptable, we have to know where to look to modify it. If we simply use VaR by itself, we will blindly rely on a single number without understanding what factors are driving the number.

VaR has much in common with scenario risk measures in that both types of measures estimate potential loss. VaR tends to do so using a model for which input parameters are created based on market returns from a particular time in history. Thus, the VaR estimate is vulnerable if correlation relationships and market volatility during the period in question are not representative of the conditions the portfolio may face in the future. VaR does, however, allow a disciplined method for stressing all factors in the portfolio. Scenario analysis allows either the risk assessment to be fully hypothetical or to be linked to a different and more extreme period of history, helping reduce some of the biases imposed by the VaR model. But there is no guarantee that the scenario chosen will be the "right" one to estimate risk for future markets. Moreover, it is particularly difficult to stress all possible risk factors in a hypothetical scenario in a way that does not embed biases similar to those that occur in VaR modeling.

Each of these measures—sensitivity risk measures, scenario risk measures, and VaR—has distinct limitations and distinct benefits. They are best used in combination because no one measure has the answer, but all provide valuable information that can help risk managers understand the portfolio and avoid unwanted outcomes and surprises.

9.1 Advantages and Limitations of Sensitivity Risk Measures and Scenario Risk Measures

Before portfolios began using risk measures based on modern portfolio theory, the very first risk measure was "position size"—the value invested in a given type of asset. Position size is a very effective risk measure for homogeneous, long-only portfolios, particularly for those familiar with the homogenous asset class in question; an experienced person can assess what the loss potential of such a portfolio is just by knowing its size. But position size is less useful for assessing interest rate risk, even less useful for summarizing the risk of a multi-asset class portfolio, and less useful still at assessing net risk in a portfolio that uses hedging instruments, short positions, and liabilities.

Sensitivity measures address some of the shortcomings of position size measures. Duration, for example, addresses the difference between a 1-year note and a 30-year note; it measures the level of interest rate risk. Option delta and duration (for fixed income) help to display net risk in a portfolio that has hedging or short positions with optionality or interest rate risk.

Sensitivities typically do not often distinguish assets by volatility, though. When measured as the sensitivity to a 1 bp or 1% move, they do not tell the user which portfolio has greater loss potential any more than position size measures do. A high-yield bond portfolio might have the same sensitivity to a 0.01% credit spread movement as an investment-grade portfolio, but they do not have the same risk because the credit spreads of the high-yield portfolio are more likely to move 0.01%, or more, than the credit spreads of the investment-grade bonds. Sensitivity measures do not distinguish by standard deviation/volatility or other higher confidence loss measures. Measuring sensitivity to a one standard deviation movement in an asset's price or yield, however, is one way to overcome this shortcoming of sensitivity.

Granularity: Too Much or Too Little?

Sensitivity measures are aggregated in categories or "buckets." (A bucket is a risk factor description such as "one- to five-year French sovereign debt.") When a number of fixed-income positions are assigned to the same bucket, the effect is an assumption of perfect correlation across the risks encompassed by that bucket. For the "one- to five-year French sovereign debt" risk factor, a short duration position in four-year French sovereign debt will be assumed to fully offset a long duration position in two-year French sovereign debt. However, this may not be true in the case of a non-parallel interest rate change; these points on the yield curve do not have a correlation coefficient of 1 to one another. The broader the buckets used, the more they can hide this kind of correlation risk; but the narrower the buckets used, the greater the complexity and thus the more difficult to portray portfolios in simple, accessible ways. The width or the narrowness of the risk-factor buckets used to portray sensitivity measures is referred to as granularity.

Scenario analysis and stress testing have well-deserved popularity, and they address many of the shortcomings of VaR described earlier. Sensitivity and scenario risk measures can complement VaR in the following ways:

- They do not need to rely on history. Sensitivity and scenario risk measures can be constructed to test the portfolio's vulnerability to a truly never-before-seen market movement. In this way, they can be free of the volatility and correlation behavior of recent market history, which may simply not be representative of stress conditions. In a scenario analysis, assets that typically have a low correlation with one another can be modeled under an assumption of perfect positive correlation simply by simulating an identical price movement for these assets. Alternatively, they can be modeled under an assumption of perfect negative correlation by simulating identical price movements (i.e., in the opposite direction). A scenario might be designed in which a market that typically exhibits an annual standard deviation of 15% moves by 20% in a single day.
- Scenarios can be designed to overcome any assumption of normal distributions; the shock used could be the equivalent of 1, 10, or 1,000 standard deviations, at the choice of the analyst—or as provided by an actual moment in history.
- Scenarios can be tailored to expose a portfolio's most concentrated positions to even worse movement than its other exposures, allowing liquidity to be taken into account.

But scenario measures are not without their own limitations:

- Historical scenarios are interesting, and illuminating, but are not going to happen in exactly the same way again, making hypothetical scenarios necessary to truly fill the gaps identified with the other risk measures listed.
- Hypothetical scenarios may incorrectly specify how assets will co-move, they may get the magnitude of movements wrong, and they may incorrectly adjust for the effects of liquidity and concentration.
- Hypothetical scenarios can be very difficult to create and maintain. Getting all factors and their relationships accurately represented in the suite of scenarios is a painstaking and possibly never-ending exercise. Accordingly, it is necessary to draw a line of "reasonableness" at which to curtail the scenario analysis, and by the very act of being curtailed, the scenario might miss the real risk.
- It is very difficult to know how to establish the appropriate limits on a scenario analysis or stress test. Because we are proposing hypothetical movements in markets and risk factors, we cannot use history to assign a probability of such a move occurring. What if rates rise instantaneously 0.50%, 1.00%, or 3.00%? How should the short end of the yield curve move versus the long end? How much should credit spreads of different qualities move? It is difficult to choose.

The more extreme the scenario, and the farther from historical experience, the less likely it is to be found believable or actionable by management of a company or a portfolio. This issue tends to lead scenario constructors to underestimate movement in order to appear credible. As an example, prior to the very large drop in real estate values that prevailed in the United States from 2008 to 2010, no similar nationwide price decline had occurred in history. Risk measurement teams at a number of firms did prepare scenarios that estimated the potential outcome if real estate prices declined meaningfully, but their scenarios in many cases were only half as large as the movements that subsequently occurred. Because these large market movements had never before occurred, there was no historical basis for estimating them, and to do so appeared irresponsible. This is an additional risk of scenario analysis: The need to keep the scenario plausible may lead to it being incorrect.

In sum, scenario analyses and stress tests have the opportunity to correct the failings of probabilistic risk measures, such as VaR and *ex ante* tracking error; however, because the version of the future they suggest may be no more accurate than that used in VaR, they may also fail to predict potential loss accurately.

As we can see, each risk measure has elements that are better than the others, and each has important failings. No one measure is the "solution" to risk management. Each is useful and necessary to answer certain questions but not sufficient to answer all possible questions—or to prevent all forms of unexpected loss. Using the measures in combination, to correct each other's failings, is as close to a solution as we come. Designing constraints by using multiple measures is the key practice used by successful risk managers. Viewing a portfolio through these multiple lenses provides a more solid framework for a risk manager or an investor to exercise judgment and can help reduce conceptual bias in portfolio management.

EXAMPLE 9

Limitations of Risk Measures

1 Which of the following is **not** a limitation of VaR?

  A It does not adjust for bonds of different durations.

  B It largely relies on recent historical correlations and volatilities.

  C It can be inaccurate if the size of positions held is large relative to available liquidity.

2 Which of the following statements about sensitivities is true?

  A When duration is measured as the sensitivity to a 1 bp change in interest rates, it can be biased by choice of the historical period preceding this measure.

  B Sensitivity measures are the best way to determine how an option can behave under extreme market movements.

  C Duration effectively assumes that the correlation between a fixed-income exposure and the risk-free rate is 1, whereas beta takes into account the historical correlation between an equity and its comparison index.

3 Which of the following is **not** a limitation of scenario measures?

  A It is difficult to ascribe probability to a given scenario.

  B Scenario measures assume a normal distribution, and market returns are not necessarily normal.

  C They risk being an infinite task; one cannot possibly measure all of the possible future scenarios.

4 Which measures are based on market returns during a particular historical period?

  A Hypothetical scenario analysis and duration sensitivity

  B Historical scenario analysis and VaR

  C Option delta and vega

Solution to 1:

A is correct. Well-executed VaR measures do adjust for bonds of differing duration, and therefore it is not a limitation of VaR. B is incorrect because VaR ordinarily uses some period of recent history as part of the calculation, and this reliance on history is one of its limitations. C is incorrect because VaR can be inaccurate and underestimate risk if portfolio positions are too large relative to the available market liquidity, and this inability to account for the illiquidity of an individual investor's position is an additional limitation of VaR.

Solution to 2:

C is correct. Duration assumes that all interest rates that affect a bond change by the same percentage (an effective correlation of 1). A is incorrect because the 1 bp change in rates is applied to current rates, not historical rates. B is incorrect because sensitivity measures are often too small to reveal the most extreme movements for option positions; the larger shocks used in scenario measures are preferable to reveal option characteristics.

Solution to 3:

B is correct. Scenario measures do not assume any given distribution, and thus this is not a limitation of scenario analysis. A is incorrect because it is in fact difficult to ascribe probability to many scenarios, and thus this is a limitation of scenario analysis. C is also incorrect because it is in fact impossible to measure all possible future scenarios, and this is a limitation of scenario analysis.

Solution to 4:

B is correct. Historical scenarios apply market returns from a particular period to the portfolio, and virtually all VaR methodologies use a historical period to underpin the VaR model (although certain methods may make adjustments if this historical period is seen to be anomalous in some way). A is incorrect because a hypothetical scenario is not based on an actual historical period, and duration sensitivity measures change in value for a given small change in rates, not for a given historical period. C is incorrect because option delta and vega measure how much an option's value will change for a given change in the price of the underlying (delta) or implied volatility (vega), and these are sensitivity measures, not measures based on a particular historical period.

USING CONSTRAINTS IN MARKET RISK MANAGEMENT 10

j explain constraints used in managing market risks, including risk budgeting, position limits, scenario limits, and stop-loss limits;

k explain how risk measures may be used in capital allocation decisions;

Designing suitable constraints to be used in market risk management is essential to managing risk effectively. Risk *measurements* in and of themselves cannot be said to be restrictive or unrestrictive: The *limits* placed on the measures drive action. VaR can be measured to a very high confidence level (for example, 99%) or to a low level (for example, 84%). But placing a loose limit on a 99% confidence VaR measure could be less of a constraint than placing a tight limit on an 84% confidence measure. It is not the confidence interval that drives conservatism as much as the limit that is placed on it.

If constraints are too tight, they may limit the pursuit of perceived opportunities and shrink returns or profitability to a sub-optimal level. If constraints are too loose, outsized losses can occur, threatening the viability of the portfolio or business. The concept of "restrictive" or "unrestrictive" relates to the risk appetite of the firm or portfolio and the sizes of losses it can tolerate. Unrestrictive limits are typically set far from current risk levels and permit larger losses than restrictive limits. As an example, for a leveraged portfolio in which insolvency could occur if cumulative daily losses exceed $10 million and the portfolio's current two week, 1% VaR measure is $3 million, an unrestrictive limit might be one set at $10 million. If the portfolio increased positions and went right up to its limit, a misestimation of VaR could result in insolvency; moreover, the fact that losses are expected to exceed the measure at least 1% of the time could mean disaster. But if the limit were set at $4 million, the portfolio might under-allocate the capital it has to invest and fail to make a high enough return on equity to thrive in a competitive environment.

Before applying constraints, particularly those involving such potential loss measures as VaR or a scenario analysis, it is worth considering how far down in the organizational hierarchy to impose them. If applied exclusively to lower level business units, the firm's aggregate risk exposure fails to take advantage of offsetting risks that may occur at higher levels of the organization. As a result, the overall company may never be able to invest according to its risk tolerance because it is "stopped out" by rules lower in the organization. For example, imagine a bank with five trading desks: It might have an overall VaR tolerance of €10 million and might set each trading desk's limit for its standalone VaR at €2 million, which seems reasonable. If there is anything lower than perfect correlation across these desks' positions, however—and particularly if one desk has a short position that to some degree serves as an offset to another desk's long position—the firm will never be able to use its €10 million risk appetite in full. The cure for this problem is over-allocation, with the caveat that a given desk might need to be cut back to its pro rata share in the event that correlations among trading desks are higher than, or the short positions across the different portfolios are not as offsetting as, the over-allocation assumes. Alternatively, some firms might use marginal VaR for each trading desk, allocating each desk a VaR budget such that the total VaR is the sum of each individual desk's marginal VaR. This approach permits each trading desk to "reinvest" the diversification benefits obtained at the aggregate level.

Among the constraints most often used in risk management are risk budgeting, position limits, scenario limits, and stop-loss limits. As is the case in risk measurement, for which multiple measures work better than any one measure alone does, so it is in risk constraints. No one approach on its own works perfectly; they are most effective in combination.

10.1 Risk Budgeting

In **risk budgeting**, the total risk appetite of the firm or portfolio is agreed on at the highest level of the entity and then allocated to sub-activities. Risk budgeting typically rests on a foundation of VaR or *ex ante* tracking error.

A bank might establish a limit on total economic capital or VaR and describe this limit as its risk appetite. Next, it might allocate this risk appetite among the basic risk types (market, credit, and operational) and different business units, geographies, and activities. It allocates to the business unit and/or risk type by specifying a limit, using its chosen measure, for that given activity. For example, it might allow its European business to use 20% of its market risk capital (the portion of its economic capital expected to be used to support market risk taking) and 40% of its credit risk capital, whereas its Asian business might have a different limit. It will set these limits based on the expected long-term profitability of the opportunity set and the demonstrated skill of a business at delivering profitable results, taking into consideration shareholders' expectations regarding the activities the bank is engaged in. As an example of potential shareholder expectations, consider a case in which a firm's shareholder disclosure suggests that the firm's predominant market risk-taking activities are in the Asian markets and that less risk-taking activity is in Europe. Shareholders will be surprised if greater losses are incurred from its European business than its Asian business. Market risk capital limits for the European business should be lower than for the Asian business to be consistent with shareholder disclosures.

A pension fund sponsor might begin with its tolerance for how much of a mismatch it is willing to tolerate overall between the total value of assets and its liabilities—its surplus at risk. Surplus at risk can be the starting point for its asset allocation decision making. Once the broad asset allocation is established, usually expressed via a set of benchmarks, the pension fund sponsor might further establish its tolerance for underperformance in a given asset class and allocate that tolerance to the asset managers selected to manage the assets by assigning each an *ex ante* tracking error budget.

A portfolio manager might have an *ex ante* tracking error budget explicitly provided by the client, or if none is provided by the client, it might instead develop a tracking error budget based on her investment philosophy and market practice. Given this budget, she will seek to optimize the portfolio's exposures relative to the benchmark to ensure that the strategies that generate the most tracking error for the portfolio are those for which she expects the greatest reward.

10.2 Position Limits

Risk budgeting follows a clear logic; but as we have noted, VaR-based measures have a number of drawbacks. One of them is that they perform poorly if portfolios are unusually concentrated, particularly with respect to market liquidity.

Position limits are limits on the market value of any given investment, or the notional principal amount for a derivatives contract. They can be expressed in currency units or as a percentage of some other value, such as net assets. Position limits do not take into account duration, volatility, and correlation, as VaR does, but they are excellent controls on overconcentration. Like risk budgeting, position limits need to be used carefully; if every asset type that a portfolio manager could invest in is constrained, he will have no room to succeed in outperforming the benchmark or generating absolute returns, assuming that is the mandate. Position limits should not be overly prescriptive but should address the event risk and single name risk that VaR handles so poorly, such as

- limits per issuer;
- limits per currency or country;
- limits on categories expected to be minimized in a given strategy, such as high-yield credit or emerging market equities;
- limits on gross size of long–short positions or derivatives activity; and
- limits on asset ownership that correspond to market liquidity measures, such as daily average trading volume.

10.3 Scenario Limits

A scenario limit is a limit on the estimated loss for a given scenario, which if exceeded, would require corrective action in the portfolio.

As discussed in Section 9, scenarios also address shortcomings of VaR, such as the potential for changes in correlation or for extreme movements that might not be predicted using a normal distribution or the historical lookback period used for the VaR measure. Just producing scenario analysis, however, without having any related action steps is not a very valuable exercise.

The action steps that generally follow a scenario analysis are to examine (1) whether the results are within risk tolerance and, in the case of asset managers, (2) whether the results are well incorporated into investor disclosures. To determine whether results are within the established risk tolerance, a tolerance level for each scenario must be developed. It is better to establish a higher tolerance for potential loss under the most extreme scenarios. If the same limit is applied to all scenarios, even extremely unlikely scenarios (e.g., "interest rates rise 1,000,000%"), then the portfolio will simply not be able to take any risk. The risk manager then observes over time whether the portfolio's sensitivity to the scenario is increasing or crosses this high-tolerance bound.

10.4 Stop-Loss Limits

A **stop-loss limit** requires a reduction in the size of a portfolio, or its complete liquidation, when a loss of a particular size occurs in a specified period.

One of the limitations of VaR described in Section 6 was "trending," in which a portfolio remains under its VaR limit each day but cumulatively loses more than expected. This trending can be managed by imposing and monitoring stop-loss limits in addition to the VaR constraints. In one form of a stop-loss limit, the portfolio's positions are unwound if its losses over a pre-specified period exceed a pre-specified level. (Those levels are typically defined to align with the overall risk tolerance.) As an example, a portfolio might have a 10-day, 1% VaR limit of $5 million, but it will be liquidated if its cumulative monthly loss ever exceeds $8 million. The relationship between the stop-loss and the VaR measure can vary depending on management preferences as well as the differing time periods with which the measures are specified.

An alternative approach to a stop-loss limit might instead be to impose a requirement to undertake hedging activity, which may include purchases of protective options, after losses of a given magnitude, with the magnitude of the hedge increasing as losses increase. This approach, called drawdown control or portfolio insurance, is more dynamic and more sophisticated than the simpler stop-loss limit.

10.5 Risk Measures and Capital Allocation

In market risk management, capital allocation is the practice of placing limits on each of a company's activities in order to ensure that the areas in which it expects the greatest reward and has the greatest expertise are given the resources needed to accomplish their goals. Allocating capital wisely ensures that an unproven strategy does not use up all of the firm's risk appetite and, in so doing, deprive the areas most likely to be successful of the capital they need to execute on their strategy.

Economic capital is often used to estimate how much of shareholders' equity could be lost by the portfolio under very unfavorable circumstances. Capital allocation may start with a measurement of economic capital (the amount of capital a firm needs to hold if it is to survive severe losses from the risks in its businesses). The company's actual, physical on-balance-sheet capital must exceed the measure of economic capital, and a minimum level of economic capital must be established to ensure that the company does not take on a risk of loss that will exceed its available capital. The company first establishes its overall risk appetite in economic capital terms, and then it subdivides this appetite among its units. This exercise is similar to risk budgeting, but in the case of corporations, banks, insurers, or hedge funds, it is more likely to be called "capital allocation." Capital allocation is often used in cases in which leverage is used by the portfolio or in which the strategy has meaningful **tail risk**, meaning that losses in extreme events could be far greater than would be expected for a portfolio of assets with a normal distribution. Economic capital is designed to measure how much shareholders' equity could be required to meet tail risk losses. Strategies that have greater-than-expected tail risk include those that sell options, sell insurance, take substantial credit risk, or have unique liquidity or exposure concentration risks. Although risk budgeting more commonly focuses on losses at the one standard deviation level, capital allocation focuses on losses at a very high confidence level in order to capture the magnitude of capital that is placed at risk by the strategy. Capital allocation seeks to understand how much of an investor's scarce resources are, or could be, used by a given portfolio, thereby making it unavailable to other portfolios.

Because a company's capital is a scarce resource and relatively expensive, it should be deployed in activities that have the best chance of earning a superior rate of return. It also should be deployed in a way that investors expect, in activities in which the company has expertise, and in strategies that investors believe the company can successfully execute.

To optimize the use of capital, the "owner" of the capital will typically establish a hurdle rate over a given time horizon; this is often expressed as the expected rate of return per unit of capital allocated. Two potential activities, Portfolio A and Portfolio B, might require different amounts of capital. Portfolio A might require €325,000, and its expected return might be €50,000 per year (15.4%). Portfolio B might have a reasonable expectation of earning €100,000 per year, but it might require €1,000,000 in capital (a 10% return). If the investor has an annualized hurdle rate of 15%, Portfolio A will exceed the hurdle rate and appear a better user of capital than Portfolio B, even though the absolute income for Portfolio B is higher.

Beyond measuring and limiting economic capital, capital allocation is sometimes used as a broad term for allocating costly resources. In some cases, the costly resource is cash; if, for instance, the portfolio has invested in options and futures trading strategies that require heavy use of margin and overcollateralization, its use of economic capital could be low and available cash may be the constraining factor. For other types of investors, such as banks or insurance companies, the capital required by regulatory bodies could be relatively large; as a result, these capital measures may be the most onerous constraint and thus the basis of capital allocation.

When the current measure of economic capital is a smaller number than the portfolio's cash or regulatory capital needs, it may not be the binding constraint. But when it is higher than other measures, it can become the binding constraint, and the one to which hurdle rates should be applied.

EXAMPLE 10

Creating Constraints with Risk Measures

1 Which of the following is **not** an example of risk budgeting?

- A Giving a foreign exchange trading desk a VaR limit of $10 million
- B Allowing a portfolio manager to have an *ex ante* tracking error up to 5% in a given portfolio
- C Reducing the positions in a portfolio after a loss of a 5% of capital has occurred in a single month

2 Which statement is true regarding risk budgeting in cases in which marginal VaR is used?

- A The total risk budget is never equal to the sum of the individual subportfolios' risk budgets.
- B The total risk budget is always equal to the sum of the individual subportfolios' risk budgets.
- C If the total risk budget is equal to the sum of the individual subportfolios' risk budgets, there is a risk that this approach may cause capital to be underutilized.

Solution to 1:

C is correct. This is an example of a stop-loss limit, not risk budgeting. The other choices are both examples of risk budgeting.

Solution to 2:

B is correct. When using marginal VaR, the total risk budget will be equal to the sum of the individual risk budgets. Choice A is not correct. C is also incorrect; it would be correct if each sub-portfolio's individual VaR measure, not adjusted for its marginal contribution, were used, which could lead to underutilization of capital.

11 APPLICATIONS OF RISK MEASURES, MARKET PARTICIPANTS AND THE DIFFERENT RISK MEASURES THEY USE

l. describe risk measures used by banks, asset managers, pension funds, and insurers.

In this section, we examine the practical applications of risk measures. First, we will look at how different types of market participants use risk measures. An understanding of how various market participants use these measures will help as we move to a discussion of their limitations.

11.1 Market Participants and the Different Risk Measures They Use

Three factors tend to greatly influence the types of risk measures used by different market participants:

- The degree to which the market participant is leveraged and the resulting need to assess minimum capitalization/maximum leverage ratios;
- The mix of risk factors to which their business is exposed (e.g., the degree of equity or fixed-income concentration in their portfolios);
- The accounting or regulatory requirements that govern their reporting.

Market participants who use a high degree of leverage typically need to assess their sensitivity to shocks to ensure that they will remain a going concern under very severe, but foreseeable, stresses. This leads them to focus on potential loss measures with a high confidence interval or to focus on rare events that might occur in a short period of time, such as two weeks. Those who use minimal (or no) leverage, such as long-only asset managers, are interested in shock sensitivity as well, but they are likely less concerned with trying to discern the difference between a 99.99% (0.01% VaR) worst case and a 99.95% (0.05% VaR) worst case. Their focus is more likely on avoiding underperformance—for example, failing to keep pace with their market benchmark when markets are doing well. For this reason, they are often more interested in lower confidence intervals—events that are more likely to occur and lead to underperformance for a given strategy. Unleveraged asset managers may also prefer to measure potential underperformance over longer periods of time, such as a quarter or a year, rather than shorter periods.

For portfolios dominated by fixed-income investments, risk managers focus on how sensitive the portfolios are to instantaneous price and yield changes in a variety of categories and typically emphasize duration, credit spread duration, and key rate duration measures. Credit spread duration measures the impact on an instrument's value if credit spreads move while risk-free rates remain unchanged. Key rate duration (sometimes called partial duration) measures the sensitivity of a bond's price to changes in specific maturities on the benchmark yield curve. Risk measurement for fixed-income portfolios is conducted using bond pricing models and by shifting each market rate assumption in the model and aggregating their portfolio's sensitivity to these market rates. Often, these factors are combined into scenarios representing expected central bank policies, inflation expectations, and/or anticipated fiscal policy changes. When portfolios are dominated by equities, risk managers typically categorize the equities by broad country markets, industries, and market capitalization levels. Also, they may additionally regress the returns of their portfolios against fundamental factor histories (such as those for growth, value, momentum, and capitalization size) to understand their exposure to such factors.

Portfolios with full fair value accounting (also called mark-to-market accounting), such as US mutual funds, European UCITS funds, and the held-for-sale portfolios of banks, are very well suited to such risk measures as VaR, economic capital (the amount of capital a firm needs to hold if it is to survive severe losses from the risks in its businesses), duration, and beta—all of which rely on measuring the changes in the fair values of assets. Asset/liability gap models are more meaningful when portfolios are subject to book value accounting in whole or in part.

11.1.1 *Banks*

Banks need to balance a number of sometimes competing aspects of risk to manage their business and meet the expectations of equity investors/equity analysts, bond investors, credit rating agencies, depositors, and regulatory entities. Some banks apply risk measures differently depending on whether the portfolio being assessed is designated as a "held-to-maturity" portfolio, which requires book value accounting, or a "held-for-sale" or "trading book" portfolio, which requires fair value accounting. Other banks will use fair value measures for all risk assessments regardless of the designation used for accounting purposes. In the following list are some of the factors that banks seek to address through their use of risk tools. In compiling this list, we have assumed that banks may treat measures differently depending on accounting treatment.

- *Liquidity gap:* The extent of any liquidity and asset/liability mismatch. The ability to raise sufficient cash for foreseeable payment needs; a view of the liquidity of assets, as well as the expected repayment date of debt.
- *VaR:* The value at risk for the held-for-sale or trading (fair value) portion of the balance sheet.
- *Leverage:* A leverage ratio is typically computed, sometimes according to a regulatory requirement or to an internally determined measure. Leverage ratios will weight risk assets using a variety of methods and rules and divide this weighted asset figure by equity. The result is that riskier assets will be assigned a greater weighting and less risky assets a lower weighting so that more equity is required to support riskier assets.
- *Sensitivities:* For the held-for-sale portion of their balance sheet, banks measure duration, key rate duration or partial duration, and credit spread duration for interest rate risk positions. Banks will also measure foreign exchange exposure and any equity or commodity exposures. All these exposure measures will include the delta sensitivities of options with any other exposures to the same

underlying asset and will also monitor gamma and vega exposures of options. Gamma and vega exposures can be broken out by term to identify how much of these risks come from long-dated versus short-dated options.

- *Economic capital:* This is measured by blending the company's market, credit, and operational risk measures to estimate the total loss the company could suffer at a very high level of confidence (e.g., 99% to 99.99%), usually in one year's time. Economic capital measures are applied to the full balance sheet, including both the held-for-sale and held-for-investment portfolios, and include market, credit, and operational risk capital.
- *Scenario analysis:* Stress tests are applied to the full balance sheet and augment economic capital and liquidity; they are used to identify whether capital is sufficient for targeted, strong negative shocks. Outside of stress testing, significant scenario analysis takes places. Scenario analysis is used to examine how the full balance sheet might be affected by different interest rate, inflation, and credit environments, such as unemployment levels for credit card lenders, home price appreciation/depreciation for mortgage lenders, and business cycle stresses for corporate lenders.

It is common for banks to compute risk measures in distinct business units and geographies and then aggregate these measures to the parent company entity.

11.1.2 *Asset Managers*

Asset managers are not typically regulated with regard to sufficient capital or liquidity; they are more commonly regulated for fair treatment of investors—that disclosures are full and accurate, that marketing is not misleading, that one client is not favored over the other. In some jurisdictions, certain market risk measures may be used to define risk limits for different fund types.

In asset management portfolios, risk management efforts are focused primarily on volatility, probability of loss, or probability of underperforming a benchmark rather than insolvency. A diversified, unleveraged, long-only fund is unlikely to see asset values decline below zero in the absence of a wholesale withdrawal of assets by the firm's clients. Although service costs and other items make insolvency a technical possibility, in practice, insolvency is a much higher threat for leveraged portfolios. Although derivatives use by asset managers can create effective leverage, these positions are often balanced by an amount of cash in the portfolio equal to the notional exposure created by the derivatives mitigating, if not fully eliminating, the impact of leverage.

Asset managers typically measure and view each portfolio separately with respect to its own constraints and limits. However, there are a few exceptions:

- Long-only asset managers: If the adviser has invested its own capital in any of the funds that it manages, these investments may need to be aggregated for the firm to assess its risk exposures across portfolios.
- Hedge funds: A hedge fund manager needs to aggregate the adviser's side-by-side investment in the various funds it advises.
- Funds of funds: Risk measures for these portfolios typically aggregate the risks of the underlying hedge funds to the master fund level.

An asset manager may choose to aggregate exposures across all funds and strategies to determine if there are unusual concentrations in individual securities or counterparties that would make management actions across all portfolios difficult to carry out (e.g., a single portfolio's holdings in a given security may not pose a liquidity risk, but if the firm were to aggregate all of its holdings in that security, it may find that the portfolio fails to meet the desired liquidity target).

It is important when observing risk measures for asset managers to determine whether the measures represent the backward-looking variability of realized returns in the portfolio as it was then constituted or use the current portfolio and measure its potential loss. Backward-looking returns-based measures (typically including standard deviation, *ex post* tracking error, Sharpe ratio, information ratio, and historical beta) have the value of showing the fund's behavior over time and help assess the skill of the manager. Only an analysis of the current holdings, however, will reveal current risk exposures. Measures that use current holdings typically include VaR, *ex ante* tracking error, duration and forward-looking beta, stress tests, and scenario analyses. All risk and performance measures can be conducted on past portfolio holdings or current portfolio holdings; it is important for the user of any measure to determine which ingredients (which set of portfolio holdings, and for market history, what length and smoothing techniques) have been used in order to use it correctly. Assessing the trends in risk exposures, including whether risk has recently risen or if other important changes have taken place in the strategy, can be accomplished by tracking the risk measures through time.

11.1.2.1 Traditional Asset Managers Asset managers that use little leverage typically find relative risk measures most meaningful and actionable. The decision to invest in a given asset class is normally the client's, not the adviser's. The adviser seeks to outperform the benchmark representative of the asset class. Exceptions include absolute return funds and asset allocation strategies, but even these can be measured relative to a benchmark. For absolute return strategies, the benchmark is typically cash or a cash-like hurdle rate. When cash is the benchmark, VaR and *ex ante* tracking error will be effectively the same if measured using the same holding period and confidence interval. (Cash has no volatility, so adding a cash benchmark into a relative VaR calculation does not affect the calculation because its zero volatility cancels out its impact; thus, the resulting calculation is the same as the VaR of the portfolio.) Asset allocation funds can use an asset allocation index as the benchmark for a relative risk measure, or they can use a custom combination of market benchmarks.

Although banks, insurers, and other market participants favor measuring VaR in currency terms relevant for the institution (e.g., dollars for a US-based insurer, yen for a Japanese bank) and measure duration and similar statistics as the value change for a 1 bp interest rate change, long-only asset managers generally prefer to express VaR in percentage terms and will divide VaR and duration by the net assets of the portfolio being analyzed. (Note that using returns as the fundamental source of data removes the last step in calculating VaR: multiplying by the size of the portfolio.)

A typical sample of risk measures used by asset managers includes the following:

- *Position limits:* Asset managers use position limits as the most frequent form of risk control for the portfolios they manage, particularly in fund offering documents that need to be understandable to a broad range of investors. Position limits include restrictions on country, currency, sector, and asset class. They may measure them in absolute terms or relative to a benchmark, and they are almost always expressed as a percentage of the portfolio's value.
- *Sensitivities:* Asset managers use the full range of sensitivity measures, including option-adjusted duration, key rate duration, and credit spread duration, and they will typically include the delta exposure of options in these measures. Measures can be expressed in absolute terms as well as relative to a benchmark.
- *Beta sensitivity:* Beta is frequently used for equity-only accounts.
- *Liquidity:* Asset managers often look at the liquidity characteristics of the assets in their portfolios. For equity portfolios, it is common to measure what percentage of daily average trading volume the portfolio holds of each equity security

and how many days it would take to liquidate a security if the manager did not want it to be too large a portion of trading volume to avoid taking a price concession.

- *Scenario analysis:* Long-only asset managers typically use stress tests or scenario analyses to verify that the risks in the portfolio are as they have been disclosed to investors and to identify any unusual behavior that could arise in stressed markets.
- *Redemption risk:* Open-end fund managers often assess what percentage of the portfolio could be redeemed at peak times and track this behavior across the funds and asset classes they manage.
- *Ex post versus ex ante tracking error:* Limits on *ex ante* tracking error are often used by traditional asset managers as a key risk metric for the portfolios they manage. It provides an estimate of the degree to which the current portfolio could underperform its benchmark. It is worth noting the distinction between *ex post* tracking error and *ex ante* tracking error: Asset managers use *ex post* tracking error to identify sources of performance and manager skill and *ex ante* tracking error to identify whether today's positions could give rise to unexpected potential performance. *Ex post* tracking error measures the historical deviation between portfolio returns and benchmark returns, and thus both the portfolio holdings and market returns are historical in this measure. *Ex ante* tracking error takes today's benchmark-relative position and exposes it to the variability of past markets to estimate what kind of benchmark-relative performance could arise from the current portfolio. *Ex post* tracking error is a useful tool for assessing manager skill and behavior. The day after a large change in portfolio strategy, *ex ante* tracking will immediately reflect the portfolio's new return profile, whereas *ex post* tracking error will not do so until the new strategy has been in place long enough to dominate the data history. (If *ex post* tracking error is computed using 200 days of history, the day after a large strategy change, only 1 of the 200 data points will reflect the current risk positioning.) Some asset managers focus on maintaining *ex ante* tracking error boundaries for the portfolios they manage to monitor and balance the potential performance impact of the active risks they are taking. **Active share** is a measure of that percentage of the portfolio that differs from the benchmark (i.e., a deviation from the benchmark). It is often monitored to help limit tracking error of the portfolio.
- *VaR:* VaR is less commonly used as a risk measure than *ex ante* tracking error by traditional asset managers, but it is used by some—particularly for portfolios that are characterized as "absolute return" strategies for which a given market benchmark may not serve as the portfolio objective.

11.1.2.2 Hedge Funds Similar to banks, hedge funds that use leverage need to observe sources and uses of cash through time, including when credit lines could be withdrawn, and need to simulate the interplay between market movements, margin calls, and the redemption rights of investors in order to understand worst-case needs for cash. A sample of the typical range of hedge fund market risk measures includes the following:

- *Sensitivities:* All hedge fund strategies will display some form of sensitivity or exposure, so the full range of sensitivity measures are useful for hedge fund risk management.
- *Gross exposure:* Long–short, market neutral, and arbitrage strategies will typically measure long exposure, short exposure, and gross exposure (the sum of the absolute value of long plus short positions) separately. Gross position risk is an important guide to the importance of correlation risk for the portfolio.

- *Leverage:* Leverage measures are common for hedge funds. It is important to understand how the measure is treating derivatives and what elements appear in the numerator versus the denominator because there are many different ways to execute the measure.
- *VaR:* Hedge funds that use VaR measures tend to focus on high confidence intervals (more than 90%) and short holding periods, and they rarely use a benchmark-relative measure.
- *Scenarios:* Hedge funds commonly use scenario/stress tests that are well tuned to the specific risks of their strategy—in merger arbitrage strategies, for example, the chance that the merger will not take place.
- *Drawdown:* In the case of the following types of hedge fund strategies, standard deviation and historical beta measures can be particularly misleading when seeking to understand what the more extreme risks can be. This is because the strategies listed frequently display decidedly non-normal return distributions, and when this is true, standard deviation is not a good guide to worst-case outcomes. For the following strategies, any historical standard deviation or historical beta measures should be supplemented by a measure of what has been the **maximum drawdown**, often defined as the worst-returning month or quarter for the portfolio or the worst peak-to-trough decline in a portfolio's returns:
    - Strategies that focus on credit risk taking, such as long–short credit, credit arbitrage, or bankruptcy investing
    - Strategies that focus on events, such as merger arbitrage
    - Strategies that make meaningful investments in non-publicly issued assets or other assets that do not reliably have a daily, independent fair value determination
    - Strategies that invest in illiquid asset classes or take large positions relative to market size in any asset class
    - Strategies that sell options or purchase bonds with embedded options
    - Strategies that are highly reliant on correlation relationships, such as equity market neutral

In addition, it is not uncommon for those investing in hedge funds to look at the returns of the hedge fund during a relevant historical period, such as the 2008 financial crisis.

PENSION FUNDS AND INSURERS

12

l. describe risk measures used by banks, asset managers, pension funds, and insurers.

A defined benefit pension plan is required to make payments to its pensioners in the future that are typically determined as a function of a retiree's final salary. This differs from a defined contribution plan, in which the plan's sponsor may be required to make contributions currently but is not responsible to ensure that they grow to a particular future amount. To meet the required payouts, defined benefit plans have significant market risk management responsibilities. This section describes the practices of defined benefit pension plans only; all mentions in this section of "pension funds" or "pension plans" refer to defined benefit pensions.

The risk management goal for pension funds is to be sufficiently funded to make future payments to pensioners. The requirements for sufficient funding vary from country to country. Different jurisdictions will have regulations concerning such items as how to compute the present value of pension liabilities (including which interest rates are permitted to be used as a discount rate) and what the sponsor of the pension plan is required to contribute when the assets in the pension fund are lower than the present value of the liabilities. In addition, some jurisdictions impose taxes when surplus—the value of the assets less the value of the liabilities—is withdrawn for other use by the plan sponsor. Although these regional differences will shape the practice of pension plan risk management in different countries, it is typically an exercise in ensuring that the plan is not likely to become significantly under- or overfunded. Overfunding occurs when the funding ratio (the assets divided by the present value of the liabilities) is greater than 100%; underfunding occurs when the funding ratio is under 100%. Overfunding may be cured over time by the plan sponsor not needing to make regular contributions to the plan because the number of employees and their salary levels, which drive the pension benefit, are growing. Underfunding, if not cured by growth in the assets in the fund over a suitable time horizon as permitted by regulation, is cured by the plan sponsor contributing to the fund. The pension plan's actions will also vary depending on its age (whether it is a new or established plan) and whether it is currently meaningfully under- or overfunded. Important market risk measures or methods for pension funds often include the following:

- *Interest rate and curve risk:* The first step of risk measurement for pension funds is the analysis of expected payments to pensioners in the future. The expected future cash flows are grouped by maturity. In the case of an international pension fund that must make future payouts in multiple currencies, they may also be grouped by currency. In cases in which the jurisdiction requires a particular fixed-income instrument or curve be used to provide the discount rate for arriving at the present value of the pension liability (such as corporate bonds in the United States, inflation-linked gilts in the United Kingdom, or government bonds in the Netherlands), the liability cash flows will be expressed as a short position at the relevant points on the curve.
- *Surplus at risk:* This measure is an application of VaR. It is computed by entering the assets in the portfolio into a VaR model as long positions and the pension liabilities as short fixed-income positions. It estimates how much the assets might underperform the liabilities, usually over one year, and pension plan sponsors may vary with respect to how high a level of confidence they choose to use (e.g., 84%, 95%, 99%). If the assets in the portfolio were invested precisely in the same fixed-income instruments to which the liabilities have been apportioned and in the same amounts, it would result in zero surplus at risk. In practice, however, it may be impossible to invest in the sizes required in the particular fixed-income instruments specified in the liability analysis, so the pension will invest in other, non-fixed-income investments, such as equities or real assets. The more volatile the investments in the pension fund and the less well correlated these assets are with the liabilities, the higher the surplus at risk. The pension fund may set a threshold level or limit on surplus at risk; when the pension fund's surplus at risk exceeds this limit, pension staff will change the fund's asset allocation to make the assets in the fund better match the liabilities. This liability-focused form of pension investing is commonly referred to as "liability driven investing."
- *Liability hedging exposures versus return generating exposures:* Although matching liabilities is an important goal of pension fund management, it is not the only goal. Pension staff may separate their investment portfolio into investments designed to match the pension liability versus those meant to generate

excess returns. The precise instruments linked to the liability cannot always be directly invested in, so a separate portion of the portfolio may be necessary and should perform the function of earning returns that can minimize the chance of having an over- or underfunded status greater than the pension fund's risk tolerance. The return-generating portion of the portfolio also helps to hedge the potential for future changes in the size of the liability that could be caused by longevity risk or by wage growth that exceeds the forecasts currently used to compute the liability.

12.1 Insurers

Insurers in the largest global economies are subject to significant regulation and accounting oversight regarding how they must retain reserves and reflect their liabilities. Regulation may also affect the pricing permitted by product line. It is common for insurers to aggregate risk from underlying business units to arrive at a firm-wide view of risk.

Insurance liabilities vary in their correlation with financial markets. The risk metrics of property and casualty insurance differ significantly from those used for life insurance and annuity products. Property and casualty insurance, including home, auto, corporate liability insurance, and health insurance, are typically not highly correlated with financial asset markets.

Insurers focus on managing a number of forms of insurance risk, for which they may use such tools as reinsurance and geographic dispersion. The market risk management measures in the property and casualty lines of business include the following:

- *Sensitivities and exposures:* Insurers often design an asset allocation for these portfolios and monitor current exposures to remain within the target ranges set forth in the target asset allocation.
- *Economic capital and VaR:* The risk measurement focus for these lines of business is capital at risk and VaR. The premiums earned in these areas are typically set to compensate for the expected payouts (usually defined as a range of possible payouts), so it is only in cases of greater-than-expected payouts that capital is tapped. The risk modeling effort is to estimate what that catastrophic loss amount could be at a given level of probability. Assessment of the risk to economic capital will include the market risks in the portfolio as well as characteristics of the insurance exposures and reinsurance coverage.
- *Scenario analysis:* Insurers use scenario analysis like other market participants that have capital at risk, such as banks and hedge funds. For the property and casualty lines, these scenarios may stress the market risks and the insurance risks in the same scenario.

Insurers do not focus on matching assets with liabilities in their property and casualty lines of business. Investment portfolios are not designed to pay out insurance claims in property and casualty insurance businesses; the premium income is primarily used for that purpose. These investments are designed to achieve a good absolute return within the constraints imposed under regulatory reserve requirements. Riskier assets are discounted relative to safer, fixed-income assets in measuring required reserves.

Life insurance and annuities have stronger ties to the financial markets, even while retaining distinct mortality-based risk profiles. Life liabilities are very long, and the reserves that insurers are required to maintain by insurance regulators are highly dependent on discount rate assumptions. Non-financial inputs include assumptions about mortality and which policyholders will either tap into options in their policy to add coverage at a given level or cancel their policy. Annuities produce returns based on financial assets, with some extra optionality driven by any life insurance elements embedded in the policy. These activities are paired with long-term investment portfolios in a variety of assets that are designed to help the insurer meet future claims.

For life portfolios, market risk measures include the following:

- *Sensitivities:* The exposures of the investment portfolio and the annuity liability are measured and monitored.
- *Asset and liability matching:* The investment portfolio is not designed to be a perfect match to the liabilities, but it is more closely matched to liabilities than is the case in property and casualty insurance.
- *Scenario analysis:* The main focus of risk measurement for the life lines of insurance are measures of potential stress losses based on the differences between the assets in which the insurance company has invested and the liabilities driven by the insurance contracts it has written to its customers. Scenario analyses need to stress both market and non-market sources of cash flow change (in which non-market changes can include changes in longevity).

EXAMPLE 11

Uses of Risk Measures by Market Participants

1 Which type of market participant is *most likely* to consistently express risk measures as a percentage of assets and relative to a benchmark?

   A Banks

   B Corporations

   C Long-only asset managers

2 How does *ex ante* tracking error differ from *ex post* tracking error?

   A *Ex ante* tracking error takes into account the behavior of options, whereas *ex post* tracking error does not.

   B *Ex post* tracking error uses a more accurate forecast of future markets than the forecast used for *ex ante* tracking error.

   C *Ex ante* tracking error uses *current* portfolio holdings exposed to the variability of historical markets, whereas *ex post* tracking error measures the variability of *historical* portfolio holdings in historical markets.

Solution to 1:

C is correct. Long-only asset managers most commonly express risk measures in percentage terms and relative to a benchmark, whereas the entities in answers A and B measure risk more commonly in currency units and in absolute terms (not relative to a benchmark). Banks occasionally express risk measures, such as economic capital, as a percentage of assets or other balance sheet measures, but bank risk measures are typically expressed in currency units.

Solution to 2:

C is correct. A is incorrect because although *ex post* tracking error accounts for the options that were in the portfolio in the past, *ex ante* tracking error might actually misstate the risk of options if it is computed using the parametric method. B is incorrect because *ex post* tracking error is not aiming to forecast the future; it is only measuring the variability of past results.

SUMMARY

This reading on market risk management models covers various techniques used to manage the risk arising from market fluctuations in prices and rates. The key points are summarized as follows:

- Value at risk (VaR) is the minimum loss in either currency units or as a percentage of portfolio value that would be expected to be incurred a certain percentage of the time over a certain period of time given assumed market conditions.
- VaR requires the decomposition of portfolio performance into risk factors.
- The three methods of estimating VaR are the parametric method, the historical simulation method, and the Monte Carlo simulation method.
- The parametric method of VaR estimation typically provides a VaR estimate from the left tail of a normal distribution, incorporating the expected returns, variances, and covariances of the components of the portfolio.
- The parametric method exploits the simplicity of the normal distribution but provides a poor estimate of VaR when returns are not normally distributed, as might occur when a portfolio contains options.
- The historical simulation method of VaR estimation uses historical return data on the portfolio's current holdings and allocation.
- The historical simulation method has the advantage of incorporating events that actually occurred and does not require the specification of a distribution or the estimation of parameters, but it is only useful to the extent that the future resembles the past.
- The Monte Carlo simulation method of VaR estimation requires the specification of a statistical distribution of returns and the generation of random outcomes from that distribution.
- The Monte Carlo simulation method is extremely flexible but can be complex and time consuming to use.
- There is no single right way to estimate VaR.
- The advantages of VaR include the following: It is a simple concept; it is relatively easy to understand and easily communicated, capturing much information in a single number. It can be useful in comparing risks across asset classes, portfolios, and trading units and, as such, facilitates capital allocation decisions. It can be used for performance evaluation and can be verified by using backtesting. It is widely accepted by regulators.
- The primary limitations of VaR are that it is a subjective measure and highly sensitive to numerous discretionary choices made in the course of computation. It can underestimate the frequency of extreme events. It fails to account for the lack of liquidity and is sensitive to correlation risk. It is vulnerable to trending or volatility regimes and is often misunderstood as a worst-case scenario. It can oversimplify the picture of risk and focuses heavily on the left tail.
- There are numerous variations and extensions of VaR, including conditional VaR (CVaR), incremental VaR (IVaR), and marginal VaR (MVaR), that can provide additional useful information.
- Conditional VaR is the average loss conditional on exceeding the VaR cutoff.
- Incremental VaR measures the change in portfolio VaR as a result of adding or deleting a position from the portfolio or if a position size is changed relative to the remaining positions.

- MVaR measures the change in portfolio VaR given a small change in the portfolio position. In a diversified portfolio, MVaRs can be summed to determine the contribution of each asset to the overall VaR.
- *Ex ante* tracking error measures the degree to which the performance of a given investment portfolio might deviate from its benchmark.
- Sensitivity measures quantify how a security or portfolio will react if a single risk factor changes. Common sensitivity measures are beta for equities; duration and convexity for bonds; and delta, gamma, and vega for options. Sensitivity measures do not indicate which portfolio has greater loss potential.
- Risk managers can use deltas, gammas, vegas, durations, convexities, and betas to get a comprehensive picture of the sensitivity of the entire portfolio.
- Stress tests apply extreme negative stress to a particular portfolio exposure.
- Scenario measures, including stress tests, are risk models that evaluate how a portfolio will perform under certain high-stress market conditions.
- Scenario measures can be based on actual historical scenarios or on hypothetical scenarios.
- Historical scenarios are scenarios that measure the portfolio return that would result from a repeat of a particular period of financial market history.
- Hypothetical scenarios model the impact of extreme movements and co-movements in different markets that have not previously occurred.
- Reverse stress testing is the process of stressing the portfolio's most significant exposures.
- Sensitivity and scenario risk measures can complement VaR. They do not need to rely on history, and scenarios can be designed to overcome an assumption of normal distributions.
- Limitations of scenario measures include the following: Historical scenarios are unlikely to re-occur in exactly the same way. Hypothetical scenarios may incorrectly specify how assets will co-move and thus may get the magnitude of movements wrong. And, it is difficult to establish appropriate limits on a scenario analysis or stress test.
- Constraints are widely used in risk management in the form of risk budgets, position limits, scenario limits, stop-loss limits, and capital allocation.
- Risk budgeting is the allocation of the total risk appetite across sub-portfolios.
- A scenario limit is a limit on the estimated loss for a given scenario, which, if exceeded, would require corrective action in the portfolio.
- A stop-loss limit either requires a reduction in the size of a portfolio or its complete liquidation (when a loss of a particular size occurs in a specified period).
- Position limits are limits on the market value of any given investment.
- Risk measurements and constraints in and of themselves are not restrictive or unrestrictive; it is the limits placed on the measures that drive action.
- The degree of leverage, the mix of risk factors to which the business is exposed, and accounting or regulatory requirements influence the types of risk measures used by different market participants.
- Banks use risk tools to assess the extent of any liquidity and asset/liability mismatch, the probability of losses in their investment portfolios, their overall leverage ratio, interest rate sensitivities, and the risk to economic capital.
- Asset managers' use of risk tools focuses primarily on volatility, probability of loss, or the probability of underperforming a benchmark.

- Pension funds use risk measures to evaluate asset/liability mismatch and surplus at risk.
- Property and casualty insurers use sensitivity and exposure measures to ensure exposures remain within defined asset allocation ranges. They use economic capital and VaR measures to estimate the impairment in the event of a catastrophic loss. They use scenario analysis to stress the market risks and insurance risks simultaneously.
- Life insurers use risk measures to assess the exposures of the investment portfolio and the annuity liability, the extent of any asset/liability mismatch, and the potential stress losses based on the differences between the assets in which they have invested and the liabilities resulting from the insurance contracts they have written.

REFERENCE

Malkiel, Burton. 2007. *A Random Walk Down Wall Street*. New York: W.W. Norton.

PRACTICE PROBLEMS

The following information relates to Questions 1–5

Randy Gorver, chief risk officer at Eastern Regional Bank, and John Abell, assistant risk officer, are currently conducting a risk assessment of several of the bank's independent investment functions. These reviews include the bank's fixed-income investment portfolio and an equity fund managed by the bank's trust department. Gorver and Abell are also assessing Eastern Regional's overall risk exposure.

Eastern Regional Bank Fixed-Income Investment Portfolio

The bank's proprietary fixed-income portfolio is structured as a barbell portfolio: About half of the portfolio is invested in zero-coupon Treasuries with maturities in the 3- to 5-year range (Portfolio P_1), and the remainder is invested in zero-coupon Treasuries with maturities in the 10- to 15-year range (Portfolio P_2). Georges Montes, the portfolio manager, has discretion to allocate between 40% and 60% of the assets to each maturity "bucket." He must remain fully invested at all times. Exhibit 1 shows details of this portfolio.

Exhibit 1 US Treasury Barbell Portfolio

	Maturity	
	P_1	P_2
	3–5 Years	10–15 Years
Average duration	3.30	11.07
Average yield to maturity	1.45%	2.23%
Market value	$50.3 million	$58.7 million

Trust Department's Equity Fund

a **Use of Options:** The trust department of Eastern Regional Bank manages an equity fund called the Index Plus Fund, with $325 million in assets. This fund's objective is to track the S&P 500 Index price return while producing an income return 1.5 times that of the S&P 500. The bank's chief investment officer (CIO) uses put and call options on S&P 500 stock index futures to adjust the risk exposure of certain client accounts that have an investment in this fund. The portfolio of a 60-year-old widow with a below-average risk tolerance has an investment in this fund, and the CIO has asked his assistant, Janet Ferrell, to propose an options strategy to bring the portfolio's delta to 0.90.

b **Value at Risk**: The Index Plus Fund has a value at risk (VaR) of $6.5 million at 5% for one day. Gorver asks Abell to write a brief summary of the portfolio VaR for the report he is preparing on the fund's risk position.

© 2019 CFA Institute. All rights reserved.

Combined Bank Risk Exposures

The bank has adopted a new risk policy, which requires forward-looking risk assessments in addition to the measures that look at historical risk characteristics. Management has also become very focused on tail risk since the subprime crisis and is evaluating the bank's capital allocation to certain higher-risk lines of business. Gorver must determine what additional risk metrics to include in his risk reporting to address the new policy. He asks Abell to draft a section of the risk report that will address the risk measures' adequacy for capital allocation decisions.

1 If Montes is expecting a 50 bp increase in yields at all points along the yield curve, which of the following trades is he *most likely* to execute to minimize his risk?

   A Sell \$35 million of P_2 and reinvest the proceeds in three-year bonds

   B Sell \$15 million of P_2 and reinvest the proceeds in three-year bonds

   C Reduce the duration of P_2 to 10 years and reduce the duration of P_1 to 3 years

2 Which of the following options strategies is Ferrell *most likely* to recommend for the client's portfolio?

   A Long calls

   B Short calls

   C Short puts

3 Which of the following statements regarding the VaR of the Index Plus Fund is correct?

   A The expected maximum loss for the portfolio is \$6.5 million.

   B Five percent of the time, the portfolio can be expected to experience a loss of at least \$6.5 million.

   C Ninety-five percent of the time, the portfolio can be expected to experience a one-day loss of no more than \$6.5 million.

4 To comply with the new bank policy on risk assessment, which of the following is the *best* set of risk measures to add to the chief risk officer's risk reporting?

   A Conditional VaR, stress test, and scenario analysis

   B Monte Carlo VaR, incremental VaR, and stress test

   C Parametric VaR, marginal VaR, and scenario analysis

5 Which of the following statements should *not* be included in Abell's report to management regarding the use of risk measures in capital allocation decisions?

   A VaR measures capture the increased liquidity risk during stress periods.

   B Stress tests and scenario analysis can be used to evaluate the effect of outlier events on each line of business.

   C VaR approaches that can accommodate a non-normal distribution are critical to understand relative risk across lines of business.

The following information relates to Questions 6–11

Hiram Life (Hiram), a large multinational insurer located in Canada, has received permission to increase its ownership in an India-based life insurance company, LICIA, from 26% to 49%. Before completing this transaction, Hiram wants to complete a risk assessment of LICIA's investment portfolio. Judith Hamilton, Hiram's chief financial officer, has been asked to brief the management committee on investment risk in its India-based insurance operations.

LICIA's portfolio, which has a market value of CAD260 million, is currently structured as shown in Exhibit 1. Despite its more than 1,000 individual holdings, the portfolio is invested predominantly in India. The Indian government bond market is highly liquid, but the country's mortgage and infrastructure loan markets, as well as the corporate bond market, are relatively illiquid. Individual mortgage and corporate bond positions are large relative to the normal trading volumes in these securities. Given the elevated current and fiscal account deficits, Indian investments are also subject to above-average economic risk.

Hamilton begins with a summary of the India-based portfolio. Exhibit 1 presents the current portfolio composition and the risk and return assumptions used to estimate value at risk (VaR).

Exhibit 1 Selected Assumptions for LICIA's Investment Portfolio

	Allocation	Average Daily Return	Daily Standard Deviation
India government securities	50%	0.015%	0.206%
India mortgage/infrastructure loans	25%	0.045%	0.710%
India corporate bonds	15%	0.025%	0.324%
India equity	10%	0.035%	0.996%

Infrastructure is a rapidly growing asset class with limited return history; the first infrastructure loans were issued just 10 years ago.

Hamilton's report to the management committee must outline her assumptions and provide support for the methods she used in her risk assessment. If needed, she will also make recommendations for rebalancing the portfolio to ensure its risk profile is aligned with that of Hiram.

Hamilton develops the assumptions shown in Exhibit 2, which will be used for estimating the portfolio VaR.

Exhibit 2 VaR Input Assumptions for Proposed CAD260 Million Portfolio

Method	Average Return Assumption	Standard Deviation Assumption
Monte Carlo simulation	0.026%	0.501%
Parametric approach	0.026%	0.501%
Historical simulation	0.023%	0.490%

Hamilton elects to apply a one-day, 5% VaR limit of CAD2 million in her risk assessment of LICIA's portfolio. This limit is consistent with the risk tolerance the committee has specified for the Hiram portfolio.

The markets' volatility during the last 12 months has been significantly higher than the historical norm, with increased frequency of large daily losses, and Hamilton expects the next 12 months to be equally volatile.

She estimates the one-day 5% portfolio VaR for LICIA's portfolio using three different approaches:

Exhibit 3 VaR Results over a One-Day Period for Proposed Portfolio

Method	5% VaR
Monte Carlo simulation	CAD2,095,565
Parametric approach	CAD2,083,610
Historical simulation	CAD1,938,874

The committee is likely to have questions in a number of key areas—the limitations of the VaR report, potential losses in an extreme adverse event, and the reliability of the VaR numbers if the market continues to exhibit higher-than-normal volatility. Hamilton wants to be certain that she has thoroughly evaluated the risks inherent in the LICIA portfolio and compares them with the risks in Hiram's present portfolio.

Hamilton believes the possibility of a ratings downgrade on Indian sovereign debt is high and not yet fully reflected in securities prices. If the rating is lowered, many of the portfolio's holdings will no longer meet Hiram's minimum ratings requirement. A downgrade's effect is unlikely to be limited to the government bond portfolio. All asset classes can be expected to be affected to some degree. Hamilton plans to include a scenario analysis that reflects this possibility to ensure that management has the broadest possible view of the risk exposures in the India portfolio.

6 Given Hamilton's expectations, which of the following models is *most appropriate* to use in estimating portfolio VaR?

- A Parametric method
- B Historical simulation method
- C Monte Carlo simulation method

7 Which risk measure is Hamilton *most likely* to present when addressing the committee's concerns regarding potential losses in extreme stress events?

- A Relative VaR
- B Incremental VaR
- C Conditional VaR

8 The scenario analysis that Hamilton prepares for the committee is *most likely* a:

- A stress test.
- B historical scenario.
- C hypothetical scenario.

9 The scenario analysis that Hamilton prepares for the committee is a valuable tool to supplement VaR *because* it:

- A incorporates historical data to evaluate the risk in the tail of the VaR distribution.
- B enables Hamilton to isolate the risk stemming from a single risk factor—the ratings downgrade.
- C allows the committee to assess the effect of low liquidity in the event of a ratings downgrade.

10 Using the data in Exhibit 2, the portfolio's annual 1% parametric VaR is *closest* to:

A CAD17 million.

B CAD31 million.

C CAD48 million.

11 What additional risk measures would be most appropriate to add to Hamilton's risk assessment?

A Delta

B Duration

C Tracking error

The following information relates to Questions 12–19

Tina Ming is a senior portfolio manager at Flusk Pension Fund (Flusk). Flusk's portfolio is composed of fixed-income instruments structured to match Flusk's liabilities. Ming works with Shrikant McKee, Flusk's risk analyst.

Ming and McKee discuss the latest risk report. McKee calculated value at risk (VaR) for the entire portfolio using the historical method and assuming a lookback period of five years and 250 trading days per year. McKee presents VaR measures in Exhibit 1.

Exhibit 1 Flusk Portfolio VaR (in $ millions)

Confidence Interval	Daily VaR	Monthly VaR
95%	1.10	5.37

After reading McKee's report, Ming asks why the number of daily VaR breaches over the last year is zero even though the portfolio has accumulated a substantial loss.

Next, Ming requests that McKee perform the following two risk analyses on Flusk's portfolio:

Analysis 1 Use scenario analysis to evaluate the impact on risk and return of a repeat of the last financial crisis.

Analysis 2 Estimate over one year, with a 95% level of confidence, how much Flusk's assets could underperform its liabilities.

Ming recommends purchasing newly issued emerging market corporate bonds that have embedded options. Prior to buying the bonds, Ming wants McKee to estimate the effect of the purchase on Flusk's VaR. McKee suggests running a stress test using a historical period specific to emerging markets that encompassed an extreme change in credit spreads.

At the conclusion of their conversation, Ming asks the following question about risk management tools: "What are the advantages of VaR compared with other risk measures?"

12 Based on Exhibit 1, Flusk's portfolio is expected to experience:

A a minimum daily loss of $1.10 million over the next year.

B a loss over one month equal to or exceeding $5.37 million 5% of the time.

C an average daily loss of $1.10 million 5% of the time during the next 250 trading days.

13 The number of Flusk's VaR breaches *most likely* resulted from:

A using a standard normal distribution in the VaR model.

B using a 95% confidence interval instead of a 99% confidence interval.

C lower market volatility during the last year compared with the lookback period.

14 To perform Analysis 1, McKee should use historical bond:

A prices.

B yields.

C durations.

15 The limitation of the approach requested for Analysis 1 is that it:

A omits asset correlations.

B precludes incorporating portfolio manager actions.

C assumes no deviation from historical market events.

16 The estimate requested in Analysis 2 is *best* described as:

A liquidity gap.

B surplus at risk.

C maximum drawdown.

17 Which measure should McKee use to estimate the effect on Flusk's VaR from Ming's portfolio recommendation?

A Relative VaR

B Incremental VaR

C Conditional VaR

18 When measuring the portfolio impact of the stress test suggested by McKee, which of the following is *most likely* to produce an accurate result?

A Marginal VaR

B Full revaluation of securities

C The use of sensitivity risk measures

19 The risk management tool referenced in Ming's question:

A is widely accepted by regulators.

B takes into account asset liquidity.

C usually incorporates right-tail events.

The following information relates to questions 20–26

Carol Kynnersley is the chief risk officer at Investment Management Advisers (IMA). Kynnersley meets with IMA's portfolio management team and investment advisers to discuss the methods used to measure and manage market risk and how risk metrics are presented in client reports.

The three most popular investment funds offered by IMA are the Equity Opportunities, the Diversified Fixed Income, and the Alpha Core Equity. The Equity Opportunities Fund is composed of two exchange-traded funds: a broadly diversified large-cap equity product and one devoted to energy stocks. Kynnersley makes the following statements regarding the risk management policies established for the Equity Opportunities portfolio:

Statement 1 IMA's preferred approach to model value at risk (VaR) is to estimate expected returns, volatilities, and correlations under the assumption of a normal distribution.

Statement 2 In last year's annual client performance report, IMA stated that a hypothetical $6 million Equity Opportunities Fund account had a daily 5% VaR of approximately 1.5% of portfolio value.

Kynnersley informs the investment advisers that the risk management department recently updated the model for estimating the Equity Opportunities Fund VaR based on the information presented in Exhibit 1.

Exhibit 1 Equity Opportunities Fund—VaR Model Input Assumptions

	Large-Cap ETF	Energy ETF	Total Portfolio
Portfolio weight	65.0%	35.0%	100.0%
Expected annual return	12.0%	18.0%	14.1%
Standard deviation	20.0%	40.0%	26.3%
Correlation between ETFs: 0.90			
Number of trading days/year: 250			

For clients interested in fixed-income products, IMA offers the Diversified Fixed-Income Fund. Kynnersley explains that the portfolio's bonds are all subject to interest rate risk. To demonstrate how fixed-income exposure measures can be used to identify and manage interest rate risk, Kynnersley distributes two exhibits featuring three hypothetical Treasury coupon bonds (Exhibit 2) under three interest rate scenarios (Exhibit 3).

Exhibit 2 Fixed-Income Risk Measure

Hypothetical Bond	Duration
Bond 1	1.3
Bond 2	3.7
Bond 3	10.2

Exhibit 3 Interest Rate Scenarios

Scenario	Interest Rate Environment
Scenario 1	Rates increase 25 bps
Scenario 2	Rates increase 10 bps
Scenario 3	Rates decrease 20 bps

One of the investment advisers comments that a client recently asked about the performance of the Diversified Fixed-Income Fund relative to its benchmark, a broad fixed-income index. Kynnersley informs the adviser as follows:

Statement 3 The Diversified Fixed-Income Fund manager monitors the historical deviation between portfolio returns and benchmark returns. The fund prospectus stipulates a target deviation from the benchmark of no more than 5 bps.

Kynnersley concludes the meeting by reviewing the constraints IMA imposes on securities included in the Alpha Core Equity Fund. The compliance department conducts daily oversight using numerous risk screens and, when indicated, notifies portfolio managers to make adjustments. Kynnersley makes the following statement:

Statement 4 It is important that all clients investing in the fund be made aware of IMA's compliance measures. The Alpha Core Equity Fund restricts the exposure of individual securities to 1.75% of the total portfolio.

20 Based on Statement 1, IMA's VaR estimation approach is *best* described as the:

A parametric method.

B historical simulation method.

C Monte Carlo simulation method.

21 In Statement 2, Kynnersley implies that the portfolio:

A is at risk of losing $4,500 each trading day.

B value is expected to decline by $90,000 or more once in 20 trading days.

C has a 5% chance of falling in value by a maximum of $90,000 on a single trading day.

22 Based *only* on Statement 2, the risk measurement approach:

A ignores right-tail events in the return distribution.

B is similar to the Sharpe ratio because it is backward looking.

C provides a relatively accurate risk estimate in both trending and volatile regimes.

23 Based on Exhibit 1, the daily 5% VaR estimate is *closest* to:

A 1.61%.

B 2.42%.

C 2.69%.

24 Based *only* on Exhibits 2 and 3, it is *most likely* that under:

A Scenario 1, Bond 2 outperforms Bond 1.

B Scenario 2, Bond 1 underperforms Bond 3.

C Scenario 3, Bond 3 is the best performing security.

25 The risk measure referred to in Statement 3 is:

A active share.

B beta sensitivity

C *ex post* tracking error.

26 In Statement 4, Kynnersley describes a constraint associated with a:

A risk budget.

B position limit.

C stop-loss limit.

SOLUTIONS

1 B is correct. Duration is a measure of interest rate risk. To reduce risk in anticipation of an increase in interest rates, Montes would seek to shorten the portfolio's duration. He is limited, however, in the amount he can shift from P_2 to P_1. Selling \$15 million of P_2 reduces that portfolio to the lower end of the permitted 40% to 60% range. By reinvesting the proceeds at the shortest maturities allowed, Montes substantially reduces the portfolio duration.

2 B is correct. An index-tracking portfolio without options has a delta of 1. To achieve a delta of 0.9, the delta of the options position must be negative. Of the three choices, only short calls have a negative delta. Long call options have deltas ranging from 0 to 1. Short calls, therefore, have deltas ranging from 0 to −1. The short call position lowers the portfolio's overall delta as desired.

3 B is correct. VaR measures the frequency of losses of a given minimum magnitude. Here the VaR indicates that on 5% of trading days, the portfolio will experience a loss of at least \$6.5 million. (Although C may appear to say the same thing as B, it actually implies that the portfolio will experience a loss on 95% of trading days.) The correct interpretation is that returns will be equal to or greater than −\$6.5 million on 95% of trading days; those returns include gains as well as losses.

4 A is correct. The bank policy requires the addition of forward-looking risk assessments, and management is focused on tail risk. Conditional VaR measures tail risk, and stress tests and scenario analysis subject current portfolio holdings to historical or hypothetical stress events.

5 A is correct. VaR measures do *not* capture liquidity risk. "If some assets in a portfolio are relatively illiquid, VaR could be understated, even under normal market conditions. Additionally, liquidity squeezes are frequently associated with tail events and major market downturns, thereby exacerbating the risk."

6 C is correct. The Monte Carlo simulation method can accommodate virtually any distribution, an important factor given the increased frequency of large daily losses. This method can also more easily accommodate the large number of portfolio holdings. The Monte Carlo method allows the user to develop her own forward-looking assumptions about the portfolio's risk and return characteristics, unlike the historical simulation method, which uses the current portfolio and re-prices it using the actual historical changes in the key factors experienced during the lookback period. Given the limited return history for infrastructure investments and Hamilton's expectations for higher-than-normal volatility, the historical simulation method would be a suboptimal choice.

7 C is correct. Conditional VaR is a measure of tail risk that provides an estimate of the average loss that would be incurred if the VaR cutoff is exceeded.

8 C is correct. A hypothetical scenario analysis allows the risk manager to estimate the likely effect of the scenario on a range of portfolio risk factors. A sovereign ratings downgrade would affect Hiram's India equity and corporate bond exposures as well as the government bond exposure. In addition, the assumptions used in constructing the scenario analysis can specifically address the effect of a need to sell large position sizes under decreased liquidity conditions resulting from a ratings downgrade. VaR alone does not accurately reflect the risk of large position sizes, which may be difficult to trade.

9 C is correct. A hypothetical scenario analysis allows Hamilton to estimate the direct effect of a ratings downgrade on the portfolio's government bond holdings and the resulting need to sell a number of the portfolio's holdings because they no longer meet the ratings guidelines. VaR alone does not accurately reflect the risk of large position sizes, which may be difficult to trade. The hypothetical scenario analysis will also highlight the effect of increased economic turmoil on all of the portfolio's exposures, not only the government bond exposures.

10 B is correct. The VaR is derived as follows:

$$\text{VaR} = \{[E(Rp) - 2.33\sigma_p](-1)\}(\text{Portfolio value}),$$

where

$E(R_p)$ = Annualized daily return = (0.00026 × 250) = 0.065

250 = Number of trading days annually

2.33 = Number of standard deviations to attain 1% VaR

σ_p = Annualized standard deviation = $\left(0.00501 \times \sqrt{250}\right)$ = 0.079215

Portfolio value = CAD260,000,000

VaR = −(0.065 − 0.184571) × CAD260,000,000

= CAD31,088,460.

11 B is correct. Given the large fixed-income exposure in the LICIA portfolio, examining the portfolio duration more closely would be prudent. Duration is the primary sensitivity exposure measure for fixed-income investments.

12 B is correct. VaR is the minimum loss that would be expected a certain percentage of the time over a specified period of time given the assumed market conditions. A 5% VaR is often expressed as its complement—a 95% level of confidence. Therefore, the monthly VaR in Exhibit 5 indicates that \$5.37 million is the minimum loss that would be expected to occur over one month 5% of the time. Alternatively, 95% of the time, a loss of more than \$5.37 million would not be expected.

13 C is correct. Flusk experienced zero daily VaR breaches over the last year yet incurred a substantial loss. A limitation of VaR is its vulnerability to different volatility regimes. A portfolio might remain under its VaR limit every day but lose an amount approaching this limit each day. If market volatility during the last year is lower than in the lookback period, the portfolio could accumulate a substantial loss without technically breaching the VaR constraint.

A is incorrect because VaR was calculated using historical simulation, so the distribution used was based on actual historical changes in the key risk factors experienced during the lookback period. Thus, the distribution is not characterized using estimates of the mean return, the standard deviation, or the correlations among the risk factors in the portfolio. In contrast, the parametric method of estimating VaR generally assumes that the distribution of returns for the risk factors is normal.

B is incorrect because a specification with a higher confidence level will produce a higher VaR. If a 99% confidence interval was used to calculate historical VaR, the VaR would be larger (larger expected minimum loss). During the last year, none of Flusk's losses were substantial enough to breach the 5% VaR number (95% confidence interval); therefore, if McKee used a 1% VaR (99% confidence interval), the number of VaR breaches would not change.

14 B is correct. In order to simulate the impact of the latest financial crisis on the current bond portfolio holdings, McKee's valuation model for bonds should use the historical yields of bonds with similar maturity. Historical yields drive the pricing of bonds more than the price history or the current duration. Historical prices for the fixed-income positions currently held in the portfolio may not exist, and even when historical prices do exist, they may not be relevant to the current characteristics (e.g., maturity) of the instrument. Even if the same bonds existed at the time of the latest financial crisis, their durations would change because of the passage of time.

A is incorrect because using a bond's past price history would mischaracterize the risk of the current portfolio holdings. For this reason, the historical yields are more important in explaining the risks. Historical prices for the fixed-income positions currently held in the portfolio may not exist, and even when historical prices do exist, they may not be relevant to the current characteristics (e.g., maturity) of the instrument.

C is incorrect because historical bond durations would not capture the current characteristics of the bonds in the portfolio. Duration is a sensitivity measure and is the weighted-average time to maturity of a bond. Even if the same bonds existed at the time of the latest financial crisis, their remaining time to maturity and durations would change because of the passage of time.

15 C is correct. Ming suggested in Analysis 1 to use a historical scenario that measures the hypothetical portfolio return that would result from a repeat of a particular period of financial market history. Historical scenarios are complementary to VaR but are not going to happen in exactly the same way again, and they require additional measures to overcome the shortcomings of the VaR.

16 B is correct. Analysis 2 describes surplus at risk. Surplus at risk is an application of VaR; it estimates how much the assets might underperform the liabilities with a given confidence level, usually over a year.

17 B is correct. Incremental VaR measures the change in a portfolio's VaR as a result of adding or removing a position from the portfolio or if a position size is changed relative to the remaining positions.

18 B is correct. McKee suggests running a stress test using a historical scenario specific to emerging markets that includes an extreme change in credit spreads. Stress tests, which apply extreme negative stress to a particular portfolio exposure, are closely related to scenario risk measures. A scenario risk measure estimates the portfolio return that would result from a hypothetical change in markets (hypothetical scenario) or a repeat of a historical event (historical scenario). When the historical simulation fully revalues securities under rate and price changes that occurred during the scenario period, the results should be highly accurate.

A is incorrect because marginal VaR measures the change in portfolio VaR given a very small change in a portfolio position (e.g., change in VaR for a $1 or 1% change in the position). Therefore, marginal VaR would not allow McKee to estimate how much the value of the option-embedded bonds would change under an extreme change in credit spreads.

C is incorrect because sensitivity risk measures use sensitivity exposure measures, such as first-order (delta, duration) and second-order (gamma, convexity) sensitivity, to assess the change in the value of a financial instrument. Although gamma and convexity can be used with delta and duration to estimate the impact of extreme market movements, they are not suited for scenario analysis related to option-embedded bonds.

19 A is correct. VaR has emerged as one of the most popular risk measures because global banking regulators require or encourage the use of it. VaR is also frequently found in annual reports of financial firms and can be used for comparisons.

20 A is correct. VaR is an estimate of the loss that is expected to be exceeded with a given level of probability over a specified time period. The parametric method typically assumes that the return distributions for the risk factors in the portfolio are normal. It then uses the expected return and standard deviation of return for each risk factor and correlations to estimate VaR.

21 B is correct. Value at risk is the minimum loss that would be expected a certain percentage of the time over a certain period of time. Statement 2 implies that there is a 5% chance the portfolio will fall in value by \$90,000 (= \$6,000,000 × 1.5%) or more in a single day. If VaR is measured on a daily basis and a typical month has 20–22 business days, then 5% of the days equates to about 1 day per month or once in 20 trading days.

22 A is correct. Statement 2 indicates that the Equity Opportunities Fund reported a daily VaR value. One of the limitations of VaR is that it focuses so heavily on left-tail events (the losses) that right-tail events (potential gains) are often ignored.

B is incorrect because VaR is viewed as forward looking in that it uses the current portfolio holdings and measures its potential loss. The Sharpe ratio represents a backward-looking, return-based measure and is used to assess the skill of the manager.

C is incorrect because VaR does not provide an accurate risk estimate in either trending or volatile regimes. A portfolio might remain under its VaR limit every day but lose an amount approaching this limit each day. Under such circumstances, the portfolio could accumulate substantial losses without technically breaching the VaR constraint. Also, during periods of low volatility, VaR will appear quite low, underestimating the losses that could occur when the environment returns to a normal level of volatility.

23 C is correct. Measuring VaR at a 5% threshold produces an estimated value at risk of 2.69%.

From Exhibit 6, the expected annual portfolio return is 14.1% and the standard deviation is 26.3%. Annual values need to be adjusted to get their daily counterparts. Assuming 250 trading days in a year, the expected annual return is adjusted by dividing by 250 and the standard deviation is adjusted by dividing by the square root of 250.

Thus, the daily expected return is 0.141/250 = 0.000564, and volatility is $0.263/\sqrt{250} = 0.016634$.

5% daily VaR = $E(R_p) - 1.65\sigma_p = 0.000564 - 1.65(0.016634) = -0.026882$. The portfolio is expected to experience a potential minimum loss in percentage terms of 2.69% on 5% of trading days.

24 C is correct. The change in value of a bond is inversely related to a change in yield. Given a bond priced at B with duration D and yield change of Δy, the rate of return or percentage price change for the bond is approximately given as follows: $\Delta B/B \approx -D\Delta y/(1 + y)$. Under Scenario 3, interest rates decrease by 20 bps. In an environment of decreasing interest rates, the bond with the highest duration will have the greatest positive return. Bond 3 has a duration of 10.2, which is greater than that of both Bond 1 (duration = 1.3) and Bond 2 (duration = 3.7).

25 C is correct. A traditional asset manager uses *ex post* tracking error when analyzing backward-looking returns. The Diversified Fixed-Income Fund prospectus stipulates a target benchmark deviation of no more than 5 bps. Tracking error is a measure of the degree to which the performance of a given investment deviates from its benchmark.

26 B is correct. Position limits are limits on the market value of any given investment; they are excellent controls on overconcentration. Position limits can be expressed in currency units or as a percentage of net assets. The Alpha Core Equity Fund restricts the exposure of individual securities to 1.75% of the total portfolio.

READING

41

Backtesting and Simulation

by Yin Luo, CFA, and Sheng Wang

Yin Luo, CPA, PStat, CFA, is at Wolfe Research LLC (USA). Sheng Wang is at Wolfe Research LLC (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe objectives in backtesting an investment strategy;
☐	**b.** describe and contrast steps and procedures in backtesting an investment strategy;
☐	**c.** interpret metrics and visuals reported in a backtest of an investment strategy;
☐	**d.** identify problems in a backtest of an investment strategy;
☐	**e.** evaluate and interpret a historical scenario analysis;
☐	**f.** contrast Monte Carlo and historical simulation approaches;
☐	**g.** explain inputs and decisions in simulation and interpret a simulation; and
☐	**h.** demonstrate the use of sensitivity analysis.

INTRODUCTION

1

a describe objectives in backtesting an investment strategy

Sarah Koh heads the quantitative research team at Newton Research Pte. SWF Fund, one of Newton's biggest clients, has asked Koh to help develop new investment strategies by rigorously and independently evaluating their risk and return profiles. SWF Fund would like Koh to evaluate the merits of a "value" equity strategy—does owning "cheap" stocks and avoiding (or short-selling) "expensive" stocks add alpha?—as well as two multifactor

© 2021 CFA Institute. All rights reserved.

fundamental strategies that incorporate several other factors besides value. SWF Fund's Investment Committee will use Koh's findings in its decision-making on whether to begin using these strategies.

Koh's work and findings for SWF Fund will be illustrated throughout the reading.

This reading provides an overview of four techniques used to evaluate investment strategies. The first technique, known as **backtesting**, tests a strategy in a historical environment, usually over long periods, answering the question "How would this strategy have performed if it were implemented in the past?" The second technique, **historical scenario analysis**, also known as **historical stress testing**, examines the efficacy of a strategy in discrete historical environments, such as during recessions or periods of high inflation. The third technique, **simulation**, explores how a strategy would perform in a hypothetical environment specified by the user, rather than a historical setting; it is a useful complement to other methods because the past may not recur and only a limited number of all possible future observations for important variables (e.g., interest rates, return correlations, economic growth) is represented in history. Finally, we explore **sensitivity analysis**, which is often combined with simulation to uncover the impact of changing key assumptions.

Increasingly powerful off-the-shelf software has moved these techniques from the realm of specialists to generalists. In a CFA Institute survey of nearly 250 analysts, portfolio managers, and private wealth managers, 50% of respondents reported that they had performed backtesting analysis on an investment strategy in the past 12 months. Although performing these analyses now has fewer technical challenges than before, understanding the steps and procedures, the implicit assumptions, the pitfalls, and the interpretation of results have only increased in importance given the proliferation of these tools. This reading is a starting point on the journey to building this core professional competency.

2 THE OBJECTIVES OF BACKTESTING

Backtesting approximates the real-life investment process by using historical data to assess whether a strategy would have produced desirable results. Although not all strategies that perform well in a backtest will produce excess returns in the future, backtesting can offer investors insight and rigor to the investment process. Conversely, a strategy that does not show efficacy in backtesting could deliver excess returns in the future, but such a strategy is unlikely to be accepted by portfolio managers and investors alike. As a result, backtesting can be employed as a rejection or acceptance criterion for an investment strategy, depending on the investment manager's process.

Backtesting has been widely used in the investment community for many years. Although it fits quantitative and systematic investment styles more naturally, it is also widely used by fundamental managers. Before using a criterion to screen for stocks (such as a valuation metric, for example), a backtest can uncover the historical efficacy of that criterion by determining if its use would have added incremental excess return.

The implicit assumption in backtesting is that the future will at least somewhat resemble history. The reality, however, is more complicated. We attempt to account for the randomness of the future using complementary techniques discussed later in the reading.

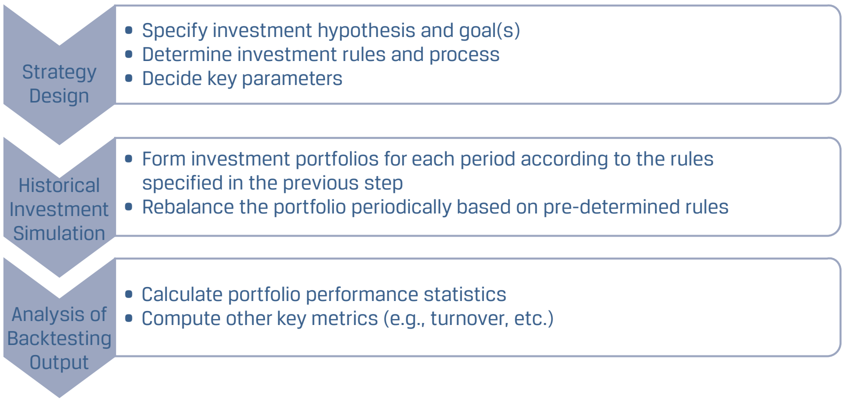
3 THE BACKTESTING PROCESS

b describe and contrast steps and procedures in backtesting an investment strategy

c interpret metrics and visuals reported in a backtest of an investment strategy

Backtesting consists of three steps: strategy design, historical investment simulation, and analysis of backtesting output. Exhibit 1 illustrates these steps and component procedures. We will discuss each step and illustrate them with example backtests of two investment strategies.

Exhibit 1 Backtesting Flowchart



Source: Wolfe Research Luo's QES.

3.1 Step 1: Strategy Design

The first step is to identify the investment goals and hypothesis. For active strategies, the goal is typically to achieve excess returns over the relevant benchmark or superior risk-adjusted absolute return. An investment hypothesis is a method—a trading rule, security selection criterion, a portfolio, etc.—aimed at achieving the goal.

The next step is to translate the hypothesis into rules and processes and to specify several key parameters, so that the hypothesis can be backtested. The key parameters include the investment universe, specific definition of returns, frequency of portfolio rebalancing, and start and end dates.

3.1.1 *Investment Universe*

The investment universe refers to all of the securities in which we can potentially invest. Although academic researchers and specialists typically use the union of Compustat/Worldscope and CRSP,[1] many practitioners use the constituents of well-known broad market indexes as their investment universe. In this reading, unless specified otherwise, we use the Russell 3000 Index, S&P/TSX Composite Index, MSCI China A, and S&P Global Broad Market Index (BMI) for the investment universe for equity strategies in the United States, Canada, mainland China, and all other markets, respectively.

1 CRSP (the Center for Research in Security Prices) provides high-quality data and security returns. The CRSP data series of New York Stock Exchange–listed stocks begins on 31 December 1925.

3.1.2 *Return Definition*

As we extend our investment universe from a single country to a global context, multiple complexities arise, such as currency, trading, and regulatory considerations. For example, we need to decide in what currency the return should be computed. The two most frequent choices are either to translate all investment returns into one single currency—typically the home country currency—or to denominate returns in local currencies. The choice of currency in backtesting often depends on whether the portfolio manager hedges their currency exposures. Managers who do not hedge their exchange rate risk often choose to backtest using single-currency-denominated returns.

If the goal of the investment strategy is excess return, a benchmark must also be specified. The benchmark used is often the benchmark for the client mandate or fund for which the investment strategy under study is applicable. The benchmark should relate to the investment universe; for example, the MSCI China A Index is a logical choice for a strategy that uses the constituents of that index as its universe.

3.1.3 *Rebalancing Frequency and Transaction Cost*

Practitioners often use a monthly frequency for portfolio rebalancing, although higher or lower frequencies are also common. Note that daily or higher frequency rebalancing typically incurs higher transaction costs, and price data will likely be biased by bid–ask spreads, asynchronous trading across different parts of the world, and missing days because of holidays in different countries. Consideration of transaction costs is critical, because many market anomalies simply disappear once they are included. As such, the analyst should explicitly communicate whether transaction costs are included or not in any presentation of the output.

3.1.4 *Start and End Date*

All else equal, investment managers prefer to backtest investment strategies using as long a history as possible, because a larger sample imparts greater statistical confidence in the results. Conversely, however, because financial data are likely to be non-stationary, performance over a long data history should be supplemented with examinations of discrete regimes within the long history (e.g., periods of high and low inflation, recessions and expansions, etc.) using historical scenario analysis, which we will discuss later in the reading.

EXAMPLE 1 STRATEGY DESIGN

After an initial conversation with the investment committee at SWF Fund, Sarah Ko notes the following:

Goal:

Superior risk-adjusted absolute return.

Hypothesis:

"Cheap" stocks—those with lower relative valuations—will outperform "expensive" stocks. In other words, exposure to the "value" factor will lead to outperformance.

Koh must now further specify the hypothesis to allow backtesting, as well as define key parameters for the backtest.

The value factor can be described using almost any combination of market price and fundamental performance measures, on a historical (called trailing) or forward-looking basis. Koh selects a simple valuation metric—trailing earnings yield, the inverse of the P/E—to quantify the "cheapness" of a stock. Although P/E is more commonly understood than earnings yield, a serious flaw is that it cannot be computed or logically interpreted if EPS is zero or negative. Earnings yield, on the other hand, can be computed for any stock so long as EPS and price data are available.

$$\text{Trailing earnings yield} = \frac{\text{Trailing 12-month EPS}}{\text{Current share price}}. \quad (1)$$

Specification of Key Parameters:

- Investment universe: Russell 3000 for the US market and S&P Europe BMI for the European market. Total returns will be hedged back into US dollars.
- Start and end date: Because data required for this strategy are widely available, Koh will use a long time period: January 1986–May 2019.
- Rebalancing frequency: monthly, including transaction costs, but returns on a 12-month moving average basis will be computed.

1 Given the backtesting strategy design outlined here, which of the following is a concern about which the investment committee of SWF Fund should be aware?
   - A The strategy assumes that the US dollar will appreciate against the euro.
   - B The historical period of the data includes recessions, currency regime changes, and periods of varying interest rates.
   - C There are serious issues with computing earnings yield for many stocks.

2 Which of the following describes the relationship between rebalancing frequency and transaction costs?
   - A Changing the rebalancing frequency from monthly to weekly would likely increase transaction costs.
   - B Changing the rebalancing frequency from monthly to quarterly would likely increase transaction costs.
   - C Rebalancing frequency has no effect on transaction costs.

3 Which of the following is not a potential concern of using a short time period for a backtest?
   - A The backtest will cover a limited number of business cycle, inflation, and interest rate regimes.
   - B The backtest may not be useful because the findings may apply only under the conditions present in the time frame.
   - C The backtest is likely to cover multiple business cycle, inflation, and interest rate regimes.

Solution:

1. B is correct. The portfolio manager is using a long data history that includes regime changes in inflation, currencies, and interest rates, so the data is non-stationary. Consequently, backtesting performance results should be supplemented with examinations of performance during the discrete regimes.

A is incorrect because the analysis makes no assumption about exchange rates. C is incorrect because earnings yield can be computed as long as EPS and price data are available.

2 A is correct. Rebalancing frequency refers to how often a portfolio is updated to reflect current data, such as (in this case) changes in earnings yields across the investment universe. Typically, the more frequently rebalancing is done, the more trading is required, which incurs more transaction costs. B is incorrect because it describes a decrease in rebalancing frequency, which would decrease transaction costs. C is incorrect because rebalancing frequency is the primary driver of trading volume, which incurs transaction costs.

3 C is correct. Covering multiple macroeconomic regimes *is not* a concern associated with using a short time period for a backtest, because macroeconomic regimes tend to be multi-year in length. A and B are incorrect because they *are* concerns associated with using a short time period: The backtest may capture only a limited experience, and thus the findings may be relevant for only that experience.

3.2 Step 2: Historical Investment Simulation

The next step is constructing the portfolio to be tested and ensuring that it is rebalanced based on the pre-determined frequency.

The portfolio construction process depends primarily on the investment hypothesis under consideration (e.g., whether it is an entire portfolio, a trading strategy, or a modification of an existing strategy), the investment manager's capabilities and style, and the client's investment mandate for which the potential strategy is relevant (e.g., are there geographical limitations? Are there size and liquidity constraints? Can the manager short stocks?). Although our examples use fundamental factor-based, quantitative equity strategies in which we assume stocks can be shorted, backtesting can be applied to any kind of investment strategy.

To simulate rebalancing, analysts typically use **rolling windows**, in which a portfolio or strategy is constituted at the beginning of a period using data from a historical in-sample period, followed by testing on a subsequent, out-of-sample period. The process is repeated as time moves forward. This approach replicates the live investing process, because investment managers adjust their positions as new information arrives. For example, assume we backtest a value strategy by measuring its performance each month from December 2011 to May 2012. The process begins on 30 November 2011 by compiling every stock's trailing 12-month earnings yield using EPS reported in the previous 12 months (i.e., from December 2010 to November 2011, the in-sample months) divided by stock prices as of 30 November 2011. We then execute the investment strategy—for example, buying stocks with high earnings yields and shorting stocks with low earnings yields—as of that date. Then, we record the investment results for the month of December (i.e., the out-of-sample, OOS, month). The process is repeated at the end of each subsequent month, by rebalancing the portfolio with refreshed trailing 12-month earnings yield data and measuring the results over the ensuing (OOS) month. Exhibit 2 illustrates this process is illustrated.

Exhibit 2 Rolling Window Backtesting of the Earnings Yield Factor

	2010:12	2011:01	2011:02	2011:03	2011:04	2011:05	2011:06	2011:07	2011:08	2011:09	2011:10	2011:11	2011:12	2012:01	2012:02	2012:03	2012:04	2012:05
11/30/2011												In-Sample (Last 12M EPS/Price)	OOS					
12/31/2011													In-Sample (Last 12M EPS/Price)	OOS				
1/31/2012														In-Sample (Last 12M EPS/Price)	OOS			
2/29/2012															In-Sample (Last 12M EPS/Price)	OOS		
3/31/2012																In-Sample (Last 12M EPS/Price)	OOS	
4/30/2012																	In-Sample (Last 12M EPS/Price)	OOS

Source: Wolfe Research Luo's QES.

3.3 Step 3: Analysis of Backtesting Output

The final step in backtesting is generating results for presentation and interpretation. We care about not only the average return of the portfolio but also the risk profile (e.g., volatility and downside risk). Therefore, analysts often use metrics such as the Sharpe ratio, the Sortino ratio, volatility, and **maximum drawdown**. Maximum drawdown is the maximum loss from a peak to a trough for an asset or portfolio.

Beyond these measures, other key performance outputs are visual: for example, time series of returns as well as distributions of returns plotted against a well-known distribution, such as the normal distribution. Visuals are an intuitive way of summarizing many datapoints that often reveal more than a single number summary measure.

It is also useful to examine the backtested cumulative performance of an investment strategy over an extended history. We recommend plotting performance using a logarithmic scale, wherein equal percentage changes are presented as the same vertical distance on the y-axis. Using these cumulative performance graphs, one can readily identify downside risk, performance decay, and structural breaks. Structural breaks, or regime changes, are the result of many exogenous factors and are one reason why the past is not always a good guide to the future. The following are examples of structural breaks:

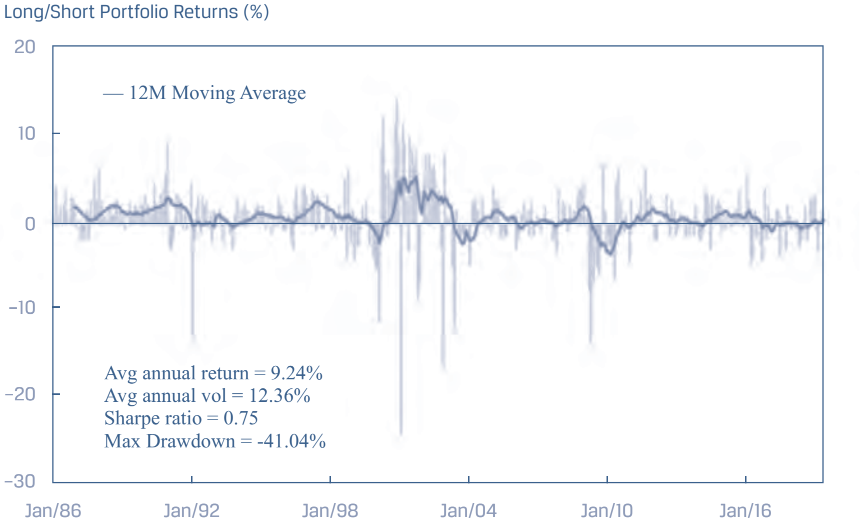
- Depressions and recessions, such as the 2008–09 global financial crisis;
- Geopolitical events, such as changing trade relationships involving countries representing important global equity and bond markets, as well as key countries exiting or entering major trading blocs;
- Major shifts in monetary and fiscal policies, such as the prolonged period of quantitative easing (QE) adopted by major central banks in the aftermath of the 2008 global financial crisis; and
- Major technological changes and advances, such as those that fueled the dot-com bubble and the proliferation of machine learning and artificial intelligence.

We implement our earnings yield-based value strategy as a long–short hedged portfolio, a widely used approach pioneered by Fama and French (1993). In this approach, the analyst sorts the investable stock universe by the relevant metric—trailing earnings yield, in this case—and divides the universe into quantiles (typically into quintiles or deciles) based on those metrics. A long–short hedged portfolio is then formed by going long the top quantile (i.e., the group of stocks with the highest earnings yield) and shorting the bottom quantile (i.e., the group of stocks with the lowest earnings yield). Individual stocks are either equally weighted or market capitalization weighted within each quantile. Although the quantiles may not have equal beta exposure, and they may have exposures to other common factors, the difference in the average earnings yield metric between quantiles is a reasonable and straightforward characterization of underlying performance.

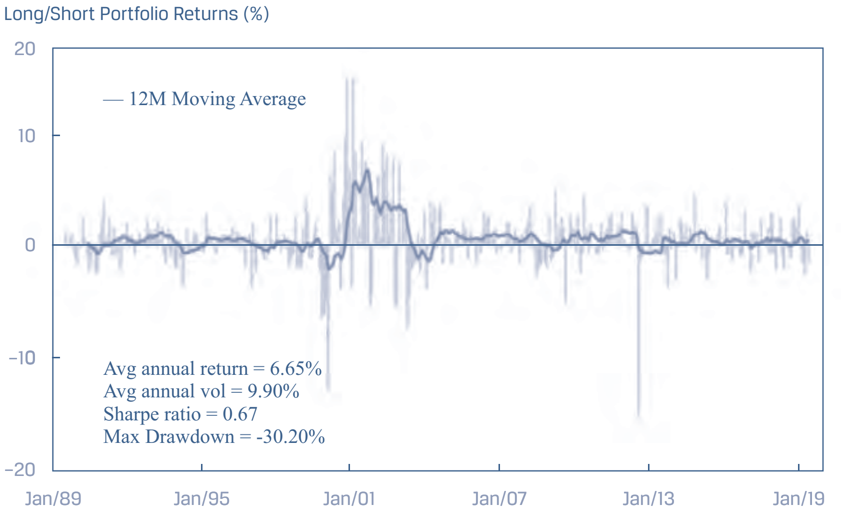
We used quintiles (e.g., top and bottom 20%) and monthly rolling windows and measured the results in several ways for the strategy in the US and European markets, shown in Exhibit 3.

Exhibit 3 Earnings Yield Factor, Long–Short Hedged Quintile Portfolio Returns (January 1986–May 2019)

A. US: Trailing Earnings Yield



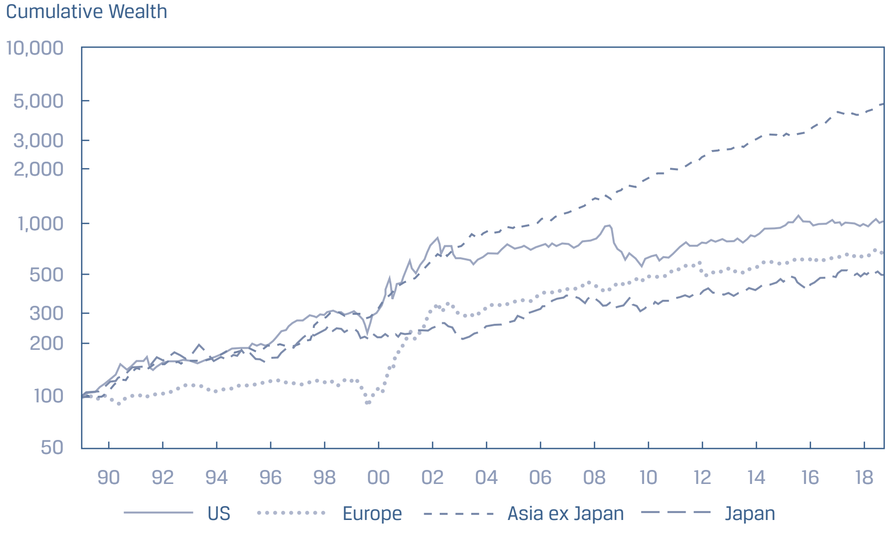
B. Europe: Trailing Earnings Yield



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

We also examined the cumulative performance of the strategy in four different stock markets: the United States, Europe, Asia ex-Japan, and Japan, with the results shown in Exhibit 4.

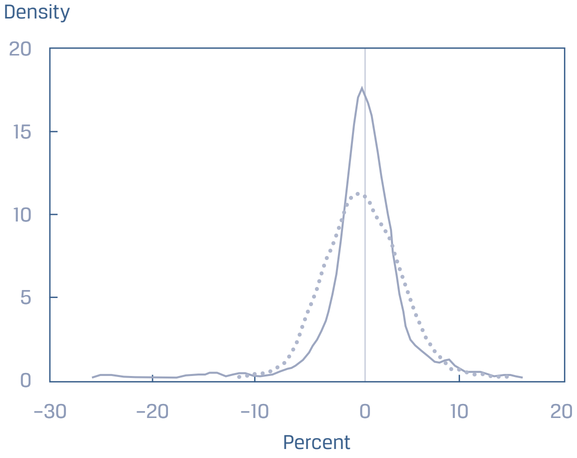
Exhibit 4 Earnings Yield Factor, Long–Short Hedged Quintile Portfolio Returns (January 1986–May 2019) in Several Markets



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, and Wolfe Research Luo's QES.

Finally, we show the distribution of the strategy's returns in the US market against the normal distribution in Exhibit 5.

Exhibit 5 Distribution of Earnings Yield Returns, United States (1986–2019)



Source: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

EXAMPLE 2 HISTORICAL INVESTMENT SIMULATION AND OUTPUT ANALYSIS

1 Describe how the backtest performance of value investing, based on the earnings yield factor, in Europe compares with that in the United States over the 1986–2019 period, as shown in Exhibit 3.

2 Describe the cumulative performance of value investing across the different markets shown in Exhibit 4 and the distributions of returns in the United States from this strategy in Exhibit 5.

Solution to 1:

In the United States, the average annual return from the value investing strategy is about 9.2%, with a Sharpe ratio of 0.75, over the backtesting period (January 1986–May 2019), as seen in Exhibit 3, Panel A. In Europe, the same investment strategy generated a significantly lower (by 250 bps) average annual return, about 6.7%, but with significantly lower volatility (Panel B). Hence, the Sharpe ratio for the European strategy, 0.67, is close to that of the US strategy. In both markets, the maximum drawdown is just over three times the volatility of the strategy. Therefore, as a long-term value strategy, the earnings yield factor offers slightly better performance in the United States than in Europe.

Solution to 2:

The value strategy has delivered strong performance over the long run across the several markets, especially in Asia ex-Japan (Exhibit 4). Performance has flattened since 2016, however, in the United States, Europe, and Japan after first leveling off in all geographies except Asia ex-Japan after 2002. Significant drawdowns and potential structural breaks can also be observed in late 1990s (i.e., during the tech bubble) and in March–May 2009 (i.e., the risk rally during the global financial crisis) in most regions.

More problematically, the strategy in the United States seems to suffer from excess kurtosis (i.e., fat tails) and negative skewness (Exhibit 5). The excess kurtosis implies that this strategy is more likely to generate surprises—that is, extreme returns—whereas the negative skewness suggests that those surprises are more likely to be negative (than positive).

4 BACKTESTING MULTIFACTOR MODELS

b describe and contrast steps and procedures in backtesting an investment strategy

c interpret metrics and visuals reported in a backtest of an investment strategy

Few investment managers use a single signal, such as earnings yield, in an investment strategy. In practice, most quantitative stock selection models use a multifactor structure, with a linear combination of factors being the dominant framework. Similarly, most fundamental managers use multiple filters in their stock screening tools.

In this section, we introduce two multifactor equity portfolio strategies to more richly illustrate backtesting: a benchmark (BM) factor portfolio, which equally weights multiple fundamental factors, and a **risk parity** (RP) factor portfolio, which weights factors based on equal risk contribution. We chose these two approaches because their weighting schemes—equal weights and equal risk weights, respectively—are objective. We will continue to use these two portfolios throughout the reading to discuss other evaluation techniques.

To backtest these two portfolios, we follow the same three steps described previously: strategy design, historical investment simulation, and output analysis.

4.1 Step 1: Strategy Design

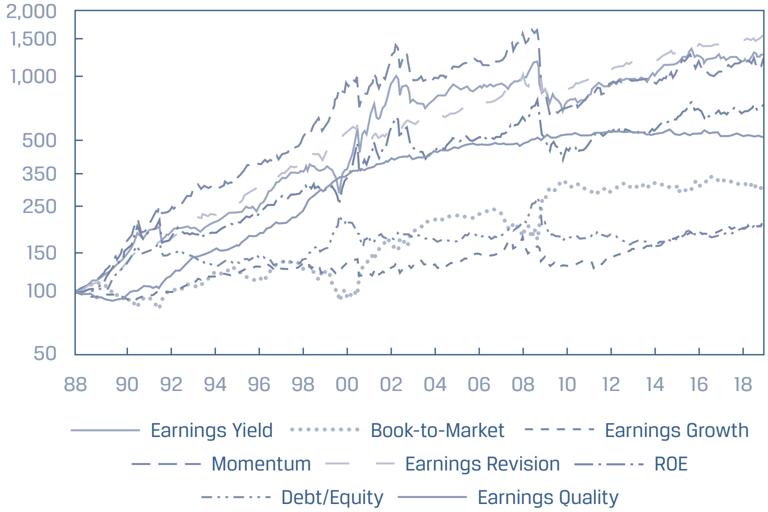
We chose eight fundamental factors from common investment styles:

1. Defensive value: Trailing earnings yield
2. Cyclical value: Book-to-market ratio
3. Growth: Consensus FY1/FY0 EPS growth
4. Price momentum: 12-month total return, excluding the most recent month
5. Analyst sentiment: 3-month EPS revision
6. Profitability: Return on equity (ROE)
7. Leverage: Debt-to-equity ratio
8. Earnings quality: Non-cash earnings (proportion of accruals in earnings)

For each factor, we form a portfolio by buying the top 20% of stocks and shorting the bottom 20% of stocks ranked by the factor. Stocks within both long and short buckets are equally weighted. For illustration purposes, we do not account for transaction costs or other portfolio implementation constraints.

As shown in Exhibit 6 (which uses a logarithm scale on the *y*-axis), all eight factor portfolios have delivered positive returns over the long term (1988–2019) in the United States. Earnings revision, earnings yield, and price momentum factors produced the highest returns, and the earnings growth and debt/equity factors lagged far behind. The eight factor portfolios appear to share some commonalities. Upon visual inspection, returns seem to fall into three clusters: (1) earnings revision, earnings yield, and price momentum; (2) ROE and earnings quality; and (3) book-to-market ratio, earnings growth, and debt/equity. They also show significant dispersions at times.

Exhibit 6 Cumulative Return of Eight Factor Portfolios, United States (1988–2019)



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

For our benchmark portfolio, we combine these eight factor portfolios by equally weighting each one. Researchers have found that such an equally weighted portfolio either outperforms or performs in line with portfolios constructed using more sophisticated optimization techniques (e.g., DeMiguel, Garlappi, and Uppal 2007).

For our risk parity (RP) portfolio, we combine the eight factor portfolios by equally weighting them by their risk contribution. Risk parity is a popular alternative portfolio construction technique that accounts for the volatility of each factor and the correlations of returns among all factors in the portfolio. The objective is for each factor to make an equal (hence "parity") risk contribution to the overall risk of the portfolio.

We backtested our two portfolios in each of the following markets: the United States, Canada, Latin America (LATAM), Europe, the United Kingdom, emerging Europe, Middle East, and Africa (EMEA), Asia ex-Japan, Japan, Australia and New Zealand (ANZ), and mainland China. Both portfolios are rebalanced monthly to maintain equal factor weights or equal factor risk contributions (i.e., risk parity). Although each of the eight underlying factor portfolios is a long–short portfolio, our BM and RP multifactor portfolios are long only, meaning the weights allocated to each factor portfolio are restricted to be non-negative, such that weights for each of the underlying portfolios are all positive and sum to 100%.

4.2 Step 2: Historical Investment Simulation

Backtesting a multifactor strategy is similar to the method introduced earlier, but the rolling-window procedure is implemented twice, once at each portfolio "layer."

First, we form eight factor portfolios at each given point in time (i.e., monthly) from 1988 until May 2019 using the rolling-window procedure discussed previously. We then combine these factor portfolios into two multifactor portfolios, each with different weights: equal weighted (BM portfolio) and equal risk weighted (RP portfolio).

A second rolling-window procedure over the same time span is required to avoid look-ahead bias. At each month end, the previous five years of monthly data are used to estimate the variance–covariance matrix for the eight factor portfolios. This is the most important ingredient to form the RP portfolio. Once the covariance matrix is estimated, we can optimize and compute the weights (i.e., weights for equal risk contribution) for each of the eight factor portfolios and then form the RP portfolio.

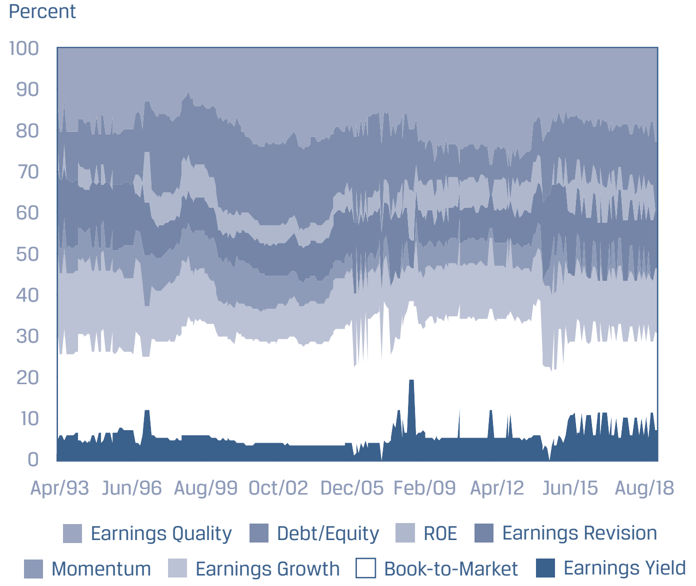
Finally, we compute the returns of the two multifactor portfolios (BM and RP) during each "out-of-sample" month from 1988 to May 2019.

4.3 Step 3: Output Analysis

Exhibit 7, Panel A, shows that the weights of the eight factor portfolios in the RP portfolio are relatively stable over time (1993–2019) in the United States, but they are certainly not equal—so we should expect the RP portfolio's risk and return profile to differ from that of the BM portfolio. Notably, book-to-market and earnings quality factor portfolios receive the largest allocations, whereas ROE and price momentum factor portfolios have the lowest weights. Although the RP portfolio appears to deliver a lower cumulative return than does the BM portfolio (Panel B), Panel C shows that the RP portfolio's volatility is less than half the volatility of the BM portfolio. As a result, the RP portfolio's Sharpe ratio is nearly twice that of the BM portfolio (Panel D).

Exhibit 7 Backtesting Multifactor Strategies: Equally Weighted Benchmark Portfolio vs. Risk Parity Weighted Portfolio

A. RP Portfolio Allocation Weights in the US



B. Cumulative Return

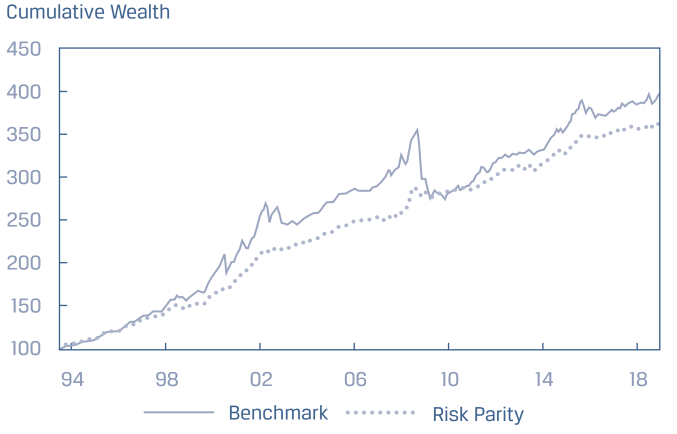
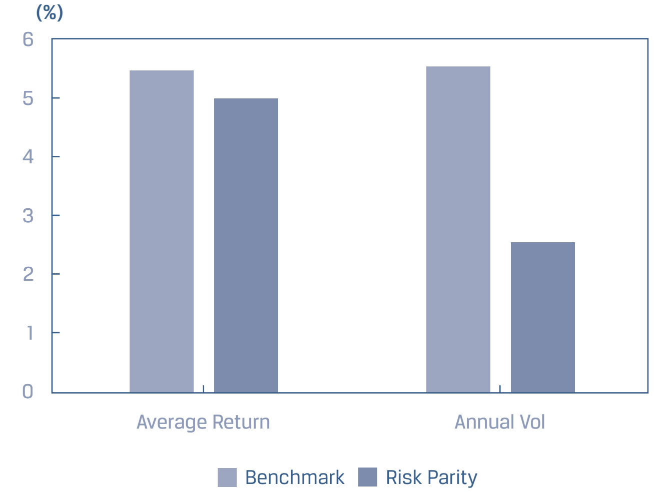
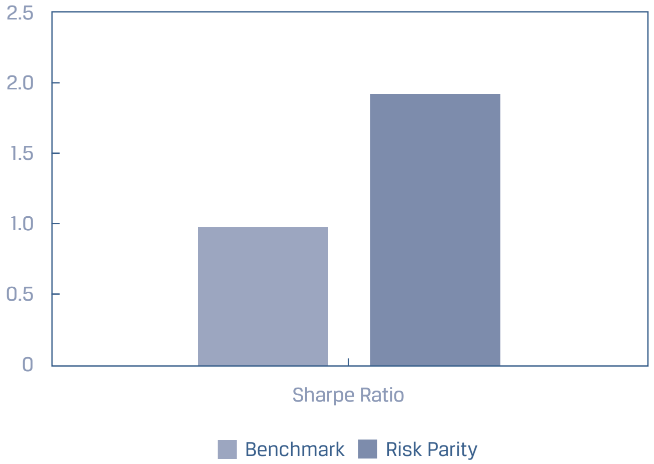


Exhibit 7 (Continued)

C. Average Return and Volatility



D. Sharpe Ratio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

Exhibit 8 presents statistics for the return distributions of the eight factor portfolios and the equally weighted BM and RP weighted multifactor portfolios from 1993 to 2019. Six of the eight factor portfolios have negative skewness (the BM portfolio does as well), and all factors and factor allocation portfolios show excess kurtosis (i.e., kurtosis exceeding 3.0). The downside risk (i.e., minimum monthly return) is clearly greater in magnitude than the maximum upside for most factor strategies. The two factor allocation strategy portfolios—BM and RP—both display moderate mean returns (0.5% and 0.4% per month, respectively) and low standard deviations (1.6% and 0.7% per month, respectively) compared with the eight underlying factor portfolios, highlighting the diversification benefits from factor allocation decisions.

Exhibit 8 Monthly Return Distributions: Factor, BM, and RP Portfolios (1993–2019)

	Earnings Yield	Book-to-Market	Earnings Growth	Momentum	Earnings Revision	ROE	Debt/Equity	Earnings Quality	Benchmark	Risk Parity
Mean	0.7%	0.4%	0.2%	0.6%	0.7%	0.5%	0.1%	0.4%	0.5%	0.4%
Median	0.6%	0.1%	0.4%	0.8%	0.8%	0.6%	0.1%	0.4%	0.5%	0.4%
Maximum	14.5%	28.9%	6.2%	11.7%	9.1%	10.8%	11.9%	5.3%	4.3%	3.7%
Minimum	(24.0%)	(12.1%)	(15.8%)	(32.7%)	(18.7%)	(28.0%)	(17.1%)	(2.6%)	(10.9%)	(2.5%)
Std. Dev	3.8%	3.7%	2.1%	4.6%	2.4%	3.9%	2.5%	1.2%	1.6%	0.7%
Skewness	(1.00)	2.82	(2.46%)	(2.36)	(2.39)	(1.92)	(0.58)	0.41	(2.40)	0.51
Kurtosis	11.06	23.61	17.80	16.56	20.76	14.96	11.55	3.87	17.78	5.37

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

Exhibit 9 compares the various downside risk measures for the eight factor portfolios and the BM and RP portfolios from 1993 to 2019. The three downside risk measures—value at risk (VaR), conditional value at risk (CVaR), and maximum drawdown—suggest that the price momentum factor, followed by the ROE factor, has the largest downside risk. The smallest downside risk is observed for the earnings quality factor. The risk parity portfolio shows considerably less downside risk than any of the eight underlying factors and the benchmark portfolio. This evidence suggests that the RP strategy benefits greatly from risk diversification, at least in the United States for the period under investigation.

Exhibit 9 Downside Risk Using Monthly Returns: Factor, BM, and RP Portfolios (1993–2019)

	Earnings Yield	Book-to-Market	Earnings Growth	Momentum	Earnings Revision	ROE	Debt/Equity	Earnings Quality	Benchmark	Risk Parity
VaR(95%)	(5.9%)	(0.7%)	(3.9%)	(8.4%)	(3.7%)	(6.8%)	(4.0%)	(1.3%)	(2.6%)	(0.7%)
CVaR(95%)	(14.3%)	(11.1%)	(10.9%)	(22.9%)	(12.8%)	(18.7%)	(8.4%)	(1.7%)	(7.9%)	(0.9%)
Max Drawdown	41.0%	35.3%	27.2%	59.7%	23.9%	47.5%	41.8%	8.3%	22.6%	3.8%

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

EXAMPLE 3 RISK AND RETURN BEYOND NORMAL DISTRIBUTION

Compare return profiles for the BM and RP strategy multifactor portfolios and explain which investment strategy offers the more attractive statistical properties for risk-averse investors (refer to Exhibits 8 and 9).

Solution:

The BM and RP portfolios have nearly the same mean monthly returns, at 0.5% and 0.4%, respectively (Exhibit 8). Although the maximum returns are similar, the RP portfolio has a much smaller minimum return (–2.5%) and a significantly lower standard deviation (0.7%) compared with those of the BM portfolio (–10.9%

and 1.6%, respectively). The RP portfolio is also slightly positively skewed (0.51%) and has moderate kurtosis (5.37), in contrast to the negative skew (−2.40%) and high kurtosis (17.78) of the BM portfolio.

The RP portfolio offers similar returns, less downside risk (confirmed by its superior VaR, CVaR, and maximum drawdown results in Exhibit 9), lower volatility, and slightly higher probability of positive returns (i.e., positive skew) compared with the BM portfolio. It is also less fat tailed (i.e., moderate kurtosis, meaning lower probability of extreme negative surprises) than the BM portfolio. Therefore, the RP portfolio has the more attractive distribution properties for risk-averse investors.

EXAMPLE 4 BACKTESTING THE PERFORMANCE OF FACTOR ALLOCATION STRATEGIES

During the presentation of her backtesting results to SWF's investment committee, Koh is asked the following questions:

1 Regarding rolling-window backtesting, which one of the following statements is *inaccurate*?

   A The data are divided into just two samples.

   B Out-of-sample data become part of the next period's in-sample data.

   C Repeated in-sample training and out-of-sample testing allow managers to adjust security positions on the basis of the arrival over time of new information.

2 Which of the following is a drawback of the long–short hedged portfolio approach for implementing factor-based portfolios?

   A The hedged portfolio is formed by going long the top quantile (with the best factor scores) and shorting the bottom quantile (with the worst factor scores).

   B Securities must be ranked by the factor being scrutinized and then grouped into quantiles based on their factor scores.

   C Not every manager can short stocks.

3 Which one of the following is *not* a metric or visual used in assessing backtesting of a factor-based investment strategy?

   A Distribution plots of factor returns

   B A word cloud of text describing the characteristics of the factor

   C Maximum drawdown

4 Regarding the use of rolling-window backtesting in assessing factor allocation to a risk parity–based strategy, which statement is correct?

   A The procedure is used once for estimating factor returns over the rolling window.

   B The procedure is used once for dividing the data into just two samples.

   C The procedure is used twice—once for estimating factor returns over the rolling window, and a second time for estimating the covariance matrix of factor returns (for deriving risk parity weights) over the rolling window.

Solution to 1:

A is correct, because the statement is inaccurate. B and C are incorrect, because they accurately describe the rolling-window backtesting technique.

Solution to 2:

C is correct, because it best describes a drawback of the long–short hedged portfolio approach. A and B are incorrect because they describe the approach itself.

Solution to 3:

B is correct, because a word cloud is not a visual used in assessing backtesting of a factor-based investment strategy. A and C are incorrect, because they are visuals and metrics, respectively, used to assess backtests of factor-based strategies.

Solution to 4:

C is correct, because the procedure must be used a second time for estimating the covariance matrix of factor returns (for deriving risk parity weights) over the rolling window. A is incorrect because the procedure must be done twice: once for estimating factor returns over the rolling window and a second time for estimating the covariance matrix of factor returns (for deriving risk parity weights). B is incorrect because the rolling-window procedure divides the sample into many samples.

5 COMMONS PROBLEMS IN BACKTESTING

d identify problems in a backtest of an investment strategy

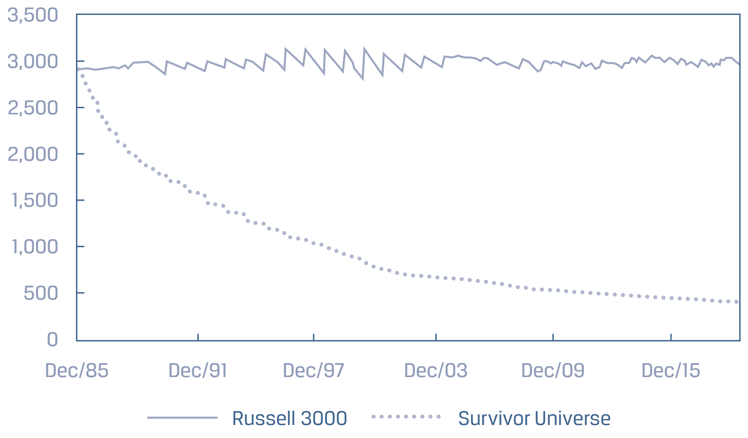
In this section, we discuss some of the most common mistakes investors make when they conduct backtests. Although backtesting is the subject of the discussion, all of these mistakes are relevant to and commonly found in quantitative research generally.

5.1 Survivorship Bias

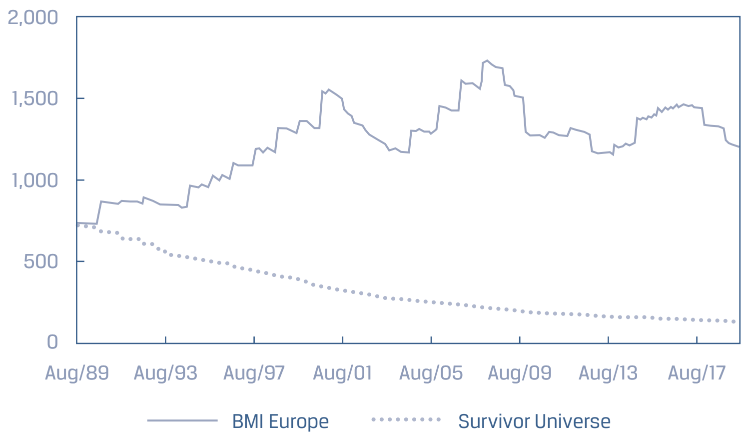
Companies continually appear and disappear from market indexes. New firms appear via IPOs, spin-offs, and outperformance. Companies disappear for many reasons, including privatization, acquisition, bankruptcy, and prolonged under- or outperformance that results in a change in market capitalization from large to mid/small and vice versa. As shown in Panel A of Exhibit 10, fewer than 400 of the constituents of the Russell 3000 Index in 1985 (less than 13%) are still included in the index as of 31 May 2019. Similarly, the S&P BMI Europe Index, which tracks the broad European market, started with about 720 stocks in 1989 and now contains around 1,200 companies. Among the 720 stocks in the index at inception, only 142 (or about 20%) were still in the index as of May 2019 (Panel B of Exhibit 10). Stocks that have remained in the index over time are referred to as "survivors."

Exhibit 10 Number of Stocks in Index vs. Survivors

A. US (Russell 3000)



B. Europe (S&P BMI)



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

Survivorship bias refers to deriving conclusions from data that reflects only those entities that have survived to that date. It is one of the most obvious but, interestingly, also one of the most common mistakes that investors make when conducting backtests. Although the problem is widely covered in the academic literature, relatively few practitioners, whether investing in equities, fixed income, hedge funds, or other asset classes, bother to quantify the implications of survivorship bias in their backtesting.

Some investors contend that because you can invest only in companies that exist today, there is nothing wrong with backtesting strategies using only the current index constituents. The problem is, however, that in the past, one could not know which companies would survive in the future, which companies would disappear, and which companies would be created and become successful enough to be added to the index. Moreover, the list of surviving firms is likely biased in one way or another—for example, it could represent primarily multinational firms, or highly innovative firms, or the most successful firms.

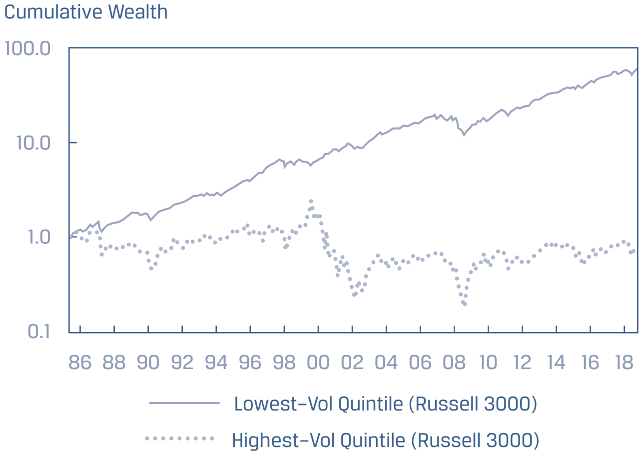
Although it is straightforward (but definitely not recommended) to backtest an investment strategy using only the survivors, tracking all companies that have ever existed in a correct point-in-time fashion (i.e., the casualties as well as the survivors) is strongly recommended, especially as such data becomes more available from data vendors. **Point-in-time data** allow analysts to use the most complete data for any given prior time period, thereby enabling the construction (and backtesting) of the most realistic investment strategies.

The difference between backtesting with current index constituents and point-in-time data is illustrated using the low-volatility anomaly, a popular investment strategy that argues that stocks with low volatility tend to outperform high-volatility stocks. A proper backtesting methodology using the point-in-time Russell 3000 universe in Panel A of Exhibit 11 confirms this view; low-volatility stocks have significantly outperformed high-volatility stocks over the three decades up to 2019.

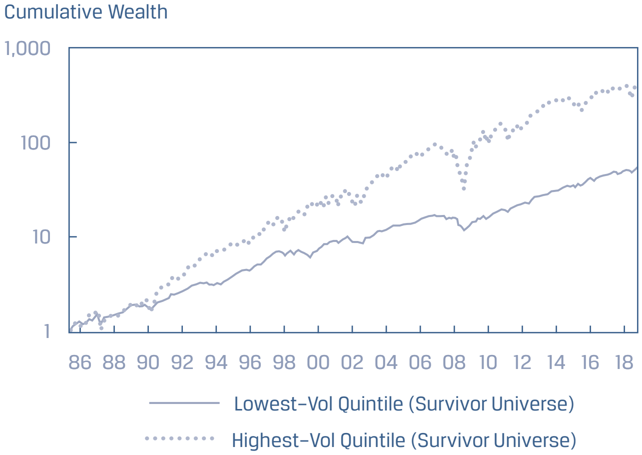
Importantly, however, if we repeat the backtesting exercise using only survivors, then the result is the opposite: high-volatility stocks outperformed low-volatility stocks by about 5.5 times (see Panel B of Exhibit 11). This example underscores the importance of accounting for survivorship bias in backtesting by using point-in-time index constituent stocks and not just the current survivors.

Exhibit 11 Survivorship Bias and the Low-Volatility Anomaly

A. Using a Point-in-Time Universe



B. Using the Survived Companies



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

5.2 Look-Ahead Bias

Another common mistake investors make in backtesting is failing to recognize and account for **look-ahead bias**. This form of bias is created by using information that was unknown or unavailable during the historical periods over which the backtest is conducted. Survivorship bias is actually a type of look-ahead bias, because the question of whether a stock will survive or be added to an index in the future is unknown during the earlier periods over which the backtesting occurs. Look-ahead bias is likely the most common mistake that practitioners make when performing backtesting. It can be overcome by using point-in-time data, which, again, might not be available. Look-ahead bias has several common forms: reporting lags, revisions, and index additions.

The first common form of look-ahead bias derives from reporting lags. For example, in conducting a backtest for year-end 2018, we would not have EPS results for the quarter ending 31 December 2018 for all publicly traded companies until some point around 31 March 2019, although many larger-cap companies might report by 31 January 2019. So, to avoid look-ahead bias, analysts typically compensate by adding several months of reporting lag for every company. This process can also introduce stale information, however. If we continue the example, by 31 January 2019 many larger-cap companies will have already reported earnings, but others, especially mid- and small-cap companies, will not have done so. By using a uniform lag assumption across all companies, the analyst will use stale financial data for some larger-cap companies.

A second problem is data revisions: Macroeconomic data are often revised multiple times, and companies often re-state their financial statements. Many databases keep only the latest numbers, replacing the past figures with the revised ones, although the revised figures were obviously not available at the original release date. By using such revised data, an analyst trying to build realistic investment scenarios going back in time would be using information that was unavailable at that point in time.

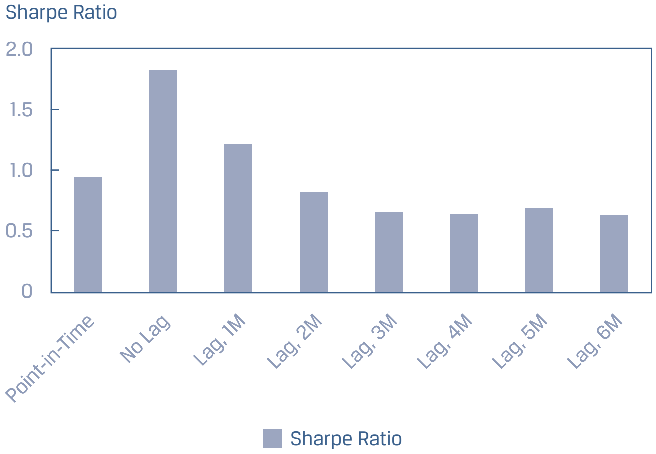
Another form of look-ahead bias arises when data vendors add new companies to their databases. When doing so, they often add several years of historical financial statements into the system. Thus, an analyst backtesting with the current database would be using information on companies that were not actually in the database during the backtesting period. The consequence of this look-ahead bias is often overly optimistic results.

To demonstrate the impact of look-ahead bias and the reporting lag assumption, we conduct monthly backtesting using the earnings yield strategy discussed previously. We compared the backtest results using a proper point-in-time database with the actual EPS data as of each month end, against reporting lag assumptions ranging from zero to six months (a zero lag assumption would suffer from full look-ahead bias).

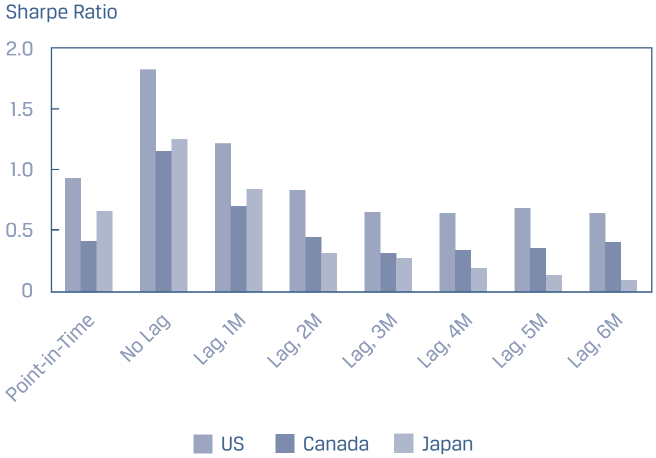
As shown in Exhibit 12, Panel A, it is clear from the backtesting results of the point-in-time scenario against the no-lag scenario that look-ahead bias inflates the performance of our value strategy in the United States by almost 100%. The impact of look-ahead bias is evident in all regions. In the United States, Canada, and Japan (Panel B), it appears that a reporting lag of between one and two months produces backtest results that are consistent with those of the proper point-in-time data. In Europe, the United Kingdom, and ANZ (Panel C), a lag assumption of between two and three months appears appropriate, whereas for Asia ex-Japan, LATAM, and emerging EMEA (Panel D), the point-in-time consistent lag assumption increases to three months. These different lag assumptions reflect the timeliness with which companies in each region report their earnings.

Exhibit 12 Look-Ahead Bias: Impact on Backtesting of Reporting Lag Assumptions (1986–2016)

A. US



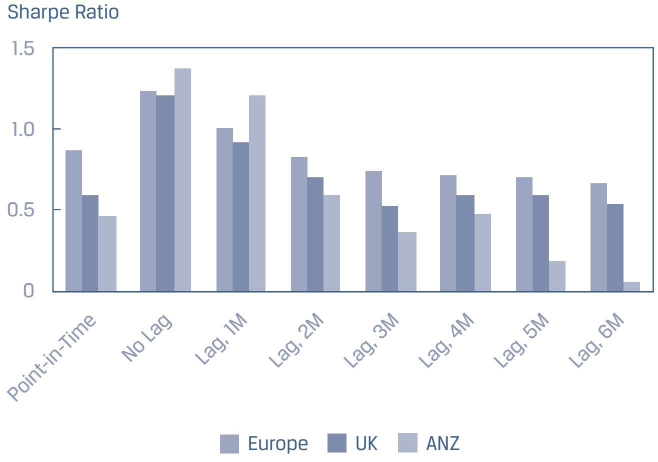
B. US, Canada, and Japan



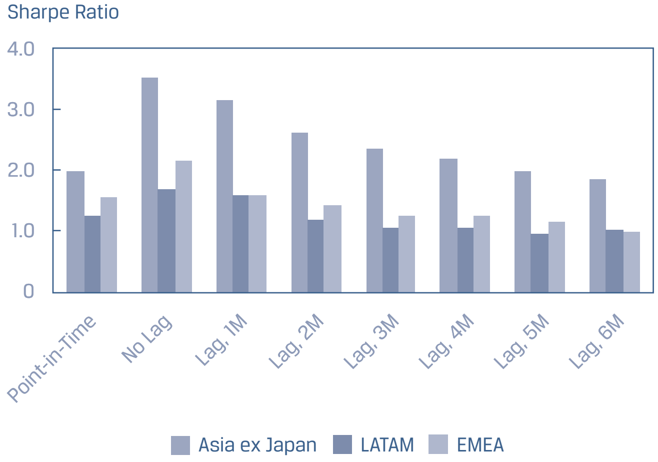
(continued)

Exhibit 12 (Continued)

C. Europe, UK, and ANZ



D. Asia ex Japan, LATAM, and EMEA



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

5.3 Data Snooping

There is often a temptation to substitute sound portfolio construction by simply backtesting many strategies and picking the best-performing strategy. This bias is called **data snooping**—making an inference after looking at statistical results rather than testing a prior inference. Otherwise known as "p-hacking," data snooping occurs when an analyst selects data or performs analyses until a significant result is found. It can take many forms, including performing interim analyses to decide whether to continue collecting data, using many variables and deciding which to report later; dropping outliers only after performing analyses; and so on. The ultimate results are often false positives.

Data snooping may be mitigated by setting a much higher hurdle than typical—for example, a t-statistic greater than 3.0—for assessing whether a newly discovered factor is indeed adding incremental value (i.e., is statistically significant). Another technique to detect and mitigate data snooping is cross validation, in which the analyst partitions the dataset into training data and testing data (i.e., "validation data") and tests a model built from the training data on the validation data. Rolling window backtesting is a form of cross-validation, albeit in a deterministic and non-random manner, as past periods (i.e., in-sample periods) are used to train a model that is applied to the next (i.e., out-of-sample) period.

EXAMPLE 5 DATA SNOOPING IN INVESTMENT MANAGEMENT

A research analyst has just presented her risk factor–based quantitative/systematic investment model for the UK market to you and several other portfolio managers. She reports the development and backtesting of several different models: The number of factors ranged from 5 to 10, rebalancing periods were monthly and quarterly, and rolling windows were implemented for 5, 15, and 25 years of historical data. She recommends the 10-factor model (with monthly rebalancing) because backtesting of 15 years of data generated the following annualized performance metrics: Sharpe ratio of 3.0 and realized volatility of 1.0%. She also reports a *t*-statistic of 2.5 and a *p*-value of 1.3% for this model of UK market returns, which were the highest and lowest statistics, respectively, of all the models.

Describe the concerns you should raise around the issue of data snooping for this seemingly very attractive strategy.

Solution:

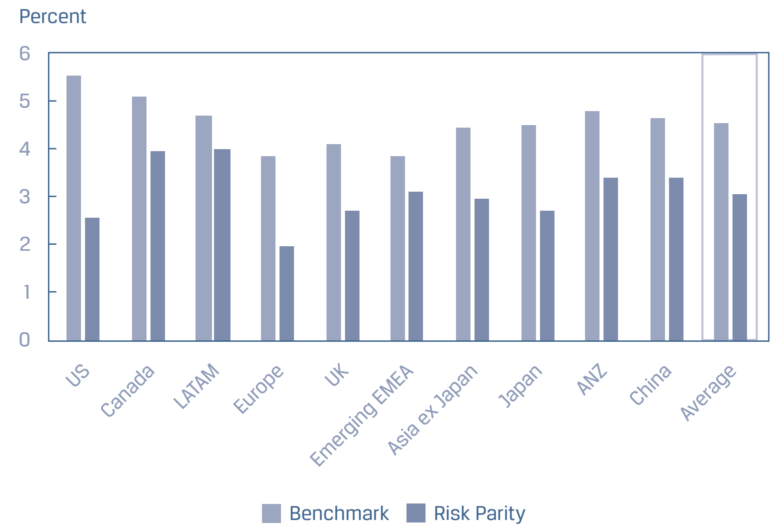
As a portfolio manager, you must be careful in assessing these performance results in light of how the analyst developed and backtested her model. For example, it is critical to know whether backtesting has incorporated transaction costs and trading liquidity. More importantly, however, you need to understand whether data snooping was involved in developing this model/strategy. Given the many variations of models developed and tested by the analyst, it is highly likely that her process suffers from model selection bias. Recommending the model with the highest *t*-statistic and lowest *p*-value also points to data snooping. One way to mitigate the problem is to raise the hurdle for an acceptable model to a *t*-statistic exceeding 3.0 (thereby lowering the *p*-value). The analyst should also consider other techniques that can be used to better understand the true performance of this model/strategy (i.e., cross-validation).

A common way to perform cross validation is to use data from different geographic regions. For example, if the risk parity strategy is developed and tested initially using US equities, the same strategy can be tested in other markets globally to assess whether risk parity is a robust factor allocation strategy.

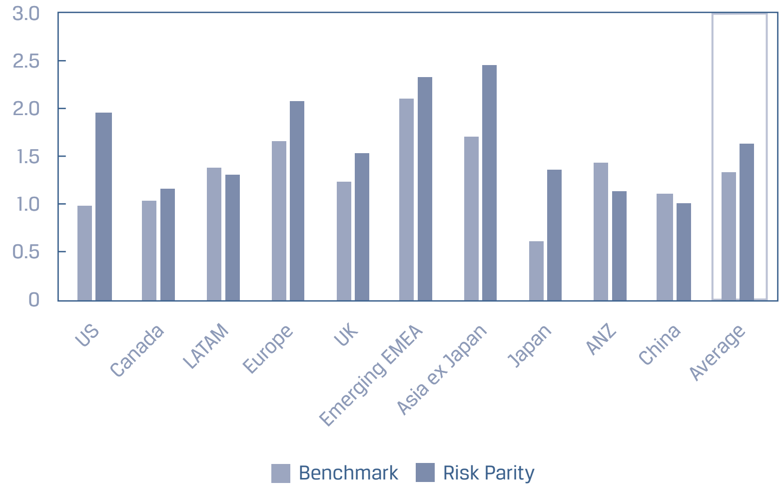
As shown in Exhibit 13, Panel A, as a risk-based factor allocation technique, the RP strategy does indeed deliver a lower realized volatility (i.e., standard deviation of returns) than the benchmark (i.e., equal-weighted factor) strategy in all 10 global markets over 1993–2019. Similarly, the RP portfolios also outperform the BM portfolios in terms of Sharpe ratio (Panel B) in 7 of the 10 global markets.

Exhibit 13 Global Cross-Validation, Equally Weighted Benchmark Portfolio vs. Risk Parity Weighted Portfolio (1993–2019)

A. Realized Volatility



B. Sharpe Ratio



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

EXAMPLE 6 COMMONS PROBLEMS IN BACKTESTING

1 An analyst develops an investment strategy by picking the strategy with the highest t-statistic and lowest p-value after backtesting dozens of different strategies. This approach is an example of which common problem in backtesting?

A Reporting lag

B Survivorship bias

C Data snooping

2 Point-in-time data are useful for avoiding the following problems that may affect backtesting *except*:

A data snooping.

B survivorship bias.

C look-ahead bias.

3 The fact that GDP figures for a quarter are not released by government statistical agencies until approximately 30 days after the quarter ends and often undergo several revisions thereafter creates a problem known as:

A data snooping.

B survivorship bias.

C reporting lag.

4 Which of the following is an example of cross-validation?

A Maximum drawdown

B Backtesting with out-of-sample data

C Incorporating point-in-time data

5 An analyst performed a backtest on an investment strategy in June 2019, selecting the constituents of the Russell 3000 Index as the investment universe, and December 1985 and May 2019 as the start and end dates, respectively. While discussing the results with some colleagues, the analyst was shown lists of the Russell 3000 Index constituents as of December 2005 and December 1995. She noticed that the lists included only 2,250 and 1,500 companies, respectively, of the Russell 3000 companies at May 2019. The analyst must correct her backtest for which problem?

A Data snooping

B Reporting lag

C Look-ahead bias

Solution:

1 C is correct. Data snooping refers to making an inference—such as formulating an investment strategy—*after* looking at statistical results rather than testing a prior inference. A is incorrect because reporting lag refers to the fact that data describing a period is often available only after the period ends and is often subject to revision. B is incorrect because survivorship bias is a form of look-ahead bias in which results are based on a limited, biased sample of subjects (e.g., only surviving companies).

2 A is correct. An analyst can still use a point-in-time dataset to make an inference based on statistical results rather than testing a prior inference. B and C are incorrect, because point-in-time data are useful for avoiding look-ahead bias and survivorship bias (a special case of look-ahead bias). Point-in-time data explicitly corrects for what is not known at a given point in time.

3 C is correct. Reporting lag refers to the fact that data describing a period is often available only after the period ends and is often subject to revision, which certainly is true of GDP data.

4 B is correct. Cross-validation is a technique that involves testing a hypothesis on a different set of data than that which was used to form the inference or initially test the hypothesis. Choice B is the definition of cross-validation.

5 C is correct. The dataset the analyst uses assumes that the Russell 3000 Index constituents as of May 2019 are the same companies that constituted the index throughout the entire backtesting period. The backtest suffers from look-ahead bias, so conclusions drawn from it will be erroneous because it includes companies that did not exist (or were not index members) over the period starting in December 1985. To correct this problem, the analyst should use a dataset of point-in-time constituents of the Russell 3000 Index.

6 HISTORICAL SCENARIO ANALYSIS

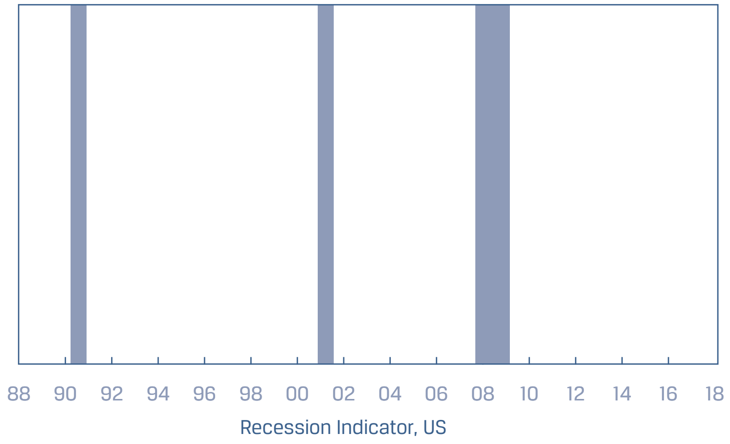
e evaluate and interpret a historical scenario analysis;

Rather than simply acknowledging or even ignoring structural breaks evident in backtesting results, an analyst should pay careful attention to different structural regimes and impacts to a strategy during regime changes. Historical scenario analysis is a type of backtesting that explores the performance and risk of an investment strategy in different structural regimes and at structural breaks. Two common examples of regime changes are from economic expansions to recessions and from low-volatility to high-volatility environments (and vice versa):

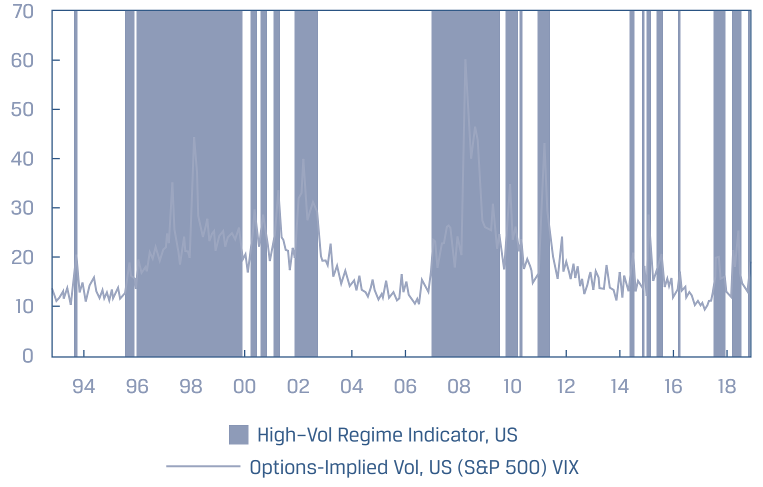
- *Expansions and recessions.* In the United States, since the start of our risk parity allocation strategy in 1993, the National Bureau of Economic Research (NBER) has recognized two official recessions: March 2001–November 2001 and December 2007–June 2009. These recessions are shown in Panel A of Exhibit 14. Although we ignore look-ahead bias in this brief example, it is important to note that business cycle inflection points—the beginning and end of expansions and recessions—are observed only in hindsight. For example, NBER did not identify December 2007 as the beginning of a recession in the United States until December 2008, and it did not identify June 2009 as the end of that recession until September 2010.
- *High- and low-volatility regimes.* The Chicago Board Options Exchange (CBOE) computes the VIX index, which gauges options-implied volatility on the S&P 500 Index. To transform the VIX into a volatility regime indicator, a five-year moving average is computed. Then, the periods when the VIX is above (below) its five-year moving average are defined as high-volatility (low-volatility) regime periods, as shown in Panel B of Exhibit 14 for 1993–2019.

Exhibit 14 Regime Changes

A. Recession Indicator



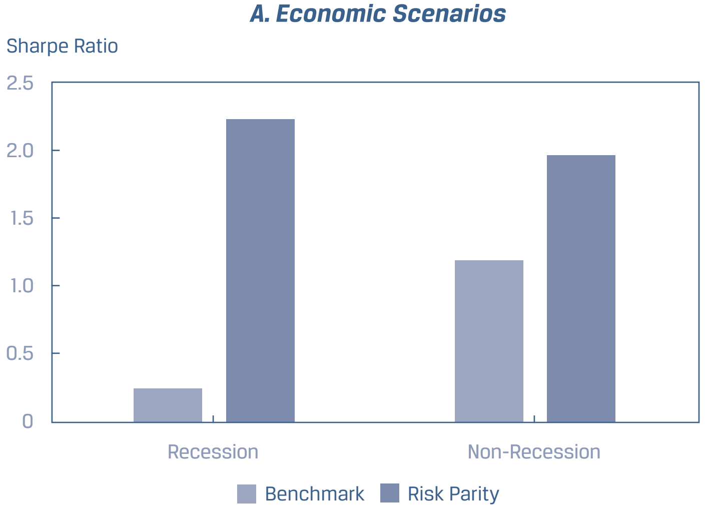
B. VIX: High- vs. Low-Volatility Regimes

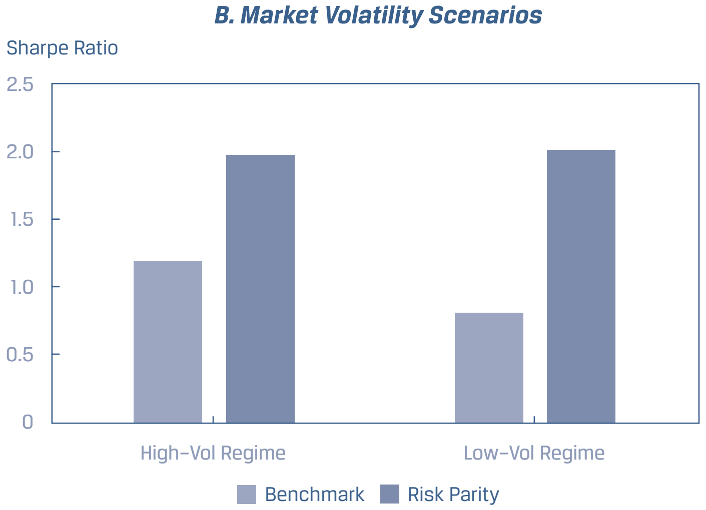


Sources: Bloomberg Finance LLP, FTSE Russell, Haver, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

We can examine the benchmark and risk parity factor portfolios with respect to these two regimes—recession versus expansion and high volatility versus low volatility. As shown in Panel A of Exhibit 15, in terms of the Sharpe ratio, the RP strategy is quite robust to recession and the BM strategy struggles in recessions. Panel B of Exhibit 15 reveals that the BM strategy's performance is slightly worse in low-volatility regimes than in high-volatility regimes, whereas the RP strategy performs equally well in both volatility environments.

Exhibit 15 Sharpe Ratio for BM and RP Portfolios in Different Macro Scenarios (1993–2019)

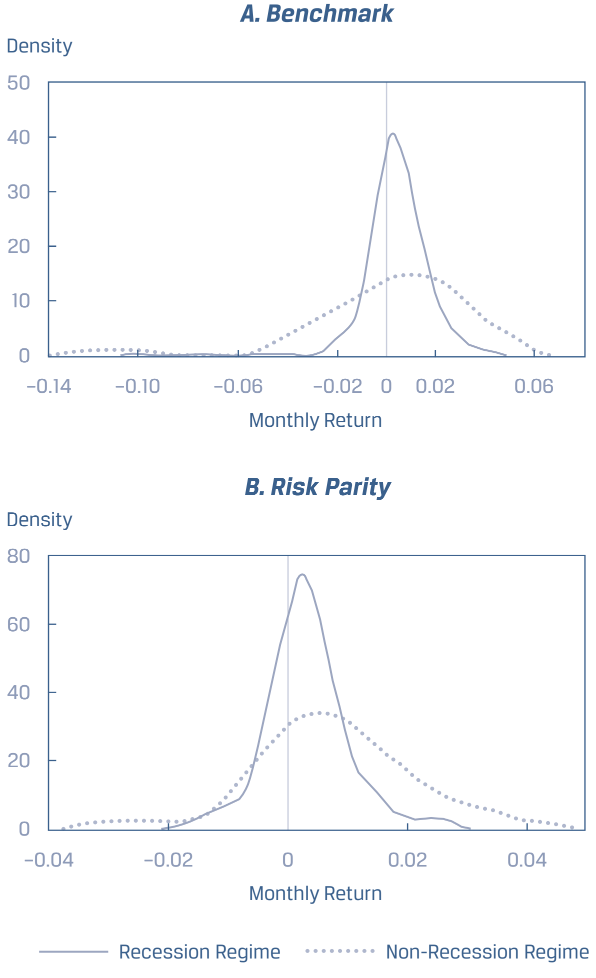




Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

In addition to the Sharpe ratio, a probability density plot can reveal additional information about the sensitivity of the return distributions of these investment strategies—for example, during recession versus non-recession periods. As shown in Exhibit 16, the distribution of returns for both the BM and RP strategies is flatter in a non-recession environment, which implies higher standard deviations during these regimes. The BM strategy suffers from negative skewness and excess kurtosis (i.e., fat tails to the left), regardless of the recession regime, but its average return is clearly lower in a recession environment (Panel A). The RP strategy also has a lower average return in the recession regime (Panel B), but its volatility and kurtosis are both also much lower compared with those of the BM strategy.

Exhibit 16 Distribution of Returns for Factor Allocation Strategies: Recession and Non-Recession Regimes



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

SIMULATION ANALYSIS

7

f contrast Monte Carlo and historical simulation approaches;

In backtesting, we essentially assume that we can go back in time, apply our investment strategies, rebalance our portfolio(s), and measure performance. This idea is intuitive because it mimics how investing is done in reality—that is, forming our ideas, implementing our strategies, and incorporating new information as it arrives.

Backtesting implicitly assumes that the past is likely to repeat itself, however, and this assumption does not fully account for the dynamic nature of financial markets, which may include extreme upside and downside risks that have never occurred before. We now explore how simulation can provide a more complete picture.

There are two basic types of simulation: historical and Monte Carlo. In **historical simulation**, rather than assuming we implemented a strategy at some past date and collecting results as the strategy runs over time, we instead construct results by selecting returns at random from many different historical periods (windows) without regard to time-ordering. Although this approach does assume, like rolling-window backtesting, that past asset returns provide guidance about future asset returns, it relaxes a key restriction by randomly changing the sequencing of historical periods from which factor returns are drawn. As a result, historical simulation is essentially a non-deterministic rolling-window backtest. Historical simulation is widely used in investment management, particularly by banks for market risk analysis.

The problem with historical time-series data (such as factor returns) is that there is only one set of realized data to draw from—the past happened only one way. A critical assumption behind classical time-series analysis—that the data are stationary—is simply not true of most financial variables. **Monte Carlo simulation** overcomes many of these issues. In Monte Carlo simulation, each key variable is assigned a statistical distribution, and observations are drawn at random from the assigned distribution.

The Monte Carlo approach is popular because it is highly flexible; an array of different distributions can be used across a variety of key variables. Rather than using historical distributions or, for example, the normal distribution (that may only roughly approximate a particular variable's return distribution), the analyst can incorporate non-normality, fat tails, tail dependence, and so on, to model key variables. The downside is that it is complex and computationally intensive.

An important goal of simulation is to verify the investment performance obtained from backtesting by accounting for randomness. Simulation is especially useful in measuring the downside risk of investment strategies.

A properly designed simulation analysis is typically implemented in the following eight steps:

1. Determine what we want to understand: the target variable. This variable is typically the return on an investment strategy or $r_{p,t}$ (the return on portfolio p at time t) and its distribution.
2. Specify key decision variables. Key decision variables are often the returns of each underlying asset, $r_{i,t}$ (the return on asset i at time t), in the overall portfolio and the weight, $\omega_{i,t}$ (the weight of asset i at time t), allocated to each asset in the portfolio. Once we know the returns and weights of all (K) underlying assets, we can readily compute the return of the portfolio as $r_{p,t} = \sum_{i=1}^{K}\left(\omega_{i,t} \times r_{i,t}\right)$. Recall that the weight of each underlying asset is determined by the investment strategy being tested.
3. Specify the number of trials (N) to run. In practice, researchers typically choose between 1,000 and 10,000 simulation runs. The greater the number of trials, the more stable the predictions of performance and variance of performance. In theory, determining the optimal number of iterations is a complex topic (for an example, see Ritter, Schoelles, Quigley, and Klein 2011).
4. Define the distributional properties of the key decision variables. At this point, historical and Monte Carlo simulations diverge. In historical simulation, we draw from historical data. Conversely, in Monte Carlo simulation, we must specify a statistical distribution for each key decision variable. Although it is up to the user, the choice of distribution should be guided by how well it has described historical observations. It might be appropriate to specify different functions (e.g., normal, lognormal, binomial) for different variables to account for the impact of correlations and tail dependence.
5. Use a random number generator to draw N random numbers for each key decision variable.

6 For each set of simulated key decision variables, compute the value of the target variable. The value of the target variable is then saved for later analysis.

7 Repeat the same processes from Steps 5 and 6 until completing the desired number of trials (N).

8 Now we have a set of N values of the target variable. In this context, it is N returns of the investment strategy. The analyst can now calculate the typical metrics, such as mean return, volatility, Sharpe ratio, and the various downside risk metrics. For simulations, analysts typically use CVaR and maximum drawdown to characterize downside risk.

7.1 Historical Simulation

g explain inputs and decisions in simulation and interpret a simulation;

Although backtesting and historical simulation rely on history to understand the future, they are different in that rolling-window backtesting is deterministic, whereas historical simulation incorporates randomness by randomly drawing returns from historical data rather than following each period chronologically.

First, a decision must be made about whether to sample from the historical returns with replacement or without replacement. Random sampling with replacement, also known as **bootstrapping**, is often used in investment research because the number of simulations needed is often larger than the size of the historical dataset.

Using the factor allocation strategies (BM and RP) for the eight factor portfolios as an example, we can perform a historical simulation as follows:

1 The target variables are the returns for the BM and RP multifactor portfolios.

2 The key decision variables are the returns of the eight underlying factor-based portfolios (the weights allocated to the eight factors are already known).

3 The simulation will be performed for N = 1,000 trials.

4 The historical simulation will be implemented using bootstrapped sampling. In this case, we will randomly draw a number from a uniform distribution (so there is equal probability of being selected) between 0 and 1.[2] Once a random number is generated, it is assigned to a specific historical month. Note that we have a total of 374 months of historical factor return data (April 1988–May 2019). We assign random numbers to specific months by dividing the span of the uniform distribution by the number of months (1.0/374 = 0.00267). Therefore, if the random number is between 0 and 0.00267, the first month is selected. Similarly, if the random number generator draws a number between 0.00267 and 0.00535 (= 2 × 0.00267), the second month is chosen, and so on.

5 The random number generator will then randomly draw 1,000 numbers from the uniform distribution between 0 and 1, and, as mentioned, sampling of the historical return data is with replacement. For example, as shown in Exhibit 17, the first five numbers generated are 0.59163, 0.32185, 0.76485, 0.89474, and 0.45431, which are then mapped to Months 222 (September 2006), 121 (April 1998), 287 (February 2012), 335 (February 2016), and 170 (May 2002), respectively. To be clear, months are mapped by dividing the random number by 0.00267, so Month 222 is determined as 0.59163/0.00267, Month 121 is 0.32185/0.00267, and so on.

2 Technically, the random number generator will draw a random number that equals or is greater than 0 but is less than 1.

Exhibit 17 Factor Returns for the First Five Randomly Selected Months

Simulation #	Month	Random #	Month #	Earnings Yield	Book-to-Market	Earnings Growth	Momentum
1	9/30/2006	0.59163	222	2.5%	0.3%	(0.8%)	(0.0%)
2	4/30/1998	0.32185	121	0.1%	0.8%	(0.2%)	(0.5%)
3	2/29/2012	0.76485	287	(1.9%)	0.5%	1.7%	1.8%
4	2/29/2016	0.89474	335	2.5%	2.4%	(0.4%)	(1.5%)
5	5/31/2002	0.45431	170	6.3%	(3.3%)	1.8%	2.4%

Simulation #	Month	Random #	Month #	Earnings Revision	ROE	Debt/Equity	Earnings Quality
1	9/30/2006	0.59163	222	(0.8%)	2.5%	0.5%	(0.5%)
2	4/30/1998	0.32185	121	(0.1%)	(0.1%)	0.3%	1.6%
3	2/29/2012	0.76485	287	1.8%	(0.5%)	(2.1%)	(0.8%)
4	2/29/2016	0.89474	335	(1.5%)	1.2%	(1.2%)	1.3%
5	5/31/2002	0.45431	170	2.4%	6.4%	(0.7%)	(1.2%)

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

6 Once a given month is selected, the returns of the corresponding eight factor portfolios represent one possible set of outcomes that we use to compute the values of our target variables—the returns of the BM and RP portfolios—using the prespecified factor weights. For example, the first trial picks the month of September 2006. The return of the benchmark portfolio is the equally weighted average of the eight factor returns, or 0.46% (= 0.125 × 2.5% + 0.125 × 0.3% + 0.125 × –0.8% + 0.125 × 0.0% + 0.125 × –0.8% + 0.125 × 2.5% + 0.125 × 0.5% + 0.125 × –0.5%).

To compute the return on the risk parity portfolio, we use the weights allocated to each of the eight factors for the final month (May 2019). As shown in Exhibit 18, for the first trial, September 2006, the weighted average return of the risk parity portfolio is 0.17%. It should be clear that each trial in the historical simulation assumes the simulated returns of the eight factors follow the same patterns observed in the sampled month—in this case, September 2006.

Exhibit 18 How to Compute the Return of the Risk Parity Portfolio, Historical Simulation

Asset (Factor)	September 2006 Return	May 2019 Weight	Weighted Return
Earnings yield	2.5%	6.0%	0.2%
Book-to-market	0.3%	30.3%	0.1%
Earnings growth	(0.8%)	11.7%	(0.1%)
Momentum	(0.0%)	5.2%	(0.0%)
Earnings revision	(0.8%)	10.4%	(0.1%)
ROE	2.5%	6.3%	0.2%
Debt/equity	0.5%	9.6%	0.0%

Exhibit 18 (Continued)

Asset (Factor)	September 2006 Return	May 2019 Weight	Weighted Return
Earnings quality	(0.5%)	20.4%	(0.1%)
Risk Parity Portfolio			**0.17%**

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

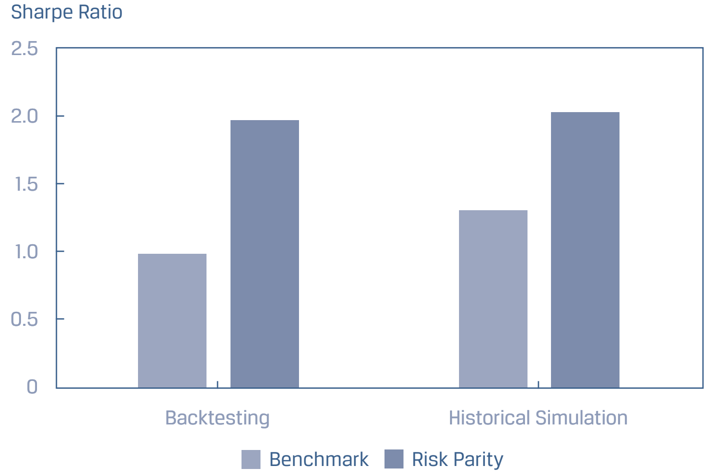
7 The same simulation process (from Steps 5 to 6) is repeated for all 1,000 trials, generating a collection of 1,000 simulated returns for the benchmark and risk parity portfolios.

8 Finally, equipped with these 1,000 return scenarios, we can calculate performance metrics of interest (Sharpe ratio, CVaR, etc.) and plot the distributions of the *simulated* benchmark and risk parity portfolio returns.

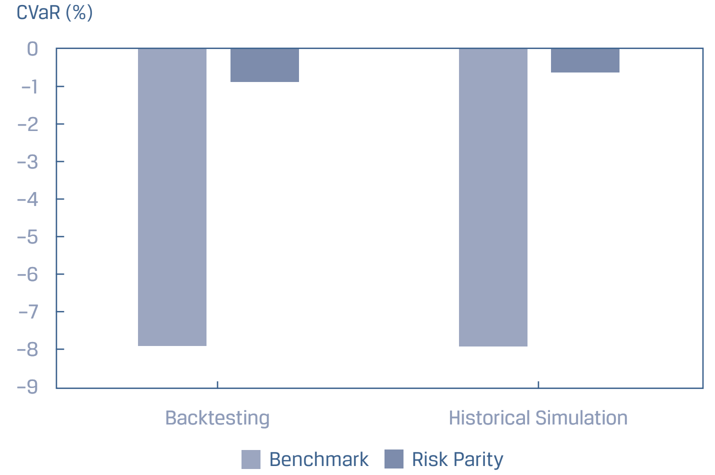
As shown in Panel A of Exhibit 19, the results of the historical simulation (over the 1,000 iterations) suggest that the Sharpe ratios of the BM and RP strategies are largely in line with the rolling-window backtesting method demonstrated previously. In particular, the RP portfolio outperforms the BM portfolio in terms of Sharpe ratio according to both methodologies. Similarly, as shown in Panel B, both methodologies indicate that the RP portfolio carries substantially less downside risk, measured by CVaR, than the BM portfolio.

Exhibit 19 Comparing Historical Simulation with Backtesting

A. Sharpe Ratio



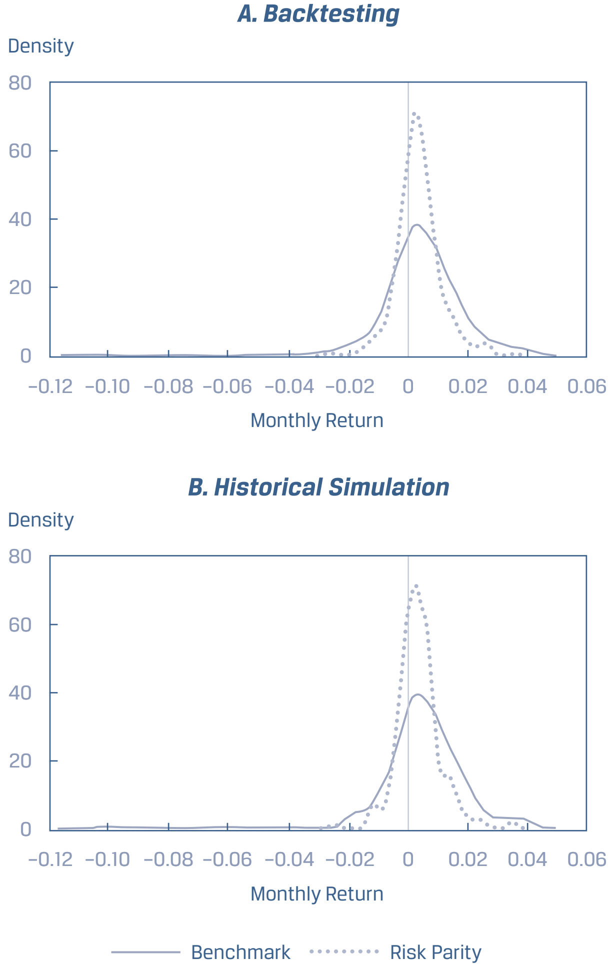
B. Conditonal Value-at-Risk



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

In addition to capturing downside risk with a single number (e.g., CVaR), we can also plot the estimated probability distribution of returns for our two investment strategies. Panel A of Exhibit 20 plots the estimated probability distribution of returns for the BM and RP portfolios using backtested returns, whereas Panel B shows the estimated return distribution plots using the historical simulated returns. We can observe a broadly similar pattern between them. Both the backtesting and historical simulation approaches suggest that the RP portfolio returns are less volatile and more skewed to the right with lower downside risk (i.e., lower standard deviation and thinner tails) than the BM portfolio returns.

Exhibit 20 Estimated Distribution Plots: Backtesting and Historical Simulation



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

7.2 Monte Carlo Simulation

g explain inputs and decisions in simulation and interpret a simulation;

An important issue with historical simulation is that the data are limited to historical observations, which may not represent the future. This deficiency can be addressed with Monte Carlo simulation, which follows similar steps as historical simulation but with a few key differences.

First, we need to specify a functional form for each key decision variable. Exploratory data analysis—focusing on moments (i.e., mean, standard deviation, skewness, kurtosis) and tail dependence—is often crucial here. The usefulness of the Monte Carlo simulation technique critically depends on whether the functional form of the statistical distribution that we specify accurately reflects the true distribution of the underlying data. Because the data's true distribution is unknown, we need to be aware of the fact that our model, like all models, only provides guidance and will not be perfect.

Regression and distribution-fitting techniques are used to estimate the parameters (i.e., mean, standard deviation, skewness, kurtosis) underlying the statistical distributions of the key decision variables. This step is typically called model calibration. Although it may sound difficult, R, Python, Matlab, and similar tools can readily perform this task with a few lines of code (for example, see the fMultivar package in R).

Before finalizing our choice of the functional form of the statistical distribution, we need to account for the following considerations:

- The distribution should reasonably describe the key empirical patterns of the underlying data. For example, asset returns roughly follow a bell curve pattern; therefore, the normal distribution and Student's t-distribution are often used as first-cut approximations.
- It is equally critical to account for the correlations between multiple key decision variables. In the case of asset or factor allocation strategies, as shown previously, the returns from multiple factors are clearly correlated; therefore, we need to specify a multivariate distribution rather than modeling each factor or asset on a standalone basis.
- The complexity of the functional form and number of parameters that determine the functional form are important. We can specify a highly complex model with many parameters (all of which need to be estimated/calibrated from historical data) that describe the empirical properties of the data well. Given limited historical data, however, we may be unable to estimate all the underlying parameters with sufficient precision. Such models tend to have low specification errors, but they suffer from large estimation errors. At the other extreme, overly simplistic models require fewer parameters (therefore, they might have low estimation errors), but they may not fit the data well (because they are misspecified). You should recognize this phenomenon as the bias–variance tradeoff, introduced in earlier readings on machine learning and big data projects.

For simulation of asset or factor allocation strategies, the distribution of asset or factor returns is typically modeled as a multivariate normal distribution—as a first-cut approximation—which captures some of the key properties of the underlying data reasonably well. More importantly, a multivariate normal distribution can be fully specified with only a few key parameters—the mean, the standard deviation, and the covariance matrix. For K assets, we need to estimate K mean returns, K standard deviations, and $[K \times (K—1)]/2$ correlations.

We have to be aware, however, that the multivariate normal distribution does not fully account for the empirical characteristics of (negative) skewness, excess kurtosis, and tail dependence apparent in the data. We will address these non-normal distribution properties shortly, when we cover sensitivity analysis.

Continuing with the same BM and RP strategies, the Monte Carlo simulation is performed as follows:

1. Our target variables are the returns for the BM and RP multifactor portfolios.
2. The key decision variables are the returns of the eight underlying factor-based portfolios.
3. We will perform the simulation using 1,000 trials.
4. We choose the multivariate normal distribution as our initial functional form. We calibrate the model—calculate the eight factor portfolio mean returns, the eight standard deviations, and the 28 elements of the covariance matrix—using the 374 months of historical factor return data (April 1988–May 2019).
5. The calibrated multivariate normal distribution is then used to simulate the future factor returns. The process by which this simulation occurs in the context of a multivariate normal distribution of eight random variables,

corresponding to our eight factor portfolios, is complex. Suffice it to say, in this case, eight randomly generated numbers from the uniform distribution are mapped onto a point on the joint cumulative probability distribution function, and this point jointly determines the values of the eight factor returns in this trial.

Exhibit 21 shows the first five sets of Monte Carlo simulated returns for the eight underlying factor-based portfolios.

Exhibit 21 Monte Carlo Simulation: First Five Simulations of Factor Returns Using a Multivariate Normal Distribution

Simulation #	Earnings Yield	Book-to-Market	Earnings Growth	Momentum	Earnings Revision	ROE	Debt/ Equity	Earnings Quality
1	(3.2%)	(3.1%)	(0.2%)	0.7%	2.3%	(3.3%)	(1.7%)	1.9%
2	(0.0%)	3.5%	0.9%	(0.4%)	0.9%	(2.4%)	(3.5%)	(0.2%)
3	0.7%	(1.8%)	2.9%	3.8%	2.5%	1.3%	(0.8%)	(0.0%)
4	9.7%	(0.5%)	1.2%	3.8%	(0.9%)	7.6%	(3.7%)	1.6%
5	1.7%	0.2%	2.9%	(0.2%)	3.0%	0.2%	(0.9%)	0.2%

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

6 Once the returns of the eight factor portfolios are simulated, we can compute the values of our target variables—the returns of the BM and RP portfolios. For example, for the first simulated set of returns, the benchmark portfolio (with equally weighted factor returns) delivers a monthly return of –0.83% (= 0.125 × –3.2% + 0.125 × –3.1% + 0.125 × –0.2% + 0.125 × 0.7% + 0.125 × 2.3% + 0.125 × –3.3% + 0.125 × –1.7% + 0.125 × 1.9%).

Similarly, using the RP allocation factor weights for the final month, May 2019 (see Exhibit 18), the simulated risk parity portfolio return is –0.86% (= 0.06 × –3.2% + 0.303 × –3.1% + 0.117 × –0.2% + 0.052 × 0.7% + 0.104 × 2.3% + 0.063 × –3.3% + 0.096 × –1.7% + 0.204 × 1.9%).

7 Next, we repeat Steps 5 and 6 for all 1,000 trials to generate a collection of 1,000 returns for the benchmark and risk parity portfolios.

8 Finally, we assess the performance and risk profiles of our two investment strategies from the 1,000 simulated returns.

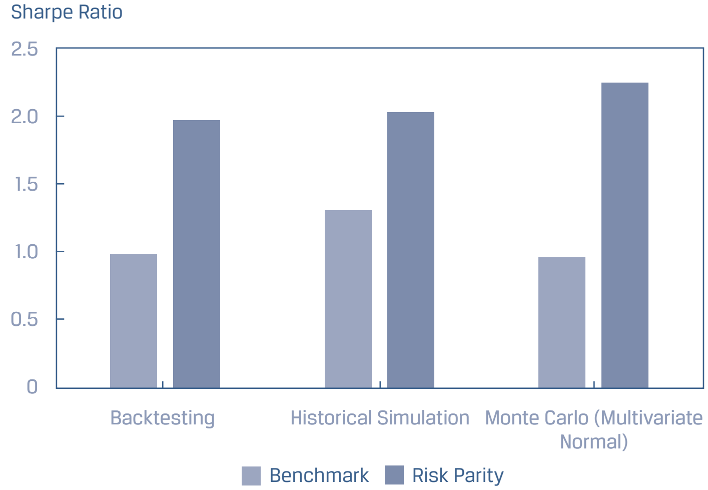
EXAMPLE 7 HOW TO INTERPRET RESULTS FROM HISTORICAL AND MONTE CARLO SIMULATIONS

Exhibit 22 shows the Sharpe ratios (Panel A) and downside risk measures, CVaRs (Panel B), for the returns of the benchmark and risk parity portfolios based on rolling-window backtesting, historical simulation, and Monte Carlo simulation of the returns on the eight underlying factor portfolios.

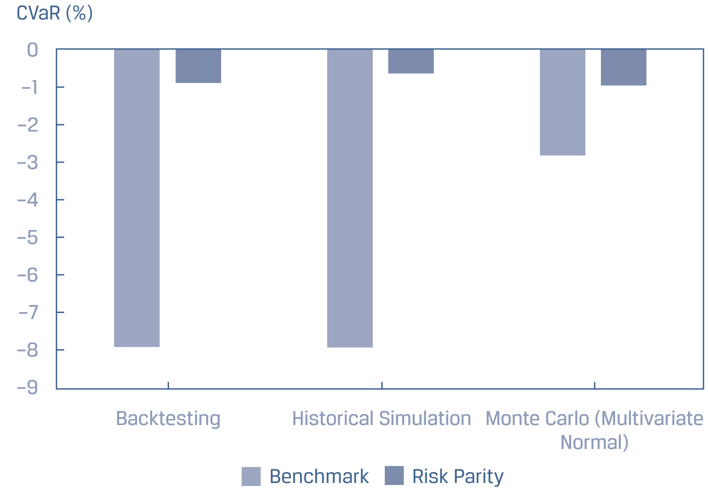
Discuss similarities and differences among the three approaches for simulated performance of the benchmark and risk parity portfolios.

Exhibit 22 Comparing Backtesting, Historical Simulation, and Monte Carlo Simulation-Based Performance for the BM and RP Portfolios

A. Sharpe Ratio



B. Conditional Value-at-Risk



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

Solution:

Note that the backtesting approach provides realistic performance metrics assuming investors have been following the same trading rules throughout the past periods under investigation. The two simulation analyses are complementary to backtesting and deliver additional insights. In particular, they account for the random nature of investment data in different ways. Historical simulation randomly samples (with replacement) from the past record of asset returns, in a manner that each set of past monthly returns is equally likely to be selected. Monte Carlo simulation randomly samples from an assumed multivariate joint probability distribution (e.g., normal or another type of distribution), in a manner that the past record of asset returns is used to calibrate the parameters of the multivariate distribution. Therefore, these simulation methods are used to independently verify the results from the rolling-window backtesting.

As shown in Panel A of Exhibit 22, the Sharpe ratio appears relatively insensitive to the simulation and backtesting methods used, with the RP strategy outperforming the BM strategy by nearly the same margin for each method. In contrast, CVaR seems to be sensitive to how randomness is treated. In particular, the Monte Carlo simulation appears to understate the downside risk of the BM strategy compared with both rolling-window backtesting and historical simulation methods (Panel B). Because the factor returns are negatively skewed with fat tails (i.e., excess kurtosis), the multivariate normal distribution assumption is likely to be underestimating the true downside risk of the BM strategy. This underestimation of risk appears only for the BM strategy because factor risks and correlations are not properly accounted for in the naive (equal) weighting scheme. Conversely, in this case, the risk parity strategy is robust to a non-normal factor return distribution, resulting in a portfolio with considerably lower downside risk.

SENSITIVITY ANALYSIS

8

h demonstrate the use of sensitivity analysis

In addition to simulation, sensitivity analysis—a technique for exploring how a target variable is affected by changes in input variables (e.g., the distribution of asset or factor returns)—can be implemented to help managers further understand the potential risks and returns of their investment strategies.

The Monte Carlo simulation just described fits a multivariate normal distribution to the factor returns—a sensible first approximation because it requires relatively few parameters to be estimated from historical data. Despite the simplicity and wide adoption in practice, the multivariate normal distribution assumption fails to account for various empirical properties in the factor return distributions, including negative skewness and fat tails. Because the value of the simulation results depends crucially on whether the selected functional form is a reasonable proxy for the true distribution, we should conduct a sensitivity analysis by fitting our factor return data to a different distribution and repeating the Monte Carlo simulation accordingly. One alternative to test is a multivariate skewed Student's t-distribution.

The Student's t-distribution is a natural extension of the multivariate normal distribution, because it has the ability to account for the skewness and the excess kurtosis often observed in factor and asset return data. It is mathematically more complex, however, and requires estimating a larger number of parameters than a normal distribution.

With the goal of determining the sensitivity of our target variables (the returns of the benchmark and the risk parity portfolios) to the new factor return distribution assumption, the procedure for the new Monte Carlo simulation process is almost identical to the one performed previously. The only two exceptions are Steps 4 and 5. In Step 4, instead of fitting the data to a multivariate normal distribution, we calibrate our model to a multivariate skewed t-distribution. In Step 5, we simulate 1,000 sets of factor returns from this new distribution function. Then, as before, we assess the performance and risk profiles of our investment strategies from the 1,000 simulated returns.

Exhibit 23 shows the first five sets of simulated factor returns from this new model. As previously, we compute the values of our target variables for each set of simulated factor returns and then assess their performance and risk characteristics. For the first set of factor returns, the equal-weighted (i.e., 0.125 for each factor) benchmark portfolio achieves a simulated monthly return of 1.21%, and the risk parity portfolio (using May 2019 factor weights in Exhibit 18) delivers a simulated return of 0.75%.

Exhibit 23 First Five Simulations of Factor Returns Using Multivariate Skewed *t*-Distribution

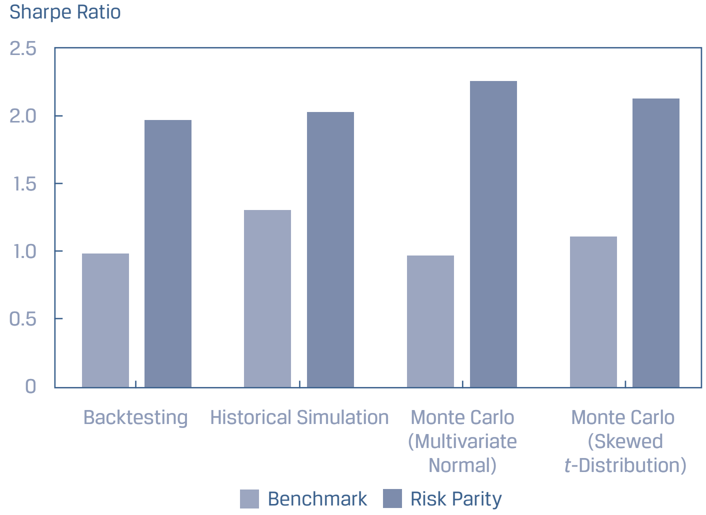
Simulation #	Earnings Yield	Book-to-Market	Earnings Growth	Momentum	Earnings Revision	ROE	Debt/ Equity	Earnings Quality
1	2.0%	0.3%	1.7%	3.1%	2.0%	0.9%	0.2%	(0.5%)
2	1.8%	(1.4%)	0.2%	4.9%	1.8%	2.7%	0.4%	(0.1%)
3	(0.6%)	0.2%	(1.0%)	(0.1%)	0.4%	1.5%	1.6%	0.9%
4	11.2%	2.6%	1.8%	1.5%	2.2%	9.6%	(2.9%)	(1.9%)
5	(3.9%)	(1.3%)	0.9%	0.9%	0.8%	(3.5%)	2.9%	0.2%

Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

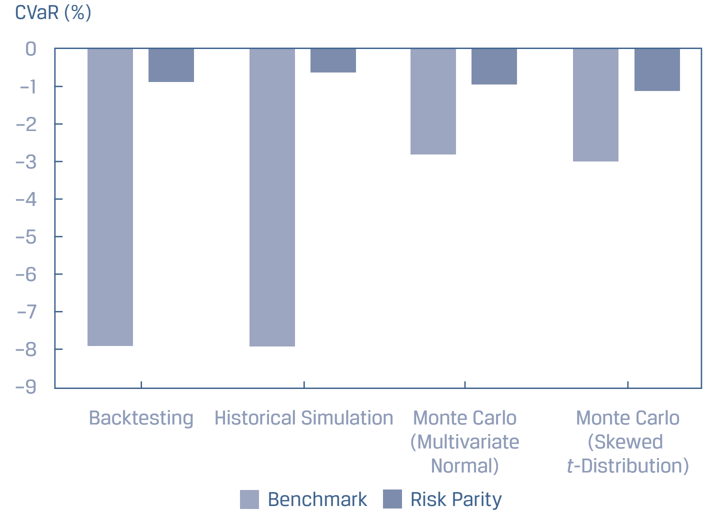
Turning to the performance and risk profiles of our investment strategies, shown in Panel A of Exhibit 24, we note that the Sharpe ratio appears insensitive to any of the particular simulation methods used by consistently suggesting that the risk parity allocation strategy outperforms the benchmark strategy. Downside risk (expressed as CVaR), however, appears quite sensitive to the choice of simulation approach for the BM strategy, but not very sensitive for the RP strategy (Panel B). If we focus on the BM strategy, the CVaR results from historical simulation and rolling-window backtesting resemble each other very closely. The CVaR results of both (multivariate skewed *t*- and multivariate normal) Monte Carlo simulations are also very similar: Both underestimate the downside risk of the BM strategy. This finding suggests that additional sensitivity analyses should be run with different functional forms for the factor return distributions.

Exhibit 24 Comparing Simulation Methods with Backtesting

A. Sharpe Ratio



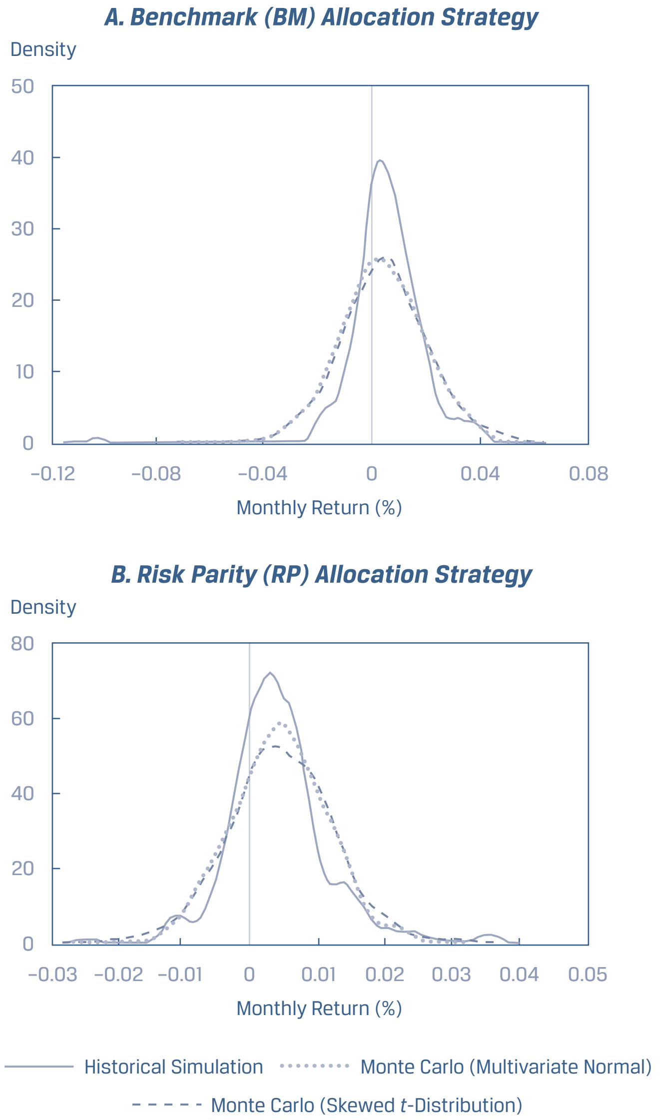
B. Conditional Value-at-Risk



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

Estimated probability density plots, in Panel A of Exhibit 25, show that the difference between the historical simulation and the two Monte Carlo methods is rather large for the BM strategy. Given the negative skewness and excess kurtosis of the BM strategy's returns, which is apparent from the shape of the historical simulation return distribution, it is not surprising that the two Monte Carlo simulations fail to account sufficiently for this left-tail risk property. Conversely, because the distribution of the RP strategy's returns is relatively symmetric and without much excess kurtosis, all three simulation methods provide a fairly similar picture (Panel B).

Exhibit 25 Estimated Distribution Plots for BM and RP Strategies Using Three Different Simulations



Sources: Bloomberg Finance LLP, FTSE Russell, S&P Capital IQ, Thomson Reuters, Wolfe Research Luo's QES.

EXAMPLE 8 SIMULATING THE PERFORMANCE OF FACTOR ALLOCATION STRATEGIES

Earlier, Sarah Koh presented her team's backtesting results for the factor-based allocation strategies being considered by an important client, SWF Fund. Now, while presenting the simulation results for these same strategies, SWF Fund's investment committee asks Koh the following questions:

1 The following are caveats regarding the use of rolling-window backtesting in assessing investment strategies *except*:

A this technique implicitly assumes that the same pattern of past performance is likely to repeat itself over time.

B this technique may not fully account for the dynamic nature of financial markets and potentially extreme downside risks.

C this technique is intuitive, because it mimics how investing is done in reality—that is, forming ideas, testing strategies, and implementing periodically.

2 Which of the following situations is *most likely* to involve data snooping?

A A researcher performs rolling-window backtesting of a new momentum strategy using 20 years of point-in-time (PIT) data from the United States. She cross-validates results by similarly analyzing PIT data from the following markets: mainland China, Asia ex-Japan, Europe, the United Kingdom, and Canada.

B A researcher tries many different modeling techniques, backtesting each of them, and then picking the best-performing model without accounting for model selection bias.

C A researcher sets a relatively high hurdle, a t-statistic greater than 3.0, for assessing whether a newly discovered factor is statistically significant.

3 Which of the following situations is *least likely* to involve scenario analysis?

A Simulating the performance and risk of investment strategies by first using stocks in the Nikkei 225 Index and then using stocks in the TOPIX 1000 Index.

B Simulating the performance and risk of investment strategies in both "trade agreement" and "no-trade-agreement" environments.

C Simulating the performance and risk of investment strategies in both high-volatility and low-volatility environments.

4 Which one of the following statements concerning historical simulation and Monte Carlo simulation is *false*?

A Historical simulation randomly samples (with replacement) from the past record of asset returns, where each set of past monthly returns is equally likely to be selected.

B Neither historical simulation nor Monte Carlo simulation makes use of a random number generator.

C Monte Carlo simulation randomly samples from an assumed multivariate joint probability distribution in which the past record of asset returns is used to calibrate the parameters of the multivariate distribution.

5 Which one of the following statements concerning Monte Carlo simulation is *false*?

A When simulating multiple assets (factors) whose returns are correlated, it is crucial to specify a multivariate distribution rather than modeling each asset on a standalone basis.

B Regression and distribution-fitting techniques are used to estimate the parameters underlying the statistical distributions of the key decision variables.

C The Monte Carlo simulation process is deterministic and non-random in nature.

6 Which of the following situations concerning simulation of a multifactor asset allocation strategy is *most likely* to involve sensitivity analysis?

A Changing the specified multivariate distribution assumption from a normal to a skewed *t*-distribution to better account for skewness and fat tails

B Splitting the rolling window between periods of recession and non-recession

C Splitting the rolling window between periods of high volatility and low volatility

Solution to 1:

C is correct, because it is not a caveat in using rolling-window backtesting. A and B are incorrect because they are caveats in the use of this technique.

Solution to 2:

B is correct, because this situation most likely involves data snooping. A and C are incorrect because these are approaches to avoiding data snooping.

Solution to 3:

A is correct, because there is no structural break or different structural regime. B and C are incorrect because they involve structural breaks/different structural regimes and thus represent different scenarios.

Solution to 4:

B is correct, because this statement is false. A and C are incorrect because they are true statements about historical and Monte Carlo simulation, respectively.

Solution to 5:

C is correct, because this statement is false. A and B are incorrect because they are true statements about Monte Carlo simulation.

Solution to 6:

A is correct, because this choice represents sensitivity analysis. B and C are incorrect because these choices represent scenario analysis.

SUMMARY

In this reading, we have discussed how to perform rolling-window backtesting—a widely used technique in the investment industry. We also described how to use scenario analysis and simulation along with sensitivity analysis to supplement backtesting, so investors can better account for the randomness in data that may not be fully captured by backtesting.

- The main objective of backtesting is to understand the risk–return trade-off of an investment strategy by approximating the real-life investment process.
- The basic steps in rolling-window backtesting are specifying the investment hypothesis and goals, determining the rules and processes behind an investment strategy, forming an investment portfolio according to those rules, rebalancing the portfolio periodically, and computing the performance and risk profiles of the strategy.

- In the rolling-window backtesting methodology, researchers use a rolling-window (or walk-forward) framework, fit/calibrate factors or trade signals based on the rolling window, rebalance the portfolio periodically, and then track the performance over time. Thus, rolling-window backtesting is a proxy for actual investing.
- Analysts need to pay attention to several behavioral issues in backtesting, including survivorship bias and look-ahead bias.
- Asset (and factor) returns are often negatively skewed and exhibit excess kurtosis (fat tails) and tail dependence compared with a normal distribution. As a result, standard rolling-window backtesting may be unable to fully account for the randomness in asset returns, particularly on downside risk.
- Financial data often face structural breaks. Scenario analysis can help investors understand the performance of an investment strategy in different structural regimes.
- Historical simulation is relatively straightforward to perform but shares pros and cons similar to those of rolling-window backtesting. For example, a key assumption these methods share is that the distribution pattern from the historical data is sufficient to represent the uncertainty in the future. Bootstrapping (or random draws with replacement) is often used in historical simulation.
- Monte Carlo simulation is a more sophisticated technique than historical simulation. In Monte Carlo simulation, the most important decision is the choice of functional form of the statistical distribution of decision variables/return drivers. Multivariate normal distribution is often used in investment research, owing to its simplicity. However, a multivariate normal distribution cannot account for negative skewness and fat tails observed in factor and asset returns.
- Sensitivity analysis, a technique for exploring how a target variable and risk profiles are affected by changes in input variables, can further help investors understand the limitations of conventional Monte Carlo simulation (which typically assumes a multivariate normal distribution as a starting point). A multivariate skewed *t*-distribution considers skewness and kurtosis but requires estimation of more parameters and thus is more likely to suffer from larger estimation errors.

REFERENCES

DeMiguel, V., L. Garlappi, and R. Uppal. 2007. "Optimal versus Naïve Diversification: How Inefficient Is the 1/N Portfolio Strategy?" London Business School Working Paper. http://faculty.london.edu/avmiguel/DeMiguel-Garlappi-Uppal-RFS.pdf.

Fama, E., and K. R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." Journal of Financial Economics 33 (1): 3–56. doi:10.1016/0304-405X(93)90023-5

Ritter, F. E., M. J. Schoelles, K. S. Quigley, and L. C. Klein. 2011. "Determining the Number of Simulation Runs: Treating Simulations as Theories by Not Sampling Their Behavior." In Human-in-the-Loop Simulations: Methods and Practice, ed. L. Rothrock and S. Narayanan, 97–116. Berlin: Springer. doi:10.1007/978-0-85729-883-6_5

PRACTICE PROBLEMS

The following information relates to Questions 1–8

Emily Yuen is a senior analyst for a consulting firm that specializes in assessing equity strategies using backtesting and simulation techniques. She is working with an assistant, Cameron Ruckey, to develop multifactor portfolio strategies based on nine factors common to the growth style of investing. To do so, Yuen and Ruckey plan to construct nine separate factor portfolios and then use them to create factor-weighted allocation portfolios.

Yuen tasks Ruckey with specifying the investment universe and determining the availability of appropriate reporting data in vendor databases. Ruckey selects a vendor database that does not provide point-in-time data, so he adjusts the database to include point-in-time constituent stocks and a reporting lag of four months.

Next, Yuen and Ruckey run initial backtests on the nine factor portfolios, calculating performance statistics and key metrics for each. For backtesting purposes, the portfolios are rebalanced monthly over a 30-year time horizon using a rolling-window procedure.

Yuen and Ruckey consider a variety of metrics to assess the results of the factor portfolio backtests. Yuen asks Ruckey what can be concluded from the data for three of the factor strategies in Exhibit 1:

Exhibit 1 Backtest Metrics for Factor Strategies

	Factor 1	Factor 2	Factor 3
VaR (95%)	(3.9%)	(1.3%)	(8.4%)
Standard deviation of returns	2.1%	1.2%	4.6%
Maximum drawdown	27.2%	8.3%	59.7%

Ruckey tells Yuen the following:

Statement 1 We do not need to consider maximum drawdown, because standard deviation sufficiently characterizes risk.

Statement 2 Factor 2 has the highest downside risk.

From her professional experience Yuen knows that benchmark and risk parity factor portfolios, in which factors are equally weighted and equally risk weighted, respectively, are popular with institutional and high-net-worth clients. To gain a more complete picture of these investment strategies' performance, Yuen and Ruckey design a Benchmark Portfolio (A) and a Risk Parity Portfolio (B), and then run two simulation methods to generate investment performance data based on the underlying factor portfolios, assuming 1,000 simulation trials for each approach:

Approach 1 Historical simulation

Approach 2 Monte Carlo simulation

© 2021 CFA Institute. All rights reserved.

Yuen and Ruckey discuss the differences between the two approaches and then design the simulations, making key decisions at various steps. During the process, Yuen expresses a number of concerns:

Concern 1: Returns from six of the nine factors are correlated.

Concern 2: The distribution of Factor 1 returns exhibits excess kurtosis and negative skewness.

Concern 3: The number of simulations needed for Approach 1 is larger than the size of the historical dataset.

For each approach, Yuen and Ruckey run 1,000 trials to obtain 1,000 returns for Portfolios A and B. To help understand the effect of the skewness and excess kurtosis observed in the Factor 1 returns on the performance of Portfolios A and B, Ruckey suggests simulating an additional 1,000 factor returns using a multivariate skewed Student's *t*-distribution, then repeating the Approach 2 simulation.

1 Following Ruckey's adjustments to the initial vendor database, backtested returns will *most likely* be subject to:

- **A** stale data.
- **B** data snooping
- **C** p-hacking

2 Based on Exhibit 1, Ruckey should conclude that:

- **A** Factor Strategy 3 has the highest portfolio turnover.
- **B** Factor Strategy 2 has less downside risk than Strategy 3.
- **C** Factor Strategy 2 has the highest returns.

3 Which of Ruckey's statements about Exhibit 1 is <u>in</u>correct?

- **A** Only Statement 1
- **B** Only Statement 2
- **C** Both Statement 1 and Statement 2

4 Simulation Approach 1 (historical simulation) differs from Approach 2 (Monte Carlo simulation) in that:

- **A** it is deterministic.
- **B** a functional form of the statistical distribution for each decision variable needs to be specified.
- **C** it assumes that sampling the returns from the actual data provides sufficient guidance about future asset returns.

5 To address Concern 1 when designing Approach 2, Yuen should:

- **A** model each factor or asset on a standalone basis.
- **B** calculate the 15 covariance matrix elements needed to calibrate the model.
- **C** specify a multivariate distribution rather than modeling each factor or asset on a standalone basis.

6 Based on Concern 2, the Factor 1 strategy is *most likely* to:

- **A** be favored by risk-averse investors.
- **B** generate surprises in the form of negative returns.
- **C** have return data that line up tightly around a trend line.

7 To address Concern 3 when designing Approach 1, Yuen should:

- **A** add monthly return observations to the dataset using interpolation.
- **B** randomly sample from the historical returns with replacement.
- **C** choose the multivariate normal distribution as the initial functional form.

8 The process Ruckey suggests to better understand how the performance of Portfolios A and B using Approach 2 is affected by the distribution of Factor 1 returns is *best* described as:

A data snooping.

B sensitivity analysis.

C inverse transformation.

The following information relates to Questions 9–16

Kata Rom is an equity analyst working for Gimingham Wealth Partners (GWP), a large investment advisory company. Rom meets with Goran Galic, a Canadian private wealth client, to explain investment strategies used by GWP to generate portfolio alpha for its clients.

Rom states that GWP is recognized in the Canadian investment industry as a leading factor-based value portfolio manager and describes how GWP creates relevant investment strategies and explains GWP's backtesting process. Rom notes the following:

Statement 1 Using historical data, backtesting approximates a real-life investment process to illustrate the risk–return tradeoff of a particular proposed investment strategy.

Statement 2 Backtesting is used almost exclusively by quantitative investment managers and rarely by fundamental investment managers, who are more concerned with information such as forward estimates of company earnings, macroeconomic factors, and intrinsic values.

Galic, who is 62 years old, decides to allocate C$2 million (representing 10% of his net worth) to an account with GWP and stipulates that portfolio assets be restricted exclusively to domestic securities. Although GWP has not backtested its strategies with such a restriction, it has backtested its strategies using a global index that includes domestic securities. Rom shows the following risk measures to Galic for three factor portfolios.

Exhibit 1 Downside Risk Measures for Model Factors

Risk Measure	Factor 1	Factor 2	Factor 3
Value at risk (VaR) (95%)	(6.49%)	(0.77%)	(2.40%)
Conditional VaR (CVaR) (95%)	(15.73%)	(4.21%)	(3.24%)
Maximum drawdown	35.10%	38.83%	45.98%

Galic asks Rom, "What happens if the future is different from the past?" Rom gives the following replies:

Statement 3: Although backtesting can offer some comfort, you are correct that it does have a weakness: Backtesting generally does not capture the dynamic nature of financial markets and in particular may not capture extreme downside risk.

Statement 4: As a result, we have captured extreme downside risk and the dynamic nature of financial markets by using the Value-at-Risk and Conditional Value-at-Risk measures.

In an effort to make Galic fully aware of the risks inherent in GWP's strategies, Rom describes a recent study that investigated the return distributions of value and momentum factors that GWP uses to construct portfolios. The study found that these distributions were non-normal based on their negative skewness, excess kurtosis, and tail dependence. Rom indicated that investment strategies based on this type of data are prone to significantly higher downside risk. Rom informs Galic that GWP also uses a technique commonly referred to as scenario analysis to examine how strategies perform in different structural regimes. Exhibit 2 compares the performance of two of GWP's factor allocation strategies in different regimes:

Exhibit 2 Scenario Analysis Using the Sharpe Ratio

Strategy/Regime	High Volatility	Low Volatility	Recession	Non-recession
Strategy I	0.88	0.64	0.20	1.00
Strategy II	1.56	1.60	1.76	1.52

Galic is surprised to see that some of the backtest results are unfavorable. He asks, "Why has GWP not considered strategies that perform better in backtesting?" Galic recently met with Fastlane Wealth Managers, who showed much better performance results. The portfolio manager at Fastlane told Galic that the company selects the top-performing strategies after performing thousands of backtests.

9 Which of Rom's statements concerning backtesting is correct?

- **A** Only Statement 1
- **B** Only Statement 2
- **C** Both Statement 1 and Statement 2

10 Which key parameter needs to be changed for a new backtest that includes Galic's restrictions?

- **A** Start and end dates
- **B** Consideration of transaction costs
- **C** Investment universe

11 Galic's concern embedded in the question "What happens if the future is different from the past?" is a problem most relevant for which investment strategy evaluation technique?

- **A** Sensitivity analysis
- **B** Backtesting
- **C** Monte Carlo simulation

12 Which of the following conclusions of Exhibit 1 is *least* likely to be true?

- **A** 5% of the time, losses from Factor 1 would be at least 6.49%.
- **B** When the VaR is exceeded in Factor 1, we should expect an average loss of 15.73%.
- **C** 5% of the time, losses from Factor 2 are likely to be worse than losses from Factor 1.

13 Based on the statistical study performed by GWP, which of the following represents a suggested course of action if GWP were to conduct Monte Carlo simulation analyses on the factor strategies?

A Inverse transformation

B Bootstrapping

C Sensitivity analysis

14 Based on Exhibit 1, which factor has the smallest downside risk as measured by the weighted average of all losses that exceed a threshold?

A Factor 1

B Factor 2

C Factor 3

15 The approach used by Fastlane Wealth Managers *most likely* incorporates:

A risk parity.

B data snooping.

C cross-validation.

16 Comparing the two strategies in Exhibit 2, the *best* risk-adjusted performance is demonstrated by:

A Strategy II in periods of low volatility and recession.

B Strategy I in periods of high volatility and non-recession.

C Strategy II in periods of high volatility and non-recession.

SOLUTIONS

1 A is correct. A reporting lag of four months is likely to introduce stale data into the backtest because many large-capitalization companies report earnings within 30–50 days of quarter end. Although assuming four months (120 days) of reporting lag will eliminate a source of look-ahead bias, it introduces a new problem (i.e., stale data).

B and C are incorrect. Data snooping and p-hacking refer to the same problem: a flawed approach to using data to make decisions. Data snooping and p-hacking are not characteristics of data, nor can they be added to a dataset by making an adjustment.

2 B is correct. Both VaR and maximum drawdown are downside risk measures, and both measures are lower for Strategy 2 than Strategy 3.

A is incorrect. We cannot deduce portfolio turnover from the metrics provided in Exhibit 1.

C is incorrect. We cannot deduce returns from the metrics provided in Exhibit 1.

3 C is correct. Both statements are incorrect. Statement 1 is incorrect because maximum drawdown and standard deviation are different measures. Maximum drawdown is typically used to represent downside risk, because it is the minimum cumulative return observed. Standard deviation is a measure of volatility. Although the two measures may be correlated, they are not substitutes for each other. Statement 2 is incorrect because two downside risk measures are presented: VaR and maximum drawdown. Factor Strategy 2 has the lowest reading for both measures, indicating that it has the *least* downside risk among the three strategies presented in Exhibit 1.

4 C is correct. Approach 1 is a historical simulation and assumes that past asset returns provide sufficient guidance about future asset returns.

A is incorrect because both approaches are non-deterministic and random in nature. Approach 1 is a historical simulation, and Approach 2 is a Monte Carlo simulation.

B is incorrect because Approach 1 is a historical simulation and each random variable of interest (key driver and/or decision variable) is randomly drawn from historical data. A functional form of the statistical distribution of returns for each decision variable needs to be specified for a Monte Carlo simulation, which is Approach 2.

5 C is correct. Approach 2 is a Monte Carlo simulation. The returns of Portfolios A and B are driven by the returns of the nine underlying factor portfolios (based on nine common growth factors). In the case of asset or factor allocation strategies, the returns from six of the nine factors are correlated, and therefore it is necessary to specify a multivariate distribution rather than modeling each factor or asset on a standalone basis.

A is incorrect because Approach 2 is a Monte Carlo simulation to generate investment performance data for the nine underlying factor portfolios. The returns of six of the nine factors are correlated, which means specifying a multivariate distribution rather than modeling each factor or asset on a standalone basis.

B is incorrect because the analyst should calculate the elements of the covariance matrix for all factors, not only the correlated factors. Doing so entails calculating 36, not 15, elements of the covariance matrix. Approach 2 is a

Monte Carlo simulation using the factor allocation strategies for Portfolios A and B for the nine factor portfolios, the returns of which are correlated, which means specifying a multivariate distribution. To calibrate the model, a few key parameters need to be calculated: the mean, the standard deviation, and the covariance matrix. For 9 assets, we need to estimate 9 mean returns, 9 standard deviations, and $\frac{9 \times (9-1)}{2} = 36$ elements of the covariance matrix. Assuming just the 6 correlated assets, the calculation is $\frac{6 \times (6-1)}{2} = 15$.

6 B is correct. The distribution of Factor 1 returns exhibits excess kurtosis and negative skewness (relative to the normal distribution). The excess kurtosis implies that these strategies are more likely to generate surprises, meaning extreme returns, whereas the negative skewness suggests those surprises are more likely to be negative (than positive).

A is incorrect because risk-averse investors are more likely to prefer distribution properties such as positive skew (higher probability of positive returns) and lower to moderate kurtosis (lower probability of extreme negative surprises). The distribution of Factor 1 returns exhibits excess kurtosis and negative skewness.

C is incorrect because the distribution of Factor 1 returns exhibits excess kurtosis and negative skewness. The joint distribution of such returns is rarely multivariate normal—so, typically the means and variances of these returns and the correlations between them are insufficient to describe the joint return distribution. In other words, the return data do not line up tightly around a trend line because of fat tails and outliers.

7 B is correct. Random sampling with replacement, also known as bootstrapping, is often used in historical simulations because the number of simulations needed is often larger than the size of the historical dataset. Because Approach 1 is a historical simulation and Concern 3 notes that the number of simulations needed is larger than the size of the historical dataset, bootstrapping should be used.

A is incorrect because this approach would result in creating observations that do not exist in the historical record. Doing so would violate the assumption and procedures of historical simulation.

C is incorrect because choosing the multivariate normal distribution as the initial functional form is typically done in a Monte Carlo simulation (Approach 2), not in a historical simulation (Approach 1). Historical simulation randomly samples from the historical dataset by drawing a number from a uniform distribution so that there is equal probability of being selected. Choice of distribution would not address the concern about the size of the dataset.

8 B is correct. Sensitivity analysis can be implemented to help managers understand how the target variable (portfolio returns) and risk profiles are affected by changes in input variables. Approach 2 is a Monte Carlo simulation, and the results depend on whether the multivariate normal distribution is the correct functional form or a reasonable proxy for the true distribution. Because this information is almost never known, sensitivity analysis using a multivariate skewed Student's *t*-distribution helps to account for empirical properties such as the skewness and the excess kurtosis observed in the underlying factor return data.

A is incorrect. Data snooping is the subconscious or conscious manipulation of data in a way that produces a statistically significant result (i.e., a p-value that is sufficiently small or a t-statistic that is sufficiently large to indicate statistically significance).

C is incorrect. The inverse transformation method is the process of converting a randomly generated number into a simulated value of a random variable.

9 A is correct. Statement 1 is correct because the main objective of backtesting is to understand the risk–return tradeoff of an investment strategy by approximating the real-life investment process.

B is incorrect because Statement 2 is inaccurate. Although backtesting fits quantitative and systematic investment styles more naturally, it has also been heavily used by fundamental managers.

C is incorrect because Statement 2 is not accurate. Backtesting, used in quantitative and systematic investment styles, is also heavily used by fundamental managers.

10 C is correct. Investment universe represents the securities in which a strategy can potentially invest. Galic's restriction to exclusively own domestic securities means the investment universe of a backtest for a strategy for Galic's account should use a domestic rather than global investment universe.

A is incorrect. Galic's restriction to domestic securities does not affect the start and end dates for a backtest.

B is incorrect. Galic's restriction to domestic securities does not change the inclusion of transaction costs in the study.

11 B is correct. An implicit assumption of backtesting is that past returns are a guide to future asset returns.

A is incorrect. Sensitivity analysis refers to modifying assumptions such as probability distributions of key variables in a Monte Carlo simulation, which is a non-deterministic evaluation technique that does not use historical data.

C is incorrect. Monte Carlo simulation is a non-deterministic evaluation technique that does not use historical data.

12 C is correct. The VaR metrics in Exhibit 1 show that 5% of the time, losses will be at least 6.49% and 0.77%, respectively, for Factor 1 and Factor 2. The CVaR metrics in Exhibit 1 show that the weighted average of all loss outcomes that exceed the VaR loss are 15.73% and 4.21% for Factor 1 and Factor 2, respectively. Thus, A is true because it correctly defines VaR, and B is true because it correctly defines CVaR, whereas C is untrue because both VaR and CVaR are lower for Factor 2 than Factor 1.

13 C is correct. Performing sensitivity analysis represents best practice given these characteristics, because the user could test different probability distributions that relax the assumptions of the normal distribution, for example.

A is incorrect. Inverse transformation is a method of random observation generation, often used in simulation.

B is incorrect. Bootstrapping refers to random sampling with replacement, often used in historical simulation.

14 C is correct. Exhibit 1 presents three downside risk measures: VaR, CVaR, and maximum drawdown. Conditional VaR is defined as the weighted average of all loss outcomes in the return distribution that exceed the VaR loss. Thus, CVaR is a more comprehensive measure of tail loss than VaR. Based on Exhibit 1, the factor with the smallest downside risk based on CVaR is Factor 3.

15 B is correct. The fact that the two firms' investment performance results differ over similar time horizons using the same data and factors may be the result of selection bias. Data snooping is a type of selection bias. Fastlane Wealth Managers is most likely selecting the best-performing modeling approach and publishing its results (i.e., data snooping).

A is incorrect because risk parity is a portfolio construction technique that accounts for the volatility of each factor and the correlations of returns among all factors to be combined in the portfolio. It is not regarded as selection bias.

C is incorrect because cross-validation is a technique used in the machine learning field, as well as in backtesting investment strategies, to partition data for model training and testing. It is not considered selection bias.

16 A is correct. Using the Sharpe ratio, the best risk-adjusted relative performance can be determined by comparing the sensitivity of the two strategies under differing macroeconomic regimes: recession versus non-recession and high volatility versus low volatility. The best risk-adjusted return will exhibit the highest Sharpe ratio. Strategy II demonstrates higher risk-adjusted returns compared with Strategy I under all four macroeconomic conditions, particularly in periods of low volatility, when the Sharpe ratio outperformance is 0.96, and recessions, when the Sharpe ratio outperformance is 1.56.

Scenario Analysis Using Sharpe Ratio

Strategy/Regime	High Volatility	Low Volatility	Recession	Non-recession
Strategy I	0.88	0.64	0.20	1.00
Strategy II	1.56	1.60	1.76	1.52
Difference (II – I)	0.68	0.96	1.56	0.52

Glossary

Abandonment option The ability to terminate a project at some future time if the financial results are disappointing.

Abnormal earnings See *residual income.*

Abnormal return The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

Absolute convergence The idea that developing countries, regardless of their particular characteristics, will eventually catch up with the developed countries and match them in per capita output.

Absolute valuation model A model that specifies an asset's intrinsic value.

Absolute version of PPP An extension of the law of one price whereby the prices of goods and services will not differ internationally once exchange rates are considered.

Accounting estimates Estimates used in calculating the value of assets or liabilities and in the amount of revenue and expense to allocate to a period. Examples of accounting estimates include, among others, the useful lives of depreciable assets, the salvage value of depreciable assets, product returns, warranty costs, and the amount of uncollectible receivables.

Accumulated benefit obligation The actuarial present value of benefits (whether vested or non-vested) attributed, generally by the pension benefit formula, to employee service rendered before a specified date and based on employee service and compensation (if applicable) before that date. The accumulated benefit obligation differs from the projected benefit obligation in that it includes no assumption about future compensation levels.

Accuracy The percentage of correctly predicted classes out of total predictions. It is an overall performance metric in classification problems.

Acquirer The company in a merger or acquisition that is acquiring the target.

Acquiring company See *acquirer*.

Acquisition The purchase of some portion of one company by another; the purchase may be for assets, a definable segment of another entity, or the entire company.

Activation function A functional part of a neural network's node that transforms the total net input received into the final output of the node. The activation function operates like a light dimmer switch that decreases or increases the strength of the input.

Active factor risk The contribution to active risk squared resulting from the portfolio's different-than-benchmark exposures relative to factors specified in the risk model.

Active return The return on a portfolio minus the return on the portfolio's benchmark.

Active risk The standard deviation of active returns.

Active risk squared The variance of active returns; active risk raised to the second power.

Active share A measure of how similar a portfolio is to its benchmark. A manager who precisely replicates the benchmark will have an active share of zero; a manager with no holdings in common with the benchmark will have an active share of one.

Active specific risk The contribution to active risk squared resulting from the portfolio's active weights on individual assets as those weights interact with assets' residual risk.

Adjusted funds from operations (AFFO) Funds from operations adjusted to remove any non-cash rent reported under straight-line rent accounting and to subtract maintenance-type capital expenditures and leasing costs, including leasing agents' commissions and tenants' improvement allowances.

Adjusted present value As an approach to valuing a company, the sum of the value of the company, assuming no use of debt, and the net present value of any effects of debt on company value.

Adjusted R^2 A measure of goodness-of-fit of a regression that is adjusted for degrees of freedom and hence does not automatically increase when another independent variable is added to a regression.

Administrative regulations or administrative law Rules issued by government agencies or other regulators.

Advanced set An arrangement in which the reference interest rate is set at the time the money is deposited.

Advanced settled An arrangement in which a forward rate agreement (FRA) expires and settles at the same time, at the FRA expiration date.

Agency costs Costs associated with the conflict of interest present when a company is managed by non-owners. Agency costs result from the inherent conflicts of interest between managers and equity owners.

Agency costs of equity The smaller the stake managers have in the company, the less their share in bearing the cost of excessive perquisite consumption—consequently, the less their desire to give their best efforts in running the company.

Agency issues Conflicts of interest that arise when the agent in an agency relationship has goals and incentives that differ from the principal to whom the agent owes a fiduciary duty. Also called *agency problems* or *principal–agent problems.*

Agglomerative clustering A bottom-up hierarchical clustering method that begins with each observation being treated as its own cluster. The algorithm finds the two closest clusters, based on some measure of distance (similarity), and combines them into one new larger cluster. This process is repeated iteratively until all observations are clumped into a single large cluster.

Allowance for loan losses A balance sheet account; it is a contra asset account to loans.

Alpha The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

American Depositary Receipt A negotiable certificate issued by a depositary bank that represents ownership in a non-US company's deposited equity (i.e., equity held in custody by the depositary bank in the company's home market).

Analysis of variance (ANOVA) The analysis that breaks the total variability of a dataset (such as observations on the dependent variable in a regression) into components representing different sources of variation. With reference to regression, ANOVA provides the inputs for an *F*-test of

the significance of the regression as a whole, as well as the inputs for the coefficient of determination and the standard error of the estimate.

Application programming interface (API) A set of well-defined methods of communication between various software components and typically used for accessing external data.

Arbitrage (1) The simultaneous purchase of an undervalued asset or portfolio and sale of an overvalued but equivalent asset or portfolio in order to obtain a riskless profit on the price differential. Taking advantage of a market inefficiency in a risk-free manner. (2) The condition in a financial market in which equivalent assets or combinations of assets sell for two different prices, creating an opportunity to profit at no risk with no commitment of money. In a well-functioning financial market, few arbitrage opportunities are possible. (3) A risk-free operation that earns an expected positive net profit but requires no net investment of money.

Arbitrage-free models Term structure models that project future interest rate paths that emanate from the existing term structure. Resulting prices are based on a no-arbitrage condition.

Arbitrage-free valuation An approach to valuation that determines security values consistent with the absence of any opportunity to earn riskless profits without any net investment of money.

Arbitrage opportunity An opportunity to conduct an arbitrage; an opportunity to earn an expected positive net profit without risk and with no net investment of money.

Arbitrage portfolio The portfolio that exploits an arbitrage opportunity.

Ask price The price at which a trader will sell a specified quantity of a security. Also called *ask*, *offer price*, or *offer*.

Asset-based approach Approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities.

Asset-based valuation An approach to valuing natural resource companies that estimates company value on the basis of the market value of the natural resources the company controls.

Asset beta The unlevered beta; reflects the business risk of the assets; the asset's systematic risk.

Asset purchase An acquisition in which the acquirer purchases the target company's assets and payment is made directly to the target company.

Asymmetric information The differential of information between corporate insiders and outsiders regarding the company's performance and prospects. Managers typically have more information about the company's performance and prospects than owners and creditors.

At market contract When a forward contract is established, the forward price is negotiated so that the market value of the forward contract on the initiation date is zero.

Authorized participants (APs) A special group of institutional investors who are authorized by the ETF issuer to participate in the creation/redemption process. APs are large broker/dealers, often market makers.

Autocorrelations The correlations of a time series with its own past values.

Autoregressive model (AR) A time series regressed on its own past values in which the independent variable is a lagged value of the dependent variable.

Backtesting The process that approximates the real-life investment process, using historical data, to assess whether an investment strategy would have produced desirable results.

Backward integration A merger involving the purchase of a target ahead of the acquirer in the value or production chain; for example, to acquire a supplier.

Backward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving backward through the network's layers.

Backwardation A condition in futures markets in which the spot price exceeds the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is higher than the longer-term futures contract price.

Bag-of-words (BOW) A collection of a distinct set of tokens from all the texts in a sample dataset. BOW does not capture the position or sequence of words present in the text.

Bankruptcy A declaration provided for by a country's laws that typically involves the establishment of a legal procedure that forces creditors to defer their claims.

Barbell portfolio Fixed-income portfolio that combines short and long maturities.

Base error Model error due to randomness in the data.

Basic earnings per share (EPS) Net earnings available to common shareholders (i.e., net income minus preferred dividends) divided by the weighted average number of common shares outstanding during the period.

Basis The difference between the spot price and the futures price. As the maturity date of the futures contract nears, the basis converges toward zero.

Basis trade A trade based on the pricing of credit in the bond market versus the price of the same credit in the CDS market. To execute a basis trade, go long the "underpriced" credit and short the "overpriced" credit. A profit is realized as the implied credit prices converge.

Bear hug A tactic used by acquirers to circumvent target management's objections to a proposed merger by submitting the proposal directly to the target company's board of directors.

Bearish flattening Term structure shift in which short-term bond yields rise more than long-term bond yields, resulting in a flatter yield curve.

Benchmark value of the multiple In using the method of comparables, the value of a price multiple for the comparison asset; when we have comparison assets (a group), the mean or median value of the multiple for the group of assets.

Best ask The offer to sell with the lowest ask price. Also called *best offer* or *inside ask*.

Best bid The offer to buy with the highest bid price. Also called the *inside bid*.

Best offer See *best ask*.

Bias error Describes the degree to which a model fits the training data. Algorithms with erroneous assumptions produce high bias error with poor approximation, causing underfitting and high in-sample error.

Bid–ask spread The ask price minus the bid price.

Bid price The price at which a trader will buy a specified quantity of a security. Also called *bid*.

Bill-and-hold basis Sales on a bill-and-hold basis involve selling products but not delivering those products until a later date.

Blockage factor An illiquidity discount that occurs when an investor sells a large amount of stock relative to its trading volume (assuming it is not large enough to constitute a controlling ownership).

Bond indenture A legal contract specifying the terms of a bond issue.

Bond risk premium The expected excess return of a default-free long-term bond less that of an equivalent short-term bond.

Bond yield plus risk premium method An estimate of the cost of common equity that is produced by summing the before-tax cost of debt and a risk premium that captures the additional yield on a company's stock relative to its bonds. The additional yield is often estimated using historical spreads between bond yields and stock yields.

Bonding costs Costs borne by management to assure owners that they are working in the owners' best interest (e.g., implicit cost of non-compete agreements).

Bonus issue of shares *See stock dividend.*

Book value Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value of equity Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value per share The amount of book value (also called carrying value) of common equity per share of common stock, calculated by dividing the book value of shareholders' equity by the number of shares of common stock outstanding.

Bootstrap aggregating (or bagging) A technique whereby the original training dataset is used to generate *n* new training datasets or bags of data. Each new bag of data is generated by random sampling with replacement from the initial training set.

Bootstrapping The use of a forward substitution process to determine zero-coupon rates by using the par yields and solving for the zero-coupon rates one by one, from the shortest to longest maturities.

Bottom-up approach With respect to forecasting, an approach that usually begins at the level of the individual company or a unit within the company.

Breakup value The value derived using a sum-of-the-parts valuation.

Breusch–Pagan test A test for conditional heteroskedasticity in the error term of a regression.

Bullet portfolio A fixed-income portfolio concentrated in a single maturity.

Bullish flattening Term structure change in which the yield curve flattens in response to a greater decline in long-term rates than short-term rates.

Bullish steepening Term structure change in which short-term rates fall by more than long-term yields, resulting in a steeper term structure.

Buy-side analysts Analysts who work for investment management firms, trusts, bank trust departments, and similar institutions.

Buyback *See share repurchase.*

Callable bond Bond that includes an embedded call option that gives the issuer the right to redeem the bond issue prior to maturity, typically when interest rates have fallen or when the issuer's credit quality has improved.

Canceled shares Shares that were issued, subsequently repurchased by the company, and then retired (cannot be reissued).

Cannibalization Cannibalization occurs when an investment takes customers and sales away from another part of the company.

Capital charge The company's total cost of capital in money terms.

Capital deepening An increase in the capital-to-labor ratio.

Capital rationing A capital rationing environment assumes that the company has a fixed amount of funds to invest.

Capital structure The mix of debt and equity that a company uses to finance its business; a company's specific mixture of long-term financing.

Capitalization of earnings method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capitalization rate The divisor in the expression for the value of perpetuity. In the context of real estate, it is the divisor in the direct capitalization method of estimating value. The cap rate equals net operating income divided by value.

Capitalized cash flow method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity. Also called *capitalized cash flow model.*

Capitalized income method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capped floater Floating-rate bond with a cap provision that prevents the coupon rate from increasing above a specified maximum rate. It protects the issuer against rising interest rates.

Carry arbitrage model A no-arbitrage approach in which the underlying instrument is either bought or sold along with an opposite position in a forward contract.

Carry benefits Benefits that arise from owning certain underlyings; for example, dividends, foreign interest, and bond coupon payments.

Carry costs Costs that arise from owning certain underlyings. They are generally a function of the physical characteristics of the underlying asset and also the interest forgone on the funds tied up in the asset.

Cash available for distribution See *adjusted funds from operations.*

Cash-generating unit The smallest identifiable group of assets that generates cash inflows that are largely independent of the cash inflows of other assets or groups of assets.

Cash offering A merger or acquisition that is to be paid for with cash; the cash for the merger might come from the acquiring company's existing assets or from a debt issue.

Cash settlement A procedure used in certain derivative transactions that specifies that the long and short parties settle the derivative's difference in value between them by making a cash payment.

Catalyst An event or piece of information that causes the marketplace to re-evaluate the prospects of a company.

Categorical dependent variables An alternative term for qualitative dependent variables.

CDS spread A periodic premium paid by the buyer to the seller that serves as a return over a market reference rate required to protect against credit risk.

Ceiling analysis A systematic process of evaluating different components in the pipeline of model building. It helps to understand what part of the pipeline can potentially improve in performance by further tuning.

Centroid The center of a cluster formed using the *k*-means clustering algorithm.

Chain rule of forecasting A forecasting process in which the next period's value as predicted by the forecasting equation is substituted into the right-hand side of the equation to give a predicted value two periods ahead.

Cheapest-to-deliver The debt instrument that can be purchased and delivered at the lowest cost yet has the same seniority as the reference obligation.

Classification and regression tree A supervised machine learning technique that can be applied to predict either a categorical target variable, producing a classification tree, or a continuous target variable, producing a regression tree. CART is commonly applied to binary classification or regression.

Clean surplus relation The relationship between earnings, dividends, and book value in which ending book value is equal to the beginning book value plus earnings less dividends, apart from ownership transactions.

Club convergence The idea that only rich and middle-income countries sharing a set of favorable attributes (i.e., are members of the "club") will converge to the income level of the richest countries.

Cluster A subset of observations from a dataset such that all the observations within the same cluster are deemed "similar."

Clustering The sorting of observations into groups (clusters) such that observations in the same cluster are more similar to each other than they are to observations in other clusters.

Cobb–Douglas production function A function of the form $Y = K^{\alpha} L^{1-\alpha}$ relating output (Y) to labor (L) and capital (K) inputs.

Coefficient of determination The percentage of the variation of the dependent variable that is explained by the independent variable. Also referred to as the "*R*-squared" or "R^2."

Cointegrated Describes two time series that have a long-term financial or economic relationship such that they do not diverge from each other without bound in the long run.

Collateral return The component of the total return on a commodity futures position attributable to the yield for the bonds or cash used to maintain the futures position. Also called *collateral yield*.

Collection frequency (CF) The number of times a given word appears in the whole corpus (i.e., collection of sentences) divided by the total number of words in the corpus.

Commercial real estate properties Income-producing real estate properties; properties purchased with the intent to let, lease, or rent (in other words, produce income).

Commodity swap A type of swap involving the exchange of payments over multiple dates as determined by specified reference prices or indexes relating to commodities.

Common size statements Financial statements in which all elements (accounts) are stated as a percentage of a key figure, such as revenue for an income statement or total assets for a balance sheet.

Company fundamental factors Factors related to the company's internal performance, such as factors relating to earnings growth, earnings variability, earnings momentum, and financial leverage.

Company share-related factors Valuation measures and other factors related to share price or the trading characteristics of the shares, such as earnings yield, dividend yield, and book-to-market value.

Comparables Assets used as benchmarks when applying the method of comparables to value an asset. Also called *comps*, *guideline assets*, or *guideline companies*.

Competition laws A law that promotes or maintains market competition by regulating anti-competitive conduct. Known as "antitrust law" in the United States, "anti-monopoly law" in China and Russia, and often referred to as "trade practices law" in the United Kingdom and Australia.

Compiled financial statements Financial statements that are not accompanied by an auditor's opinion letter.

Complexity A term referring to the number of features, parameters, or branches in a model and to whether the model is linear or non-linear (non-linear is more complex).

Composite variable A variable that combines two or more variables that are statistically strongly related to each other.

Comprehensive income All changes in equity other than contributions by, and distributions to, owners; income under clean surplus accounting; includes all changes in equity during a period except those resulting from investments by owners and distributions to owners. Comprehensive income equals net income plus other comprehensive income.

Comps Assets used as benchmarks when applying the method of comparables to value an asset.

Concentrated ownership Ownership structure consisting of an individual shareholder or a group (controlling shareholders) with the ability to exercise control over the corporation.

Conditional convergence The idea that convergence of per capita income is conditional on the countries having the same savings rate, population growth rate, and production function.

Conditional heteroskedasticity Heteroskedasticity in the error variance that is correlated with the values of the independent variable(s) in the regression.

Conditional VaR (CVaR) The weighted average of all loss outcomes in the statistical (i.e., return) distribution that exceed the VaR loss. Thus, CVaR is a more comprehensive measure of tail loss than VaR is. Sometimes referred to as the *expected tail loss* or *expected shortfall*.

Confusion matrix A grid used for error analysis in classification problems, it presents values for four evaluation metrics including true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Conglomerate discount The discount possibly applied by the market to the stock of a company operating in multiple, unrelated businesses.

Conglomerate merger A merger involving companies that are in unrelated businesses.

Consolidation The combining of the results of operations of subsidiaries with the parent company to present financial statements as if they were a single economic unit. The assets, liabilities, revenues, and expenses of the subsidiaries are combined with those of the parent company, eliminating intercompany transactions.

Constant dividend payout ratio policy A policy in which a constant percentage of net income is paid out in dividends.

Constant returns to scale The condition that if all inputs into the production process are increased by a given percentage, then output rises by that same percentage.

Contango A condition in futures markets in which the spot price is lower than the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is lower than the longer-term futures contract price.

Contingent consideration Potential future payments to the seller that are contingent on the achievement of certain agreed-on occurrences.

Continuing earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *persistent earnings*, or *underlying earnings*.

Continuing residual income Residual income after the forecast horizon.

Continuing value The analyst's estimate of a stock's value at a particular point in the future.

Control premium An increment or premium to value associated with a controlling ownership interest in a company.

Convergence The property by which as expiration approaches, the price of a newly created forward or futures contract will approach the price of a spot transaction. At expiration, a forward or futures contract is equivalent to a spot transaction in the underlying.

Conversion period For a convertible bond, the period during which bondholders have the right to convert their bonds into shares.

Conversion price For a convertible bond, the price per share at which the bond can be converted into shares.

Conversion rate (or ratio) For a convertible bond, the number of shares of common stock that a bondholder receives from converting the bond into shares.

Conversion value For a convertible bond, the value of the bond if it is converted at the market price of the shares. Also called *parity value*.

Convertible bond Bond with an embedded conversion option that gives bondholders the right to convert their bonds into the issuer's common stock during a pre-determined period at a pre-determined price.

Convexity A measure of how interest rate sensitivity changes with a change in interest rates.

Core earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *persistent earnings*, or *underlying earnings*.

Core real estate investment style Investing in high-quality, well-leased, core property types with low leverage (no more than 30% of asset value) in the largest markets with strong, diversified economies. It is a conservative strategy designed to avoid real estate–specific risks, including leasing, development, and speculation in favor of steady returns. Hotel properties are excluded from the core categories because of the higher cash flow volatility resulting from single-night leases and the greater importance of property operations, brand, and marketing.

Corpus A collection of text data in any form, including list, matrix, or data table forms.

Cost approach An approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities. In the context of real estate, this approach estimates the value of a property based on what it would cost to buy the land and construct a new property on the site that has the same utility or functionality as the property being appraised.

Cost of carry model A model that relates the forward price of an asset to the spot price by considering the cost of carry (also referred to as future-spot parity model).

Cost of debt The cost of debt financing to a company, such as when it issues a bond or takes out a bank loan.

Cost of equity The required rate of return on common stock.

Covariance stationary Describes a time series when its expected value and variance are constant and finite in all periods and when its covariance with itself for a fixed number of periods in the past or future is constant and finite in all periods.

Covered bonds A senior debt obligation of a financial institution that gives recourse to the originator/issuer and a predetermined underlying collateral pool.

Covered interest rate parity The relationship among the spot exchange rate, the forward exchange rate, and the interest rates in two currencies that ensures that the return on a hedged (i.e., covered) foreign risk-free investment is the same as the return on a domestic risk-free investment. Also called *interest rate parity*.

Cox-Ingersoll-Ross model A general equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is directly related to the level of interest rates.

Creation basket The list of securities (and share amounts) the authorized participant (AP) must deliver to the ETF manager in exchange for ETF shares. The creation basket is published each business day.

Creation/redemption The process in which ETF shares are created or redeemed by authorized participants transacting with the ETF issuer.

Creation units Large blocks of ETF shares transacted between the authorized participant (AP) and the ETF manager that are usually but not always equal to 50,000 shares of the ETF.

Credit correlation The correlation of credit (or default) risks of the underlying single-name CDS contained in an index CDS.

Credit curve The credit spreads for a range of maturities of a company's debt.

Credit default swap A derivative contract between two parties in which the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit derivative A derivative instrument in which the underlying is a measure of the credit quality of a borrower.

Credit event The event that triggers a payment from the credit protection seller to the credit protection buyer.

Credit protection buyer One party to a credit default swap; the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit protection seller One party to a credit default swap; the seller makes a promise to pay compensation for credit losses resulting from the default.

Credit risk The risk that the borrower will not repay principal and interest. Also called *default risk*.

Credit valuation adjustment The value of the credit risk of a bond in present value terms.

Cross-validation A technique for estimating out-of-sample error directly by determining the error in validation samples.

Current exchange rate For accounting purposes, the spot exchange rate on the balance sheet date.

Current rate method Approach to translating foreign currency financial statements for consolidation in which all assets and liabilities are translated at the current exchange rate. The current rate method is the prevalent method of translation.

Curvature One of the three factors (the other two are level and steepness) that empirically explain most of the changes in the shape of the yield curve. A shock to the curvature factor affects mid-maturity interest rates, resulting in the term structure becoming either more or less hump-shaped.

Curve trade Buying a CDS of one maturity and selling a CDS on the same reference entity with a different maturity.

Cyclical businesses Businesses with high sensitivity to business- or industry-cycle influences.

Data mining The practice of determining a model by extensive searching through a dataset for statistically significant patterns.

Data preparation (cleansing) The process of examining, identifying, and mitigating (i.e., cleansing) errors in raw data.

Data snooping The subconscious or conscious manipulation of data in a way that produces a statistically significant result (i.e., the *p*-value is sufficiently small or the *t*-statistic sufficiently large to indicate statistical significance), such as by running multiple simulations and naively accepting the best result. Also known as p-hacking.

Data wrangling (preprocessing) This task performs transformations and critical processing steps on cleansed data to make the data ready for ML model training (i.e., preprocessing), and includes dealing with outliers, extracting useful variables from existing data points, and scaling the data.

"Dead-hand" provision A poison pill provision that allows for the redemption or cancellation of a poison pill provision only by a vote of continuing directors (generally directors who were on the target company's board prior to the takeover attempt).

Debt rating An objective measure of the quality and safety of a company's debt based upon an analysis of the company's ability to pay the promised cash flows. It includes an analysis of any indentures.

Deep learning Algorithms based on deep neural networks, ones with many hidden layers (more than two), that address highly complex tasks, such as image classification, face recognition, speech recognition, and natural language processing.

Deep neural networks Neural networks with many hidden layers—at least 2 but potentially more than 20—that have proven successful across a wide range of artificial intelligence applications.

Default risk See *credit risk*.

Defined benefit pension plans Plan in which the company promises to pay a certain annual amount (defined benefit) to the employee after retirement. The company bears the investment risk of the plan assets.

Defined contribution pension plans Individual accounts to which an employee and typically the employer makes contributions, generally on a tax-advantaged basis. The amounts of contributions are defined at the outset, but the future value of the benefit is unknown. The employee bears the investment risk of the plan assets.

Definitive merger agreement A contract signed by both parties to a merger that clarifies the details of the transaction, including the terms, warranties, conditions, termination details, and the rights of all parties.

Delay costs Implicit trading costs that arise from the inability to complete desired trades immediately. Also called *slippage.*

Delta The relationship between the option price and the underlying price, which reflects the sensitivity of the price of the option to changes in the price of the underlying. Delta is a good approximation of how an option price will change for a small change in the stock.

Dendrogram A type of tree diagram used for visualizing a hierarchical cluster analysis; it highlights the hierarchical relationships among the clusters.

Dependent variable The variable whose variation about its mean is to be explained by the regression; the left-side variable in a regression equation. Also referred to as the *explained variable*.

Depository Trust and Clearinghouse Corporation A US-headquartered entity providing post-trade clearing, settlement, and information services.

Descriptive statistics The study of how data can be summarized effectively.

Diluted earnings per share (Diluted EPS)Net income, minus preferred dividends, divided by the weighted average number of common shares outstanding considering all dilutive securities (e.g., convertible debt and options); the EPS that would result if all dilutive securities were converted into common shares.

Dilution A reduction in proportional ownership interest as a result of the issuance of new shares.

Dimension reduction A set of techniques for reducing the number of features in a dataset while retaining variation across observations to preserve the information contained in that variation.

Diminishing marginal productivity When each additional unit of an input, keeping the other inputs unchanged, increases output by a smaller increment.

Direct capitalization method In the context of real estate, this method estimates the value of an income-producing property based on the level and quality of its net operating income.

Discount To reduce the value of a future payment in allowance for how far away it is in time; to calculate the present value of some future amount. Also, the amount by which an instrument is priced below its face value.

Discount factor The present value or price of a risk-free single-unit payment when discounted using the appropriate spot rate.

Discount for lack of control An amount or percentage deducted from the pro rata share of 100% of the value of an equity interest in a business to reflect the absence of some or all of the powers of control.

Discount for lack of marketability An amount of percentage deducted from the value of an ownership interest to reflect the relative absence of marketability.

Discount function Discount factors for the range of all possible maturities. The spot curve can be derived from the discount function and vice versa.

Discount rate Any rate used in finding the present value of a future cash flow.

Discounted abnormal earnings model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Discounted cash flow (DCF) analysis In the context of merger analysis, an estimate of a target company's value found by discounting the company's expected future free cash flows to the present.

Discounted cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows. In the context of real estate, this method estimates the value of an income-producing property based on discounting future projected cash flows.

Discounted cash flow model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Discriminant analysis A multivariate classification technique used to discriminate between groups, such as companies that either will or will not become bankrupt during some time frame.

Dispersed ownership Ownership structure consisting of many shareholders, none of which has the ability to individually exercise control over the corporation.

Divestiture The sale, liquidation, or spin-off of a division or subsidiary.

Dividend A distribution paid to shareholders based on the number of shares owned.

Dividend coverage ratio The ratio of net income to dividends.

Dividend discount model (DDM) A present value model of stock value that views the intrinsic value of a stock as present value of the stock's expected future dividends.

Dividend displacement of earnings The concept that dividends paid now displace earnings in all future periods.

Dividend imputation tax system A taxation system that effectively assures corporate profits distributed as dividends are taxed just once and at the shareholder's tax rate.

Dividend index point A measure of the quantity of dividends attributable to a particular index.

Dividend payout ratio The ratio of cash dividends paid to earnings for a period.

Dividend policy The strategy a company follows with regard to the amount and timing of dividend payments.

Dividend rate The annualized amount of the most recent dividend.

Dividend yield Annual dividends per share divided by share price.

Divisive clustering A top-down hierarchical clustering method that starts with all observations belonging to a single large cluster. The observations are then divided into two clusters based on some measure of distance (similarity). The algorithm then progressively partitions the intermediate clusters into smaller ones until each cluster contains only one observation.

Document frequency (DF) The number of documents (texts) that contain a particular token divided by the total number of documents. It is the simplest feature selection method and often performs well when many thousands of tokens are present.

Document term matrix (DTM) A matrix where each row belongs to a document (or text file), and each column represents a token (or term). The number of rows is equal to the number of documents (or text files) in a sample text dataset. The number of columns is equal to the number of tokens from the BOW built using all the documents in the sample dataset. The cells typically contain the counts of the number of times a token is present in each document.

Dominance An arbitrage opportunity when a financial asset with a risk-free payoff in the future must have a positive price today.

Double taxation system Corporate earnings are taxed twice when paid out as dividends. First, corporate pretax earnings are taxed regardless of whether they will be distributed as dividends or retained at the corporate level. Second, dividends are taxed again at the individual shareholder level.

Downstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary) such that the investor company records a profit on its income statement. An example is a sale of inventory by the investor company to the associate or by a parent to a subsidiary company.

Dual-class shares Shares that grant one share class superior or even sole voting rights, whereas the other share class has inferior or no voting rights.

Due diligence Investigation and analysis in support of a recommendation; the failure to exercise due diligence may sometimes result in liability according to various securities laws.

Dummy variable A type of qualitative variable that takes on a value of 1 if a particular condition is true and 0 if that condition is false.

Duration A measure of the approximate sensitivity of a security to a change in interest rates (i.e., a measure of interest rate risk).

Dutch disease A situation in which currency appreciation driven by strong export demand for resources makes other segments of the economy (particularly manufacturing) globally uncompetitive.

Earnings surprise The difference between reported EPS and expected EPS. Also referred to as *unexpected earnings*.

Earnings yield EPS divided by price; the reciprocal of the P/E.

Economic profit See *residual income*.

Economic sectors Large industry groupings.

Economic value added (EVA®) A commercial implementation of the residual income concept; the computation of EVA® is the net operating profit after taxes minus the cost of capital, where these inputs are adjusted for a number of items.

Economies of scale A situation in which average costs per unit of good or service produced fall as volume rises. In reference to mergers, the savings achieved through the consolidation of operations and elimination of duplicate resources.

Edwards–Bell–Ohlson model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Effective convexity Sensitivity of duration to changes in interest rates.

Effective duration Sensitivity of the bond's price to a 100 bps parallel shift of the benchmark yield curve, assuming no change in the bond's credit spread.

Effective spread Two times the difference between the execution price and the midpoint of the market quote at the time an order is entered.

Eigenvalue A measure that gives the proportion of total variance in the initial dataset that is explained by each eigenvector.

Eigenvector A vector that defines new mutually uncorrelated composite variables that are linear combinations of the original features.

Embedded options Contingency provisions found in a bond's indenture or offering circular representing rights that enable their holders to take advantage of interest rate movements. They can be exercised by the issuer, by the bondholder, or automatically depending on the course of interest rates.

Ensemble learning A technique of combining the predictions from a collection of models to achieve a more accurate prediction.

Ensemble method The method of combining multiple learning algorithms, as in ensemble learning.

Enterprise value Total company value (the market value of debt, common equity, and preferred equity) minus the value of cash and investments.

Enterprise value multiple A valuation multiple that relates the total market value of all sources of a company's capital (net of cash) to a measure of fundamental value for the entire company (such as a pre-interest earnings measure).

Equilibrium The condition in which supply equals demand.

Equity carve-out A form of restructuring that involves the creation of a new legal entity and the sale of equity in it to outsiders.

Equity charge The estimated cost of equity capital in money terms.

Equity REITs REITs that own, operate, and/or selectively develop income-producing real estate.

Equity swap A swap transaction in which at least one cash flow is tied to the return on an equity portfolio position, often an equity index.

Error autocorrelations The autocorrelations of the error term.

Error term The difference between an observation and its expected value, where the expected value is based on the true underlying population relation between the dependent and independent variables. Also known simply as the *error*.

ESG integration An ESG investment approach that focuses on systematic consideration of material ESG factors in asset allocation, security selection, and portfolio construction decisions for the purpose of achieving the product's stated investment objectives.

Estimated parameters With reference to a regression analysis, the estimated values of the population intercept and population slope coefficients in a regression.

***Ex ante* tracking error** A measure of the degree to which the performance of a given investment portfolio might be expected to deviate from its benchmark; also known as *relative VaR*.

***Ex ante* version of PPP** The hypothesis that expected changes in the spot exchange rate are equal to expected differences in national inflation rates. An extension of relative purchasing power parity to expected future changes in the exchange rate.

Ex-dividend Trading ex-dividend refers to shares that no longer carry the right to the next dividend payment.

Ex-dividend date The first date that a share trades without (i.e., "ex") the right to receive the declared dividend for the period.

Excess earnings method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Exchange ratio The number of shares that target stockholders are to receive in exchange for each of their shares in the target company.

Exercise date The date when employees actually exercise stock options and convert them to stock.

Exercise value The value of an option if it were exercised. Also sometimes called *intrinsic value*.

Expanded CAPM An adaptation of the CAPM that adds to the CAPM a premium for small size and company-specific risk.

Expectations approach A procedure for obtaining the value of an option derived from discounting at the risk-free rate its expected future payoff based on risk neutral probabilities.

Expected exposure The projected amount of money an investor could lose if an event of default occurs, before factoring in possible recovery.

Expected holding-period return The expected total return on an asset over a stated holding period; for stocks, the sum of the expected dividend yield and the expected price appreciation over the holding period.

Expected shortfall See *conditional VaR*.

Expected tail loss See *conditional VaR*.

Exploratory data analysis (EDA) The preliminary step in data exploration, where graphs, charts, and other visualizations (heat maps and word clouds) as well as quantitative methods (descriptive statistics and central tendency measures) are used to observe and summarize data.

Exposure to foreign exchange risk The risk of a change in value of an asset or liability denominated in a foreign currency due to a change in exchange rates.

Extendible bond Bond with an embedded option that gives the bondholder the right to keep the bond for a number of years after maturity, possibly with a different coupon.

External growth Company growth in output or sales that is achieved by buying the necessary resources externally (i.e., achieved through mergers and acquisitions).

Extra dividend *See special dividend.*

F1 score The harmonic mean of precision and recall. F1 score is a more appropriate overall performance metric (than accuracy) when there is unequal class distribution in the dataset and it is necessary to measure the equilibrium of precision and recall.

Factor A common or underlying element with which several variables are correlated.

Factor betas An asset's sensitivity to a particular factor; a measure of the response of return to each unit of increase in a factor, holding all other factors constant.

Factor portfolio See *pure factor portfolio*.

Factor price The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors.

Factor risk premium The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors. Also called *factor price*.

Factor sensitivity See *factor betas*.

Failure to pay When a borrower does not make a scheduled payment of principal or interest on any outstanding obligations after a grace period.

Fair market value The market price of an asset or liability that trades regularly.

Fair value The amount at which an asset (or liability) could be bought (or incurred) or sold (or settled) in a current transaction between willing parties, that is, other than in a forced or liquidation sale. As defined in IFRS and US GAAP, it is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date.

Feature engineering A process of creating new features by changing or transforming existing features.

Feature selection A process whereby only pertinent features from the dataset are selected for model training. Selecting fewer features decreases model complexity and training time.

Features The independent variables (X's) in a labeled dataset.

Financial contagion A situation in which financial shocks spread from their place of origin to other locales. In essence, a faltering economy infects other, healthier economies.

Financial distress Heightened uncertainty regarding a company's ability to meet its various obligations because of lower or negative earnings.

Financial transaction A purchase involving a buyer having essentially no material synergies with the target (e.g., the purchase of a private company by a company in an unrelated industry or by a private equity firm would typically be a financial transaction).

First-differencing A transformation that subtracts the value of the time series in period $t - 1$ from its value in period t.

First-order serial correlation Correlation between adjacent observations in a time series.

Fitting curve A curve which shows in- and out-of-sample error rates (E_{in} and E_{out}) on the y-axis plotted against model complexity on the x-axis.

Fixed price tender offer Offer made by a company to repurchase a specific number of shares at a fixed price that is typically at a premium to the current market price.

Fixed-rate perpetual preferred stock Non-convertible, non-callable preferred stock with a specified dividend rate that has a claim on earnings senior to the claim of common stock, and no maturity date.

Flight to quality During times of market stress, investors sell higher-risk asset classes such as stocks and commodities in favor of default-risk-free government bonds.

Flip-in pill A poison pill takeover defense that dilutes an acquirer's ownership in a target by giving other existing target company shareholders the right to buy additional target company shares at a discount.

Flip-over pill A poison pill takeover defense that gives target company shareholders the right to purchase shares of the acquirer at a significant discount to the market price, which has the effect of causing dilution to all existing acquiring company shareholders.

Float Amounts collected as premium and not yet paid out as benefits.

Floored floater Floating-rate bond with a floor provision that prevents the coupon rate from decreasing below a specified minimum rate. It protects the investor against declining interest rates.

Flotation cost Fees charged to companies by investment bankers and other costs associated with raising new capital.

Forced conversion For a convertible bond, when the issuer calls the bond and forces bondholders to convert their bonds into shares, which typically happens when the underlying share price increases above the conversion price.

Foreign currency transactions Transactions that are denominated in a currency other than a company's functional currency.

Forward curve The term structure of forward rates for loans made on a specific initiation date.

Forward dividend yield A dividend yield based on the anticipated dividend during the next 12 months.

Forward integration A merger involving the purchase of a target that is farther along the value or production chain; for example, to acquire a distributor.

Forward P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Forward price The fixed price or rate at which the transaction, scheduled to occur at the expiration of a forward contract, will take place. This price is agreed to at the initiation date of the forward contract.

Forward pricing model The model that describes the valuation of forward contracts.

Forward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving forward through the network's layers.

Forward rate An interest rate determined today for a loan that will be initiated in a future period.

Forward rate agreement An over-the-counter forward contract in which the underlying is an interest rate on a deposit. A forward rate agreement (FRA) calls for one party to make a fixed interest payment and the other to make an interest payment at a rate to be determined at contract expiration.

Forward rate model The forward pricing model expressed in terms of spot and forward interest rates.

Forward rate parity The proposition that the forward exchange rate is an unbiased predictor of the future spot exchange rate.

Forward value The monetary value of an existing forward contract.

Franking credit A tax credit received by shareholders for the taxes that a corporation paid on its distributed earnings.

Free cash flow The actual cash that would be available to the company's investors after making all investments necessary to maintain the company as an ongoing enterprise (also referred to as free cash flow to the firm); the internally generated funds that can be distributed to the company's investors (e.g., shareholders and bondholders) without impairing the value of the company.

Free cash flow hypothesis The hypothesis that higher debt levels discipline managers by forcing them to make fixed debt service payments and by reducing the company's free cash flow.

Free cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows.

Free cash flow to equity The cash flow available to a company's common shareholders after all operating expenses, interest, and principal payments have been made and necessary investments in working and fixed capital have been made.

Free cash flow to equity model A model of stock valuation that views a stock's intrinsic value as the present value of expected future free cash flows to equity.

Free cash flow to the firm The cash flow available to the company's suppliers of capital after all operating expenses (including taxes) have been paid and necessary investments in working and fixed capital have been made.

Free cash flow to the firm model A model of stock valuation that views the value of a firm as the present value of expected future free cash flows to the firm.

Frequency analysis The process of quantifying how important tokens are in a sentence and in the corpus as a whole. It helps in filtering unnecessary tokens (or features).

Friendly transaction A potential business combination that is endorsed by the managers of both companies.

Functional currency The currency of the primary economic environment in which an entity operates.

Fundamental factor models A multifactor model in which the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices.

Fundamentals Economic characteristics of a business, such as profitability, financial strength, and risk.

Funds available for distribution (FAD) See *adjusted funds from operations*.

Funds from operations (FFO) Net income (computed in accordance with generally accepted accounting principles) *plus* (1) gains and losses from sales of properties and (2) depreciation and amortization.

Futures price The price at which the parties to a futures contract agree to exchange the underlying (or cash). In commodity markets, the price agreed on to deliver or receive a defined quantity (and often quality) of a commodity at a future date.

Futures value The monetary value of an existing futures contract.

FX carry trade An investment strategy that involves taking long positions in high-yield currencies and short positions in low-yield currencies.

Gamma A measure of how sensitive an option's delta is to a change in the underlying. The change in a given instrument's delta for a given small change in the underlying's value, holding everything else constant.

Generalize When a model retains its explanatory power when predicting out-of-sample (i.e., using new data).

Generalized least squares A regression estimation technique that addresses heteroskedasticity of the error term.

Going-concern assumption The assumption that the business will maintain its business activities into the foreseeable future.

Going-concern value A business's value under a going-concern assumption.

Goodwill An intangible asset that represents the excess of the purchase price of an acquired company over the value of the net identifiable assets acquired.

Grant date The day that stock options are granted to employees.

Green bond Bonds in which the proceeds are designated by issuers to fund a specific project or portfolio of projects that have environmental or climate benefits.

Greenmail The purchase of the accumulated shares of a hostile investor by a company that is targeted for takeover by that investor, usually at a substantial premium over market price.

Greenwashing The risk that a green bond's proceeds are not actually used for a beneficial environmental or climate-related project.

Grid search A method of systematically training a model by using various combinations of hyperparameter values, cross validating each model, and determining which combination of hyperparameter values ensures the best model performance.

Gross domestic product A money measure of the goods and services produced within a country's borders over a stated period.

Gross lease A lease under which the tenant pays a gross rent to the landlord, who is responsible for all operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Ground truth The known outcome (i.e., target variable) of each observation in a labelled dataset.

Growth accounting equation The production function written in the form of growth rates. For the basic Cobb–Douglas production function, it states that the growth rate of output equals the rate of technological change plus α multiplied by the growth rate of capital plus $(1 - \alpha)$ multiplied by the growth rate of labor.

Growth capital expenditures Capital expenditures needed for expansion.

Growth option The ability to make additional investments in a project at some future time if the financial results are strong. Also called *expansion option*.

Guideline assets Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline companies Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline public companies Public-company comparables for the company being valued.

Guideline public company method A variation of the market approach; establishes a value estimate based on the observed multiples from trading activity in the shares of public companies viewed as reasonably comparable to the subject private company.

Guideline transactions method A variation of the market approach; establishes a value estimate based on pricing multiples derived from the acquisition of control of entire public or private companies that were acquired.

Harmonic mean A type of weighted mean computed by averaging the reciprocals of the observations and then taking the reciprocal of that average.

Hazard rate The probability that an event will occur, given that it has not already occurred.

Hedonic index Unlike a repeat-sales index, a hedonic index does not require repeat sales of the same property. It requires only one sale. The way it controls for the fact that different properties are selling each quarter is to include variables in the regression that control for differences in the characteristics of the property, such as size, age, quality of construction, and location.

Heteroskedastic With reference to the error term of regression, having a variance that differs across observations.

Heteroskedasticity The property of having a nonconstant variance; refers to an error term with the property that its variance differs across observations.

Heteroskedasticity-consistent standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Hierarchical clustering An iterative unsupervised learning procedure used for building a hierarchy of clusters.

Highest and best use The concept that the best use of a vacant site is the use that would result in the highest value for the land. Presumably, the developer that could earn the highest risk-adjusted profit based on time, effort, construction and development cost, leasing, and exit value would be the one to pay the highest price for the land.

Historical exchange rates For accounting purposes, the exchange rates that existed when the assets and liabilities were initially recorded.

Historical scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Historical simulation A simulation method that uses past return data and a random number generator that picks observations from the historical series to simulate an asset's future returns.

Historical simulation method The application of historical price changes to the current portfolio.

Historical stress testing The process that tests how investment strategies would perform under some of the most negative (i.e., adverse) combinations of events and scenarios.

Ho–Lee model The first arbitrage-free term structure model. The model is calibrated to market data and uses a binomial lattice approach to generate a distribution of possible future interest rates.

Holding period return The return that an investor earns during a specified holding period; a synonym for total return.

Holdout samples Data samples that are not used to train a model.

Homoskedasticity The property of having a constant variance; refers to an error term that is constant across observations.

Horizontal merger A merger involving companies in the same line of business, usually as competitors.

Horizontal ownership Companies with mutual business interests (e.g., key customers or suppliers) that have cross-holding share arrangements with each other.

Hostile transaction An attempt to acquire a company against the wishes of the target's managers.

Human capital The accumulated knowledge and skill that workers acquire from education, training, or life experience.

Hybrid approach With respect to forecasting, an approach that combines elements of both top-down and bottom-up analyses.

Hyperparameter A parameter whose value must be set by the researcher before learning begins.

I-spreads Shortened form of "interpolated spreads" and a reference to a linearly interpolated yield.

Illiquidity discount A reduction or discount to value that reflects the lack of depth of trading or liquidity in that asset's market.

Impairment Diminishment in value as a result of carrying (book) value exceeding fair value and/or recoverable value.

Impairment of capital rule A legal restriction that dividends cannot exceed retained earnings.

Implementation shortfall The difference between the money return (or value) on a notional or paper portfolio and the actual portfolio return (or value).

Implied volatility The standard deviation that causes an option pricing model to give the current option price.

In-sample forecast errors The residuals from a fitted time-series model within the sample period used to fit the model.

iNAVs "Indicated" net asset values are intraday "fair value" estimates of an ETF share based on its creation basket.

Income approach A valuation approach that values an asset as the present discounted value of the income expected from it. In the context of real estate, this approach estimates the value of a property based on an expected rate of return. The estimated value is the present value of the expected future income from the property, including proceeds from resale at the end of a typical investment holding period.

Incremental VaR (IVaR) A measure of the incremental effect of an asset on the VaR of a portfolio by measuring the difference between the portfolio's VaR while including a specified asset and the portfolio's VaR with that asset eliminated.

Indenture A written contract between a lender and borrower that specifies the terms of the loan, such as interest rate, interest payment schedule, or maturity.

Independent board directors Directors with no material relationship with the company with regard to employment, ownership, or remuneration.

Independent regulators Regulators recognized and granted authority by a government body or agency. They are not government agencies per se and typically do not rely on government funding.

Independent variable A variable used to explain the dependent variable in a regression; a right-side variable in a regression equation. Also referred to as the *explanatory variable*.

Index CDS A type of credit default swap that involves a combination of borrowers.

Indicator variable A variable that takes on only one of two values, 0 or 1, based on a condition. In simple linear regression, the slope is the difference in the dependent variable for the two conditions. Also referred to as a *dummy variable*.

Industry structure An industry's underlying economic and technical characteristics.

Information gain A metric which quantifies the amount of information that the feature holds about the response. Information gain can be regarded as a form of non-linear correlation between Y and X.

Information ratio (IR) Mean active return divided by active risk; or alpha divided by the standard deviation of diversifiable risk.

Informational frictions Forces that restrict availability, quality, and/or flow of information and its use.

Inside ask See *best ask*.

Inside bid See *best bid*.

Inside spread The spread between the best bid price and the best ask price. Also called the *market bid-ask spread*, *inside bid-ask spread*, or *market spread*.

Insiders Corporate managers and board directors who are also shareholders of a company.

Inter-temporal rate of substitution The ratio of the marginal utility of consumption *s* periods in the future (the numerator) to the marginal utility of consumption today (the denominator).

Intercept The expected value of the dependent variable when the independent variable in a simple linear regression is equal to zero.

Interest rate risk The risk that interest rates will rise and therefore the market value of current portfolio holdings will fall so that their current yields to maturity then match comparable instruments in the marketplace.

Interlocking directorates Corporate structure in which individuals serve on the board of directors of multiple corporations.

Internal rate of return Abbreviated as IRR. Rate of return that discounts future cash flows from an investment to the exact amount of the investment; the discount rate that makes the present value of an investment's costs (outflows) equal to the present value of the investment's benefits (inflows).

International Fisher effect The proposition that nominal interest rate differentials across currencies are determined by expected inflation differentials.

Intrinsic value The value of an asset given a hypothetically complete understanding of the asset's investment characteristics; the value obtained if an option is exercised based on current conditions. The difference between the spot exchange rate and the strike price of a currency.

Inverse price ratio The reciprocal of a price multiple—for example, in the case of a P/E, the "earnings yield" E/P (where P is share price and E is earnings per share).

Investment value The value to a specific buyer, taking account of potential synergies based on the investor's requirements and expectations.

ISDA Master Agreement A standard or "master" agreement published by the International Swaps and Derivatives Association. The master agreement establishes the terms for each party involved in the transaction.

Judicial law Interpretations of courts.

Justified (fundamental) P/E The price-to-earnings ratio that is fair, warranted, or justified on the basis of forecasted fundamentals.

Justified price multiple The estimated fair value of the price multiple, usually based on forecasted fundamentals or comparables.

***K*-fold cross-validation** A technique in which data (excluding test sample and fresh data) are shuffled randomly and then are divided into k equal sub-samples, with $k - 1$ samples used as training samples and one sample, the kth, used as a validation sample.

***K*-means** A clustering algorithm that repeatedly partitions observations into a fixed number, k, of non-overlapping clusters.

***K*-nearest neighbor** A supervised learning technique that classifies a new observation by finding similarities ("nearness") between this new observation and the existing data.

Kalotay–Williams–Fabozzi (KWF) model An arbitrage-free term structure model that describes the dynamics of the log of the short rate and assumes constant drift, no mean reversion, and constant volatility.

Key rate durations Sensitivity of a bond's price to changes in specific maturities on the benchmark yield curve. Also called *partial durations*.

***k*th-order autocorrelation** The correlation between observations in a time series separated by k periods.

Labeled dataset A dataset that contains matched sets of observed inputs or features (X's) and the associated output or target (Y).

Labor force Everyone of working age (ages 16 to 64) who either is employed or is available for work but not working.

Labor force participation rate The percentage of the working age population that is in the labor force.

Labor productivity The quantity of real GDP produced by an hour of labor. More generally, output per unit of labor input.

Labor productivity growth accounting equation States that potential GDP growth equals the growth rate of the labor input plus the growth rate of labor productivity.

Lack of marketability discount An extra return to investors to compensate for lack of a public market or lack of marketability.

LASSO Least absolute shrinkage and selection operator is a type of penalized regression which involves minimizing the sum of the absolute values of the regression coefficients. LASSO can also be used for regularization in neural networks.

Latency The elapsed time between the occurrence of an event and a subsequent action that depends on that event.

Law of one price A principle that states that if two investments have the same or equivalent future cash flows regardless of what will happen in the future, then these two investments should have the same current price.

Leading dividend yield Forecasted dividends per share over the next year divided by current stock price.

Leading P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Learning curve A curve that plots the accuracy rate (= 1 – error rate) in the validation or test samples (i.e., out-of-sample) against the amount of data in the training sample, which is thus useful for describing under- and overfitting as a function of bias and variance errors.

Learning rate A parameter that affects the magnitude of adjustments in the weights in a neural network.

Level One of the three factors (the other two are steepness and curvature) that empirically explain most yield curve shape changes. A shock to the level factor changes the yield for all maturities by an almost identical amount.

Leveraged buyout A transaction whereby the target company management team converts the target to a privately held company by using heavy borrowing to finance the purchase of the target company's outstanding shares.

Leveraged recapitalization A post-offer takeover defense mechanism that involves the assumption of a large amount of debt that is then used to finance share repurchases. The effect is to dramatically change the company's capital structure while attempting to deliver a value to target shareholders in excess of a hostile bid.

Libor–OIS spread The difference between Libor and the overnight indexed swap rate.

Limit order book The book or list of limit orders to buy and sell that pertains to a security.

Lin-log model A regression model in which the independent variable is in logarithmic form.

Linear classifier A binary classifier that makes its classification decision based on a linear combination of the features of each data point.

Linear regression Regression that models the straight-line relationship between the dependent and independent variables. Also known as *least squares regression* and *ordinary least squares regression*.

Linear trend A trend in which the dependent variable changes at a constant rate with time.

Liquidating dividend A dividend that is a return of capital rather than a distribution from earnings or retained earnings.

Liquidation To sell the assets of a company, division, or subsidiary piecemeal, typically because of bankruptcy; the form of bankruptcy that allows for the orderly satisfaction of creditors' claims after which the company ceases to exist.

Liquidation value The value of a company if the company were dissolved and its assets sold individually.

Liquidity preference theory A term structure theory that asserts liquidity premiums exist to compensate investors for the added interest rate risk they face when lending long term.

Liquidity premium The premium or incrementally higher yield that investors demand for lending long term.

Local currency The currency of the country where a company is located.

Local expectations theory A term structure theory that contends the return for all bonds over short periods is the risk-free rate.

Log-lin model A regression model in which the dependent variable is in logarithmic form.

Log-linear model With reference to time-series models, a model in which the growth rate of the time series as a function of time is constant.

Log-log model A regression model in which both the dependent and independent variables are in logarithmic form. Also known as the *double-log model*.

Log-log regression model A regression that expresses the dependent and independent variables as natural logarithms.

Logistic regression (logit model) A qualitative-dependent-variable multiple regression model based on the logistic probability distribution.

Long/short credit trade A credit protection seller with respect to one entity combined with a credit protection buyer with respect to another entity.

Look-ahead bias The bias created by using information that was unknown or unavailable in the time periods over which backtesting is conducted, such as company earnings and macroeconomic indicator values.

Lookback period The time period used to gather a historical data set.

Loss given default The amount that will be lost if a default occurs.

Macroeconomic factor model A multifactor model in which the factors are surprises in macroeconomic variables that significantly explain equity returns.

Macroeconomic factors Factors related to the economy, such as the inflation rate, industrial production, or economic sector membership.

Maintenance capital expenditures Capital expenditures needed to maintain operations at the current level.

Majority shareholders Shareholders that own more than 50% of a corporation's shares.

Majority-vote classifier A classifier that assigns to a new data point the predicted label with the most votes (i.e., occurrences).

Managerialism theories Theories that posit that corporate executives are motivated to engage in mergers to maximize the size of their company rather than shareholder value (a form of agency cost).

Marginal VaR (MVaR) A measure of the effect of a small change in a position size on portfolio VaR.

Market approach Valuation approach that values an asset based on pricing multiples from sales of assets viewed as similar to the subject asset.

Market conversion premium per share For a convertible bond, the difference between the market conversion price and the underlying share price, which allows investors to identify the premium or discount payable when buying a convertible bond rather than the underlying common stock.

Market conversion premium ratio For a convertible bond, the market conversion premium per share expressed as a percentage of the current market price of the shares.

Market efficiency A finance perspective on capital markets that deals with the relationship of price to intrinsic value. The **traditional efficient markets formulation** asserts that an asset's price is the best available estimate of its intrinsic value. The **rational efficient markets formulation** asserts that investors should expect to be rewarded for the costs of information gathering and analysis by higher gross returns.

Market fragmentation Trading the same instrument in multiple venues.

Market impact The effect of the trade on transaction prices. Also called *price impact.*

Market timing Asset allocation in which the investment in the market is increased if one forecasts that the market will outperform T-bills.

Market value of invested capital The market value of debt and equity.

Mature growth rate The earnings growth rate in a company's mature phase; an earnings growth rate that can be sustained long term.

Maximum drawdown The worst cumulative loss ever sustained by an asset or portfolio. More specifically, maximum drawdown is the difference between an asset's or a portfolio's maximum cumulative return and its subsequent lowest cumulative return.

Mean reversion The tendency of a time series to fall when its level is above its mean and rise when its level is below its mean; a mean-reverting time series tends to return to its long-term mean.

Mean square error (MSE) The sum of squares error divided by the degrees of freedom, $n - k - 1$; in a simple linear regression, $n - k - 1 = n - 2$.

Mean square regression (MSR) The sum of squares regression divided by the number of independent variables k; in a simple linear regression, $k = 1$.

Merger The absorption of one company by another; two companies become one entity and one or both of the pre-merger companies ceases to exist as a separate entity.

Metadata Data that describes and gives information about other data.

Method based on forecasted fundamentals An approach to using price multiples that relates a price multiple to forecasts of fundamentals through a discounted cash flow model.

Method of comparables An approach to valuation that involves using a price multiple to evaluate whether an asset is relatively fairly valued, relatively undervalued, or relatively overvalued when compared to a benchmark value of the multiple.

Midquote price The average, or midpoint, of the prevailing bid and ask prices.

Minority interest The proportion of the ownership of a subsidiary not held by the parent (controlling) company.

Minority shareholders Shareholders that own less than 50% of a corporation's shares.

Mispricing Any departure of the market price of an asset from the asset's estimated intrinsic value.

Mixed offering A merger or acquisition that is to be paid for with cash, securities, or some combination of the two.

Model specification With reference to regression, the set of variables included in the regression and the regression equation's functional form.

Molodovsky effect The observation that P/Es tend to be high on depressed EPS at the bottom of a business cycle and tend to be low on unusually high EPS at the top of a business cycle.

Momentum indicators Valuation indicators that relate either price or a fundamental (such as earnings) to the time series of their own past values (or in some cases to their expected value).

Monetary assets and liabilities Assets and liabilities with value equal to the amount of currency contracted for, a fixed amount of currency. Examples are cash, accounts receivable, accounts payable, bonds payable, and mortgages payable. Inventory is not a monetary asset. Most liabilities are monetary.

Monetary/non-monetary method Approach to translating foreign currency financial statements for consolidation in which monetary assets and liabilities are translated at the current exchange rate. Non-monetary assets and liabilities are translated at historical exchange rates (the exchange rates that existed when the assets and liabilities were acquired).

Monetizing Unwinding a position to either capture a gain or realize a loss.

Monitoring costs Costs borne by owners to monitor the management of the company (e.g., board of director expenses).

Monte Carlo simulation A technique that uses the inverse transformation method for converting a randomly generated uniformly distributed number into a simulated value of a random variable of a desired distribution. Each key decision variable in a Monte Carlo simulation requires an assumed statistical distribution; this assumption facilitates incorporating non-normality, fat tails, and tail dependence as well as solving high-dimensionality problems.

Mortgages Loans with real estate serving as collateral for the loans.

Multicollinearity A regression assumption violation that occurs when two or more independent variables (or combinations of independent variables) are highly but not perfectly correlated with each other.

Multiple linear regression Linear regression involving two or more independent variables.

Multiple linear regression model A linear regression model with two or more independent variables.

Mutual information Measures how much information is contributed by a token to a class of texts. MI will be 0 if the token's distribution in all text classes is the same. MI approaches 1 as the token in any one class tends to occur more often in only that particular class of text.

Mutually exclusive projects Mutually exclusive projects compete directly with each other. For example, if Projects A and B are mutually exclusive, you can choose A or B, but you cannot choose both.

N-grams A representation of word sequences. The length of a sequence varies from 1 to n. When one word is used, it is a unigram; a two-word sequence is a bigram; and a 3-word sequence is a trigram; and so on.

***n*-Period moving average** The average of the current and immediately prior $n - 1$ values of a time series.

Naked credit default swap A position where the owner of the CDS does not have a position in the underlying credit.

Name entity recognition An algorithm that analyzes individual tokens and their surrounding semantics while referring to its dictionary to tag an object class to the token.

Negative serial correlation Serial correlation in which a positive error for one observation increases the chance of a negative error for another observation, and vice versa.

Net asset balance sheet exposure When assets translated at the current exchange rate are greater in amount than liabilities translated at the current exchange rate. Assets exposed to translation gains or losses exceed the exposed liabilities.

Net asset value The difference between assets and liabilities, all taken at current market values instead of accounting book values.

Net asset value per share Net asset value divided by the number of shares outstanding.

Net lease A lease under which the tenant pays a net rent to the landlord and an additional amount based on the tenant's pro rata share of the operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Net liability balance sheet exposure When liabilities translated at the current exchange rate are greater assets translated at the current exchange rate. Liabilities exposed to translation gains or losses exceed the exposed assets.

Net operating income Gross rental revenue minus operating costs but before deducting depreciation, corporate overhead, and interest expense. In the context of real estate, a measure of the income from the property after deducting operating expenses for such items as property taxes, insurance, maintenance, utilities, repairs, and insurance but before deducting any costs associated with financing and before deducting federal income taxes. It is similar to EBITDA in a financial reporting context.

Net regulatory burden The private costs of regulation less the private benefits of regulation.

Network externalities The impact that users of a good, a service, or a technology have on other users of that product; it can be positive (e.g., a critical mass of users makes a product more useful) or negative (e.g., congestion makes the product less useful).

Neural networks Highly flexible machine learning algorithms that have been successfully applied to a variety of supervised and unsupervised tasks characterized by nonlinearities and interactions among features.

No-arbitrage approach A procedure for obtaining the value of an option based on the creation of a portfolio that replicates the payoffs of the option and deriving the option value from the value of the replicating portfolio.

No-growth company A company without positive expected net present value projects.

No-growth value per share The value per share of a no-growth company, equal to the expected level amount of earnings divided by the stock's required rate of return.

Non-cash rent An amount equal to the difference between the average contractual rent over a lease term (the straight-line rent) and the cash rent actually paid during a period. This figure is one of the deductions made from FFO to calculate AFFO.

Non-convergence trap A situation in which a country remains relatively poor, or even falls further behind, because it fails to implement necessary institutional reforms and/or adopt leading technologies.

Non-monetary assets and liabilities Assets and liabilities that are not monetary assets and liabilities. Non-monetary assets include inventory, fixed assets, and intangibles, and non-monetary liabilities include deferred revenue.

Non-renewable resources Finite resources that are depleted once they are consumed; oil and coal are examples.

Non-residential properties Commercial real estate properties other than multi-family properties, farmland, and timberland.

Nonearning assets Cash and investments (specifically cash, cash equivalents, and short-term investments).

Nonstationarity With reference to a random variable, the property of having characteristics, such as mean and variance, that are not constant through time.

Normal EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normalized EPS*.

Normalized earnings The expected level of mid-cycle earnings for a company in the absence of any unusual or temporary factors that affect profitability (either positively or negatively).

Normalized EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normal EPS*.

Normalized P/E P/E based on normalized EPS data.

Notional amount The amount of protection being purchased in a CDS.

NTM P/E Next 12-month P/E: current market price divided by an estimated next 12-month EPS.

Off-the-run A series of securities or indexes that were issued/created prior to the most recently issued/created series.

On-the-run The most recently issued/created series of securities or indexes.

One hot encoding The process by which categorical variables are converted into binary form (0 or 1) for machine reading. It is one of the most common methods for handling categorical features in text data.

One-sided durations Effective durations when interest rates go up or down, which are better at capturing the interest rate sensitivity of bonds with embedded options that do not react symmetrically to positive and negative changes in interest rates of the same magnitude.

One-tier board Board structure consisting of a single board of directors, composed of executive (internal) and non-executive (external) directors.

Opportunity cost The value that investors forgo by choosing a particular course of action; the value of something in its best alternative use.

Optimal capital structure The capital structure at which the value of the company is maximized.

Option-adjusted spread (OAS) Constant spread that, when added to all the one-period forward rates on the interest rate tree, makes the arbitrage-free value of the bond equal to its market price.

Orderly liquidation value The estimated gross amount of money that could be realized from the liquidation sale of an asset or assets, given a reasonable amount of time to find a purchaser or purchasers.

Organic growth Company growth in output or sales that is achieved by making investments internally (i.e., excludes growth achieved through mergers and acquisitions).

Other comprehensive income Changes to equity that bypass (are not reported in) the income statement; the difference between comprehensive income and net income.

Other post-employment benefits Promises by the company to pay benefits in the future, such as life insurance premiums and all or part of health care insurance for its retirees.

Out-of-sample forecast errors The differences between actual and predicted values of time series outside the sample period used to fit the model.

Overfitting When a model fits the training data too well and so does not generalize well to new data.

Overnight indexed swap (OIS) rate An interest rate swap in which the periodic floating rate of the swap equals the geometric average of a daily unsecured overnight rate (or overnight index rate).

Pairs trading An approach to trading that uses pairs of closely related stocks, buying the relatively undervalued stock and selling short the relatively overvalued stock.

Par curve A hypothetical yield curve for coupon-paying Treasury securities that assumes all securities are priced at par.

Par swap A swap in which the fixed rate is set so that no money is exchanged at contract initiation.

Parametric method A method of estimating VaR that uses the historical mean, standard deviation, and correlation of security price movements to estimate the portfolio VaR. Generally assumes a normal distribution but can be adapted to non-normal distributions with the addition of skewness and kurtosis. Sometimes called the *variance–covariance method* or the *analytical method*.

Partial regression coefficients The slope coefficients in a multiple regression. Also called *partial slope coefficients*.

Partial slope coefficients The slope coefficients in a multiple regression. Also called *partial regression coefficients*.

Parts of speech An algorithm that uses language structure and dictionaries to tag every token in the text with a corresponding part of speech (i.e., noun, verb, adjective, proper noun, etc.).

Payout amount The loss given default times the notional.

Payout policy The principles by which a company distributes cash to common shareholders by means of cash dividends and/or share repurchases.

Payouts Cash dividends and the value of shares repurchased in any given year.

Pecking order theory The theory that managers consider how their actions might be interpreted by outsiders and thus order their preferences for various forms of corporate financing. Forms of financing that are least visible to outsiders (e.g., internally generated funds) are most preferable to managers and those that are most visible (e.g., equity) are least preferable.

PEG ratio The P/E-to-growth ratio, calculated as the stock's P/E divided by the expected earnings growth rate.

Penalized regression A regression that includes a constraint such that the regression coefficients are chosen to minimize the sum of squared residuals *plus* a penalty term that increases in size with the number of included features.

Pension obligation The present value of future benefits earned by employees for service provided to date.

Perfect capital markets Markets in which, by assumption, there are no taxes, transaction costs, or bankruptcy costs and in which all investors have equal ("symmetric") information.

Perpetuity A perpetual annuity, or a set of never-ending level sequential cash flows, with the first cash flow occurring one period from now.

Persistent earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *continuing earnings*, or *underlying earnings*.

Pet projects Projects in which influential managers want the corporation to invest. Often, unfortunately, pet projects are selected without undergoing normal capital budgeting analysis.

Physical settlement Involves actual delivery of the debt instrument in exchange for a payment by the credit protection seller of the notional amount of the contract.

Point-in-time data Data consisting of the exact information available to market participants as of a given point in time. Point-in-time data is used to address look-ahead bias.

Poison pill A pre-offer takeover defense mechanism that makes it prohibitively costly for an acquirer to take control of a target without the prior approval of the target's board of directors.

Poison puts A pre-offer takeover defense mechanism that gives target company bondholders the right to sell their bonds back to the target at a pre-specified redemption price, typically at or above par value; this defense increases the need for cash and raises the cost of the acquisition.

Portfolio balance approach A theory of exchange rate determination that emphasizes the portfolio investment decisions of global investors and the requirement that global investors willingly hold all outstanding securities denominated in each currency at prevailing prices and exchange rates.

Positive serial correlation Serial correlation in which a positive error for one observation increases the chance of a positive error for another observation; a negative error for one observation increases the chance of a negative error for another observation.

Potential GDP The maximum amount of output an economy can sustainably produce without inducing an increase in the inflation rate. The output level that corresponds to full employment with consistent wage and price expectations.

Precision In error analysis for classification problems it is ratio of correctly predicted positive classes to all predicted positive classes. Precision is useful in situations where the cost of false positives (FP), or Type I error, is high.

Preferred habitat theory A term structure theory that contends that investors have maturity preferences and require yield incentives before they will buy bonds outside of their preferred maturities.

Premise of value The status of a company in the sense of whether it is assumed to be a going concern or not.

Premium leg The series of payments the credit protection buyer promises to make to the credit protection seller.

Premiums Amounts paid by the purchaser of insurance products.

Present value model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Present value of growth opportunities The difference between the actual value per share and the no-growth value per share. Also called *value of growth*.

Presentation currency The currency in which financial statement amounts are presented.

Price improvement When trade execution prices are better than quoted prices.

Price momentum A valuation indicator based on past price movement.

Price multiples The ratio of a stock's market price to some measure of value per share.

Price-setting option The operational flexibility to adjust prices when demand varies from what is forecast. For example, when demand exceeds capacity, the company could benefit from the excess demand by increasing prices.

Price-to-earnings ratio (P/E) The ratio of share price to earnings per share.

Priced risk Risk for which investors demand compensation for bearing (e.g., equity risk, company-specific factors, macroeconomic factors).

Principal components analysis (PCA) An unsupervised ML technique used to transform highly correlated features of data into a few main, uncorrelated composite variables.

Principle of no arbitrage In well-functioning markets, prices will adjust until there are no arbitrage opportunities.

Prior transaction method A variation of the market approach; considers actual transactions in the stock of the subject private company.

Private market value The value derived using a sum-of-the-parts valuation.

Probability of default The probability that a bond issuer will not meet its contractual obligations on schedule.

Probability of survival The probability that a bond issuer will meet its contractual obligations on schedule.

Procedural law The body of law that focuses on the protection and enforcement of the substantive laws.

Production-flexibility option The operational flexibility to alter production when demand varies from forecast. For example, if demand is strong, a company may profit from employees working overtime or from adding additional shifts.

Project sequencing To defer the decision to invest in a future project until the outcome of some or all of a current project is known. Projects are sequenced through time, so that investing in a project creates the option to invest in future projects.

Projection error The vertical (perpendicular) distance between a data point and a given principal component.

Prospective P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Protection leg The contingent payment that the credit protection seller may have to make to the credit protection buyer.

Protection period Period during which a bond's issuer cannot call the bond.

Provision for loan losses An income statement expense account that increases the amount of the allowance for loan losses.

Proxy fight An attempt to take control of a company through a shareholder vote.

Proxy statement A public document that provides the material facts concerning matters on which shareholders will vote.

Prudential supervision Regulation and monitoring of the safety and soundness of financial institutions to promote financial stability, reduce system-wide risks, and protect customers of financial institutions.

Pruning A regularization technique used in CART to reduce the size of the classification or regression tree—by pruning, or removing, sections of the tree that provide little classifying power.

Purchasing power gain A gain in value caused by changes in price levels. Monetary liabilities experience purchasing power gains during periods of inflation.

Purchasing power loss A loss in value caused by changes in price levels. Monetary assets experience purchasing power loss during periods of inflation.

Purchasing power parity (PPP) The idea that exchange rates move to equalize the purchasing power of different currencies.

Pure expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *unbiased expectations theory*.

Pure factor portfolio A portfolio with sensitivity of 1 to the factor in question and a sensitivity of 0 to all other factors.

Putable bond Bond that includes an embedded put option, which gives the bondholder the right to put back the bonds to the issuer prior to maturity, typically when interest rates have risen and higher-yielding bonds are available.

Qualitative dependent variables Dummy variables used as dependent variables rather than as independent variables.

Quality of earnings analysis The investigation of issues relating to the accuracy of reported accounting results as reflections of economic performance. Quality of earnings analysis is broadly understood to include not only earnings management but also balance sheet management.

Random forest classifier A collection of a large number of decision trees trained via a bagging method.

Random walk A time series in which the value of the series in one period is the value of the series in the previous period plus an unpredictable random error.

Rational efficient markets formulation See *market efficiency.*

Readme files Text files provided with raw data that contain information related to a data file. They are useful for understanding the data and how they can be interpreted correctly.

Real estate investment trusts Tax-advantaged entities (companies or trusts) that own, operate, and—to a limited extent—develop income-producing real estate property.

Real estate operating companies Regular taxable real estate ownership companies that operate in the real estate industry in countries that do not have a tax-advantaged REIT regime in place or that are engage in real estate activities of a kind and to an extent that do not fit in their country's REIT framework.

Real interest rate parity The proposition that real interest rates will converge to the same level across different markets.

Real options Options that relate to investment decisions such as the option to time the start of a project, the option to adjust its scale, or the option to abandon a project that has begun.

Rebalance return A return from rebalancing the component weights of an index.

Recall Also known as *sensitivity*, in error analysis for classification problems it is the ratio of correctly predicted positive classes to all actual positive classes. Recall is useful in situations where the cost of false negatives (FN), or Type II error, is high.

Reconstitution When dealers recombine appropriate individual zero-coupon securities and reproduce an underlying coupon Treasury.

Recovery rate The percentage of the loss recovered.

Redemption basket The list of securities (and share amounts) the authorized participant (AP) receives when it redeems ETF shares back to the ETF manager. The redemption basket is published each business day.

Reference entity The borrower (debt issuer) covered by a single-name CDS.

Reference obligation A particular debt instrument issued by the borrower that is the designated instrument being covered.

Regime With reference to a time series, the underlying model generating the times series.

Regression analysis A tool for examining whether a variable is useful for explaining another variable.

Regression coefficients The intercept and slope coefficient(s) of a regression.

Regular expression (regex) A series of texts that contains characters in a particular order. Regex is used to search for patterns of interest in a given text.

Regularization A term that describes methods for reducing statistical variability in high-dimensional data estimation problems.

Regulatory arbitrage Entities identify and use some aspect of regulations that allows them to exploit differences in economic substance and regulatory interpretation or in foreign and domestic regulatory regimes to their (the entities') advantage.

Regulatory burden The costs of regulation for the regulated entity.

Regulatory capture Theory that regulation often arises to enhance the interests of the regulated.

Regulatory competition Regulators may compete to provide a regulatory environment designed to attract certain entities.

Reinforcement learning Machine learning in which a computer learns from interacting with itself or data generated by the same algorithm.

Relative-strength indicators Valuation indicators that compare a stock's performance during a period either to its own past performance or to the performance of some group of stocks.

Relative valuation models A model that specifies an asset's value relative to the value of another asset.

Relative VaR See *ex ante tracking error*.

Relative version of PPP The hypothesis that changes in (nominal) exchange rates over time are equal to national inflation rate differentials.

Renewable resources Resources that can be replenished, such as a forest.

Rental price of capital The cost per unit of time to rent a unit of capital.

Repeat sales index As the name implies, this type of index relies on repeat sales of the same property. In general, the idea supporting this type of index is that because it is the same property that sold twice, the change in value between the two sale dates indicates how market conditions have changed over time.

Replacement cost In the context of real estate, the value of a building assuming it was built today using current construction costs and standards.

Reporting unit For financial reporting under US GAAP, an operating segment or one level below an operating segment (referred to as a component).

Required rate of return The minimum rate of return required by an investor to invest in an asset, given the asset's riskiness.

Residential properties Properties that provide housing for individuals or families. Single-family properties may be owner-occupied or rental properties, whereas multi-family properties are rental properties even if the owner or manager occupies one of the units.

Residual The difference between an observation and its predicted value, where the predicted value is based on the estimated linear relation between the dependent and independent variables using sample data.

Residual autocorrelations The sample autocorrelations of the residuals.

Residual income Earnings for a given period, minus a deduction for common shareholders' opportunity cost in generating the earnings. Also called *economic profit* or *abnormal earnings*.

Residual income method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Residual income model (RIM) A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share. Also called *discounted abnormal earnings model* or *Edwards–Bell–Ohlson model*.

Residual loss Agency costs that are incurred despite adequate monitoring and bonding of management.

Restructuring Reorganizing the capital structure of a firm.

Return on capital employed Operating profit divided by capital employed (debt and equity capital).

Return on invested capital A measure of the after-tax profitability of the capital invested by the company's shareholders and debtholders.

Reverse carry arbitrage A strategy involving the short sale of the underlying and an offsetting opposite position in the derivative.

Reverse stock split A reduction in the number of shares outstanding with a corresponding increase in share price but no change to the company's underlying fundamentals.

Reverse stress testing A risk management approach in which the user identifies key risk exposures in the portfolio and subjects those exposures to extreme market movements.

Reviewed financial statements A type of non-audited financial statements; typically provide an opinion letter with representations and assurances by the reviewing accountant that are less than those in audited financial statements.

Rho The change in a given derivative instrument for a given small change in the risk-free interest rate, holding everything else constant. Rho measures the sensitivity of the option to the risk-free interest rate.

Risk budgeting The allocation of an asset owner's total risk appetite among groups or divisions (in the case of a trading organization) or among strategies and managers (in the case of an institutional or individual investor).

Risk decomposition The process of converting a set of holdings in a portfolio into a set of exposures to risk factors.

Risk factors Variables or characteristics with which individual asset returns are correlated. Sometimes referred to simply as *factors*.

Risk parity A portfolio allocation scheme that weights stocks or factors based on an equal risk contribution.

Robust standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Roll When an investor moves its investment position from an older series to the most current series.

Roll return The component of the return on a commodity futures contract attributable to rolling long futures positions forward through time. Also called *roll yield*.

Rolling down the yield curve A maturity trading strategy that involves buying bonds with a maturity longer than the intended investment horizon. Also called *riding the yield curve*.

Rolling windows A backtesting method that uses a rolling-window (or walk-forward) framework, rebalances the portfolio after each period, and then tracks performance over time. As new information arrives each period, the investment manager optimizes (revises and tunes) the model and readjusts stock positions.

Root mean squared error (RMSE) The square root of the average squared forecast error; used to compare the out-of-sample forecasting performance of forecasting models.

Sale-leaseback A situation in which a company sells the building it owns and occupies to a real estate investor and the company then signs a long-term lease with the buyer to continue to occupy the building. At the end of the lease, use of the property reverts to the landlord.

Sales comparison approach In the context of real estate, this approach estimates value based on what similar or comparable properties (comparables) transacted for in the current market.

Scaled earnings surprise Unexpected earnings divided by the standard deviation of analysts' earnings forecasts.

Scaling The process of adjusting the range of a feature by shifting and changing the scale of the data. Two of the most common ways of scaling are normalization and standardization.

Scatter plot A chart in which two variables are plotted along the axis and points on the chart represent pairs of the two variables. In regression, the dependent variable is plotted on the vertical axis and the independent variable is plotted along the horizontal axis. Also known as a scattergram and a *scatter diagram*.

Scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Scree plots A plot that shows the proportion of total variance in the data explained by each principal component.

Screening The application of a set of criteria to reduce a set of potential investments to a smaller set having certain desired characteristics.

Seasonality A characteristic of a time series in which the data experience regular and predictable periodic changes; for example, fan sales are highest during the summer months.

Secured overnight financing rate (SOFR) A daily volume-weighted index of rates on qualified cash borrowings collateralized by US Treasuries that is expected to replace Libor as a floating reference rate for swaps.

Securities offering A merger or acquisition in which target shareholders are to receive shares of the acquirer's common stock as compensation.

Security selection risk See *active specific risk*.

Segmented markets theory A term structure theory that contends yields are solely a function of the supply and demand for funds of a particular maturity.

Self-regulating organizations (SROs) Self-regulating bodies that are given recognition and authority, including enforcement power, by a government body or agency.

Self-regulatory bodies Private, non-governmental organizations that both represent and regulate their members. Some self-regulating organizations are also independent regulators.

Sell-side analysts Analysts who work at brokerages.

Sensitivity analysis A technique for exploring how a target variable (e.g., portfolio returns) and risk profiles are affected by changes in input variables (e.g., the distribution of asset or factor returns).

Sentence length The number of characters, including spaces, in a sentence.

Serially correlated With reference to regression errors, errors that are correlated across observations.

Service period For employee stock options, usually the period between the grant date and the vesting date.

Settled in arrears An arrangement in which the interest payment is made (i.e., settlement occurs) at the maturity of the underlying instrument.

Settlement In the case of a credit event, the process by which the two parties to a CDS contract satisfy their respective obligations.

Shaping risk The sensitivity of a bond's price to the changing shape of the yield curve.

Share repurchase A transaction in which a company buys back its own shares. Unlike stock dividends and stock splits, share repurchases use corporate cash.

Shareholder activism Strategies used by shareholders to attempt to compel a company to act in a desired manner.

Shareholders' equity Total assets minus total liabilities.

Shark repellents A pre-offer takeover defense mechanism involving the corporate charter (e.g., staggered boards of directors and supermajority provisions).

Simple linear regression (SLR) A regression that summarizes the relation between the dependent variable and a single independent variable.

Simulation A technique for exploring how a target variable (e.g. portfolio returns) would perform in a hypothetical environment specified by the user, rather than a historical setting.

Single-name CDS Credit default swap on one specific borrower.

Sinking fund bond A bond that requires the issuer to set aside funds over time to retire the bond issue, thus reducing credit risk.

Slope coefficient The coefficient of an independent variable that represents the average change in the dependent variable for a one-unit change in the independent variable.

Soft margin classification An adaptation in the support vector machine algorithm which adds a penalty to the objective function for observations in the training set that are misclassified.

Special dividend A dividend paid by a company that does not pay dividends on a regular schedule or a dividend that supplements regular cash dividends with an extra payment.

Spin-off A form of restructuring in which shareholders of a parent company receive a proportional number of shares in a new, separate entity; shareholders end up owning stock in two different companies where there used to be one.

Split-off A form of restructuring in which shareholders of the parent company are given shares in a newly created entity in exchange for their shares of the parent company.

Split-rate tax system In reference to corporate taxes, a split-rate system taxes earnings to be distributed as dividends at a different rate than earnings to be retained. Corporate profits distributed as dividends are taxed at a lower rate than those retained in the business.

Spot curve The term structure of spot rates for loans made today.

Spot price The current price of an asset or security. For commodities, the current price to deliver a physical commodity to a specific location or purchase and transport it away from a designated location.

Spot rate The interest rate that is determined today for a risk-free, single-unit payment at a specified future date.

Spot yield curve The term structure of spot rates for loans made today.

Stabilized NOI In the context of real estate, the expected NOI when a renovation is complete.

Stable dividend policy A policy in which regular dividends are paid that reflect long-run expected earnings. In contrast to a constant dividend payout ratio policy, a stable dividend policy does not reflect short-term volatility in earnings.

Standard error of the estimate A measure of the fit of a regression line, calculated as the square root of the mean square error. Also known as the *standard error of the regression* and the *root mean square error*.

Standard error of the forecast A measure of the uncertainty associated with a forecasted value of the dependent variable that depends on the standard error of the estimate, the variability of the independent variable, the deviation of the forecasted independent variable from the mean in the regression, and the number of observations.

Standard error of the slope coefficient The standard error of the slope, which in a simple linear regression is the ratio of the model's standard error of the estimate (s_e) to the square root of the variation of the independent variable.

Standardized beta With reference to fundamental factor models, the value of the attribute for an asset minus the average value of the attribute across all stocks, divided by the standard deviation of the attribute across all stocks.

Standardized unexpected earnings Unexpected earnings per share divided by the standard deviation of unexpected earnings per share over a specified prior time period.

Static trade-off theory of capital structure A theory pertaining to a company's optimal capital structure. The optimal level of debt is found at the point where additional debt would cause the costs of financial distress to increase by a greater amount than the benefit of the additional tax shield.

Statistical factor model A multifactor model in which statistical methods are applied to a set of historical returns to determine portfolios that best explain either historical return covariances or variances.

Statutes Laws enacted by legislative bodies.

Statutory merger A merger in which one company ceases to exist as an identifiable entity and all its assets and liabilities become part of a purchasing company.

Steady-state rate of growth The constant growth rate of output (or output per capita) that can or will be sustained indefinitely once it is reached. Key ratios, such as the capital–output ratio, are constant on the steady-state growth path.

Steepness The difference between long-term and short-term yields that constitutes one of the three factors (the other two are level and curvature) that empirically explain most of the changes in the shape of the yield curve.

Stock dividend A type of dividend in which a company distributes additional shares of its common stock to shareholders instead of cash.

Stock purchase An acquisition in which the acquirer gives the target company's shareholders some combination of cash and securities in exchange for shares of the target company's stock.

Stop-loss limit Constraint used in risk management that requires a reduction in the size of a portfolio, or its complete liquidation, when a loss of a particular size occurs in a specified period.

Straight bond An underlying option-free bond with a specified issuer, issue date, maturity date, principal amount and repayment structure, coupon rate and payment structure, and currency denomination.

Straight-line rent The average annual rent under a multi-year lease agreement that contains contractual increases in rent during the life of the lease.

Straight-line rent adjustment See *non-cash rent.*

Straight voting Voting structure in which shareholders are granted the right of one vote for each share owned.

Stranded assets Assets that are obsolete or not economically viable.

Strategic transaction A purchase involving a buyer that would benefit from certain synergies associated with owning the target firm.

Stress tests A risk management technique that assesses the portfolio's response to extreme market movements.

Stripping A dealer's ability to separate a bond's individual cash flows and trade them as zero-coupon securities.

Subsidiary merger A merger in which the company being purchased becomes a subsidiary of the purchaser.

Substantive law The body of law that focuses on the rights and responsibilities of entities and relationships among entities.

Succession event A change of corporate structure of the reference entity, such as through a merger, a divestiture, a spinoff, or any similar action, in which ultimate responsibility for the debt in question is unclear.

Sum of squares error (SSE) The sum of the squared deviations of (1) the value of the dependent variable and (2) the value of the dependent variable based on the estimated regression line. Also referred to as the *residual sum of squares.*

Sum of squares regression (SSR) The sum of the squared deviations of (1) the value of the dependent variable based on the estimated regression line and (2) the mean of the dependent variable.

Sum of squares total (SST) The sum of the squared deviations of the dependent variable from its mean; the variation of the dependent variable. Also referred to as the *total sum of squares.*

Sum-of-the-parts valuation A valuation that sums the estimated values of each of a company's businesses as if each business were an independent going concern.

Summation operator A functional part of a neural network's node that multiplies each input value received by a weight and sums the weighted values to form the total net input, which is then passed to the activation function.

Supernormal growth Above-average or abnormally high growth rate in earnings per share.

Supervised learning Machine learning where algorithms infer patterns between a set of inputs (the *X*'s) and the desired output (*Y*). The inferred pattern is then used to map a given input set into a predicted output.

Support vector machine A linear classifier that determines the hyperplane that optimally separates the observations into two sets of data points.

Survivorship bias The bias that results when data as of a given date reflects only those entities that have survived to that date. Entities can include any element of an index or list that is constituted through time: stocks, investment funds, etc. Survivorship bias is a form of look-ahead bias.

Sustainable growth rate The rate of dividend (and earnings) growth that can be sustained over time for a given level of return on equity, keeping the capital structure constant and without issuing additional common stock.

Swap curve The term structure of swap rates.

Swap rate The "price" that swap traders quote among one another. It is the rate at which the present value of all the expected floating-rate payments received over the life of the floating-rate bond equal the present value of all the expected fixed-rate payments made over the life of the fixed-rate bond.

Swap rate curve The term structure of swap rates.

Swap spread The difference between the fixed rate on an interest rate swap and the rate on a Treasury note with equivalent maturity; it reflects the general level of credit risk in the market.

Systematic risk Risk that affects the entire market or economy; it cannot be avoided and is inherent in the overall market. Systematic risk is also known as non-diversifiable or market risk.

Systemic risk The risk of failure of the financial system.

Tail risk The risk that losses in extreme events could be greater than would be expected for a portfolio of assets with a normal distribution.

Takeover A merger; the term may be applied to any transaction but is often used in reference to hostile transactions.

Takeover premium The amount by which the takeover price for each share of stock must exceed the current stock price in order to entice shareholders to relinquish control of the company to an acquirer.

Tangible book value per share Common shareholders' equity minus intangible assets reported on the balance sheet, divided by the number of shares outstanding.

Target In machine learning, the dependent variable (*Y*) in a labeled dataset; the company in a merger or acquisition that is being acquired.

Target capital structure A company's chosen proportions of debt and equity.

Target company See *target.*

Target payout ratio A strategic corporate goal representing the long-term proportion of earnings that the company intends to distribute to shareholders as dividends.

Taxable REIT subsidiaries Subsidiaries that pay income taxes on earnings from non-REIT-qualifying activities like merchant development or third-party property management.

Technical indicators Momentum indicators based on price.

TED spread A measure of perceived credit risk determined as the difference between Libor and the T-bill yield of matching maturity.

Temporal method A variation of the monetary/non-monetary translation method that requires not only monetary assets and liabilities, but also non-monetary assets and liabilities that are measured at their current value on the balance sheet date to be translated at the current exchange rate. Assets and liabilities are translated at rates consistent with the timing of their measurement value. This method is typically used when the functional currency is other than the local currency.

Tender offer A public offer whereby the acquirer invites target shareholders to submit ("tender") their shares in return for the proposed payment.

Term frequency (TF) Ratio of the number of times a given token occurs in all the texts in the dataset to the total number of tokens in the dataset.

Term premium The additional return required by lenders to invest in a bond to maturity net of the expected return from continually reinvesting at the short-term rate over that same time horizon.

Terminal price multiples The price multiple for a stock assumed to hold at a stated future time.

Terminal share price The share price at a particular point in the future.

Terminal value of the stock The analyst's estimate of a stock's value at a particular point in the future. Also called *continuing value of the stock*.

Test sample A data sample that is used to test a model's ability to predict well on new data.

Theta The change in a derivative instrument for a given small change in calendar time, holding everything else constant. Specifically, the theta calculation assumes nothing changes except calendar time. Theta also reflects the rate at which an option's time value decays.

Time series A set of observations on a variable's outcomes in different time periods.

Tobin's *q* The ratio of the market value of debt and equity to the replacement cost of total assets.

Token The equivalent of a word (or sometimes a character).

Tokenization The process of splitting a given text into separate tokens. Tokenization can be performed at the word or character level but is most commonly performed at word level.

Top-down approach With respect to forecasting, an approach that usually begins at the level of the overall economy. Forecasts are then made at more narrowly defined levels, such as sector, industry, and market for a specific product.

Total factor productivity (TFP) A multiplicative scale factor that reflects the general level of productivity or technology in the economy. Changes in total factor productivity generate proportional changes in output for any input combination.

Total invested capital The sum of market value of common equity, book value of preferred equity, and face value of debt.

Tracking error The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking risk*.

Tracking risk The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking error*.

Trailing dividend yield The reciprocal of current market price divided by the most recent annualized dividend.

Trailing P/E A stock's current market price divided by the most recent four quarters of EPS (or the most recent two semi-annual periods for companies that report interim data semi-annually). Also called *current P/E*.

Training sample A data sample that is used to train a model.

Tranche CDS A type of credit default swap that covers a combination of borrowers but only up to pre-specified levels of losses.

Transaction exposure The risk of a change in value between the transaction date and the settlement date of an asset of liability denominated in a foreign currency.

Treasury shares/stock Shares that were issued and subsequently repurchased by the company.

Trend A long-term pattern of movement in a particular direction.

Triangular arbitrage An arbitrage transaction involving three currencies that attempts to exploit inconsistencies among pairwise exchange rates.

Trimming Also called truncation, it is the process of removing extreme values and outliers from a dataset.

Triple-net leases Common leases in the United States and Canada that require each tenant to pay its share of the following three operating expenses: common area maintenance and repair expenses; property taxes; and building insurance costs. Also known as *NNN leases*.

Two-tier board Board structure consisting of a supervisory board that oversees a management board.

Unbiased expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *pure expectations theory*.

Unconditional heteroskedasticity Heteroskedasticity of the error term that is not correlated with the values of the independent variable(s) in the regression.

Uncovered interest rate parity The proposition that the expected return on an uncovered (i.e., unhedged) foreign currency (risk-free) investment should equal the return on a comparable domestic currency investment.

Underlying earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *core earnings*, or *persistent earnings*.

Unexpected earnings The difference between reported EPS and expected EPS. Also referred to as an *earnings surprise*.

Unit root A time series that is not covariance stationary is said to have a unit root.

Unsupervised learning Machine learning that does not make use of labeled data.

Upfront payment The difference between the credit spread and the standard rate paid by the protection buyer if the standard rate is insufficient to compensate the protection seller. Also called *upfront premium*.

Upfront premium See *upfront payment*.

Upstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary company) such that the associate company records a profit on its income statement. An example is a sale of inventory by the associate to the investor company or by a subsidiary to a parent company.

Validation sample A data sample that is used to validate and tune a model.

Valuation The process of determining the value of an asset or service either on the basis of variables perceived to be related to future investment returns or on the basis of comparisons with closely similar assets.

Value additivity An arbitrage opportunity when the value of the whole equals the sum of the values of the parts.

Value at risk (VaR) The minimum loss that would be expected a certain percentage of the time over a certain period of time given the assumed market conditions.

Value of growth The difference between the actual value per share and the no-growth value per share.

Variance error Describes how much a model's results change in response to new data from validation and test samples. Unstable models pick up noise and produce high variance error, causing overfitting and high out-of-sample error.

Vasicek model A partial equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is constant.

Vega The change in a given derivative instrument for a given small change in volatility, holding everything else constant. A sensitivity measure for options that reflects the effect of volatility.

Venture capital investors Private equity investors in development-stage companies.

Vertical merger A merger involving companies at different positions of the same production chain; for example, a supplier or a distributor.

Vertical ownership Ownership structure in which a company or group that has a controlling interest in two or more holding companies, which in turn have controlling interests in various operating companies.

Vested benefit obligation The actuarial present value of vested benefits.

Vesting date The date that employees can first exercise stock options.

Visibility The extent to which a company's operations are predictable with substantial confidence.

Voting caps Legal restrictions on the voting rights of large share positions.

Web spidering (scraping or crawling) programs Programs that extract raw content from a source, typically web pages.

Weighted average cost of capital (WACC) A weighted average of the after-tax required rates of return on a company's common stock, preferred stock, and long-term debt, where the weights are the fraction of each source of financing in the company's target capital structure.

Weighted harmonic mean See *harmonic mean*.

White-corrected standard errors A synonym for robust standard errors.

White knight A third party that is sought out by the target company's board to purchase the target in lieu of a hostile bidder.

White squire A third party that is sought out by the target company's board to purchase a substantial minority stake in the target—enough to block a hostile takeover without selling the entire company.

Winner's curse The tendency for the winner in certain competitive bidding situations to overpay, whether because of overestimation of intrinsic value, emotion, or information asymmetries.

Winsorization The process of replacing extreme values and outliers in a dataset with the maximum (for large value outliers) and minimum (for small value outliers) values of data points that are not outliers.

Write-down A reduction in the value of an asset as stated in the balance sheet.

Yield curve factor model A model or a description of yield curve movements that can be considered realistic when compared with historical data.

Zero A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

Zero-coupon bond A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

CFA Institute®
CFA Program

PORTFOLIO MANAGEMENT AND ETHICAL AND PROFESSIONAL STANDARDS

CFA® Program Curriculum
2022 • LEVEL II • VOLUME 6

WILEY

CONTENTS

O indicates an optional segment

How to Use the CFA Program Curriculum

Congratulations on your decision to enter the Chartered Financial Analyst (CFA®) Program. This exciting and rewarding program of study reflects your desire to become a serious investment professional. You are embarking on a program noted for its high ethical standards and the breadth of knowledge, skills, and abilities (competencies) it develops. Your commitment should be educationally and professionally rewarding.

The credential you seek is respected around the world as a mark of accomplishment and dedication. Each level of the program represents a distinct achievement in professional development. Successful completion of the program is rewarded with membership in a prestigious global community of investment professionals. CFA charterholders are dedicated to life-long learning and maintaining currency with the ever-changing dynamics of a challenging profession. CFA Program enrollment represents the first step toward a career-long commitment to professional education.

The CFA exam measures your mastery of the core knowledge, skills, and abilities required to succeed as an investment professional. These core competencies are the basis for the Candidate Body of Knowledge (CBOK™). The CBOK consists of four components:

- A broad outline that lists the major CFA Program topic areas (www.cfainstitute.org/programs/cfa/curriculum/cbok);
- Topic area weights that indicate the relative exam weightings of the top-level topic areas (www.cfainstitute.org/programs/cfa/curriculum);
- Learning outcome statements (LOS) that advise candidates about the specific knowledge, skills, and abilities they should acquire from readings covering a topic area (LOS are provided in candidate study sessions and at the beginning of each reading); and
- CFA Program curriculum that candidates receive upon exam registration.

Therefore, the key to your success on the CFA exams is studying and understanding the CBOK. The following sections provide background on the CBOK, the organization of the curriculum, features of the curriculum, and tips for designing an effective personal study program.

BACKGROUND ON THE CBOK

CFA Program is grounded in the practice of the investment profession. CFA Institute performs a continuous practice analysis with investment professionals around the world to determine the competencies that are relevant to the profession, beginning with the Global Body of Investment Knowledge (GBIK®). Regional expert panels and targeted surveys are conducted annually to verify and reinforce the continuous feedback about the GBIK. The practice analysis process ultimately defines the CBOK. The CBOK reflects the competencies that are generally accepted and applied by investment professionals. These competencies are used in practice in a generalist context and are expected to be demonstrated by a recently qualified CFA charterholder.

© 2021 CFA Institute. All rights reserved.

The CFA Institute staff—in conjunction with the Education Advisory Committee and Curriculum Level Advisors, who consist of practicing CFA charterholders—designs the CFA Program curriculum in order to deliver the CBOK to candidates. The exams, also written by CFA charterholders, are designed to allow you to demonstrate your mastery of the CBOK as set forth in the CFA Program curriculum. As you structure your personal study program, you should emphasize mastery of the CBOK and the practical application of that knowledge. For more information on the practice analysis, CBOK, and development of the CFA Program curriculum, please visit www.cfainstitute.org.

ORGANIZATION OF THE CURRICULUM

The Level II CFA Program curriculum is organized into 10 topic areas. Each topic area begins with a brief statement of the material and the depth of knowledge expected. It is then divided into one or more study sessions. These study sessions should form the basic structure of your reading and preparation. Each study session includes a statement of its structure and objective and is further divided into assigned readings. An outline illustrating the organization of these study sessions can be found at the front of each volume of the curriculum.

The readings are commissioned by CFA Institute and written by content experts, including investment professionals and university professors. Each reading includes LOS and the core material to be studied, often a combination of text, exhibits, and in-text examples and questions. End of Reading Questions (EORQs) followed by solutions help you understand and master the material. The LOS indicate what you should be able to accomplish after studying the material. The LOS, the core material, and the EORQs are dependent on each other, with the core material and EORQs providing context for understanding the scope of the LOS and enabling you to apply a principle or concept in a variety of scenarios.

The entire readings, including the EORQs, are the basis for all exam questions and are selected or developed specifically to teach the knowledge, skills, and abilities reflected in the CBOK.

You should use the LOS to guide and focus your study because each exam question is based on one or more LOS and the core material and practice problems associated with the LOS. As a candidate, you are responsible for the entirety of the required material in a study session.

We encourage you to review the information about the LOS on our website (www.cfainstitute.org/programs/cfa/curriculum/study-sessions), including the descriptions of LOS "command words" on the candidate resources page at www.cfainstitute.org.

FEATURES OF THE CURRICULUM

End of Reading Questions/Solutions *All End of Reading Questions (EORQs) as well as their solutions are part of the curriculum and are required material for the exam.* In addition to the in-text examples and questions, these EORQs help demonstrate practical applications and reinforce your understanding of the concepts presented. Some of these EORQs are adapted from past CFA exams and/or may serve as a basis for exam questions.

Glossary For your convenience, each volume includes a comprehensive Glossary. Throughout the curriculum, a **bolded** word in a reading denotes a term defined in the Glossary.

Note that the digital curriculum that is included in your exam registration fee is searchable for key words, including Glossary terms.

LOS Self-Check We have inserted checkboxes next to each LOS that you can use to track your progress in mastering the concepts in each reading.

Source Material The CFA Institute curriculum cites textbooks, journal articles, and other publications that provide additional context or information about topics covered in the readings. As a candidate, you are not responsible for familiarity with the original source materials cited in the curriculum.

Note that some readings may contain a web address or URL. The referenced sites were live at the time the reading was written or updated but may have been deactivated since then.

Some readings in the curriculum cite articles published in the *Financial Analysts Journal®*, which is the flagship publication of CFA Institute. Since its launch in 1945, the *Financial Analysts Journal* has established itself as the leading practitioner-oriented journal in the investment management community. Over the years, it has advanced the knowledge and understanding of the practice of investment management through the publication of peer-reviewed practitioner-relevant research from leading academics and practitioners. It has also featured thought-provoking opinion pieces that advance the common level of discourse within the investment management profession. Some of the most influential research in the area of investment management has appeared in the pages of the *Financial Analysts Journal*, and several Nobel laureates have contributed articles.

Candidates are not responsible for familiarity with *Financial Analysts Journal* articles that are cited in the curriculum. But, as your time and studies allow, we strongly encourage you to begin supplementing your understanding of key investment management issues by reading this, and other, CFA Institute practice-oriented publications through the Research & Analysis webpage (www.cfainstitute.org/en/research).

Errata The curriculum development process is rigorous and includes multiple rounds of reviews by content experts. Despite our efforts to produce a curriculum that is free of errors, there are times when we must make corrections. Curriculum errata are periodically updated and posted by exam level and test date online (www.cfainstitute.org/en/programs/submit-errata). If you believe you have found an error in the curriculum, you can submit your concerns through our curriculum errata reporting process found at the bottom of the Curriculum Errata webpage.

DESIGNING YOUR PERSONAL STUDY PROGRAM

Create a Schedule An orderly, systematic approach to exam preparation is critical. You should dedicate a consistent block of time every week to reading and studying. Complete all assigned readings and the associated problems and solutions in each study session. Review the LOS both before and after you study each reading to ensure that

you have mastered the applicable content and can demonstrate the knowledge, skills, and abilities described by the LOS and the assigned reading. Use the LOS self-check to track your progress and highlight areas of weakness for later review.

Successful candidates report an average of more than 300 hours preparing for each exam. Your preparation time will vary based on your prior education and experience, and you will probably spend more time on some study sessions than on others.

You should allow ample time for both in-depth study of all topic areas and additional concentration on those topic areas for which you feel the least prepared.

CFA INSTITUTE LEARNING ECOSYSTEM (LES)

As you prepare for your exam, we will email you important exam updates, testing policies, and study tips. Be sure to read these carefully.

Your exam registration fee includes access to the CFA Program Learning Ecosystem (LES). This digital learning platform provides access, even offline, to all of the readings and End of Reading Questions found in the print curriculum organized as a series of shorter online lessons with associated EORQs. This tool is your one-stop location for all study materials, including practice questions and mock exams.

The LES provides the following supplemental study tools:

Structured and Adaptive Study Plans The LES offers two ways to plan your study through the curriculum. The first is a structured plan that allows you to move through the material in the way that you feel best suits your learning. The second is an adaptive study plan based on the results of an assessment test that uses actual practice questions.

Regardless of your chosen study path, the LES tracks your level of proficiency in each topic area and presents you with a dashboard of where you stand in terms of proficiency so that you can allocate your study time efficiently.

Flashcards and Game Center The LES offers all the Glossary terms as Flashcards and tracks correct and incorrect answers. Flashcards can be filtered both by curriculum topic area and by action taken—for example, answered correctly, unanswered, and so on. These Flashcards provide a flexible way to study Glossary item definitions.

The Game Center provides several engaging ways to interact with the Flashcards in a game context. Each game tests your knowledge of the Glossary terms a in different way. Your results are scored and presented, along with a summary of candidates with high scores on the game, on your Dashboard.

Discussion Board The Discussion Board within the LES provides a way for you to interact with other candidates as you pursue your study plan. Discussions can happen at the level of individual lessons to raise questions about material in those lessons that you or other candidates can clarify or comment on. Discussions can also be posted at the level of topics or in the initial Welcome section to connect with other candidates in your area.

Practice Question Bank The LES offers access to a question bank of hundreds of practice questions that are in addition to the End of Reading Questions. These practice questions, only available on the LES, are intended to help you assess your mastery of individual topic areas as you progress through your studies. After each practice question, you will receive immediate feedback noting the correct response and indicating the relevant assigned reading so you can identify areas of weakness for further study.

Mock Exams The LES also includes access to three-hour Mock Exams that simulate the morning and afternoon sessions of the actual CFA exam. These Mock Exams are intended to be taken after you complete your study of the full curriculum and take practice questions so you can test your understanding of the curriculum and your readiness for the exam. If you take these Mock Exams within the LES, you will receive feedback afterward that notes the correct responses and indicates the relevant assigned readings so you can assess areas of weakness for further study. We recommend that you take Mock Exams during the final stages of your preparation for the actual CFA exam. For more information on the Mock Exams, please visit www.cfainstitute.org.

PREP PROVIDERS

You may choose to seek study support outside CFA Institute in the form of exam prep providers. After your CFA Program enrollment, you may receive numerous solicitations for exam prep courses and review materials. When considering a prep course, make sure the provider is committed to following the CFA Institute guidelines and high standards in its offerings.

Remember, however, that there are no shortcuts to success on the CFA exams; reading and studying the CFA Program curriculum *is* the key to success on the exam. The CFA Program exams reference only the CFA Institute assigned curriculum; no prep course or review course materials are consulted or referenced.

SUMMARY

Every question on the CFA exam is based on the content contained in the required readings and on one or more LOS. Frequently, an exam question is based on a specific example highlighted within a reading or on a specific practice problem and its solution. To make effective use of the CFA Program curriculum, please remember these key points:

1. All pages of the curriculum are required reading for the exam.
2. All questions, problems, and their solutions are part of the curriculum and are required study material for the exam. These questions are found at the end of the readings in the print versions of the curriculum. In the LES, these questions appear directly after the lesson with which they are associated. The LES provides immediate feedback on your answers and tracks your performance on these questions throughout your study.
3. We strongly encourage you to use the CFA Program Learning Ecosystem. In addition to providing access to all the curriculum material, including EORQs, in the form of shorter, focused lessons, the LES offers structured and adaptive study planning, a Discussion Board to communicate with other candidates, Flashcards, a Game Center for study activities, a test bank of practice questions, and online Mock Exams. Other supplemental study tools, such as eBook and PDF versions of the print curriculum, and additional candidate resources are available at www.cfainstitute.org.
4. Using the study planner, create a schedule and commit sufficient study time to cover the study sessions. You should also plan to review the materials, answer practice questions, and take Mock Exams.
5. Some of the concepts in the study sessions may be superseded by updated rulings and/or pronouncements issued after a reading was published. Candidates are expected to be familiar with the overall analytical framework contained in the assigned readings. Candidates are not responsible for changes that occur after the material was written.

FEEDBACK

At CFA Institute, we are committed to delivering a comprehensive and rigorous curriculum for the development of competent, ethically grounded investment professionals. We rely on candidate and investment professional comments and feedback as we work to improve the curriculum, supplemental study tools, and candidate resources.

Please send any comments or feedback to info@cfainstitute.org. You can be assured that we will review your suggestions carefully. Ongoing improvements in the curriculum will help you prepare for success on the upcoming exams and for a lifetime of learning as a serious investment professional.

Portfolio Management

STUDY SESSIONS

Study Session 15	Portfolio Management (1)
Study Session 16	Portfolio Management (2)

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to explain and demonstrate the use of portfolio theory in risk and return estimation, security selection, and other practical applications. The candidate should also be able to explain the portfolio management process.

Portfolio management and risk management are key investment activities. Incorporating investor objectives, constraints, capital market expectations, and relevant risk considerations, together with portfolio construction, execution, and evaluation represent core activities in the investment process.

© 2021 CFA Institute. All rights reserved.

PORTFOLIO MANAGEMENT

STUDY SESSION

16

Portfolio Management (2)

This study session begins by identifying and explaining the ties between the real economy and financial markets, including effects on asset values. The "fundamental pricing equation" is presented as a basic pricing framework for financial instruments. The asset prices of risk-free debt, risky debt, public equities, and real estate are shown to be affected via the business cycle's impact on risk-free rates, the yield curve, inflation, and risk premiums. Analysis of active portfolio management follows, including a discussion of active risk and active return (Sharpe, information ratios). The fundamental law of active management is presented along with several investment applications. The session concludes with an overview of how securities trading supports the investment process. This reading discusses direct and indirect costs of trading, developments in electronic trading and the effects on transaction costs and market fragmentation, and the risks posed by electronic trading and how regulators control them.

READING ASSIGNMENTS

Reading 42	Economics and Investment Markets by Andrew Clare, PhD, and Thomas F. Cosimano, PhD
Reading 43	Analysis of Active Portfolio Management by Roger G. Clarke, PhD, Harindra de Silva, PhD, CFA, and Steven Thorley, PhD, CFA
Reading 44	Trading Costs and Electronic Markets by Larry Harris, PhD, CFA

© 2021 CFA Institute. All rights reserved.

READING

42

Economics and Investment Markets

by Andrew Clare, PhD, and Thomas F. Cosimano, PhD

Andrew Clare, PhD, is at Cass Business School (United Kingdom). Thomas F. Cosimano, PhD, is Professor Emeritus in Finance at the Mendoza College of Business, University of Notre Dame, and Visiting Scholar, International Monetary Fund (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** explain the notion that to affect market values, economic factors must affect one or more of the following: 1) default-free interest rates across maturities, 2) the timing and/or magnitude of expected cash flows, and 3) risk premiums;
☐	**b.** explain the role of expectations and changes in expectations in market valuation;
☐	**c.** explain the relationship between the long-term growth rate of the economy, the volatility of the growth rate, and the average level of real short-term interest rates;
☐	**d.** explain how the phase of the business cycle affects policy and short-term interest rates, the slope of the term structure of interest rates, and the relative performance of bonds of differing maturities;
☐	**e.** describe the factors that affect yield spreads between non-inflation-adjusted and inflation-indexed bonds;
☐	**f.** explain how the phase of the business cycle affects credit spreads and the performance of credit-sensitive fixed-income instruments;
☐	**g.** explain how the characteristics of the markets for a company's products affect the company's credit quality;
☐	**h.** explain how the phase of the business cycle affects short-term and long-term earnings growth expectations;
☐	**i.** explain the relationship between the consumption hedging properties of equity and the equity risk premium;

(continued)

© 2020 CFA Institute. All rights reserved.

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**j.** describe cyclical effects on valuation multiples;
☐	**k.** describe how economic analysis is used in sector rotation strategies;
☐	**l.** describe the economic factors affecting investment in commercial real estate.

1 INTRODUCTION

The state of the economy and financial market activity are interconnected. Financial markets are the forums where savers are connected with investors. This activity enables savers to defer consumption today for consumption in the future, allows governments to raise the capital necessary to create a secure society, and permits corporations to access capital to exploit profitable investment opportunities, which, in turn, should help to generate future economic growth and employment. Furthermore, all financial instruments essentially represent claims on an underlying economy. There is, therefore, an important and fundamental connection that runs from the decisions of economic agents, as they plan their present and future consumption, to the prices of financial instruments, such as bonds and equities.

The purpose of this reading is to identify and explain the links between the real economy and financial markets and to show how economic analysis can be used to develop ways of valuing both individual financial market securities and aggregations of these securities, such as financial market indexes. We begin by reviewing what we refer to as the fundamental pricing equation for all financial instruments. Using this framework, we then move on to explore the relationship between the economy and real default-free debt. From there, we can extend the analysis to the ways in which the economy can influence the prices of the following: nominal default-free debt; credit risky debt (for example, corporate bonds); publicly traded equities; and commercial real estate.

2 FRAMEWORK FOR THE ECONOMIC ANALYSIS OF FINANCIAL MARKETS: THE PRESENT VALUE MODEL

a explain the notion that to affect market values, economic factors must affect one or more of the following: 1) default-free interest rates across maturities, 2) the timing and/or magnitude of expected cash flows, and 3) risk premiums;

The reference point for the analysis of this reading is the present value model of asset valuation. The impact of economic factors on asset values can be studied in the context of that model by examining how economic factors can affect discount rates and future cash flows. These topics are explored in more detail in the following sections.

2.1 The Present Value Model

The value of an asset must be related to the benefits that we expect to receive from holding it; for many assets (e.g., financial securities), these benefits are its future cash flows, which may be specified in the security's contract, as is the case with bonds, or be discretionary, as is the case with ordinary shares. Intuitively, a given amount of money received in the future will be valued less by individual investors than the same amount of money received today. Because an investor can use cash for present consumption, he or she needs an incentive to defer it to the future and more so as the future becomes less certain. These considerations provide an economic rationale for valuing an asset by discounting its future cash flows to derive its present value.

Equation 1 presents the fundamental present value formula for the value at time t of any financial asset i, V_t^i, which we assume equals its current market price, P_t^i. In general, we will speak of the time t as "today."

$$P_t^i = \sum_{s=1}^{N} \frac{E_t\left[\widetilde{CF}_{t+s}^i\right]}{\left(1 + l_{t,s} + \theta_{t,s} + \rho_{t,s}^i\right)^s}, \qquad (1)$$

where

P_t^i = the value of asset i at time t (today)

N = number of cash flows in the life of the asset

$\widetilde{CF}_{t+s}^i$ = the uncertain, nominal cash flow paid s periods in the future

$E_t\left[\widetilde{CF}\right]$ = the expectation of the random variable $\widetilde{CF}$ conditional on the information available to investors today (t)

$l_{t,s}$ = yield to maturity on a real default-free investment today (t), which pays one unit of currency s periods in the future

$\theta_{t,s}$ = expected inflation rate between t and $t + s$

$\rho_{t,s}^i$ = the risk premium required today (t) to pay the investor for taking on risk in the cash flow of asset i, s periods in the future

This expression is general enough to be used to value all financial instruments. The present values of all of the instrument's cash flows are summed from 1 to N. Some assets, such as a five-year zero-coupon bond, may have only one cash flow, and so N would equal five in that case, with cash flows in Periods 1–4 equal to zero. At the other extreme, dividend-paying equities produce cash flows in the form of dividends into the indefinite future, in which case N could, technically, be equal to infinity.

According to Equation 1, effects of the economy on asset prices are transmitted through some combination of influences on the numerator—the asset's expected cash flows—and denominator—the discount rate(s) applied to the asset's expected cash flows.

A factor that typically distinguishes one financial asset class from another is the degree of certainty that investors have about future cash flows. At one extreme there may be little uncertainty. For example, despite losing its AAA rating from Standard & Poor's, investors might still attach a relatively low probability to the prospect of the US Treasury not making the scheduled payments on its debts on time and in full. Investors may regard the probability of the German government defaulting on its debts to be very low, too. At the other extreme, investors may be very uncertain about the size and timing of dividend payments from an equity investment and will also have to consider the prospect of receiving no dividends at all in the event that the company declares bankruptcy.

The uncertainty about future cash flows is reflected in the discount rate in Equation 1. We can think of the discount rate as having three distinct components. The first component is $l_{t,s}$ which effectively represents the return that an investor requires on a real default-free fixed-income security at present time t for a cash flow to be paid s periods in the future. For example, readers can think of this return as being analogous to the return expected on an investment in an inflation-linked bond issued by the government of a developed economy.

The second component in the discount rate, $\theta_{t,s}$, represents the additional return required by investors, above that required from investing in a real default-free investment, for investing in a nominal default-free investment. This additional return is required even though an investor may attach a zero probability to not being paid on time and in full, because future nominal payments will be affected by inflation. In essence, this component of the discount rate represents the compensation that investors demand for the inflation that they expect to experience over the investment horizon. Compensation is demanded because investors are concerned about the real purchasing power of their investments in the future rather than in the nominal value of the future cash flows.

The third component of the discount rate in Equation 1, $\rho^i_{t,s}$, represents the additional return that investors expect for investing in financial assets because of uncertainty about the asset's future cash flows. In other words, all securities, even those issued by governments of developed economies and considered risk free in that there is negligible risk that the issuer will default (as we will see later), carry some risks for which risk-averse investors will want to be compensated. Indeed, the size of this risk premium will vary among asset classes, and this variation is largely responsible for the distinction between one asset class and another. We remind the reader of this difference among assets by placing a superscript i to indicate that this premium is specific to the asset under consideration.

As we will show throughout this reading, the size and nature of this addition to expected return, $\rho^i_{t,s}$, will depend on the characteristics of the asset or asset class in question, which, in turn, will be determined by developments and expected developments in the real economy. This means that the discount rates applied to the cash flows of financial assets will almost certainly vary over time as perceptions of expected economic growth, inflation, and cash flow risk change. In particular during recessions, the risk premium that investors demand on financial assets, especially those that are not default-free, may rise because investors in general may be less willing and able to take on heightened default risk during such periods.

The $\rho^i_{t,s}$ component may include more than just the compensation for the uncertainty related to financial cash flows that may be subject to default risk. In particular, $\rho^i_{t,s}$ may also reflect other types of risk—for example, liquidity risk. Liquidity risk refers to the possibility that a financial asset cannot be converted quickly into cash at close to its fair value; it is particularly characteristic of investments in commercial real estate and high-yield corporate bonds/loans. And as many investors learned during the 2008–2009 global financial crisis, some debt instruments, such as mortgage-backed securities, can become very illiquid at just the moment when investors become most risk averse and when they want to be holding cash rather than riskier financial assets.

In summary, the expected cash flows for any financial asset, i, can be discounted using the following general expression for the discount rate (this additive expression is an approximation of the exact multiplicative expression that includes interaction effects between the terms):

$$1 + l_{t,s} + \theta_{t,s} + \rho^i_{t,s} \text{ for } s = 1, \ldots, N. \quad (2)$$

A major purpose of this reading is to identify the relationship between these elements of the discount rate and the underlying economy and also to decompose $\rho_{t,s}^{i}$ into its component parts for each asset class.

EXPECTATIONS AND ASSET VALUES

3

b explain the role of expectations and changes in expectations in market valuation;

Examining Equation 1 reveals simple but important observations:

- Asset values depend not on past cash flows but on the expectation of future cash flows.
- These expectations are based on (conditional on) current information (indicated by the time subscript t) that may be relevant to forecasting future cash flows. Any information that may contribute to the accuracy or precision of expectations is relevant.

Because asset values are dependent on expectations of future cash flows, information that changes expectations affects asset values and realized returns. Information that has been anticipated is already reflected in asset prices, but information that is different from what was expected constitutes real news that requires expectations to adjust. The adjustment generates a holding period return that differs from the expected return. This observation about investors' anticipations is important to understanding sometimes seemingly counterintuitive market reactions to economic information releases. Investors judge economic data releases relative to their expectations for the data. Prices may fall (rise) despite "good" ("bad") news if the expectation was for better (worse) news.

Thus, for valuation, one important distinction is information that is "news" or new information and information that has been fully anticipated. Therefore, news is a surprise relative to fully anticipated information.

Although this reading's focus is the effect of economic factors on asset values, investor sentiment (e.g., enthusiasm or despair) can also affect asset values. Economic factors affect asset values through generally direct effects on cash flows and/or discount rates, but investor sentiment affects asset values through direct effects on discount rates via higher or lower risk premiums (and possibly indirect effects on future cash flows).

THE DISCOUNT RATE ON REAL DEFAULT-FREE BONDS: REAL DEFAULT-FREE INTEREST RATES

4

a explain the notion that to affect market values, economic factors must affect one or more of the following: 1) default-free interest rates across maturities, 2) the timing and/or magnitude of expected cash flows, and 3) risk premiums;

Having introduced several fundamental concepts, we can now begin the analysis of economic factors affecting asset values in detail. The first step in understanding the relationship between the economy and investment markets involves noting that the purchase of an investment involves the opportunity cost of lower consumption today. In other words, by buying a financial asset, an investor defers some current consumption. The sum of all of these individual saving and consumption decisions is thus going to have an impact on the price of financial assets.

To explain how investors' concerns for satisfying consumption needs through economic fluctuations affect asset prices, modern finance makes reference to several expressions (Equations 3 through 6) that model how investors evaluate consumption trade-offs. These expressions, and associated (non-testable) calculations, show the analysis behind statements about the relationships between real interest rates, GDP growth, and the volatility of GDP; verbal statements of the intuition behind the relationships are also given that aid in understanding the entire reading.

4.1 Real Default-Free Interest Rates

To demonstrate the importance of the aggregation of these individual saving and investment decisions, we can think through how the aggregated consumption and investment decisions of individuals might determine the real default-free interest rate in an economy—that is, $l_{t,s}$ in Equation 1. Consider a single individual who has to choose between using some portion of his or her wealth to consume today (t) or investing that wealth in default-free bonds that will pay investors one dollar when they mature s periods in the future.

Think of this bond as being issued by a highly rated, developed-economy government, so there is only a negligible prospect of default. Also think of this bond as being inflation index-linked so that investors do not need to concern themselves with the impact of future inflation on the bond's future cash flow. An alternative way of conceptualizing the bond is to assume (only for the moment) that it is issued by this government in a world that has no inflation. Later we will re-introduce inflation into the pricing problem.

What sort of return would investors require on a bond that is both default-free and unaffected by future inflation? It is tempting to say that an investor would require no return on such a bond because there is no risk of losing money over the investment period in either nominal or real terms. But the choice to invest today involves the opportunity cost of not consuming today. It is the aggregated opportunity cost of all investors that will determine the price of this asset today and its return over the investment horizon.

Think of the return on the asset as the opportunity cost (price) of consuming today. If the return increases, the investor substitutes away from current consumption to future consumption by purchasing an asset. Consequently, as with any other economic decision, an investor must consider the relative prices of the two alternatives. In this case, the investor can

- pay price $P_{t,s}$ today, t, of a default-free bond paying 1 monetary unit of income s periods in the future or
- buy goods worth $P_{t,s}$ dollars today.

The decision to purchase this bond will be determined by the willingness of individuals to substitute consumption today for consumption in the future. This trade-off is measured by the marginal utility of consumption s periods in the future relative to the marginal utility of consumption today (t). The marginal utility of consumption is the additional satisfaction or utility that a consumer derives from one additional unit of consumption. The ratio of these two marginal utilities—the ratio of the marginal utility of consumption *s* periods in the future (the numerator) to the marginal utility of consumption today (the denominator)—is known as the **inter-temporal rate of substitution**, denoted $\tilde{m}_{t,s}$.

In "good" economic times, individuals may have relatively high levels of current income so that current consumption is high. In this case, the utility derived from an additional unit of consumption today will be relatively low. Conversely, in "bad" economic times, current income and consumption will tend to be relatively low, which means that the utility derived from an additional unit of consumption today will be relatively high. In addition, the marginal utility of consumption of investors diminishes as their wealth increases because they have already satisfied fundamental needs. Thus, investors would receive a larger benefit (utility) from an asset that pays off more in bad economic times relative to one that pays off in good economic times.

The rate of substitution is a random variable because an investor will not know how much she has available in the future from other sources of income, such as salary from working. This uncertainty is present even for an investment that pays a certain amount in one period because the value of this investment is determined by how much utility the investor receives from this investment.

Given this uncertainty, the investor must make the decision today based on her expectations of future circumstances when she receives the payoff from the investment. This expectation is conditional on the information that the investor has when the decision is made. Thus, if the investor wanted to consider an investment in a zero-coupon bond at time *t* that is certain to pay off one unit of real consumption in *s* periods, then

$$P_{t,s} = E_t\left(1\tilde{m}_{t,s}\right) = E_t\left(\tilde{m}_{t,s}\right) \quad (3)$$

where $\tilde{m}_{t,s}$ is the investor's marginal willingness to trade consumption at time *t* for (real) wealth at time $t + s$.[1]

EXAMPLE 1

The Inter-Temporal Rate of Substitution (1 of 3)

Suppose the investor's willingness to trade present for future consumption can be represented as

$$\tilde{m}_{t,1} = e^{a+b\tilde{z}}$$

Here, $\tilde{z}$ is a random shock to the economy that affects the cash flows of the marginal investor; $\tilde{z}$ is what makes $\tilde{m}_{t,1}$ a random variable. The exponential form is consistent with assumptions about investor risk aversion and consumption growth often made in finance; *a* and *b* are typically negative given those same assumptions.[2] Parameters of the distribution of $\tilde{z}$ consistent with observed

1 The term $\tilde{m}_{t,s}$ is technically defined as (δ)[MU(C_{t+s})/MU(C_t)], where MU denotes marginal utility of consumption, C, and δ is a discount factor that captures the preference for consumption at t rather than later at t + s. The discount factor applied to MU (C_{t+s}) adjusts it for the time difference of s periods. The tilde on $\tilde{m}_{t,s}$ indicates it is a random (stochastic) variable, and the term may be called the stochastic inter-temporal rate of substitution. As in Example 1, this inter-temporal rate of substitution varies based on the realization of the shock to economic activity, $\tilde{z}$. In the modern theory of asset pricing, the term m is also referred to as the stochastic discount factor or pricing kernel.

2 The negative exponential function expression reflects constant relative risk aversion utility and lognormally distributed consumption growth. In e^{a+bz}, the expected value is a and standard deviation is z. For an explanation of constant relative risk aversion, see an investments text such as Elton, Gruber, Brown, and Goetzmann (2014).

market data can also be established. Suppose that $\tilde{z}$, assumed to have a mean of zero, takes on one of two values—a negative value indicating a bad state or a positive value indicating a good state. The probabilities of bad and good states are 0.4 and 0.6, respectively.

Using market-consistent values in the exponent in the expression for $\widetilde{m}_{t,1}$, we can calculate the price of a bond promising $1 for sure in one year as the expected value of the investor's willingness to trade present for future consumption:

$$E_t\left(\widetilde{m}_{t,1}\right) = 0.4e^{a+b\times(z \text{ for a bad state})} + 0.6e^{a+b\times(z \text{ for a good state})}$$
$$= 0.4 \times 0.954676 + 0.6 \times 0.954379 = 0.954498$$

In the calculation, 0.954676 and 0.954379 are the asset's prices in the bad and good states, respectively. Note that the values for the random shock are consistent with the level of the yield curve in the United States from January 1999 to January 2014. The derivation of these numbers is beyond the scope of this reading. Also note the higher value of $1 received in the bad state. Following Equation 3, the investor is willing to buy the risk-free bond today for $0.954498 in exchange for $1 in one year. Also notice that the willingness to invest is smaller for the positive shock (z for a good state), because an investor is willing to pay less for the bond in the case of a good state. Thus, the positive shock is associated with a higher level of consumption today by the investor.[3]

The investor knows that she cannot affect the price of the bond, and so she must decide whether to buy or sell the bond based on this given price, $P_{t,1}$ (0.954498 from Example 1). If this price of the bond was less than the investor's expectation of the inter-temporal rate of substitution (suppose this is 0.9560), then she would prefer to buy more of the bond today. As more bonds are purchased, today's consumption falls and marginal utility of consumption today rises, so expectations conditional on current information of the inter-temporal rate of substitution, $E_t\left(\widetilde{m}_{t,s}\right)$, fall. This process continues until the rate of substitution is equal to the bond price shown in Equation 3; that is, equivalently, 0.9560 would fall and converge on 0.954498.

It is worthwhile to emphasize this point: All investors are essentially making investment decisions using Equation 3; some will want to sell their bonds to fund additional, current consumption, whereas others will want to buy bonds and defer some additional consumption until the future. To demonstrate the link between the bond price and these consumption/investment decisions, imagine for the moment that the market price of this bond is too "low" for an individual investor. In this case, the investor with a higher initial inter-temporal rate of substitution (higher $\widetilde{m}_{t,s}$) would buy more of the bond. As a result of this purchase, the investor will consume less today, leading to an increase in today's marginal utility, but he or she would expect to have more consumption and thus lower marginal utility in the future. Consequently, the inter-temporal rate of substitution would fall.

One investor cannot influence the equilibrium price. But if a substantial group of investors responded this way, then the demand and price of the bond would rise; in the illustration earlier, it would mean that it is possible for the price of the bond, 0.954498, to rise at the same time that the individual investor's inter-temporal rate of

3 An exponential function, $f(x) = e^x$, is always increasing in the variable x and increases at an increasing rate as x gets large. For example, $f(0) = e^0 = 1$ and $f(0.05) = e^{0.05} = 1.0513$. If x is negative, the more negative x is, the smaller the value of the function. For example, compare $f(-0.05) = e^{-0.05} = 0.9512$ and $f(-0.02) = e^{-0.02} = 0.9802$; the function is still increasing in x.

substitution was falling. This process would continue until all investors' willingness to invest converges on a single equilibrium value so that Equation 3 is true for all individuals and the market price is determined.

Conversely, if the market price of the bond were too "high" for a group of investors, then the investors with a lower inter-temporal rate of substitution would buy less of the bond. They would have more consumption and lower marginal utility today, but they would expect to have less consumption and higher marginal utility in the future. As a result, the inter-temporal rate of substitution would rise and the demand and price of the bond would fall. This process would again continue until Equation 3 is true for all individuals.

EXAMPLE 2

The Inter-Temporal Rate of Substitution (2 of 3)

In Example 1, suppose the current market price of the real default-free bond is \$9,540 per \$10,000, but the investor's inter-temporal rate of substitution is \$0.954498 per \$1 promised. The investor would then value the guarantee of \$10,000 in one period more than the market, so she would purchase it. As she buys more of the bond, her future income will be higher and its marginal utility lower, leading to a fall in her marginal willingness to invest in the risk-free asset. Only if there are many investors with the willingness to trade at \$9,544.98 would the market price increase until all investors have the same marginal willingness to invest.

In summary, all investors use Equation 3 to make their investment decisions, so the equilibrium price in the market for these bonds equals the expectation of the inter-temporal rate of substitution of every single investor who participates in the bond market.

If the investment horizon for this bond is one year and the payoff then is \$1, the return on this bond can be written as the future payoff minus the current payment relative to the current payment:[4]

$$l_{t,1} = \frac{1 - P_{t,1}}{P_{t,1}} = \frac{1}{E_t\left(\tilde{m}_{t,1}\right)} - 1 \tag{4}$$

Consequently, the return is higher for lower current prices. Equation 4 implies that the one-period real risk-free rate is inversely related to the inter-temporal rate of substitution. That is, the higher the return the investor can earn, the more important current consumption becomes relative to future consumption.

EXAMPLE 3

The Inter-Temporal Rate of Substitution (3 of 3)

Following the circumstances in Example 1, the one-period real risk-free interest rate is $l_{t,1} = \dfrac{1 - 0.954498}{0.954498} = 0.047671$, or 4.7671%.

4 The step from the first expression to the second follows from rearranging the first expression, $\dfrac{1 - P_{t,1}}{P_{t,1}}$, as $\dfrac{1}{P_{t,1}} - 1$ and then substituting from Equation 3.

5 THE DISCOUNT RATE ON REAL DEFAULT-FREE BONDS: UNCERTAINTY AND RISK PREMIUMS

a explain the notion that to affect market values, economic factors must affect one or more of the following: 1) default-free interest rates across maturities, 2) the timing and/or magnitude of expected cash flows, and 3) risk premiums;

An investor's expected marginal utility associated with a given expected payoff is decreased by any increase in uncertainty of the payoff; thus, the investor must be compensated with a higher expected return. This result follows from decreasing marginal utility of wealth or income because the loss of utility from lower wealth is larger than the gain from an equivalent increase in wealth. The risk premium compensates the investor for the loss from this fluctuation in future wealth or income. An individual who requires compensation for this uncertainty is called "risk averse." This property was seen in Example 1, in which the inter-temporal rate of substitution was lower in the good state of the economy compared with the bad state.

For the valuation of cash flows under uncertainty, a second property of most investors' utility is important. In particular, an investor's absolute risk aversion is assumed to fall if he or she has higher wealth or income. Absolute risk aversion relates to the amount held in risky assets at different levels of wealth; under the assumption of decreasing absolute risk aversion made here, an investor invests larger amounts in risky assets as wealth or income increases (note that absolute risk aversion is in contrast to relative risk aversion, which relates to the fraction, not the amount, of wealth held in risky assets at different levels of wealth). Consequently, one's marginal utility is always lower as one's wealth or income increases. In this case, the risk premium for a given risk is lower for wealthier individuals because the average loss of marginal utility (slope of utility) from any risk taking is smaller, which means that relative to poorer individuals, wealthier individuals are more willing to take on a given risk. Consequently, wealthier investors are willing to buy more risky assets because they would value the asset more than poorer investors. But the expected marginal utility for wealthier investors will decline as they buy more of the risky asset. Eventually, both the wealthier and poorer investors would have the same willingness to invest in risky assets when the financial market is in equilibrium.

EXAMPLE 4

The Case of Increasing Wealth

This idea can be illustrated by raising the economic shock by a fixed amount regardless of whether the economy is good or bad, which has the effect of increasing the individual's resources and making her wealthier. For example, suppose we add 0.1 to $\tilde{z}$ and thus to the resources of the investor relative to the shock in Example 1. The expected inter-temporal rate of substitution for the investor is now lower for this safe asset (the default-free bond). The expected value of the investor's willingness to trade present for future consumption would then be

$$E_t\left(\tilde{m}_{t,1}\right) = 0.4 \times 0.954528 + 0.6 \times 0.954231 = 0.954350$$

Compare this result with Example 1. The inter-temporal rate of substitution is lower under the good and bad shock to the economy. As a result, the expected inter-temporal rate of substitution $E_t\left(\tilde{m}_{t,1}\right)$ is lower for the wealthier investor by 0.000148 (= 0.954498 − 0.954350). Thus, the wealthier investor will buy the safe bond only at a lower price, and if this lower price is not the equilibrium price, the investor will substitute away from riskless assets to risky assets. Because of decreasing absolute risk aversion with wealth and because their fundamental consumption needs are met, wealthy investors will demand a lower premium than poorer investors for holding risky assets, all else being equal.

An individual with decreasing absolute risk aversion would lower the price of safe assets (see Altug and Labadie [2008] for a derivation of this result). If the rich individuals are a large percentage of the market, then the equilibrium return on the safe asset increases with the lower price. As a result, the poorer individuals would have incentives to increase their savings with the expected higher return on the safe asset. These savings allow all investors to partially compensate for any additional losses during possible bad times. Consequently, all the investors in the financial market would increase their savings when uncertainty about their future income increases. This higher savings means that the expected marginal utility in the future is lower because the investors' future resources are higher. Thus, the equilibrium price based on Equation 3 is lower, meaning that investors are compensated with a higher expected return when uncertainty in income increases.

6 THE DISCOUNT RATE ON REAL DEFAULT-FREE BONDS: RISK PREMIUMS ON RISKY ASSETS

a explain the notion that to affect market values, economic factors must affect one or more of the following: 1) default-free interest rates across maturities, 2) the timing and/or magnitude of expected cash flows, and 3) risk premiums;

The price of other (non-default-free) financial instruments is established relative to the price of the default-free bond. This relationship can be seen by considering a default-free bond with a maturity of s periods (s is greater than or equal to two). Assume that the investor is holding the security for only one period. Its current price is $P_{t,s}$. In this case, the bond has value $\tilde{P}_{t+1,s-1}$ in one period because the term to maturity of the bond has been reduced by one period relative to its original maturity date. As a result, the investor's decision is now given by

$$P_{t,s} = E_t\left(\tilde{P}_{t+1,s-1}\tilde{m}_{t,1}\right) \quad (5)$$

The price in one period is uncertain because the s period bond is sold at the market price before it matures. Also notice that there is no interest payment because the bond promises a payment only at the terminal time. If a coupon is promised at time $t + 1$, then its value would have to be added to the right-hand side of Equation 5.

EXAMPLE 5

Pricing a Two-Period Default-Free Bond

In this example, we illustrate how the pricing formula in Equation 5 leads to a risk premium on a two-period default-free bond that is not present in the one-period default-free bond. In these calculations, we use five or six digits to the right of the decimal point because the risk premium is small for a two-period bond relative to a one-period default-free bond.

Suppose the price at Time 1 of the two-year default-free bond is given by

$$\tilde{P}_{t+1,2-1} = e^{a'+b'\tilde{z}}$$

In this case, the future price can be shown to be

$\tilde{P}_{t+1,2-1} = 0.839181$ for \$1 at Time 2 with probability p = 0.4 and

$\tilde{P}_{t+1,2-1} = 0.954840$ for \$1 at Time 2 with probability p = 0.6.

The expected price at time t + 1 of a \$1 bond maturing at time t + 2 is 0.4 × 0.839181 + 0.6 × 0.954840 = \$0.908576. Without considering the investor's willingness to invest, the current value of the two-period bond is the simple present value using the one-period real risk-free interest rate of 4.7671% (from Example 3) as the discount rate. Thus, under the assumption stated, the bond would be worth $\dfrac{E_t\left(\tilde{P}_{t+1,s-1}\right)}{1 + l_{t,1}} = \dfrac{0.908576}{1.047671}$ = \$0.867234. But the actual price in the financial markets based on Equation 5 is

$$\begin{aligned} P_{t,s} &= E_t\left(\tilde{P}_{t+1,s-1}\tilde{m}_{t,1}\right) \\ &= 0.4 \times 0.839181 \times 0.954676 + 0.6 \times 0.954840 \times 0.954379 \\ &= 0.867226, \end{aligned}$$

where 0.954676 and 0.954379 are the asset's prices in the bad and good states, as determined in Example 1. The price based on Equation 5 is smaller than the present discounted value at the risk-free rate; the difference is 0.000008 per 1 principal value (i.e., 0.867234 – 0.867226 = 0.000008). Thus, the holder of a two-year bond earns a risk premium. The reason for this result can be seen by calculating

$$\begin{aligned} E_t\left(\tilde{P}_{t+1,s-1}\right)E_t\left(\tilde{m}_{t,1}\right) &= 0.908576(0.4 \times 0.954676 + 0.6 \times 0.954379) \\ &= 0.867234, \end{aligned}$$

where 0.908576 is the Time 1 price of the bond as determined earlier. Consequently, we see that

$$E_t\left(\tilde{P}_{t+1,s-1}\right)E_t\left(\tilde{m}_{t,1}\right) > E_t\left(\tilde{P}_{t+1,s-1}\tilde{m}_{t,1}\right)$$

To summarize, the price uncertainty of the two-period bond at t = 1 gives rise to a risk premium, although the bond is default-risk free.

Example 5 showed how future price uncertainty creates a discount for risk. We now derive an alternative expression for the pricing relationship in Equation 5 that explains the nature of that discount and sheds further light on the conclusion of Example 5. In statistics texts, the following relationship between expected values and covariance is proven:

$$E_t\left(\tilde{x}\tilde{y}\right) = E_t\left(\tilde{x}\right)E_t\left(\tilde{y}\right) + \text{cov}\left(\tilde{x},\tilde{y}\right)$$

Here, $\text{cov}_t\left(\tilde{x},\tilde{y}\right)$ refers to the conditional (on information at time t) covariance of the random variable $\tilde{x}$ with $\tilde{y}$. Thus, from Equation 5,

$$P_{t,s} = E_t\left(\tilde{P}_{t+1,s-1}\tilde{m}_{t,1}\right) = E_t\left(\tilde{P}_{t+1,s-1}\right)E_t\left(\tilde{m}_{t,1}\right) + \text{cov}\left(\tilde{P}_{t+1,s-1},\tilde{m}_{t,1}\right)$$

But from Equation 4, $1 + l_{t,1} = \dfrac{1}{E_t\left(\tilde{m}_{t,1}\right)}$. So, an alternative way to view the pricing relationship in Equation 5 is

$$P_{t,s} = \frac{E_t\left(\tilde{P}_{t+1,s-1}\right)}{1 + l_{t,1}} + \text{cov}_t\left(\tilde{P}_{t+1,s-1}, \tilde{m}_{t,1}\right), \qquad (6)$$

where $\text{cov}_t\left(\tilde{P}_{t+1,s-1}, \tilde{m}_{t,1}\right)$ represents the covariance between an investor's inter-temporal rate of substitution, $\tilde{m}_{t,1}$, and the random future price of the investment at $t + 1$, $\tilde{P}_{t+1,s-1}$, based on the information available to investors today (t). The subscript is reduced by one because an investment with time to maturity s at time t becomes an investment with time to maturity $s - 1$ at time $t + 1$ (Cochrane 2005).

Equation 6 expresses the value of a risky asset as the sum of two terms. The first term is the asset's expected future price discounted at the risk-free rate. It may be called the risk-neutral present value because it represents a risky asset's value if investors did not require compensation for bearing risk (notice the parallel with the fundamental pricing equation, Equation 1, if it had one cash flow and no risk premium). In Example 5, this value is 0.867234.

The covariance term is the discount for risk. Note that with a one-period default-free bond, the covariance term is zero because the future price is a known constant ($1) and the covariance of a random quantity with a constant is zero; and intuitively, its value is given by the first term. Consequently, Equation 6 reduces to Equation 3 for the one-period default-free bond. But with the two-period default-free bond, the future price of $1 two periods in the future is known with certainty, but the price one period in the future is not. Consequently, the covariance term is not zero.

In general with risk-averse investors, the covariance term for most risky assets is expected to be negative. That is, when the expected future price of the investment is high, the marginal utility of future consumption relative to that of current consumption is low. Alternatively, during bad economic times, investors expect a smaller labor income in the future, so the marginal utility of future consumption, and hence the inter-temporal rate of substitution, is higher. This relationship leads investors to demand a higher required rate of trade-off of future for current consumption—as in bad economic times when the labor market contracts. Bad economic times also tend to be associated with declining risky asset payouts (declining earnings and dividends for ordinary shares and defaults for bonds), leading to declining asset prices. The result is that the covariance term for risky assets is typically negative, so the price of the asset is lower. This negative covariance term results in a positive risk premium, $\rho^i_{t,s}$, in Equation 1 because a lower price today leads to a higher return over time. Holding all else constant, the risk premium term and the required return for an asset should be higher, and its current market price is lower the larger the magnitude of the negative covariance term.

EXAMPLE 6

An Alternative Method to Evaluate the Price Discount for Risk

The covariance between the investor's willingness to invest and the price of the two-year bond next period can also be computed as follows (recall the standard formula for covariance is $\text{cov}(\bar{x}, \bar{y}) = \Sigma p_i[x_i - E(x)][y_i - E(y)]$:

$$\begin{aligned}\text{cov}_t\left(\tilde{m}_{t,1}, \tilde{P}_{t+1,2-1}\right) &= 0.4(0.954676 - 0.954498) \times (0.839181 - 0.908576) + \\ &\quad 0.6(0.954379 - 0.954498) \times (0.954840 - 0.908576) \\ &= -0.000008\end{aligned}$$

In the bad state of the economy, the willingness to invest (0.954676) is above its average (0.954498), yet the bond price (0.839181) is below its average (0.908576). The reverse is true in the good state. Thus, the covariance between the intertemporal rate of substitution and the price of the asset is negative. This result means the investor finds this investment inferior to one with a payoff that is independent of her willingness to invest. In particular, we have

$$P_{t,s} = -\frac{E_t\left(\tilde{P}_{t+1,s-1}\right)}{1 + l_{t,1}} = -0.000008.$$

With a lower price, the return on the two-year bond is higher.

In this example, the price discount is not too large because the risk between a one- and two-year US government bond is not that crucial. However, riskier assets, such as equity, will have a higher discount. In addition, the higher risk premium on equity still follows from the covariance between the cash flow and the investor's willingness to invest over the time horizon of the investment. Thus, a higher risk premium for stocks arises from a larger value for this covariance.

The risk premium can be computed as follows: The expected holding period return on the s period bond through time $t + 1$, using the results of Example 5, is given by

$$\begin{aligned}r_{t,s} &= \frac{E_t\left(\tilde{P}_{t+1,s-1}\right) - P_{t,s}}{P_{t,s}} \\ &= \frac{0.908576 - 0.867226}{0.867226} = 0.047681, \text{ or } 4.7681\%,\end{aligned} \tag{7}$$

so the risk premium $\rho^i_{t,s} = r_{t,s} - l_{t,1} = 0.047681 - 0.047671 = 0.00001$.

Alternatively, Equations 7 and 6 can be manipulated so that

$$\begin{aligned}r_{t,s} - l_{t,1} &= \frac{E_t\left(\tilde{P}_{t+1,s-1}\right)}{P_{t,s}} - \left(1 + l_{t,1}\right) = -\frac{\left(1 + l_{t,1}\right)}{P_{t,s}}\text{cov}_t\left(\tilde{m}_{t,1}, \tilde{P}_{t+1,s-1}\right) \\ &= -\left(1 + l_{t,s}\right)\text{cov}_t\left(\tilde{m}_{t,1}, \frac{\tilde{P}_{t+1,s-1}}{P_{t,s}}\right) = -\frac{(1 + 0.047671)}{0.867226} \times (-0.000008) \\ &= 0.00001 = \rho^i_{t,s},\end{aligned} \tag{8}$$

which is the return premium demanded by investors because of the uncertain Time 1 price of the riskless two-period bond.[5]

This relationship implies that an asset's risk premium, $\rho^{i}_{t,s}$ in Equation 1, is driven by the covariance of its returns with the inter-temporal rate of substitution for consumption and can exist even for a default-free bond because of the uncertainty of its price before maturity. Most risky assets have returns that tend to be high during good times, when the marginal value of consumption is low, and low during bad times, when the marginal value of consumption is high, and so bear a positive risk premium. Any asset that tended to have relatively high returns when the marginal utility of consumption was high would provide a type of hedge against bad times, bear a negative risk premium, and have a relatively high price and low required rate of return.

7 DEFAULT-FREE INTEREST RATES AND ECONOMIC GROWTH

c explain the relationship between the long-term growth rate of the economy, the volatility of the growth rate, and the average level of real short-term interest rates;

From the previous discussion, it is a relatively small conceptual step to understand the relationship between an economy's GDP growth and real default-free interest rates. If there is a known independent change in real GDP growth or a change that can be forecasted perfectly, then an increase in real GDP growth should lead to an increase in the real default-free rate of interest because more goods and services will be available in the future relative to today. The result is that investors' willingness to substitute over time will fall, resulting in less saving and more borrowing, so that the real default-free interest rate increases, as in Equation 4.

But GDP growth from one period to the next cannot be perfectly anticipated. Under these uncertain circumstances, interest rates will still be positively related to the expected growth rate of GDP, but additionally they will be positively related to the expected volatility of GDP growth.

EXAMPLE 7

The Effect of Volatility on Prices

One can see the effect of volatility by doubling the standard deviation of the random variable $\tilde{z}$ from what was assumed in Example 1. In this case, the price of the one-period bond in Example 1 would be

$$E_t\left(\tilde{m}_{t,1}\right) = 0.4 \times 0.954855 + 0.6 \times 0.954260 = 0.954498$$

5 Notice that simultaneously multiplying Equation 6 by $(1 + l_{t,1})$ and dividing by $P_{t,s}$ gives

$$(1 + l_{t,1}) = \frac{E_t\left(\tilde{P}_{t+1,s-1}\right)}{P_{t,s}} + \left(1 + l_{t,1}\right)\text{cov}_t\left(\tilde{m}_{t,1}, \frac{\tilde{P}_{t+1,s-1}}{P_{t,s}}\right), \text{ w}$$

so we can write $\dfrac{E_t\left(\tilde{P}_{t+1,s-1}\right)}{P_{t,s}} - \left(1 + l_{t,1}\right)$ in Equation 8 as

$$\frac{E_t\left(\tilde{P}_{t+1,s-1}\right)}{P_{t,s}} - \frac{E_t\left(\tilde{P}_{t+1,s-1}\right)}{P_{t,s}} - \left(1 + l_{t,1}\right)\text{cov}_t\left(\tilde{m}_{t,1}, \frac{\tilde{P}_{t+1,s-1}}{P_{t,s}}\right) = -\left(1 + l_{t,1}\right)\text{cov}_t\left(\tilde{m}_{t,1}, \frac{\tilde{P}_{t+1,s-1}}{P_{t,s}}\right).$$

Notice that the expected value is the same as in Example 1 but that the prices in each state are more dispersed, reflecting the doubling of the standard deviation. For the two-period default-free bond, continuing with the parameter values (a' and b') from Example 5, we would compute

$$E_t\left(\tilde{P}_{t+1,2-1}\right) = 0.4 \times 0.776625 + 0.6 \times 1.005451 = 0.913921.$$

Notice that doubling the volatility leads to a somewhat unrealistic price greater than 1 (implying a negative yield) in the good state, even though the expected price is less than 1.

Then,

$$\begin{aligned} P_{t,s} &= E_t\left(\tilde{P}_{t+1,2-1}\tilde{m}_{t,1}\right) \\ &= 0.4 \times 0.776625 \times 0.954855 + 0.6 \times 1.005451 \times 0.954260 = 0.872303. \end{aligned}$$

The risk neutral price is $\dfrac{E_t\left(\tilde{P}_{t+1,2-1}\right)}{1 + l_{t,1}} = \dfrac{0.913921}{1.047671} = 0.872336$. So, from Equation 6, $\text{cov}_t\left(\tilde{m}_{t,1}, \tilde{P}_{t+1,2-1}\right) = 0.872303 - 0.872336 = -0.000033$.

As a result, the holding period return on a two-period bond for one year is higher [i.e., (0.913921 − 0.872303)/0.872303 = 4.771%, compared with 4.768% in Example 6], and because of the higher volatility, investors require a higher premium.

There are two practical implications of this analysis for the values of real default-free interest rates:

- An economy with higher trend real economic growth, other things being equal, should have higher real default-free interest rates than an economy with lower trend growth. We should thus expect to find that real default-free interest rates in fast-growing, developing economies, such as India and China, are higher than in slower-growing, developed economies, such as Western Europe, Japan, and the United States. The higher rate of economic growth occurs for developing economies because a developing economy is typically below its steady state growth, so it grows faster to catch up. During these periods, the marginal product of capital (the additional output resulting from the addition of one unit of capital, holding all else constant) would be expected to be higher, so the real default-free interest rate should also be expected to be higher. Of course, this advantage will dissipate as the economy matures, as in the case of Japan and Western Europe from 1950 to 2000.
- Again, other things being equal, real interest rates are higher in an economy in which GDP growth is more volatile than in an economy in which growth is more stable.

EXAMPLE 8

The Present Value Model and Macroeconomic Factors

1 An asset's risk premium is high when:

 A there is no relationship between its future payoff and investors' marginal utility from future consumption.

 B there is a positive relationship between its future payoff and investors' marginal utility from future consumption.

C there is a negative relationship between its future payoff and investors' marginal utility from future consumption.

2 The relationship between the real risk-free interest rate and real GDP growth is:

A negative.

B neutral.

C positive.

3 The relationship between the real risk-free interest rate and the volatility of real GDP growth is:

A negative.

B neutral.

C positive.

4 A risky asset offers high positive returns during business downturns. A colleague argues that the nominal required rate of return on the asset may be less than the nominal risk-free rate. Is the colleague correct?

A Yes

B No, the return must be higher than the nominal risk-free rate.

C No, the relationship between the asset's nominal return and the nominal risk-free rate is indeterminate.

5 Suppose you are analyzing the expected impact of an increase in real GDP growth above trend on overall equity market valuation. Assume real growth in income of the corporate sector follows real GDP growth. Assume also that there is no impact of the increase on inflation. On the basis of theory and holding all else constant, explain why the impact of the assumed increase in real GDP growth on overall equity market valuation is ambiguous.

Solution to 1

C is correct. An asset's risk premium is determined by the relationship between its future payoff and the marginal value of consumption as given by the covariance between the two quantities. When the covariance is negative—that is, payoffs are low and expected utility from consumption is high—or equivalently, when times in the future are expected to be bad and the value of an extra unit of consumption is high, the risk premium will be high. When the covariance term is zero (there is no relationship), the asset is risk free. When the covariance term is positive, the asset is a hedge and will have a rate of return less than the risk-free rate.

Solution to 2

C is correct. The real risk-free rate is positively related to real GDP growth. An increase in real GDP growth reduces the need for investors to save for future consumption because more goods and services will be available to them in the future relative to today as a result of higher expected income in the future. A higher real rate of interest is needed to induce individuals to save for future consumption in such circumstances.

Solution to 3

C is correct. The real risk-free rate is positively related to the volatility of real GDP growth. An increase in volatility of real GDP growth means that there is greater risk that the income available for consumption will be lower than expected. Therefore, risk-averse investors will require a higher real rate of return in compensation.

Solution to 4

A is correct. For the required return to be less than the risk-free rate, the asset's risk premium would need to be negative. Because the asset supplies relatively high returns in economic conditions in which the marginal utility of consumption is relatively high, the covariance term in Equation 6 is positive and the asset thus bears a negative risk premium.

Solution to 5

Equation 1 can be applied to the overall equity market, which is the aggregate of individual equity securities. The increase in real GDP growth would be expected to affect both the numerator and denominator of Equation 1 in offsetting ways, so the overall effect on equity market value is ambiguous. The impact of an increase in real GDP growth on expected corporate earnings is positive by assumption, which by itself would suggest an increase in equity market value by a larger numerator value in Equation 1. However, the increase in real GDP should also increase the real risk-free rate, which by itself would suggest a decrease in equity market value by increasing the rate at which expected cash flows are discounted. We cannot infer which effect will dominate from the information given. The overall effect on equity market value is ambiguous, under the assumptions given.

8 REAL DEFAULT-FREE INTEREST RATES AND THE BUSINESS CYCLE

d explain how the phase of the business cycle affects policy and short-term interest rates, the slope of the term structure of interest rates, and the relative performance of bonds of differing maturities;

One of the crucial insights that macro-finance provides is that there should be a connection between the real risk-free rate of interest available in an economy and the underlying trend rate of economic growth in the same economy. We explored the roots of this relationship earlier. To recap briefly, the willingness of investors to substitute future wealth for current consumption will be inversely related to the change in real GDP growth. In a world where GDP growth could be forecasted perfectly, there will be a positive relationship between the real risk-free rate and real GDP growth. But GDP growth is not perfectly predictable. Because of this unpredictability, we also concluded that the real risk-free rate would not only be positively related to real GDP growth but also positively related to the volatility of real GDP growth.

Equation 1 shows that the real default-free required return, $l_{t,s}$, is a component of the discount rate that we apply to the cash flows generated by all financial instruments.

8.1 Economic Growth and Real Yields

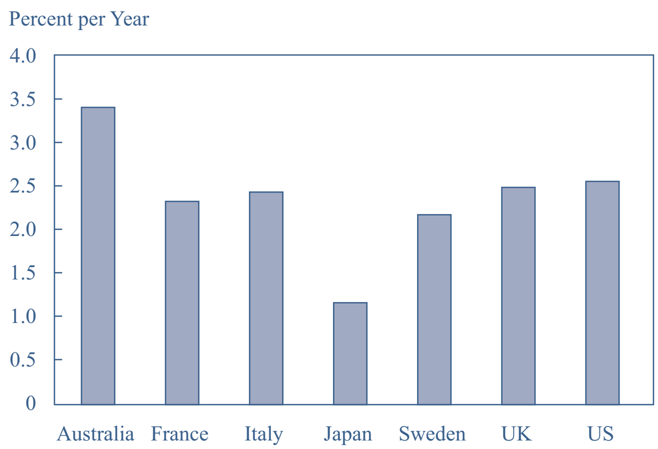
For evidence of the relationship between real interest rates and GDP growth, we could focus on the yields available from inflation-linked bonds issued by governments in developed economies. These bonds pay a "real" return (or yield) plus a return that is linked directly to an index of consumer prices. Index-linked bonds are issued by many governments in developed economies, including Canada, France, Germany, Italy, Sweden, the United Kingdom, and the United States, and also by some governments in developing countries, such as Brazil. In some markets—for example, the United Kingdom's index-linked gilt market—both the coupon and principal payments from these bonds are indexed to a measure of consumer prices. In other markets—for example, the US Treasury Inflation-Protected Securities (TIPS) market—the principal payment is indexed and the coupon is a function of the indexed principal. In both cases, any increase in the level of the consumer price index over time (that is, positive inflation) leads to an increase in both the coupon payment and eventual principal payment. Although the details of the indexation vary from bond market to bond market, for all practical purposes, we can think of these bonds as being inflation protected.

Given the earlier discussion in sections on real default-free interest rates and default-free interest rates and economic growth, other things being equal, we would expect the (real) yields on inflation-indexed bonds to be higher for those countries with high growth, such as India and China, relative to those issued by, say, the UK or US governments, where economic growth is much lower. Other things being equal, we should also expect to see real yields on short-dated index-linked bonds issued by governments of economies that are very volatile to be higher than real yields on those issued by governments of economies that are less volatile.

Although many index-linked government bond markets are relatively new, we can examine the cross-sectional relationship between economic growth and real risk-free yields relatively easily. Panel A of Exhibit 1 shows the real yields on a set of short-dated index-linked government bonds in 2007, immediately prior to the 2008–2009 global financial crisis. The real yield on short-dated Japanese government bonds at that time were lower than elsewhere, whereas Panel B shows that Japanese growth had been historically very low and not very volatile up to that point in time. Of the developed-economy bond yields in this exhibit, those issued by the Australian government offered the highest yield, perhaps reflecting the relatively strong Australian economic growth shown in Panel B.

Exhibit 1 Real Yields, GDP Growth, and Volatility for Various Countries

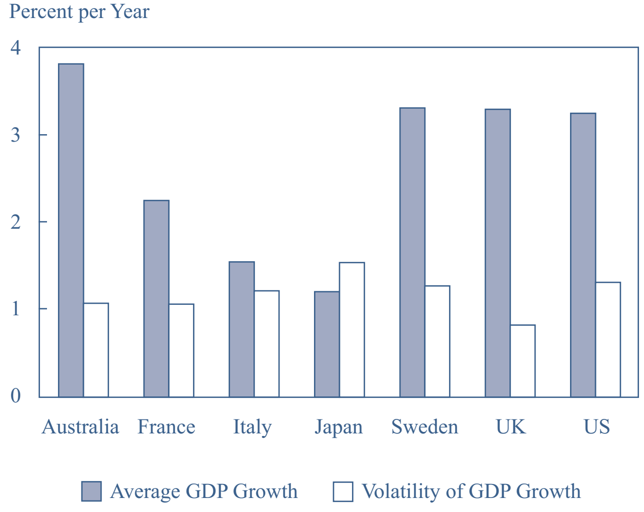
A. Real Yields, July 2007



(continued)

Exhibit 1 (Continued)

B. Growth and Volatility, 1996–2007



Sources: Based on data from Thomson Reuters and the authors' calculations.

It is difficult to discern a very clear pattern between historical economic growth, the volatility in that growth, and short-term real yields. Nonetheless, it is interesting to note that the correlation between this limited set of bond yields (Panel A of Exhibit 1) and historical growth (gray bars in Panel B of Exhibit 1) is 0.57, but the correlation with historical volatility (white bars in Panel B) is 0.74. So there does appear to be some support for the prediction of macro-theory, although this sample is of course very limited. One of the reasons why there is perhaps not a clearer relationship is that the real yield data are forward looking. The real yield data represent the required real return on these bonds based on expected future growth and volatility in that growth, whereas the GDP-based variables represent historical growth and volatility. If investors use the past as a guide for the future, they might expect a reasonably high correlation between past growth and current real yields. But the past may be a very bad guide for the future, particularly in the case of rapidly developing countries or following the sort of major shock to global economic growth that occurred following the collapse of Lehman Brothers.

EXAMPLE 9

The Evolution of Real Yields

We have seen that there is at least tentative evidence to suggest a positive relationship between economic growth and the yields on short-dated real government bonds. But how does the business cycle affect these yields over time?

The problem in gathering evidence on the drivers of real yields over time is that index-linked government bond markets are a fairly recent financial innovation, especially when compared with conventional government bonds. For example, it was only in 1997 that the first index-linked bond was issued by the US Treasury. The oldest index-linked government bond market is the United Kingdom's. The UK government issued its first index-linked gilt in 1981. To investigate the connections between the macroeconomy and the real risk-free rate over time, we can focus on the UK's index-linked gilt market.

Exhibit 2 Real Yields on UK Index-Linked Gilt and Volatility of UK and OECD GDP Growth

Notes: GDP growth is represented by a three-year moving standard deviation of the variable. Shaded areas in Panel A indicate UK recessions.
Sources: Based on data from OECD.Stat and UK National Statistics Office.

Panel A of Exhibit 2 shows the real yield on a short-dated constant maturity UK index-linked gilt. We will begin by focusing on the period from 1985 to 2007. Although fairly volatile, there is a clear downward trend in this yield from 1985 to 2007. One explanation for this decline could be a commensurate decline in expectations about UK economic growth. However, real economic growth between 1985 and 1999 averaged 2.8%, and between 2000 and 2007 it averaged 2.7%. There was very little change in average growth at a time when real yields were falling, and therefore, it is probably fair to assume that expectations of future growth were relatively stable over this time too. Panel B of Exhibit 2 shows the volatility of UK real GDP growth as represented by a three-year moving standard deviation of this variable. It shows that the volatility of UK economic growth declined quite dramatically from 1995 to 2007. This decline in UK GDP volatility was also experienced elsewhere in the global economy. The same chart shows the decline in GDP volatility for OECD countries. This decline in economic volatility has been called "the great moderation"—that is, a period when the global economy and its financial markets were characterized by relatively low levels of volatility. Therefore, one plausible explanation for the declining level of real interest rates in the United Kingdom is that they were driven down by the moderation in economic volatility between the early 1990s and 2007.

The evidence in Exhibit 2 suggests that declining levels of economic volatility led to declining levels of the real default-free interest rate in the United Kingdom between 1999 and 2007. However, the absence of such markets elsewhere over this sample period does not mean that the same phenomenon was absent or irrelevant in other developed-economy bond markets. The yield on a conventional government bond includes a number of components, one of which is the real default-free rate of return. So, in all likelihood, declining global economic volatility led to declines in the real rates of return required by investors elsewhere, which, in turn, may have contributed to the decline in conventional government bond yields. See Example 10 for an illustration of how the global financial crisis affected real short rates. We will focus on the drivers of conventional government bond yields relative to index-linked government bond yields at a later stage.

EXAMPLE 10

Post–Global Financial Crisis, 2008–2011 Real Default-Free Yields

Exhibit 3 shows the yields on short-dated index-linked bonds at the end of 2011. Compared with their pre-crisis levels shown in Exhibit 2, they had all fallen. The collapse of Lehman Brothers and the ensuing liquidity and sovereign debt crisis caused economic and financial market volatility to rise substantially, as shown in Panel B of Exhibit 2. Other things being equal, one would have expected the real yields to rise, not to fall. But other things were not equal. One explanation for the fall in these real yields is that, despite the higher volatility experienced in 2008–2011, investors believed that future real economic growth would be lower and, therefore, that the equilibrium real yield in these economies was deemed to be commensurately lower.

Exhibit 3 Real Yields on Short-Dated Index-Linked Bonds, December 2011

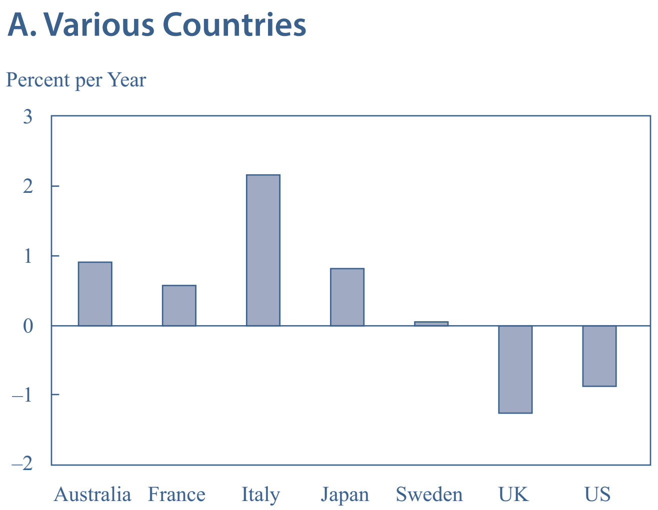
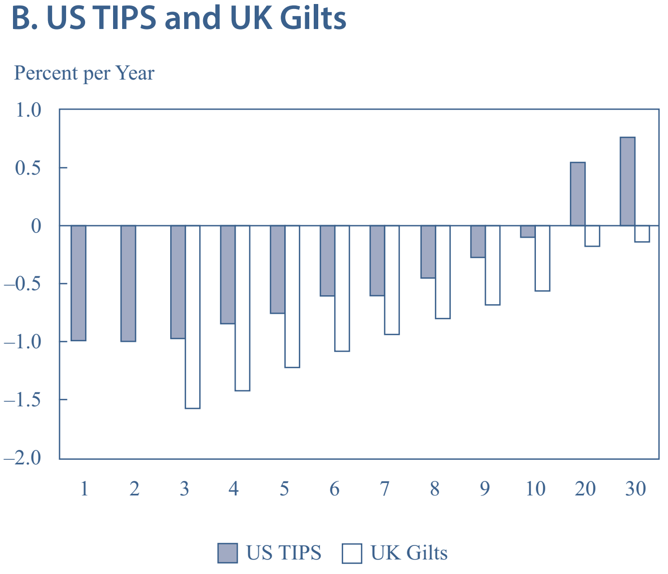


Exhibit 3 (Continued)



Sources: Based on data from Bloomberg and Thomson Reuters.

Panel A of Exhibit 3 also shows that real short-dated rates in the United Kingdom and United States were not just lower than they were in the immediate pre-crisis period but that they were negative, too. Panel B shows that most inflation-linked bonds issued by the US government (TIPS) and all inflation-linked bonds issued by the UK government (index-linked gilts) at this time were also negative (term structure will be explored in more detail later). One explanation for the lower yields may have been the fear among some investors of very high levels of inflation in the future. Arguably, the easy monetary policy, including both formal (US and UK) and informal (eurozone) quantitative easing programs, implemented because of the collapse of Lehman Brothers along with other events, may have led enough investors to believe that inflation-linked government bonds offered the inflation protection they needed despite the low and even negative historical real yields.

But it is also important to remember that index-linked bonds issued by developed-economy governments not only are quite special, given the credit and inflation protection that they provide, but also are often in very limited supply. These characteristics mean that in times of crisis and of great uncertainty, investors may see them as a safe haven for their capital, which can, in turn, drive down their yields.

8.2 Real Default-Free Interest Rate Summary

The real default-free interest rate, which we have proxied here with the yields on short-term inflation-protected government bonds, has a close connection with the business cycle via the related connection with the saving decisions of individuals. We can put this discussion in the context of the basic pricing shown in Equation 1. For a real default-free bond, Equation 1 simplifies to

$$P_t^i = \sum_{s=1}^{N} \frac{CF_{t+s}^i}{\left(1 + l_{t,s}\right)^s}. \quad (9)$$

Because it is a fixed-interest investment that is default-free, the cash flow at time $t + s$ is certain. Equation 9 implies that it is only changes in $l_{t,s}$ that will affect the price of such a bond. In turn, $l_{t,s}$ will be determined by real economic growth and the volatility in economic growth over time as a result of the aggregation of the consumption and saving decisions of individual investors.

9 THE YIELD CURVE AND THE BUSINESS CYCLE: SHORT-TERM NOMINAL INTEREST RATES

d explain how the phase of the business cycle affects policy and short-term interest rates, the slope of the term structure of interest rates, and the relative performance of bonds of differing maturities;

Earlier, we considered the determination of the price of a real default-free bond (see Equation 9). The analysis demonstrated that the saving and investment decisions of investors mean that the expected return on these bonds will be positively related to both expected real GDP growth and the expected volatility of this growth. We now move on to consider the price of a default-free bond that pays a fixed nominal (currency) amount when it matures. We will consider, for example, a bond issued by a government in a developed economy where the prospect of default is so negligible that it is ignored.

What factors would affect the price of such a bond? First, we consider a world without inflation. In this world, investors would still be giving up current consumption by investing in this bond today, in which case Equation 9 would be appropriate. But of course, deferring consumption at time t in a world with positive inflation will have an impact on the quantity of goods that can be bought at time $t + s$ when the bond matures. Investors will want to be compensated by this bond for the inflation that they expect between t and $t + s$, which we define as $\theta_{t,s}$. If investors could forecast inflation perfectly, they would demand a return given by $l_{t,s} + \theta_{t,s}$ to compensate them for the expected inflation and ensure the real level of consumption. But unless the investment horizon is very short, investors are unlikely to be very confident in their ability to forecast inflation accurately. Because we generally assume that investors are risk averse and thus need to be compensated for taking on risk and because they seek compensation for expected inflation, they will also seek compensation for taking on the uncertainty related to future inflation. We denote this risk premium by $\pi_{t,s}$, which is distinct from the risk premium in Equation 1 ($\rho^i_{t,s}$).[6] We can rewrite our basic pricing formula for a default-free nominal coupon-paying bond from Equation 9 as follows:

$$P_t^i = \sum_{s=1}^{N} \frac{CF_{t+s}^i}{\left(1 + l_{t,s} + \theta_{t,s} + \pi_{t,s}\right)^s}. \qquad (10)$$

Note that the bond's payoff is still certain in nominal terms because we are assuming that there is a negligible chance that the issuer (a developed-economy government) will default on its commitments. It is the real value of this payoff that is now uncertain, hence the need for a risk premium, $\pi_{t,s}$, on nominal bonds.

6 Even though it is a risk premium, we have suppressed the superscript i on the inflation uncertainty risk premium because it is not asset specific and applies across all asset classes.

EXAMPLE 11

The Risk Premium for Inflation Uncertainty

Suppose that an analyst estimates that the real risk-free rate is 1.25% and that average inflation over the next year will be 2.5%. If the analyst observes the price of a default-free bond with a face value of £100 and one full year to maturity as being equal to £95.92, what would be the implied premium embedded in the bond's price for inflation uncertainty?

Solution

The (approximate) implied premium can be calculated as follows:

$$\pi_{t,s} = 0.504\% = \frac{100}{95.92} - (1 + 0.0125 + 0.025).$$

Having established Equation 10, we will now focus on the relationship between short-term nominal interest rates and the business cycle, in which the nominal bond issued by a government in a developed economy has a very short maturity—for example, a US government Treasury bill.

9.1 Short-Term Nominal Interest Rates and the Business Cycle

Treasury bills (T-bills) are very short-dated nominal zero-coupon government bonds. T-bills are issued by most developed-economy governments or by their agents to help smooth the cash flow needs of the government. The short-dated nature of T-bills and the fact that they are often used to implement monetary policy means that their yields are also usually very closely related to the central bank's policy rate. Indeed, because of their short-dated nature, the uncertainty that investors would have about inflation over an investment horizon of, say, s equals three months will usually be relatively low. Therefore, for the purposes of the exposition in this section of the reading, we will assume that $\pi_{t,s}$ is so negligible that we can ignore it. So, we can modify Equation 10 to give Equation 11, which can capture all of the salient features of the pricing dynamics of a T-bill:

$$P_t^i = \frac{CF_{t+s}^i}{\left(1 + l_{t,s} + \theta_{t,s}\right)^s}. \quad (11)$$

Note that the summation term is not needed because there is only one payment from a T-bill.

We have already examined the way in which the real default-free rate of interest, $l_{t,s}$, will vary over time with the business cycle and how it may also be affected by its status as a haven in times of economic uncertainty. We now move on to consider how a central bank's policy rate, which is a short-term nominal interest rate, evolves with the business cycle.

10 TREASURY BILLS AND THE BUSINESS CYCLE

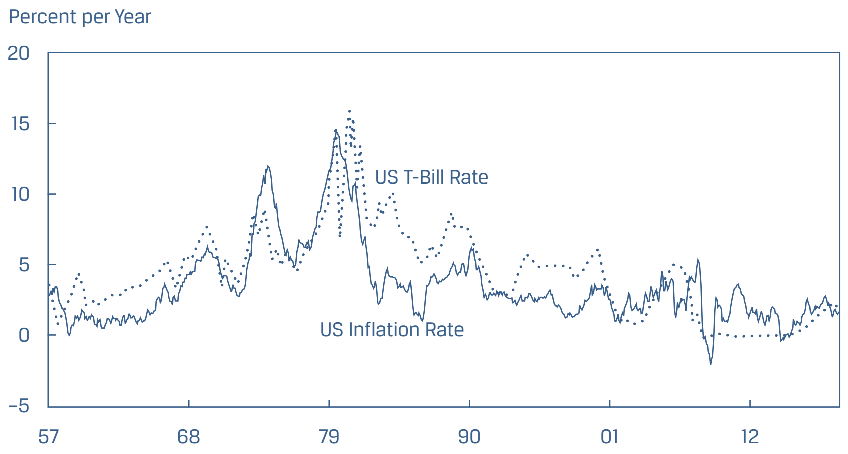
d explain how the phase of the business cycle affects policy and short-term interest rates, the slope of the term structure of interest rates, and the relative performance of bonds of differing maturities;

To summarize briefly, the nominal rate of interest will equal the real interest rate that is required to balance the requirements of savers and investors plus investors' expectations of inflation over the relevant borrowing or lending period. It follows that short-term nominal interest rates will be positively related to short-term real interest rates and to short-term inflation expectations. Other things being equal, we would also expect these interest rates to be higher in economies with higher, more volatile growth and with higher average levels of inflation over time.

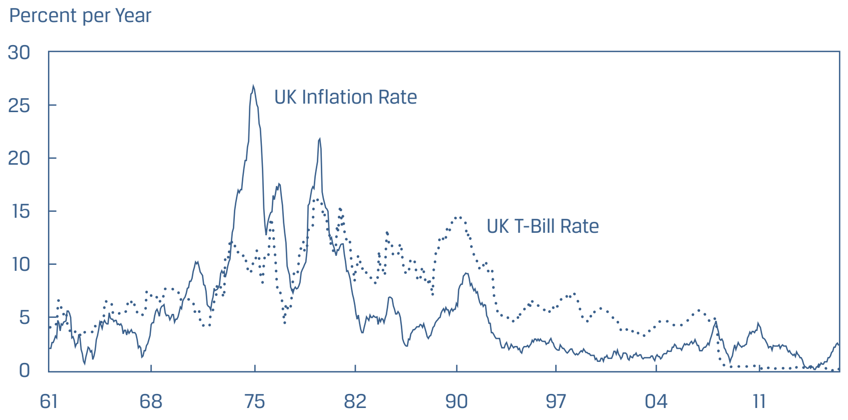
Panel A of Exhibit 4 shows the yield on a three-month US T-bill, and Panel B shows the yield on an equivalent T-bill issued by the UK government. In each panel of the exhibit, we also present the inflation rates in these two economies. There is a close correlation between measured inflation and T-bill yields in both economies. Although measured inflation is not the same as expected inflation, it is likely that current inflation plays a big role in the formation of inflation expectations, particularly over the very short investment horizon involved when investing in a T-bill.

Exhibit 4 Treasury Bill Rates and Inflation

A. United States



B. United Kingdom

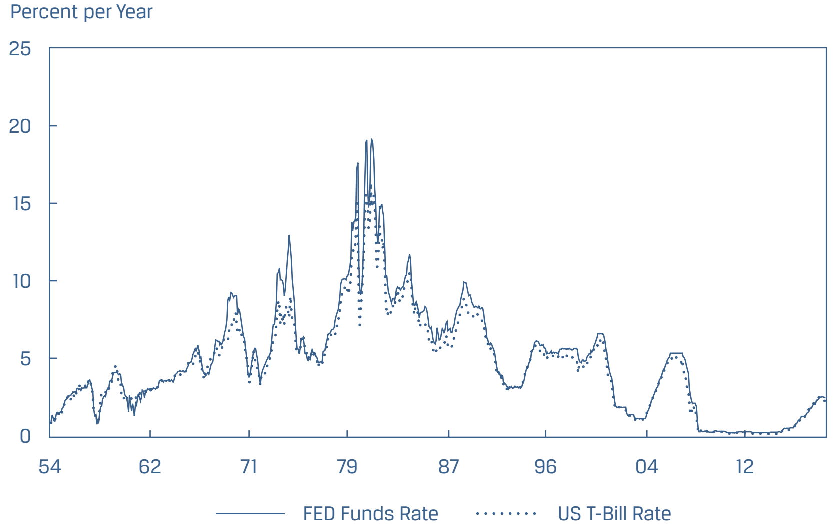


Source: Based on data from FRED.

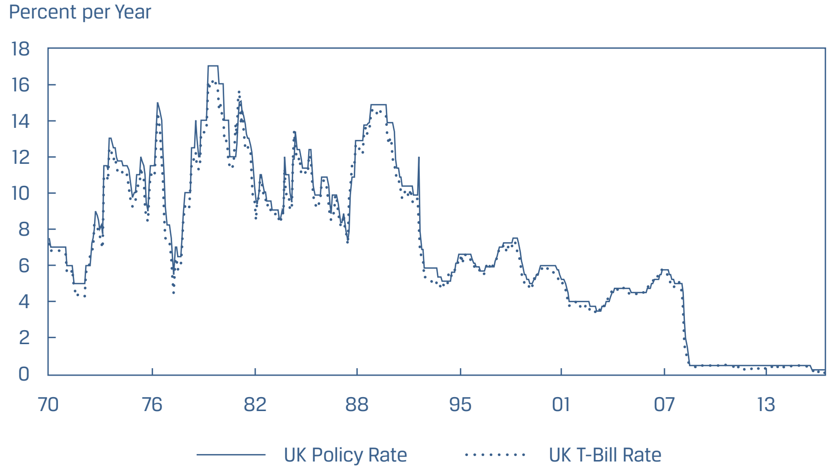
Exhibit 4 clearly shows that the inflation environment is a key driver of short-term interest rates. The central banks and monetary authorities responsible for setting interest rates in an economy do so in response to the economy's position in the business cycle—cutting their policy rates when activity and/or inflation are judged to be "too low" and raising rates when activity and/or inflation are judged to be "too high." In other words, a responsible central bank or monetary authority will usually set its policy rate with reference to the level of expected economic activity and the expected rate of increase of prices—that is, inflation. Exhibit 5 shows the close relationship between the yields on short-term default-free T-Bills in the United States and United Kingdom and the policy rates of their respective central banks.

Exhibit 5 Interest Rates and Policy Rates

A. United States



B. United Kingdom



Source: Based on data from Thomson Reuters.

A US economist, John Taylor, devised a rule for setting policy rates, a rule that could help rate setters gauge whether their policy rate is at an "appropriate" level (Taylor 1993). This rule is known as the Taylor rule, and it takes the following form:

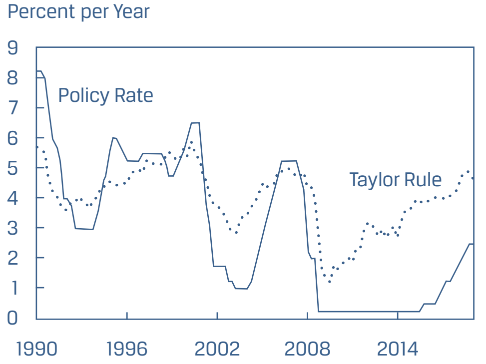
$$\begin{aligned} pr_t &= l_t + \iota_t + 0.5\left(\iota_t - \iota_t^*\right) + 0.5\left(Y_t - Y_t^*\right) \\ &= l_t + 1.5\iota_t - 0.5\iota_t^* + 0.5\left(Y_t - Y_t^*\right), \end{aligned} \tag{12}$$

where pr_t is the policy rate at time t, l_t is the level of real short-term interest rates that balances long-term savings and borrowing in the economy, ι_t is the rate of inflation, ι_t^* is the target rate of inflation, and Y_t and Y_t^* are, respectively, the logarithmic levels of actual and potential real GDP. The difference between Y_t and Y_t^* is known as the "output gap," which is essentially measured in percentage terms. When the output gap is positive, it implies that the economy is producing beyond its sustainable capacity. This situation is similar to a marathon runner who sets off way too fast at the start of a race; in the end, he will overheat and break down unless he reduces his running pace. Conversely, when the output gap is negative, it implies that the economy is producing below its sustainable capacity. This situation is similar to a marathon runner who sets off too slowly. If he wants to win the race, at some point, he will have to use up conserved energy and speed up. Positive output gaps are usually associated with high and/or rising inflation, whereas negative output gaps are usually accompanied by high levels of unemployment. Generally, the policy rule should have a larger weight on inflation (1.5) relative to the weight on output (0.5). The purpose is to stabilize inflation over the longer term near the targeted inflation rate (note that the reason for the weightings is that the inflation rate appears twice in the equation; see the first line of Equation 12). When inflation is close to the targeted or preferred rate and when the output gap is zero, the appropriate policy rate will be equal to the level of the short-term real interest rate, l_t, that balances long-term savings and borrowing in the economy plus the targeted/preferred rate of inflation. This level of the policy rate is often referred to by economists as the neutral policy rate—that is, the policy rate that neither spurs on nor impedes real economic activity. Other things being equal, when inflation is above (below) the targeted level, the policy rate should be above (below) the neutral rate, and when the output gap is positive (negative), the policy rate should also be above (below) the neutral rate. For example, if l_t is 2.0%, ι_t is 3.0%, ι_t^* is 2.0%, and the output gap is 2.0%, then the "appropriate" policy rate implied in the Taylor rule would be 6.5%.

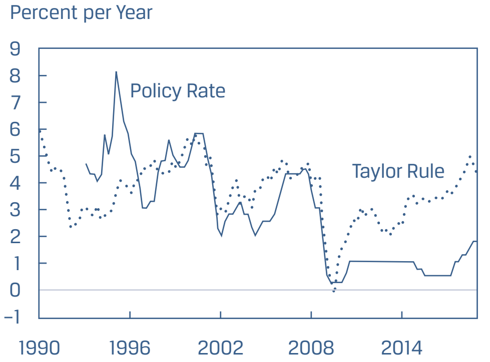
Using fairly conservative parameters, including inflation targets when they are known, and a measure of the output gap estimated by the OECD, we have calculated policy rates based on the Taylor rule for three developed economies back to 1990, as shown in Exhibit 6. The policy rates based on the Taylor rule for the United States, shown in Panel A, seem to track the Fed's actual policy rates fairly closely until the collapse of the high-tech bubble in the early 2000s. According to the Taylor rule, the Fed kept policy rates "too low for too long" between 2002 and 2005. A similar picture emerges for Canada, as shown in Panel B. There is less evidence that policy rates were kept "too low for too long" after the collapse of the high-tech bubble in the United Kingdom, as shown in Panel C. More recently, in response to the liquidity and credit crisis, all three central banks cut their policy rates sharply. According to the Taylor rule, for all three economies, policy rates were "too low" by the end of 2018.

Exhibit 6 Policy Rates and Taylor Rule Calculations

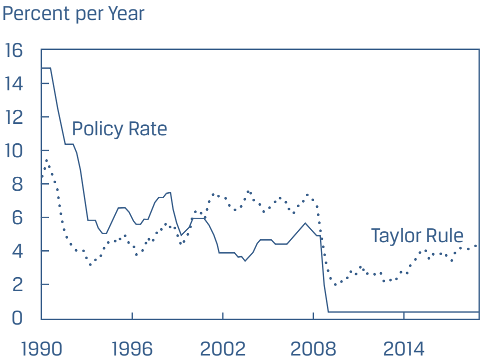
A. United States



B. Canada



C. United Kingdom



Sources: Based on data from each country's central bank and the authors' calculations.

The notion that short-term interest rates can be at too high or too low a level implies that the relationship between short-term interest rates and the business cycle are interdependent. In other words, instead of moderating the business cycle over time with deft changes to the policy rate, central bankers can exaggerate any cycle by not responding optimally to economic conditions—that is, by committing policy errors. For example, setting rates too low for too long risks creating a credit bubble, whereas setting them too high for too long could lead to recessionary or even depression-like economic conditions.

10.1 Short-Term Interest Rate Summary

Short-term default-free interest rates tend to be very heavily influenced by the inflation environment and inflation expectations over time. But they will also be influenced by real economic activity, which, in turn, is influenced by the saving and investment decisions of households. But these interest rates will also be affected by the central bank's policy rate, which, in turn, should fluctuate around the neutral policy rate as central banks respond over time to deviations in inflation from a preferred or target rate and to developments in the output gap. Finally, it is important to remember that the neutral rate will also vary with the level of real economic growth and with the expected volatility of that growth. In addition, the neutral rate might also change if the level of inflation targeted or preferred by the central bank changes. Between 1992 and 1997, the UK inflation target was between 2% and 4%; in 1997, the target became 2.5% with a 1 percentage point allowance around this target. The target was changed in 2003 to 2.0% and to a different definition of inflation, with a 1 percentage point allowance around this target.

In the next section of this reading, we will focus on the relationship between the underlying economy and longer-term nominal default-free government bonds.

11 CONVENTIONAL GOVERNMENT BONDS AND BREAK-EVEN INFLATION RATES

e describe the factors that affect yield spreads between non-inflation-adjusted and inflation-indexed bonds;

The pricing equation shown in Equation 10 can be used to highlight the key components that go into pricing conventional (coupon-paying) government bonds. We will consider the impact that time to maturity can have on the pricing formula later, but first we will focus on the impact of inflation expectations on conventional bond prices.

11.1 Break-Even Inflation Rates

The fundamental difference between the pricing formula as applied to, for example, a three-month T-bill (as in Equation 11) and its application to, for example, a default-free zero-coupon bond (as in Equation 10) relates to their investment horizons. The relative certainty about the real payoff from a three-month T-Bill and thus the relative certainty about the amount of consumption that the investor will be able to undertake with the payoff means that the investment in the T-Bill will be a good hedge against possible bad consumption outcomes. In other words, the payoff, in real terms, from a three-month T-bill is highly unlikely to fall if the investor loses his or her job during the T-bill's three-month investment horizon. The low, probably zero, correlation between the T-bill's payoff with bad consumption outcomes will mean that the risk premium needed to tempt an investor to invest in the T-bill will be close to zero (hence Equation 11).

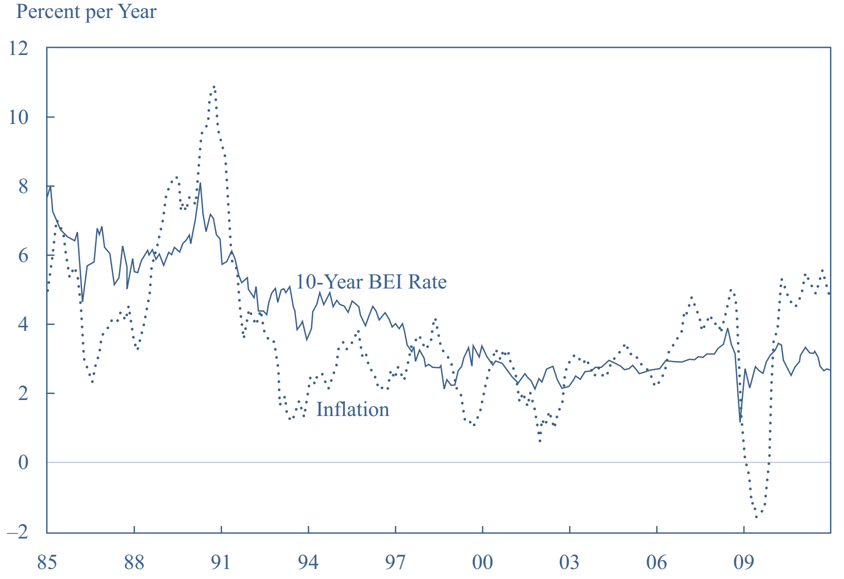
However, it is unlikely that the same level of certainty would apply, for example, to a 20-year default-free conventional government bond. For such a bond, it would seem reasonable to assume that the risk premium would be higher than that related to a one- or three-month T-bill. Note that the cash flow in Equation 10 is still certain, but only in nominal terms. Because investors will naturally have less confidence in their ability to form views about future inflation over 20 years relative to their ability to form those views over three months, the greater uncertainty about the real value of the bond's payoff will cause investors to demand a premium in compensation for this uncertainty, represented by $\pi_{t,s}$ in Equation 10.

The difference between the yield on, for example, a zero-coupon default-free nominal bond and on a zero-coupon default-free real bond of the same maturity is known as the break-even inflation (BEI) rate. It should be clear from the discussion earlier that this break-even inflation rate will incorporate the inflation expectations of investors over the investment horizon of the two bonds, $\theta_{t,s}$, plus a risk premium that will be required by investors to compensate them predominantly for uncertainty about future inflation, $\pi_{t,s}$. Although the evolution of real zero-coupon default-free yields over time should be driven mainly by the inter-temporal rate of substitution, the evolution of their nominal equivalents will, in addition, be driven by changing expectations about inflation and changing perceptions about the uncertainty of the future inflation environment. We can see this evolution by plotting the constant maturity zero-coupon break-even inflation rates over time.

Panels A, B, and C of Exhibit 7 show the 10-year break-even inflation rates derived from three government bond markets in developed economies where index-linked government bonds have been available for some time now—Australia, the United Kingdom, and the United States—along with the respective inflation rates of each economy. The UK and Australian data, which are available for longer historical periods, show the gradual decline in break-even inflation rates since the mid-1980s. This decline was probably driven by the changing inflation environment in these economies. Between 1985 and 1990, inflation averaged approximately 6.0% and 7.5% in the United Kingdom and Australia, respectively. Between 2000 and 2011, having fallen steadily during the 1990s, inflation averaged 3.0% and 3.2% in the United Kingdom and Australia, respectively. Ten-year break-even rates in the United States were only available starting in 1997, a period when US inflation was relatively low and stable. Panel D of Exhibit 7 highlights the impact of the liquidity and credit crisis of 2008–2009 on break-even rates for a range of economies. It shows that for all of these developed economies, 10-year break-even inflation rates fell in response to the weaker global economic environment and weaker inflationary backdrop. The weaker inflationary pressure arises from the lower demand for resources in an economic downturn, so that cost and prices do not rise as fast. For example, 10-year Italian break-even inflation rates fell from 2.3% to 0.8%, reflecting the effect of the eurozone crisis on the Italian economy at that time.

Exhibit 7 Break-Even Inflation Rates and Inflation

A. United Kingdom



B. Australia

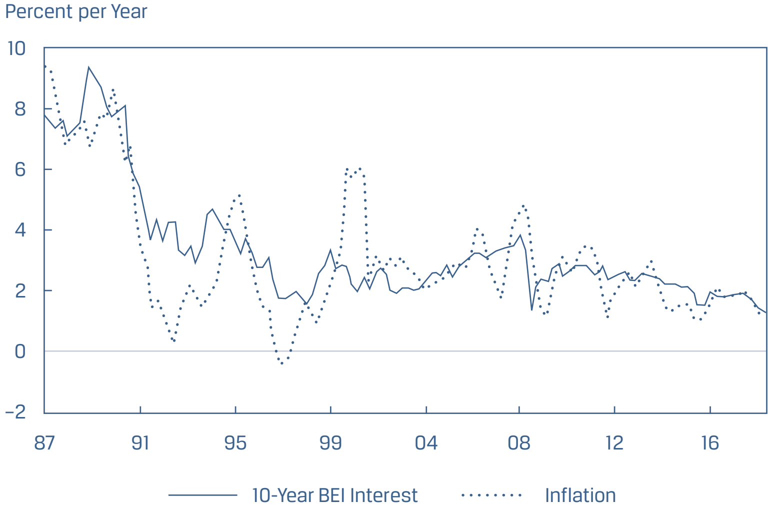
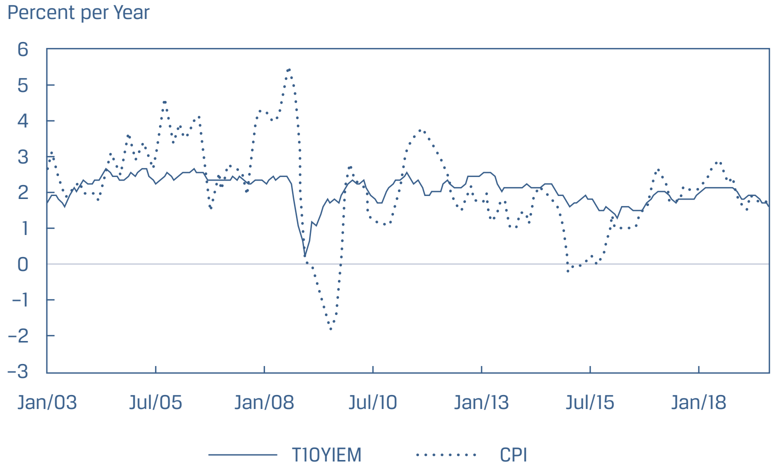
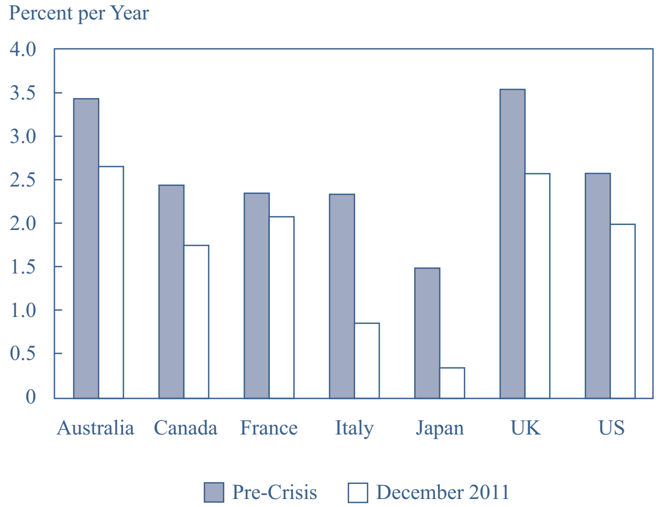


Exhibit 7 (Continued)

C. United States



D. Changes in BEI Rates



Sources: Based on data from Bloomberg and the authors' calculations.

Being able to measure financial market expectations of future inflation is of great value to central banks. Break-even inflation rates provide an independent view about future inflation that can be compared with the judgment of the central bank; although, of course, this judgment can be interdependent. However, it is important to remember that break-even inflation rates are not simply the markets' best guess of future inflation over the relevant investment horizon. Break-even inflation rates will also include a risk premium to compensate investors for their uncertainty largely about future inflation and, therefore, the uncertainty about the quantity of goods and services that they will be able to consume in the future.

12 THE DEFAULT-FREE YIELD CURVE AND THE BUSINESS CYCLE

d explain how the phase of the business cycle affects policy and short-term interest rates, the slope of the term structure of interest rates, and the relative performance of bonds of differing maturities;

So far we have discussed the fundamental pricing relationship for default-free real and nominal bonds and short-term nominal interest rates. We now elaborate on these relationships over different investment horizons. We have already indicated that the maturity of a bond will have an impact on the way that investors price it. We now focus on this relationship more specifically. But first consider Panel A in Exhibit 8, which shows the US zero-coupon Treasury curve on three different dates. From July 2007 (just prior to the wider financial crisis) to the end of 2011, the US Treasury curve shifted down by between 3 and 4 percentage points and also became steeper. The short end of the curve was clearly influenced by the reduction in the Fed's policy rate over this period, which fell from 5.25% to virtually 0%. Panel B shows that there was a similar decline in the short end of the gilt curve, as the UK central bank gradually cut its policy rate from 5.75% to 0.50% in response to the same crisis.

Panel A shows that the Treasury curve was upward sloping on each of these dates. Panel B shows that by the end of 2011 the UK government and US government curves looked very similar, but what is interesting is that the UK government curve was downward sloping in July 2007. This slope meant that the UK government could borrow 1-year money at 6.25% but 30-year money at 4.8%. In fact, on the same date, the UK government could borrow 50-year money at just over 3.0%. What economic factors could explain not only the fall in Treasury and gilt yields, as well as those elsewhere in the developed world, over this period but also the very negative slope of the gilt curve in the summer of 2007?

Exhibit 8 US and UK Government Bond Yields and Break-Even Inflation Rates for July 2007, December 2010, and December 2011

A. US Treasury Curve

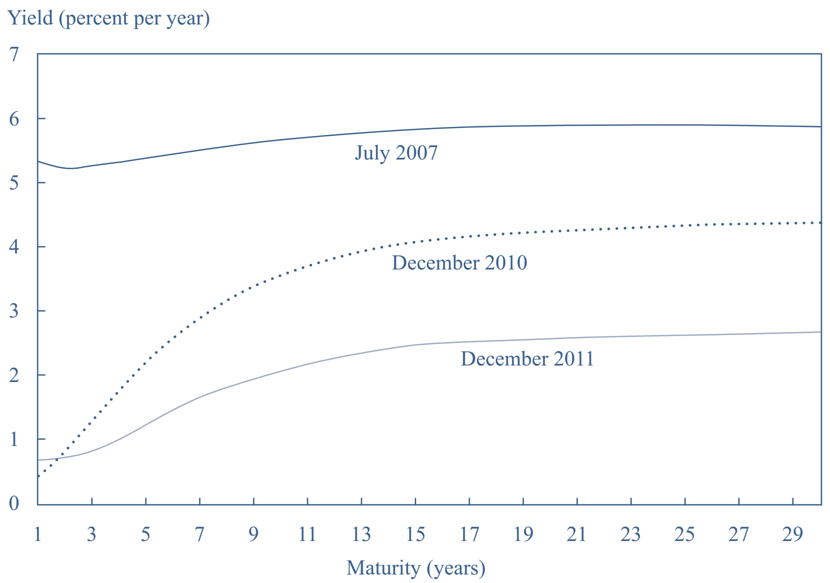
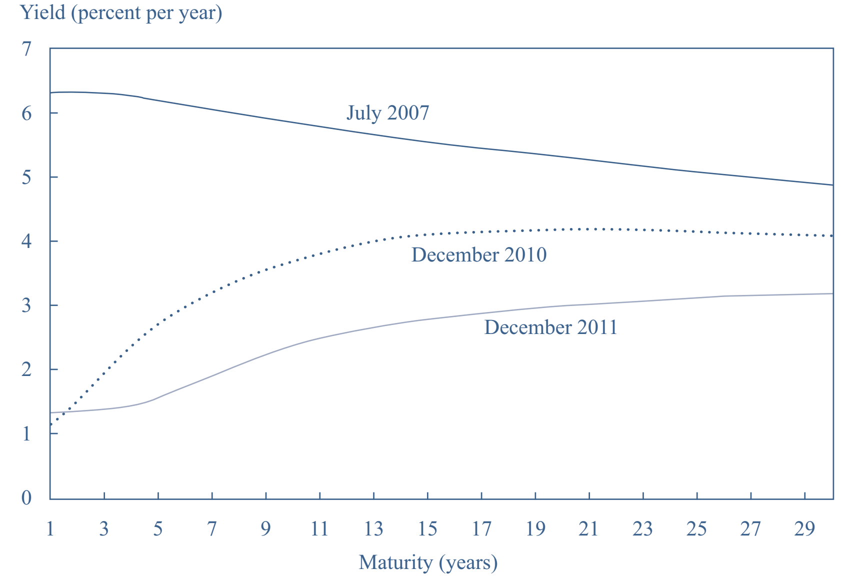


Exhibit 8 (Continued)

B. UK Gilt Curve

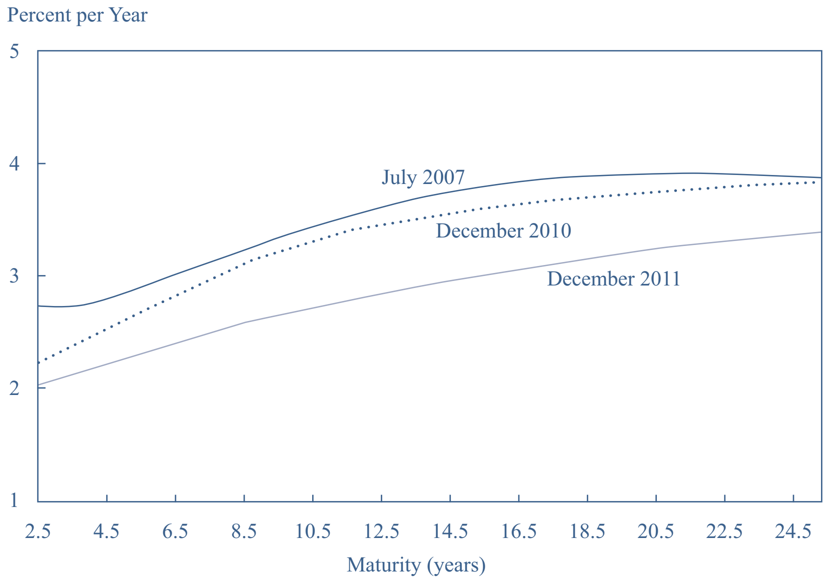


C. US BEI Rates

(continued)

Exhibit 8 (Continued)

D. UK BEI Rates



Sources: Based on data from the Bank of England and Thomson Reuters.

Panels A and B of Exhibit 8 show how the shape of the default-free yield curves shifted in response to the financial crisis precipitated by the collapse of Lehman Brothers. Panels C and D of the exhibit show how break-even inflation rates shifted over the same period. When there is a significant shift in the yield curve, it is often very informative to break down the change into the real and inflationary components. Over this period, the US and UK break-even curves both shifted by just less than a full percentage point across the entire maturity spectrum. This fact could be taken to indicate that as the global recession gathered pace, US and UK government bond investors gradually "priced in" lower and lower levels of future inflation. But crucially, Panels A and B show that the nominal curves shifted down by more than 1 percentage point. For example, the US curve fell by as much as 3 percentage points. These facts suggest that market participants saw the financial crisis as potentially having a bigger impact on economic growth than on inflation. Indeed, by December 2011 the US and UK break-even inflation curves were both upward sloping.

This dissection of default-free yield curves and the interpretations that are often made by analysts based on the relative movements of the real and break-even components can be very informative, but the analysis presupposes either that there is no risk premium embedded in investors' return expectations or that any risk premium is constant over time. But the risk premium is unlikely to be zero or constant over time.

EXAMPLE 12

Level, Slope, and Curvature of the Yield Curve

The yield curves shown in Exhibit 8 all have three distinct characteristics. These characteristics are referred to as *level*, which indicates whether rates are high or low, on average; *slope*, which is an indicator of the steepness of the curve, or how quickly or slowly rates change with maturity; and *curvature*, an indicator of how much the curve is different from a straight line. These characteristics were first noted in rates by Steeley in 1990 and Litterman and Scheinkman in 1991. All three components can change over time. We can calculate how much of the change in any yield curve is attributable to each of these factors over time. We have performed this calculation for the UK and US yield curves using data with maturities ranging from 3 months to 10 years, spanning the period from January 1999 to January 2014. The principal components analysis technique was used to perform the analysis. These results are shown in Exhibit 9. The majority of movement of the US (92.7%) and UK (95.2%) yield curves is in the level—that is, shifts up and down in the yield curve over time. The slope component accounts for a much smaller proportion of the change over time—6.9% for the US curve and 4.5% for the UK curve. Curvature changes account for only 0.3% in each case (note that we have rounded these figures to the nearest decimal point for convenience). Taken together then, changes in these three factors explain the vast majority of changes to these two yield curves over time. These results are typical for government bond yield curves in developed economies.

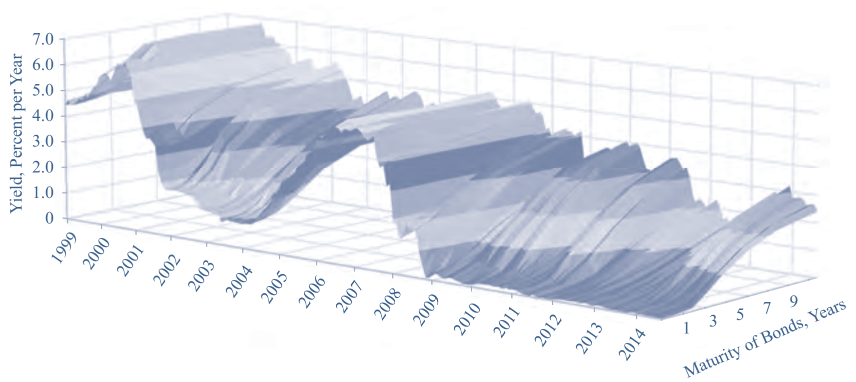
Exhibit 9 Percentage of Yield Curve Movement Explained by Three Components

	Level	Slope	Curvature
United States	92.7%	6.9%	0.3%
United Kingdom	95.2%	4.5%	0.3%

Sources: Based on data from Bloomberg and the authors' calculations.

Exhibit 10 presents another way of viewing the dynamics of the US yield curve for 15 years starting in 1999. The chart shows the way in which the level, slope, and curvature of the US Treasury curve has changed over time.

Exhibit 10 The US Yield Curve, 1999–2014



We would expect the level of economic activity to influence yield curve levels. We would also expect that views of future inflation will determine the level of these yield curves because these are nominally dominated bonds. The slope of the yield curve will be influenced by the magnitude of the risk premium in

Equation 5 between the price of the bond and the inter-temporal rate of substitution over the investor's time horizon. A positive slope would be a reflection of this risk premium. However, it is not the only variable that affects the slope of the yield curve. The policy rate of the central bank is set based on Taylor rule–like considerations (that is, consideration of the components that make up the rule) so that short rates will tend to be lower during recessions because central banks tend to lower their policy rate in these times. (Note that no suggestion is made that central banks slavishly follow the Taylor rule; the rule just neatly encapsulates two of the key macroeconomic considerations that go into the process of setting the interest rate.) But the impact of monetary policy on longer-term rates will not be as strong because the central bank will usually be expected to bring short-term rates back to normal as the recession recedes and the risk-free rates will increase as economic growth recovers. Thus, the slope of the yield curve will increase during the recession. Finally, if investors anticipate that policy rates as well as short-term risk-free rates will revert back to normal as the recession recedes, then the yield curve will become steeper for the short-term maturities but flatter for the long-term maturities so that the curvature can increase as well. As a result, the shape of the yield curve and its three factors can provide valuable information for both central banks and investors.

13 THE SLOPE OF THE YIELD CURVE AND THE TERM SPREAD

d explain how the phase of the business cycle affects policy and short-term interest rates, the slope of the term structure of interest rates, and the relative performance of bonds of differing maturities;

The required return on future default-risk-free cash flow was explained as consisting of a real interest rate, a premium for expected inflation, and a risk premium demanded by risk-averse investors for the uncertainty about what inflation will actually be (see Equation 10). Thus, referring to government yield curves, expectations of increasing or decreasing short-term interest rates might be connected to expectations related to future inflation rates and/or the maturity structure of inflation risk premiums.

Expectations of declining short-term interest rates can explain the downward-sloping UK gilt curve in the summer of 2007. If bond market participants expect interest rates to decline, then reinvestment of the principal amounts of maturing short-term bonds at declining interest rates would offset the initial yield advantage of the shorter-dated bonds. These expectations caused the United Kingdom's yield curve to be downward sloping or inverted.

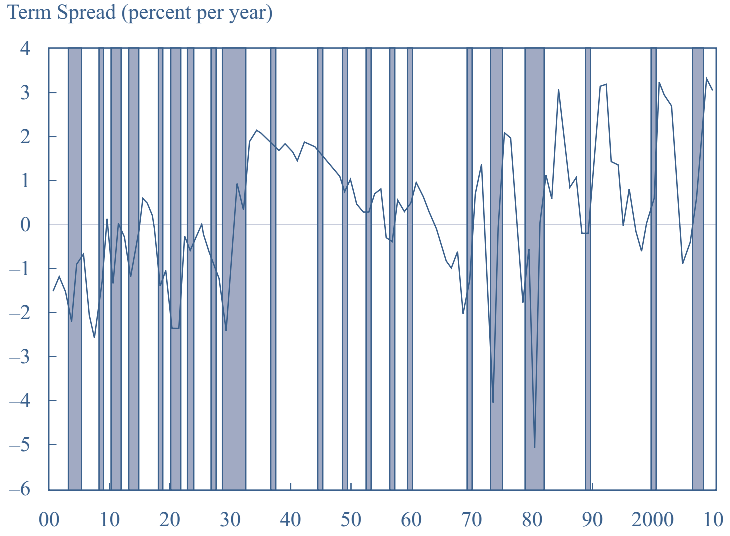
Thus, the variation in short rates over time—in particular, the central bank's policy rate—can influence the shape of the yield curve. These short rates are, in turn, driven by the positive relationship between the real rate of interest that balances investment and saving decisions over time and by the level of and volatility in GDP growth, as well as by the variation of the rate of inflation around the central bank's target, or preferred, level.

13.1 The Term Spread and the Business Cycle

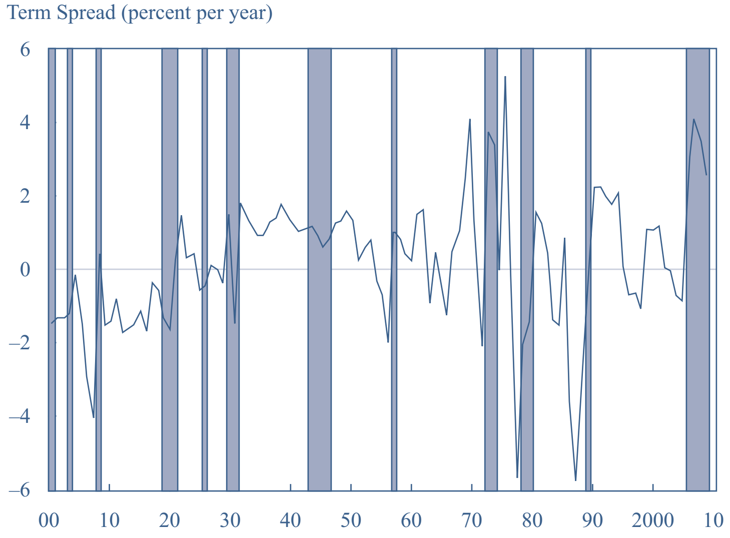
Exhibit 11 shows the time variation in the slope of the US and UK government yield curves since 1900. In both cases, the slopes have been calculated as the difference (spread) between a long-dated government bond and the yield on an equivalent one-year bond.

Exhibit 11 US and UK Government Yield Curve Spreads

A. US Treasury



B. UK Gilt



Note: Shaded areas indicate recessions.
Sources: Based on data from NBER and the Bank of England.

In both markets, there are times when the curves were very steep. For example, in the mid-1970s the steep slope implied expectations of a sharp increase in interest rates. This was a time when both inflation and inflation expectations were high, following the first oil shock of 1973.

But there are times when both curves are steeply inverted—for example, in 1979–1980. The inverted curves in these times implied an expectation of sharply falling inflation and future interest rates. In both economies, the nominal policy rates were extremely high: Policy rates peaked at 17.5% in the United States in December 1980 and at 17% in the United Kingdom in November 1979. During this period, the Fed chairman, Paul Volcker, and Margaret Thatcher had raised policy rates in their respective economies in response to the second oil shock of 1979 and administrations on both sides of the Atlantic came to the conclusion that the defeat of inflation should be the number one policy objective. The inverted curves in both markets suggest that investors expected rates to come down once the causes of high current inflation had been removed. Generally speaking, Exhibit 11 also reveals that a recession is often preceded by a flattening, or even an inversion, in the yield curve. In general, the late stages of a business expansion are often characterized by a peak in inflation and thus relatively high short-term interest rates. If longer-maturity yields reflect lower inflation rates and diminished business credit demand, the yield curve would tend to flatten or invert. An inverted yield curve, in particular, is often read as being a predictor of recession.

EXAMPLE 13

Interest Rates, the Yield Curve, and the Business Cycle

1 What financial instrument is best suited to the study of the relationship of real interest rates with the business cycle?
   - A Default-free nominal bonds
   - B Investment-grade corporate bonds
   - C Default-free inflation-indexed bonds

2 Suppose investors forecast an unanticipated increase in real GDP growth and the volatility of GDP growth for a particular country. The effect of such a forecast would be for the coupon payments of an inflation-indexed bond issued by the government of the country:
   - A to rise.
   - B to fall.
   - C to be indeterminate.

3 The yield spread between non-inflation-adjusted and inflation-indexed bonds of the same maturity is affected by:
   - A a risk premium for future inflation uncertainty only.
   - B investors' inflation expectations over the remaining maturity of the bonds.
   - C both a risk premium for future inflation uncertainty and investors' inflation expectations over the remaining maturity of the bonds.

4 State an economic reason why inverted yield may predict a recession.

Solution to 1

C is correct. These bonds' prices are sensitive to changes in real interest rates because the payments are adjusted for changes in the price of goods.

Solution to 2

A is correct. The coupon payments would be expected to increase, reflecting an increase in the real interest rate.

Solution to 3

C is correct. The difference between the yield on a zero-coupon default-free nominal bond (such as a US government STRIP) and on a zero-coupon default-free real bond of the same maturity (such as a US government TIPS) is called the break-even inflation rate. The break-even inflation rate should incorporate investors' inflation expectations over the remaining maturity plus a risk premium for uncertainty about future inflation, as in Equation 10.

Solution to 4

The late stages of business cycles are often characterized by relatively high inflation and high short-term interest rates. To the extent that longer-term yields reflect expectations of declining inflation and a slackening in demand for credit, the yield curve would be expected to flatten or invert.

14 EVIDENCE ON RISK PREMIUMS FOR DEFAULT-FREE BONDS

d explain how the phase of the business cycle affects policy and short-term interest rates, the slope of the term structure of interest rates, and the relative performance of bonds of differing maturities;

If, as seems likely, most investors want to be compensated for taking on risk, then the yield curve, as well as containing information about the interest rate expectations of investors, will also embody a risk premium.

We have already explained why investors would value investments that paid off more in bad times relative to those investments that paid off less in these times or produced negative returns. This preference tends to drive the expected return down and the price of these favored investments up relative to those with prices that are more positively correlated with bad times. The average slopes of the US and UK government curves from 1900 to 2011 were 0.24% and 0.14%, respectively; in the post-1945 period, they were 0.50% and 0.40%, respectively. This difference suggests that, on average, investors have been willing to pay a premium for shorter-dated US and UK government bonds, which, in turn, means that longer-dated bonds may not be such a good hedge against economic bad times. One interpretation of an upward-sloping yield curve is that short-dated bonds are less positively (or more negatively) correlated with bad times than are long-dated bonds.

Exhibit 12 presents information on the relationship of government bonds with a range of maturities from a selection of countries. Panel A shows that the average yield differences (longer minus shorter) between different bond maturities, with one exception, are all positive. This fact suggests that the bond risk premium generally rises with maturity, which is why it is often referred to as the term premium. Panel B presents the total return on these government bonds by maturity. Over the sample periods, the total returns achieved generally rise with maturity in each of the bond markets. But why have government bond investors generally been rewarded for holding longer-dated government bonds relative to shorter-dated bonds?

In Panel C, we present the correlation between (1) the total return on bonds with various maturities and (2) the economic growth of the relevant economy. One thing to notice is that the correlations are predominantly negative. This fact suggests that government bonds in these markets tend to pay off in bad times, which means that investors are willing to pay a relatively high price for them. Therefore, investors should be willing to accept a relatively low return from government bonds because they are at least a partial hedge against "bad" consumption outcomes.

Exhibit 12 Government Bond Spreads, Total Returns, and GDP Growth Correlations for Four Markets

Panel A: Spreads

	5 vs. 2	10 vs. 5	30 vs. 10
Canada	0.40	0.40	
France	0.46	0.52	0.60
UK	0.27	0.26	−0.05
US	0.55	0.44	0.39

Panel B: Total Returns

	2	5	10	30
Canada	5.61%	6.70%	8.08%	9.60%
France	4.44%	6.09%	7.80%	10.24%
UK	7.14%	7.82%	10.12%	11.03%
US	5.77%	6.91%	7.72%	9.07%

Panel C: Correlation with GDP

	2	5	10	30
Canada	−4.93%	−5.12%	−0.70%	2.75%
France	4.71%	−9.14%	−10.27%	−2.13%
UK	−4.66%	−2.74%	−5.11%	0.79%
US	−7.30%	−14.33%	−12.31%	−10.35%

Note: The sample period for the United Kingdom and the United States is January 1980 to December 2018. For France and Canada, it is January 1985 to December 2018.
Sources: Thomson Reuters and authors' calculations.

The results shown in Exhibit 12 suggest that government bond risk premiums

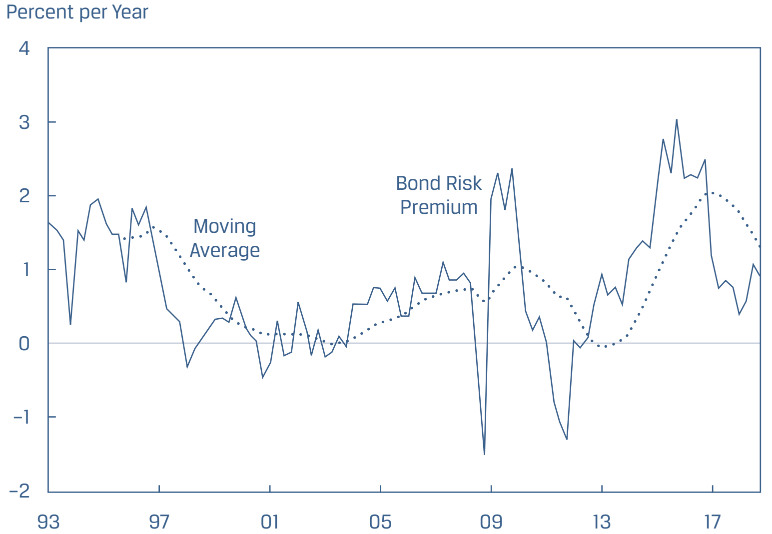
- are positive,
- are probably related to the consumption hedging benefits of government bonds, and
- are positively related to bond maturity, which means that the "normal" shape for the yield curve is upward sloping.

The last point also helps to explain why the US Treasury curve was generally upward sloping between the summer of 2007 and the end of 2011: A significant portion of the slope was probably related to the existence of a positive risk premium on US Treasuries that increased with maturity.

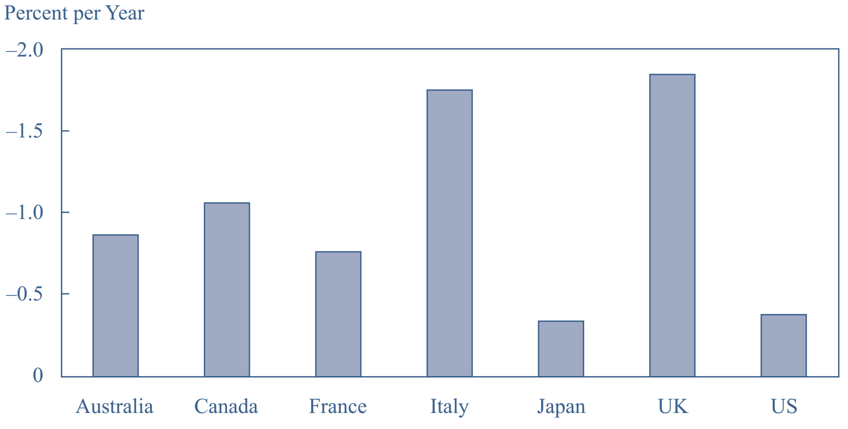
However, bond risk premiums ($\pi_{t,s}$), like other risk premiums, will not be constant over time. In times of economic uncertainty, investors will tend to more highly value assets that pay off in bad times—government bonds—which will force their prices up as the risk premium demanded falls. Unfortunately, it is impossible to say how big or small this premium really is or should be or how it evolves with the business cycle. But we can get some idea by performing the following experiment. Suppose we can assume that the real return on an index-linked government bond is a good proxy for the real rate of interest that balances savings and investment in an economy over time. If we subtract the yield on an index-linked government bond from the yield on a conventional government bond with a similar maturity and then subtract from this amount a survey-based measure of inflation expectations, what is left over is the bond risk premium. That is, the extra yield investors require for holding a conventional government bond over and above the real required return and the return in compensation for expected inflation.

Exhibit 13 Bond Risk Premiums (BRPs)

A. UK Government BRP



B. Change in Government BRP, 2007–2011



Note: The period 1993–1997 in Panel A was before the inflation-targeting regime.

Panel A of Exhibit 13 presents a calculation of this kind for the UK government bond market. There are a number of points worth noting with respect to this particular representation of the bond risk premium. First, it is certainly imperfect because it shows that the bond risk premium is negative at times. A negative bond risk premium implies that investors are willing to pay for assuming risk, which is inconsistent with risk aversion. Second, putting aside these negative values for the moment, we can see that the bond risk premium varies over time. Between 1993 and mid-1997, the risk premium on UK government bonds averages just less than 1.5%; from mid-1997 to 2005, it averages around 0.25%; and from 2013 to 2018, it averages around 1.42%.

Panel B in Exhibit 13 is based on similar estimates for the bond risk premiums over the global financial crisis for a range of markets and shows the change in the bond risk premiums from July 2007 to December 2011. For all the government bond markets for which we could make these calculations, the bond risk premium fell—in some cases, quite substantially. This fact implies that investors placed greater value on the consumption-hedging properties of government bonds as a result of the financial crisis. Other things being equal, this implies that they were willing to pay a higher price for these bonds compared with the price they were willing to pay before the crisis. This also implies that bond risk premiums will tend to rise in times when investors place less value on the consumption hedging properties of government bonds.

15 OTHER FACTORS

d explain how the phase of the business cycle affects policy and short-term interest rates, the slope of the term structure of interest rates, and the relative performance of bonds of differing maturities;

In practice, the shape of the yield curve and the relative performance of bonds with different maturities over the business cycle depend on a complex mixture of interest rate expectations and risk premium considerations. For example, a downward-sloping curve is probably largely the result of investor expectations of future declines in interest rates. The drivers of an upward-sloping curve are more ambiguous. For example, the existence of bond risk premiums that are positively related to maturity means that an upward-sloping curve may not embody expectations of future rate increases. Conversely, it could imply a combination of expected rate increases and risk premiums or even expected rate cuts that are more than offset by the existence of positive risk premiums.

We have to acknowledge that there will also be times when other factors play a part in shaping the yield curve. Developed-economy government bonds are technically default-free because these governments can, in principle, always print cash to meet the promised payments. In this sense, they are very special financial instruments. These markets can also be influenced by supply and demand factors that seem to move yields in ways that do not appear to be consistent with the business cycle. Consider the following examples:

- In the late 1990s, fiscal surpluses in the United States led some investors to take the view that the supply of Treasuries would shrink as the US government paid back its debts (leading to the Treasury scare of 2000). This perceived reduction in future supply was said to have been responsible for the decline in yields as investors bought up these bonds in anticipation of their future scarcity.
- In the early 2000s, Treasury yields were apparently being pushed down by Asian central banks that were using their growing trade surpluses to purchase US Treasuries.
- Regulatory factors can also play an important role in determining government bond yields. This influence is particularly clear in the gilt market. In 1997, the UK government passed legislation that effectively compelled UK pension funds to buy long-dated gilts. This legislation appears to have been one of the main drivers of the inversion of the long end of the United Kingdom's yield curve (see Exhibit 8, Panel B). Since that time, new accounting rules for pensions—FRS17, followed by IAS19—forced UK pension schemes to increase their demand for long-dated UK government bonds further still. These actions created a vicious circle because the new accounting rules required the schemes to discount their

liabilities using the long-dated yields as the discount rate. Consequently, UK pension funds bought long-dated bonds, forcing their yields down. The decline in yields caused the present value of pension liabilities to rise relative to the value of scheme assets, creating deficits. To achieve a better match between assets and liabilities, schemes tried to buy more long-dated gilts, causing their yields to fall further and liabilities to rise further. A similar phenomenon affected core eurozone bond and swap markets. For more than three years, the yield spread between 10- and 30-year Dutch government bonds (30 minus 10) was negative. It was argued that the negative spread was a direct result of the hedging activities of Dutch pension funds, which had over €1 trillion of pension liabilities. The buying pressure on a government bond market that was only around 60% the size of these liabilities thus caused long-dated Dutch government bond prices to rise.

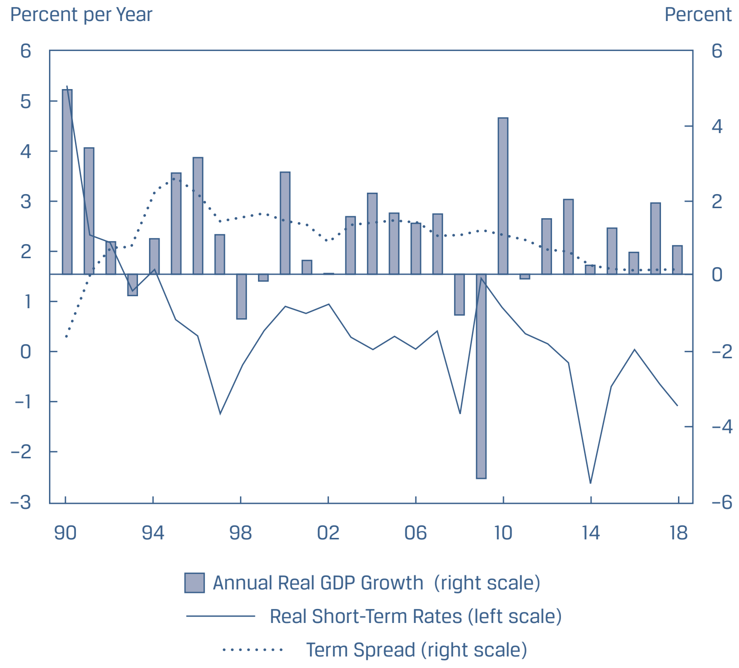
There is no doubt that supply and demand considerations along with poorly thought out regulatory or accounting rules can have an impact on government bond markets. But determining the extent of these effects is very difficult. Nevertheless, it is difficult to explain the very inverted shape of the long end of the gilt curve, which has persisted for many years now, without reference to such factors.

EXAMPLE 14

The Japanese Yield Curve and Business Cycle, 1990–2018

During the 1980s, real Japanese GDP growth averaged more than 4.60% per year. During the 1990s and 2000s, it averaged around 1.5%. The catalyst for the decline in growth was the collapse of Japan's property bubble and the stock crash in the early 1990s. Exhibit 14 shows Japan's real annual GDP growth over the post-bubble period, which incorporates the global financial crisis of 2008–2009. The exhibit also shows how real short-term policy rates fell steadily from a level of 4%–5% in 1990, finally becoming negative by 1997. Since that time, real rates have generally been negative as Japan's central bank, the Bank of Japan, has tried to stimulate its economy by cutting its nominal policy rate to (near) zero.

Exhibit 14 Japan's GDP Growth, Real Short-Term Rates, and Term Spread, 1990–2018



Sources: Based on data from Thomson Reuters and the authors' calculations.

Exhibit 14 also shows the term spread over this period—that is, the slope of the yield curve (created by subtracting the policy rate from 10-year Japanese government bonds). At the start of the period, as the Bank of Japan cut its policy rate from just over 8.0% at the end of 1990 to 0.45% by the end of 1995, the curve steepened—from −1.63% at the end of 1990 to 2.59% by the end of 1995. The sharp steepening of the curve in response to the cuts in the policy rate arguably indicated the market's view that the policy stimulus would work; that is, the yields on longer-dated Japanese government bonds embedded some expectation of positive growth and inflation in the future. However, since the mid-1990s, despite cutting its policy rate to zero and enacting programs of quantitative easing, the term spread has fallen steadily. By the end of 2018, it was virtually 0.00%. The weak economic growth and inflation environment in Japan since the collapse of the property bubble is reflected in the fall and flattening of the Japanese yield curve.

16 CREDIT PREMIUMS AND THE BUSINESS CYCLE

f explain how the phase of the business cycle affects credit spreads and the performance of credit-sensitive fixed-income instruments;

Earlier we discussed the economic drivers of what we have referred to as default-free interest rates and bond yields. But the financial crisis has caused many to question what "default-free" really means. The bonds issued by many European governments, including those issued by France and Italy as well as those issued by the Greek, Portuguese, Irish, Belgian, and Spanish governments, were all thought to be default-free before the euro financial crisis from 2010 to 2012. Even the default-free status of US Treasuries has now been questioned by both investors and credit rating agencies. Any bond that is perceived to be default-free will, by definition, not have to compensate investors for taking on default, or credit, risk. However, for any corporate or government bond that embodies the non-zero probability that the issuer may default on its obligations, bondholders will demand a risk premium, referred to as the credit premium.

In Equation 1, we emphasized that the discount rate on the cash flows of financial assets will normally include a risk premium, which we defined generically as $\rho_{t,s}^{i}$. When we considered the pricing formula for default-free government bonds, we emphasized that the risk premium attached to these bonds, $\pi_{t,s}$, is largely a function of uncertainty about future inflation and that this uncertainty is likely to be greater for longer-dated bonds relative to shorter-dated bonds (recall that this risk premium, $\pi_{t,s}$, is distinct from the addition to return that investors require on the basis of their expectations of inflation over the investment horizon, $\theta_{t,s}$, per Equation 1). But it is important to emphasize that these factors will play a role in the pricing of bonds that embody credit risk, too. In other words, the evolution of $\pi_{t,s}$ plays a role in determining the price of credit risky bonds too. We thus need to acknowledge the separate role that credit risk plays in the price of a corporate bond or, indeed, any bond with default risk. We have therefore adapted our basic pricing Equation 1 to augment the discount rate with a credit premium, $\gamma_{t,s}^{i}$, which is distinct from the inflation-based risk premium attached to default-free bonds, $\pi_{t,s}$, as shown in Equation 13.

$$P_t^i = \sum_{s=1}^{N} \frac{E_t\left[\widetilde{CF}_{t+s}^{i}\right]}{\left(1 + l_{t,s} + \theta_{t,s} + \pi_{t,s} + \gamma_{t,s}^{i}\right)^s} \tag{13}$$

Notice that Equation 13 acknowledges the uncertain nature of the cash flows, $CF_{t,s}$, on credit risky bonds; although the schedule of these payments is known, the existence of credit risk means that there will be uncertainty about whether they will be paid as scheduled. In the event of a default, the amount that the bond investor receives will depend on the recovery rate, which will also be an unknown quantity. The risk premium demanded by investors because of these uncertainties is represented by $\gamma_{t,s}^{i}$ in Equation 13.

In this section, we will focus on the credit premium, $\gamma_{t,s}^{i}$, and in particular on the relationship between the business cycle and the credit premium on corporate bonds.

EXAMPLE 15

The Credit Risk Premium

Suppose that an analyst estimates that the real risk-free rate is 1.25%, average inflation over the next year will be 2.5%, and the premium required by investors for inflation uncertainty is 0.50%. If the analyst observes the price of a corporate bond with a face value of £100, with one full year to maturity, as being equal to £94.21, what would be the implied credit premium embedded in the bond's price for inflation uncertainty?

Solution

The (approximate) implied premium can be calculated as follows:

$$\gamma_{t,s}^{i} = 1.90\% = \frac{100}{94.21} - (1 + 1.25\% + 2.50\% + 0.50\%).$$

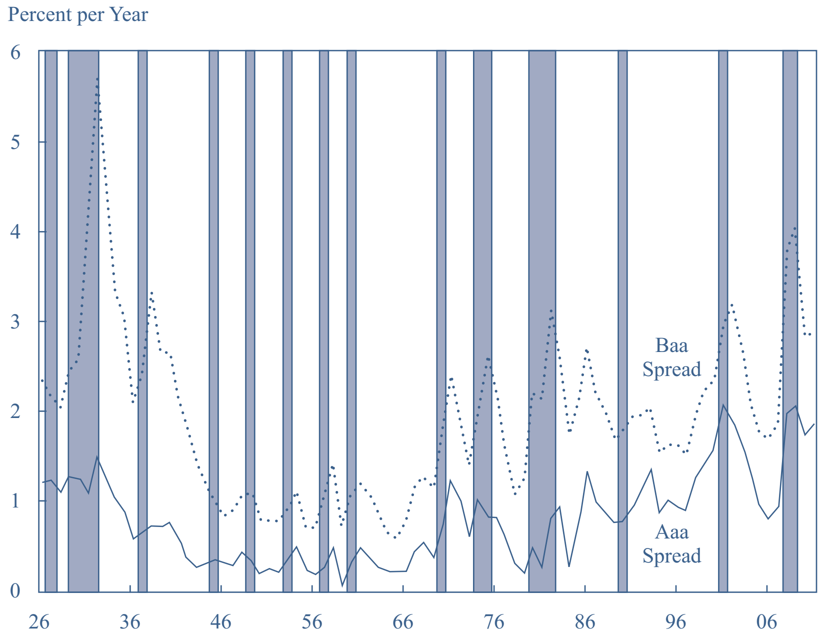
16.1 Credit Spreads and the Credit Risk Premium

The difference between the yield on a corporate bond and that on a government bond with the same currency denomination and maturity is generally referred to as the credit spread. It is demanded by investors in compensation for the additional credit risk that they bear compared with that embodied in the government bond.

As Equation 13 shows, credit risky bonds share the same risk as default-free bonds, which market participants often refer to as interest rate risk, but they also embody credit risk, $\gamma_{t,s}^{i}$. Other things being equal, a parallel shift up in the yield curve will have an almost identical proportionate impact on the prices of, say, a five-year government bond and a five-year corporate bond. And over time, again other things being equal, the interest rate component of a corporate bond will be driven by the same factors that drive government bond yields and returns. In other words, they are both subject to interest rate risk.

It is the credit risk component of a corporate bond, $\gamma_{t,s}^{i}$, and the evolution of bond spreads that will cause corporate and comparable government bond returns to diverge over time. It would seem sensible to assume that the premium demanded would tend to rise in times of economic weakness, when the probability of a corporate default and bankruptcy is highest. Exhibit 15 confirms this view. The exhibit shows a representative spread on both AAA/Aaa and BBB/Baa rated US corporate bonds over US Treasuries. (The AAA rating category in "AAA/Aaa" is the rating category used by both Standard & Poor's and Fitch Ratings; Aaa is the equivalent rating category used by Moody's Investors Service.) First, the Baa spread is always higher than the Aaa spread, reflecting the lower credit quality of Baa rated bonds relative to Aaa rated bonds. Second, the US recession periods shaded grey in the chart indicate that both low- and higher-grade corporate bond spreads do tend to rise in the lead-up to and during a recession and to decline once the economy comes out of recession.

Exhibit 15 Credit Spreads and the Business Cycle



Note: Shaded areas indicate recessions.
Source: Based on data from Moody's Investors Service.

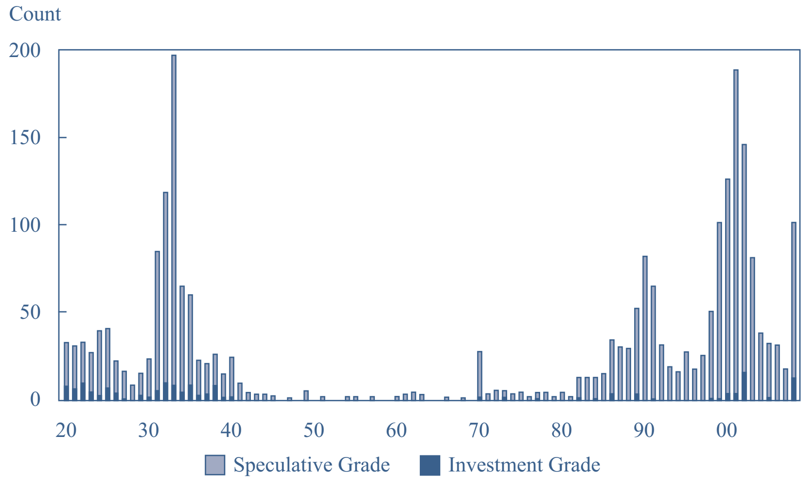
As expected, the business cycle has a profound effect on credit spreads, but what are the basic components of the credit spread? If we assume that investors are risk neutral, then they will simply demand a return (yield) on their corporate bond investments sufficient to compensate them for the possible loss that they could incur from holding a corporate bond. In turn, this expected loss will depend on the probability of default and the expected recovery rate in the event of default, as shown in Equation 14:

$$\text{Expected loss} = \text{Probability of default} \times (1 - \text{Recovery rate}). \tag{14}$$

In the instance where investors are risk neutral, the expected return on, say, a 10-year government bond would be equal to the loss-adjusted expected return on a comparable 10-year corporate bond. In practice, however, investors are risk averse, so the expected return on a corporate bond will be higher than that on a comparable government bond, even if a significant amount of the credit risk can be mitigated by holding a diversified portfolio of corporate bonds. One of the main reasons why investors continue to be exposed to considerable market risk even in a well-diversified portfolio is that defaults tend to cluster around downturns in the business cycle. Panels A and B of Exhibit 16 show this quite clearly. Panel A shows the number of US corporate defaults per year since 1920, and Panel B shows annual default rates over the same period. Both charts show that there are often long periods of time when there are very few defaults. However, the US depression of the 1930s and the recessions in the 1980s, 1990s, and 2000s were all associated with relatively high default levels and rates. The historical default rates on different ratings classes are sometimes used by analysts as a proxy for the probability of default in Equation 14 for expected loss.

Exhibit 16 US Corporate Defaults, Default Rates, Recovery Rates, and Loss Rates

A. Number of Defaults



B. Default Rates

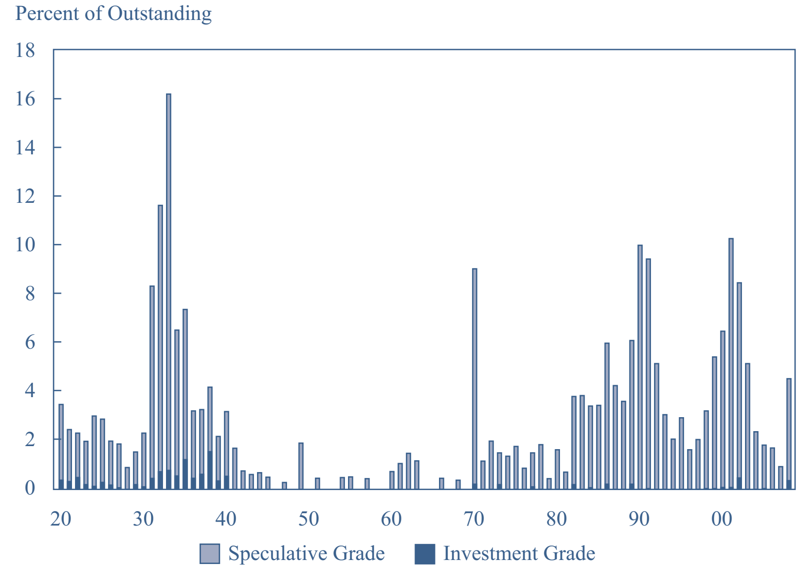
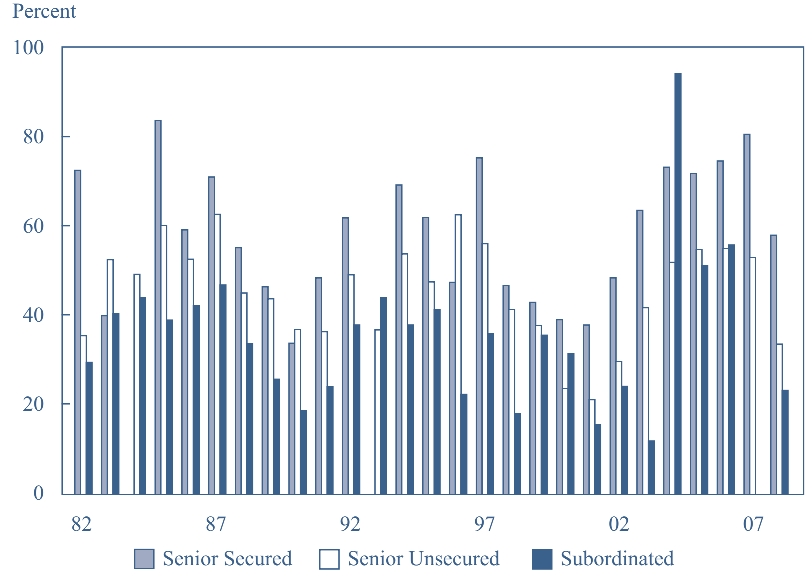
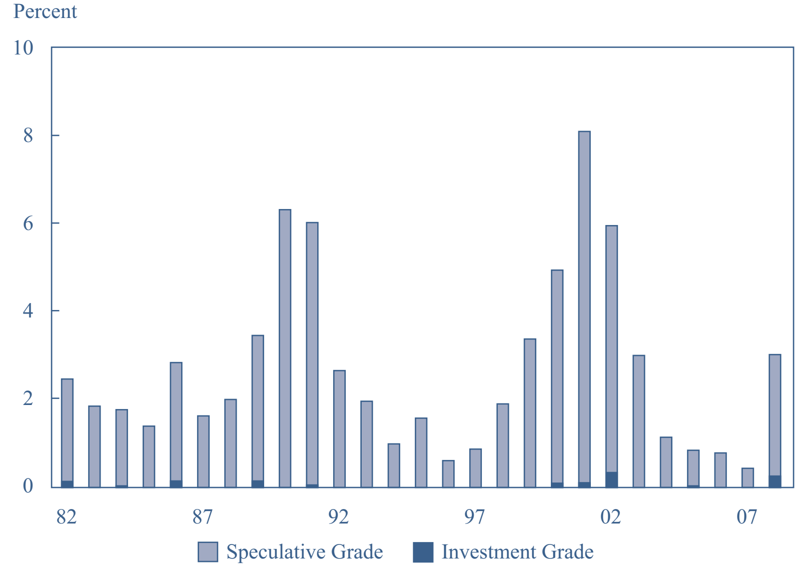


Exhibit 16 (Continued)

C. Recovery Rates



D. Loss Rates



Source: Based on data from Moody's Investors Service.

Panel C of Exhibit 16 shows the evolution of recovery rates on three types of corporate bonds: senior secured, senior unsecured, and subordinated. Senior secured debt, as the name suggests, is secured by a lien or other claim against some or all of the company's assets, whereas senior unsecured debt has no explicit claim to the company's assets in the event of bankruptcy. This explains why recovery rates are generally higher for secured as opposed to unsecured debt holders. Subordinated debt holders, as the name suggests, have an inferior claim on the company's assets compared with senior debt holders, and unsurprisingly, recovery rates are often very low. The recovery rate ranking is interesting, but we can also see from Panel C that recovery rates tend to be higher when the economy is expanding and lower when it is contracting. The reason is that assets that can be sold in order to recover value for bond holders are likely to fetch a higher price in a buoyant economic environment than in a stagnant one. Finally, Panel D shows the loss rates on US corporate debt from 1982 to 2008. These loss rates are the net result of the defaults and recovery rates over time. Unsurprisingly, these loss rates are counter-cyclical with regard to the business cycle, meaning that they tend to rise as economic activity declines.

17 INDUSTRY-SPECIFIC AND COMPANY-SPECIFIC CREDIT QUALITY

g explain how the characteristics of the markets for a company's products affect the company's credit quality;

Although spreads will evolve with the business cycle, Exhibit 17 illustrates that spreads between corporate bond sectors with different ratings will often have very different sensitivities to the business cycle. Panel A presents a shorter but finer picture of the relative performance of US corporate bonds by Moody's rating category. The graph shows that when spreads are narrowing relative to government bonds, the spreads between higher- and lower-rated bond categories also narrow. In these times, although corporate bonds will generally outperform government bonds, lower-rated corporate bonds will tend to outperform higher-rated bonds. The converse is true as spreads widen, a phenomenon that is illustrated most graphically following the collapse of Lehman Brothers in 2008. The spread on speculative, or high-yield, debt rose from a pre-Lehman Brothers collapse low of around 2.8% to a peak of just more than 20.0%. Over the same period, Baa rated debt spreads rose from around 1.1% to 8.5% and Aaa corporate bond spreads rose from 0.6% to 4.5%.

Exhibit 17 US Credit Spreads and the Business Cycle

A. Moody's Rating

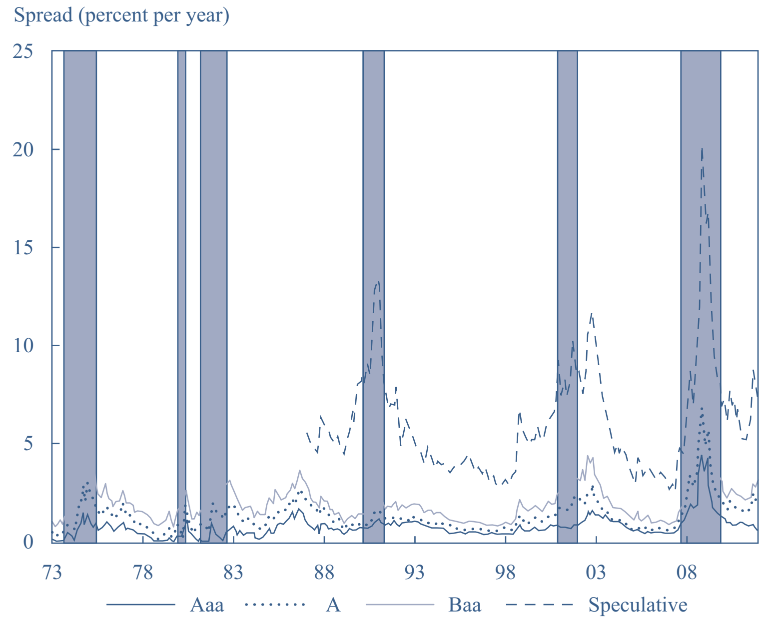
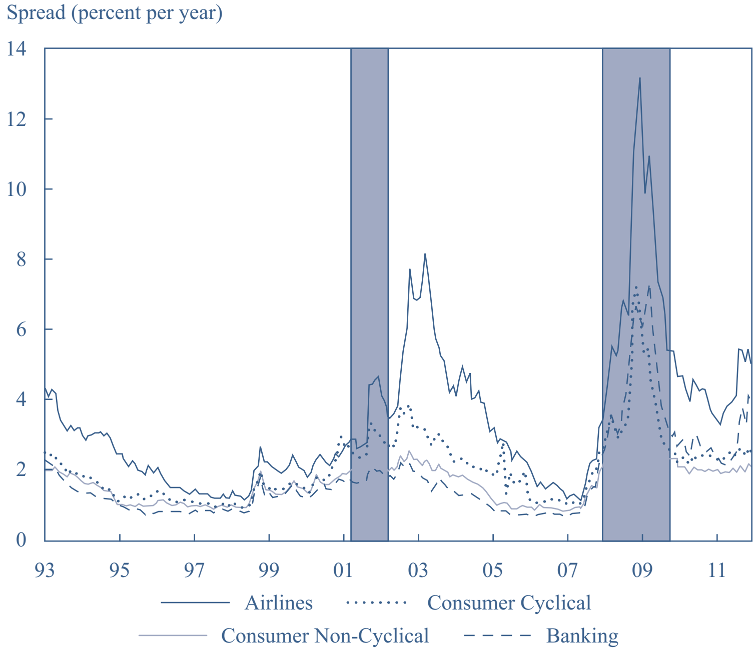


Exhibit 17 (Continued)

B. Industrial Sector



Note: Shaded areas indicate recessions.
Sources: Based on data from Thomson Reuters and the authors' calculations.

Panel B in Exhibit 17 illustrates another determinant of credit spreads: industrial sector. Analysis by industrial sector addresses the question of how the type of goods and services that individual companies produce may be related to credit quality. Some industrial sectors are more sensitive to the business cycle than others. This sensitivity can be related to the types of goods and services that they sell or to the indebtedness of the companies in the sector. Panel B shows evidence of only the divergent performance of corporate bond sectors over a relatively short period, but the performance of the four sectors shown is very different in times of economic stress. First, in both of the recessions that this period covers (indicated by the shaded areas in Panel B), the spread on the consumer cyclical sector rose more dramatically than it did for corporate bonds in the consumer non-cyclical sector. For example, the spread on the consumer cyclical sector peaked at just under 4.0% in 2003, compared with around 2.5% for the consumer non-cyclical sector. Second, the graph shows how sensitive the airline sector's credit spread is to the business cycle. The sharp widening of spreads in this sector as a result of both recessions is probably also a function of the lower credit quality, on average, of companies in the airline sector. Third, the recession of the early 2000s had only a mild impact on the spreads of banks but a much larger impact in the post-Lehman period, when the sector spread peaked at nearly 7.5%. This difference highlights the fact that the last recession and crisis were first and foremost a banking crisis. But perhaps the most interesting feature is the narrowing of sector spreads in the summers of 1998 and 2007. In both of these periods, investors were content to receive virtually the same credit spread, $\gamma^{i}_{t,s}$, on airline company debt as on debt issued by companies in the consumer non-cyclical sector.

17.1 Company-Specific Factors

Corporate bond spreads will be driven over time by the business cycle, but the impact of the economic environment on spreads will depend on issuers' industrial sector and rating. When spreads widen, the spreads on bonds issued by corporations with a low credit rating and/or that are part of a cyclical sector will tend to widen the most. Company-specific factors will also play a part in determining the difference in the yield of an individual corporate issuer and that of a government bond with the same maturity. Issuers that are profitable, have low debt interest payments, and are not heavily reliant on debt financing will tend to have a high credit rating because their ability to pay is commensurately high.

Exhibit 18 provides summary statistics on financial statements for companies across a range of Moody's rating categories. Pre-tax interest coverage is calculated by dividing total pre-tax earnings by total debt interest payments. On average, Aaa companies had \$17.60 of pre-tax earnings for every \$1 of interest payment to which they were committed. By contrast, on average, Baa companies had only \$2.50, whereas the average B and Caa rated companies could not cover their interest payments with current-period pre-tax earnings. The ratio of free operating cash flow to total debt gives another indication of the profitability and financial flexibility of a company relative to its outstanding debt. There is again a clear deterioration in this metric as average rating quality declines. Finally, the ratio of total debt to total capital gives an idea of the overall indebtedness of a company. Together these and other ratios allow analysts and credit rating agencies to determine a company's ability to meet its debt obligations as they come due. If this ability declines relative to other issuers in their sector, then the spread demanded on their debt will rise, relative to the sector average, and their rating may be lowered by the rating agency.

Exhibit 18 Ratings and Financial Ratios

	Aaa	Aa	A	Baa	Ba	B	Caa
Pre-tax interest coverage (×)	17.6	7.6	4.1	2.5	1.5	0.9	0.7
Free operating cash flow/Total debt (%)	42.3	28	13.6	6.1	3.2	1.6	0.8
Total debt/Total capital (%)	21.9	32.7	40.3	48.8	66.2	71.5	71.2

Source: Based on data from Moody's Investors Service.

18 SOVEREIGN CREDIT RISK

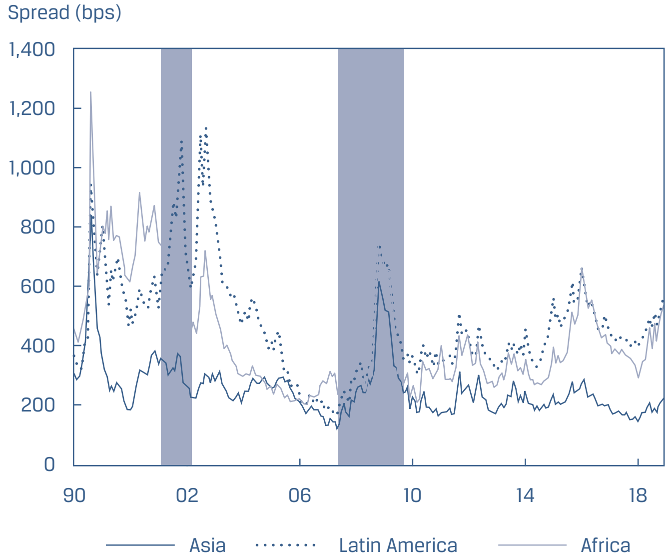
f explain how the phase of the business cycle affects credit spreads and the performance of credit-sensitive fixed-income instruments;

So far we have discussed the risk premium demanded by investors on both default-free and corporate debt. But credit premiums have always been an important component of the expected return on bonds issued by governments in developing or emerging economies. Even though many of these governments can print money to meet their debt obligations *in extremis*, meaning that they could technically avoid defaulting on these debts, many developing-economy governments have defaulted on their debts in the past. For example, the Russian government defaulted on its debt in 1998, and many others, including Argentina, Brazil, and Mexico, have also defaulted. Such defaults are often very country specific in character, but the global economic environment, oil prices, and the evolution of global trade will often play a part in precipitating such sovereign defaults.

The credit risk embodied in bonds issued by governments in emerging markets is normally expressed by comparing the yields on these bonds with the yields on bonds with comparable maturity issued by the US Treasury. Panel A of Exhibit 19 shows the evolution of this spread for three emerging market bond indexes. The impact of the credit crisis is clear; spreads rose in response to the uncertain economic environment globally. But the volatility of the spreads from 1998 to 2003 is a function of the Asian financial crisis in 1997, the Russian debt crisis in 1998, and the recession in developed economies in 2001–2002 following the collapse of the high-tech bubble. What is interesting is the decline in spreads for US Treasuries up to 2007, along with the much narrower spreads between the regions at this time. Strong global economic growth between 2003 and 2007 convinced investors that they did not need such a high reward for emerging market default risk and that they did not need to differentiate much between regions.

Exhibit 19 Sovereign Credit Spreads

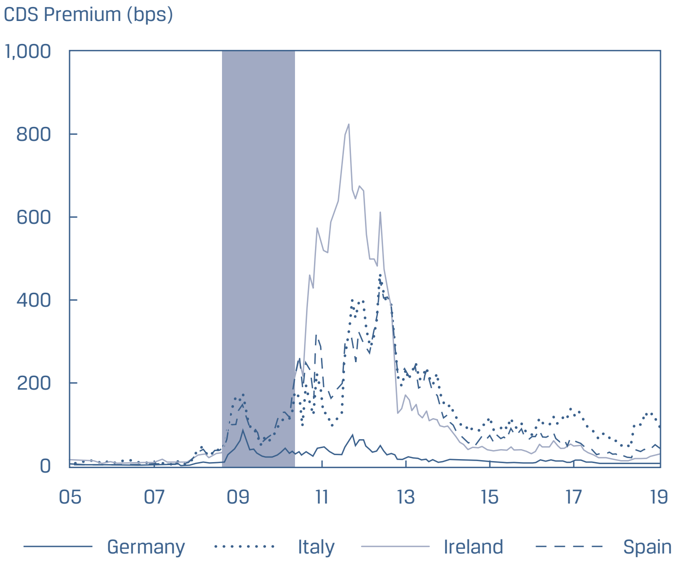
A. Emerging Markets



(continued)

Exhibit 19 (Continued)

B. Developed Economies



Note: Shaded areas indicate US recessions.
Sources: Based on data from Thomson Reuters and the authors' calculations.

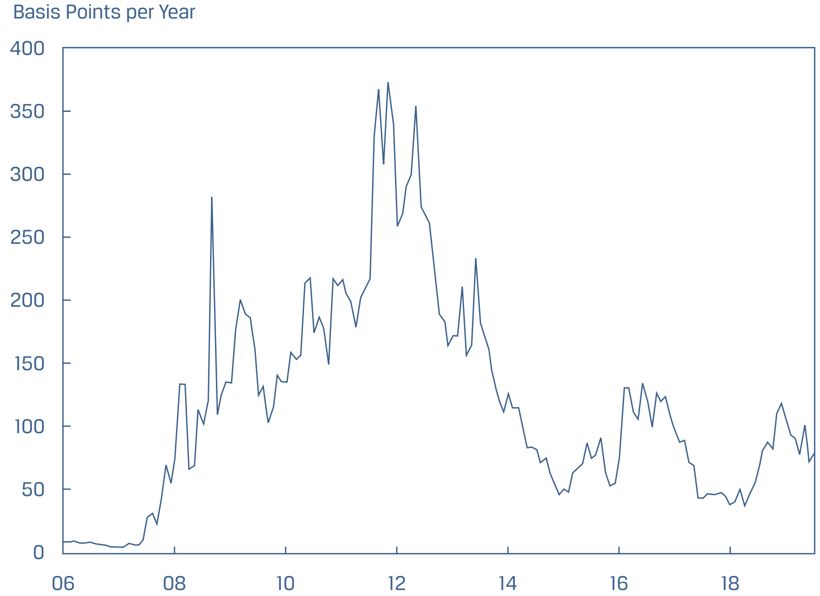
The 2008–2009 global financial crisis caused many investors to question what is meant by the term "default-free." Until that time, there was thought to be a set of developed-economy government issuers for which the likelihood of debt default was so low that it could almost be ignored. In other words, investors did not demand meaningful compensation for assuming this risk. Panel B of Exhibit 19 shows how this perception changed dramatically for the debts of a set of eurozone economies during the crisis. The chart shows the cost that an investor would have to pay for insuring themselves against a sovereign default on German, Italian, Irish, and Spanish government debt over the next five years through the purchase of credit default swaps (CDSs). For example, in January 2006 that cost was 1.8, 8.8, 2.5, and 2.8 bps, respectively. So, to insure oneself against a default on bonds issued by these governments with, say, a notional value of €10 million would have cost €1,800, €8,800, €2,500, and €2,800, respectively, per year. However, by August 2011, this insurance cost had risen to €59,830, €306,860, €825,390, and €300,610, respectively. Although the causes of this reassessment of sovereign credit risk inherent in developed-economy debt were complex, the basic reason for the increase in the credit risk premium was a reassessment by investors of these sovereign issuers' ability to pay and the likelihood that they might default. The perception of their ability to pay deteriorated dramatically as private-sector debts were absorbed onto sovereign balance sheets. And so, to some extent, the rise in this insurance cost was related to the balance sheets of these sovereign nations in much the same way that a deterioration in the quality of the balance sheet of a corporate borrower would cause its credit spread to widen.

EXAMPLE 16

The Credit Premium for the Royal Bank of Scotland

The global financial crisis had an impact on the prices of all financial assets. Exhibits 15 and 17 show the impact of the crisis on credit spreads derived from indexes, whereas Exhibit 19 demonstrates how the same crisis caused the credit spreads on some sovereign issuers to rise dramatically too. In Exhibit 20, we focus on the impact of the crisis on the CDS premium (a close proxy for the credit spread) on five-year Royal Bank of Scotland senior unsecured debt. The exhibit shows the same increase in the perception of credit risk. The eventual decline in the premium was a consequence of the UK government's nationalization of this systemically important global bank.

Exhibit 20 CDS Premium on Five-Year Royal Bank of Scotland Senior Unsecured Debt



Source: Based on data from Thomson Reuters.

18.1 Credit Premium Summary

The credit premium ($\gamma^i_{t,s}$) is the additional yield required by investors over and above the yield required on comparable default-free debt that investors demand for taking on credit risk. It will tend to rise and fall with the business cycle, mainly because credit risk will tend to rise as an economy turns down and to fall as an economy turns up. However, when credit spreads are generally narrowing, the rate of improvement will tend to be greater for those bonds issued by entities with a relatively weaker ability to pay. At these times too, investors seem to be less discerning among issuers with weak and strong credit credentials. But as the business cycle turns down, those issuers with a good credit rating tend to outperform those with lower ratings as the spread between low- and higher-quality issuers widens. This relationship between the economic cycle and defaults means that credit risky bonds (corporate or sovereign) tend to perform poorly in bad economic times, and because of this tendency, investors demand a credit premium.

19 EQUITIES AND THE EQUITY RISK PREMIUM

i. explain the relationship between the consumption hedging properties of equity and the equity risk premium;

Earlier we discussed the credit risk embedded in a bond that has been issued by either a corporation or government, which might not honor its promise to pay the coupons and principal payment in full and on time. Investors can thus not be certain that they will receive the future scheduled cash flows from credit risky bonds. However, when investors purchase bonds that embody credit risk, normally they at least know the proposed schedule of payments and how they are to be determined. But there are other financial instruments in which both the size and timing of the cash flows are uncertain and, indeed, where the cash flows may not materialize at all. The best example of a security that has cash flows with these characteristics is equity because the dividend payment is not promised, can rise and fall over time, and in the event that the issuing corporation becomes bankrupt, can cease altogether.

For equities, we can rewrite the generic pricing equation, Equation 1, as follows:

$$P_t^i = \sum_{s=1}^{\infty} \frac{E_t\left[\widetilde{CF}_{t+s}^i\right]}{\left(1 + l_{t,s} + \theta_{t,s} + \pi_{t,s} + \gamma_{t,s}^i + \kappa_{t,s}^i\right)^s} \tag{15}$$

Notice that this equation is essentially the same as that for credit risky bonds (Equation 13), but there is no maturity to the cash flows, so investors are essentially buying cash flows (dividends) into perpetuity (∞). In addition, we now have a new term in the discount rate, $\kappa_{t,s}^i$, which is the additional return that investors require for investing in equities, over and above what they require for investing in credit risky bonds $\left(l_{t,s} + \theta_{t,s} + \pi_{t,s} + \gamma_{t,s}^i\right)$.

The term $\kappa_{t,s}^i$ is essentially the equity premium relative to credit risky bonds. This is not the way the equity risk premium is usually expressed. We have expressed it this way for the moment because of the following reasons:

- If a company experiences financial difficulties because the company's debt holders have the senior claim on the company's cash flow, the equity holders will receive the residue, which could be zero, and
- in the event that a company's financial difficulties become so bad that the company is forced into bankruptcy, both bond and equity investors will lose. But depending on the quality of the company, investors in the corporate bond can usually expect to get some of their investment back. The equity investors, however, will normally lose all of their investment. Both debt and equity investors are exposed to risk, but the potential loss is greater for the equity investor.

These are the reasons why investors will require a risk premium, $\kappa_{t,s}^i$, over and above the one that they would require on the corporation's debt, $\gamma_{t,s}^i$. Equation 15 shows that both corporate bond and equity holders face what we might define as corporate risk. It is the combination of this risk that is usually referred to as the equity risk premium. Because the risk associated with equities is normally expressed relative to default-free debt of the same currency, we can rewrite Equation 15 as follows:

$$P_t^i = \sum_{s=1}^{\infty} \frac{E_t\left[\widetilde{CF}_{t+s}^i\right]}{\left(1 + l_{t,s} + \theta_{t,s} + \pi_{t,s} + \lambda_{t,s}^i\right)^s} \tag{16}$$

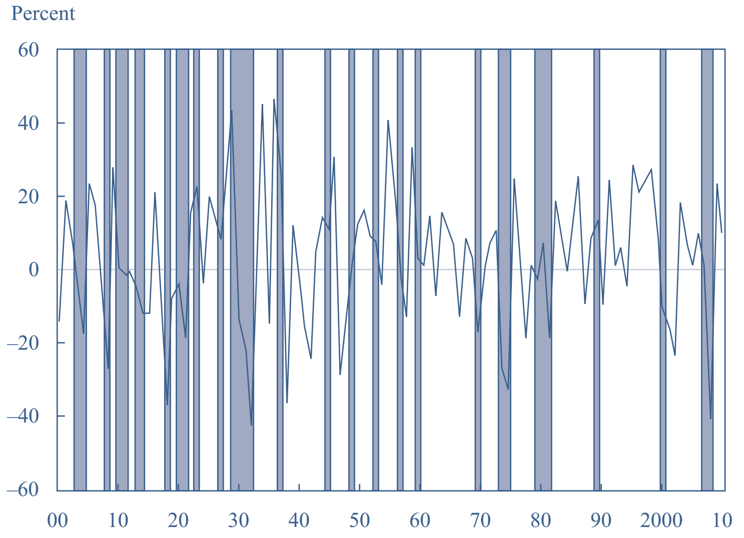
where the equity risk premium, $\lambda_{t,s}^i$, is equal to $\gamma_{t,s}^i + \kappa_{t,s}^i$. That is, it is the addition to return required by investors over and above the compensation for risk that they require for holding a default-free government bond of the same currency (technically, a very long-dated, plain-vanilla, coupon-paying, default-free bond).

19.1 Equities and Bad Consumption Outcomes

Equity investors will demand an equity risk premium if the consumption hedging properties of equities are poor—that is, if equities tend not to pay off in bad times. Our arguments earlier indicate that the equity risk premium should be positive and therefore, implicitly, that equities are a bad hedge for bad consumption outcomes. However, tying down the exact relationship between equity performance and consumption over time has proved to be very difficult. But we can get some idea of the relationship if we consider a very long history of the real returns produced by equities.

Exhibit 21 Annual Real Equity Returns, 1900–2010

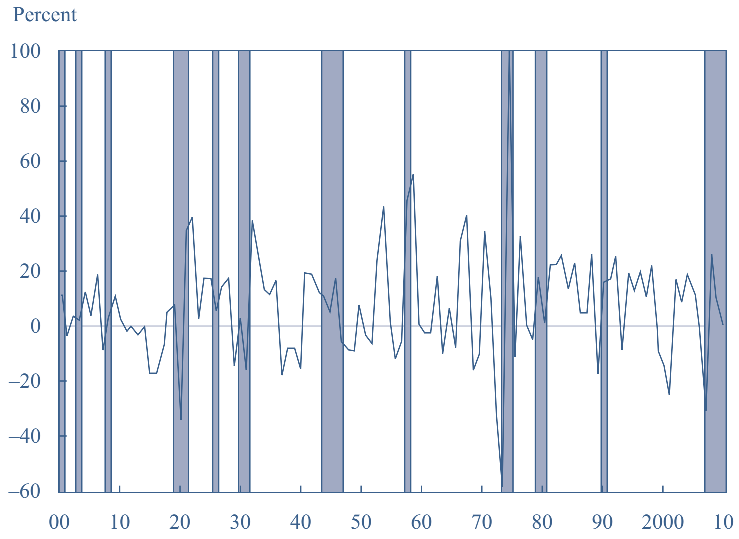
A. United States



(continued)

Exhibit 21 (Continued)

B. United Kingdom



Note: Shaded areas indicate recessions.
Sources: Based on data from Shiller (2000), the Bank of England, and the authors' calculations.

Exhibit 21 shows the annual real (inflation-adjusted) returns generated by both US (Panel A) and UK (Panel B) equities from 1900 to 2010. Generally speaking, sharp falls in equity prices are associated with recessions—bad times. For example, real UK equity prices more than halved as a recession hit the United Kingdom in 1972, and real US equity prices fell by more than 40% during the Great Depression. More recently, real UK and US equity prices fell by 30% and 40%, respectively, in 2009. Given this evidence, it is difficult to argue that equities are a good hedge for bad consumption outcomes. We would thus expect the equity risk premium to be positive, and given the scale of the declines in prices possible in bad times, we might expect it to be quite large.

Before we consider how large the equity premium should be, we will first focus on the cash flow that equities generate. It is the nature of this cash flow that leads investors to demand an equity risk premium in the first place.

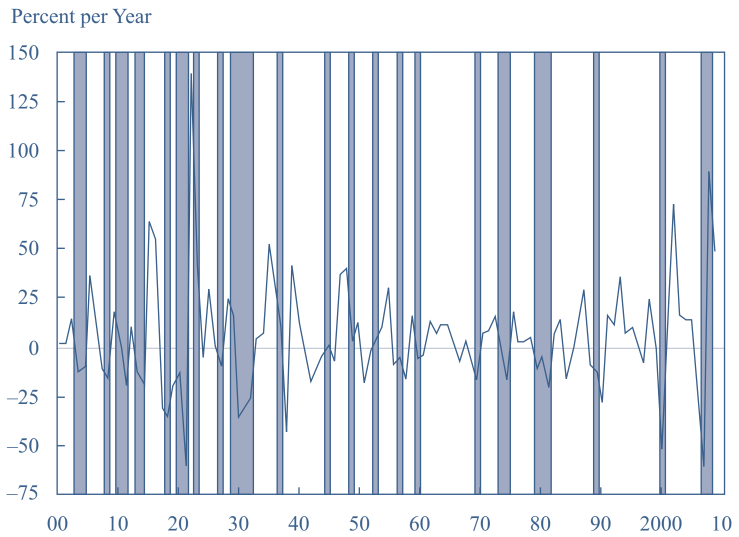
20 EARNINGS GROWTH AND THE ECONOMIC CYCLE

h explain how the phase of the business cycle affects short-term and long-term earnings growth expectations;

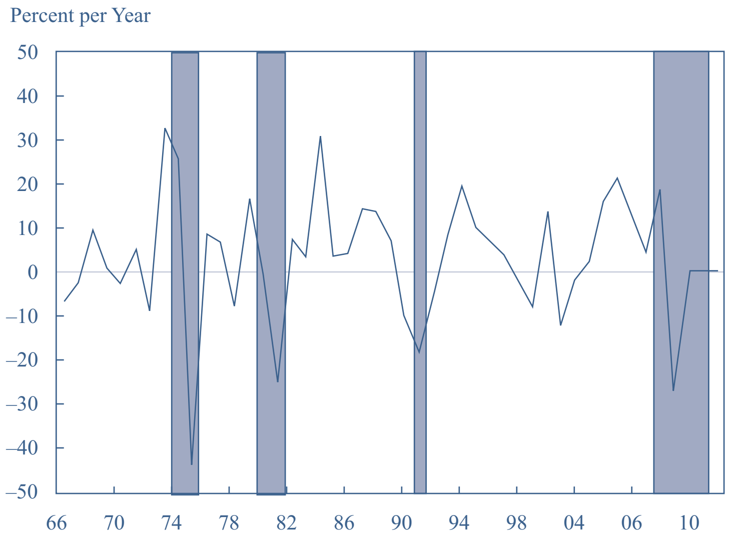
The uncertainty about—and time variation in—future dividends, as represented by the numerator in Equation 15, is a key feature of equity investment. Panels A and B of Exhibit 22 show a long history of US real earnings growth and a shorter history of UK real earnings growth, respectively. The exhibit shows that a sharp decline in real earnings nearly always coincides with a recession, which is to be expected; recessions are associated with declines in employment, incomes, output, and, subsequently, profitability. US real earnings fell dramatically during the Great Depression and by nearly 60% in 2009. Conversely, sharp increases in profit growth occur at the end of a period of recession and in some cases while recession conditions still persist. Thus, corporate profitability can lead an economy out of recession as well as into it: A negative demand shock can cause demand and corporate profits to shrink. In response, companies lay off workers, reducing their cost base and thereby adding to the recessionary backdrop. When an upturn in demand occurs, perhaps in response to monetary policy stimulus, demand growth on a lower cost base can lead to a sharp increase in corporate profits, which then leads companies to invest and hire more staff, and so on. Some analysts thus consider corporate profitability to be an important leading indicator of the business cycle and believe it provides useful information about future growth.

Exhibit 22 Real Equity Earnings Growth in the United States and the United Kingdom

A. United States



B. United Kingdom



Note: Shaded areas indicate recessions.
Sources: Based on data from Shiller (2000) and the Bank of England.

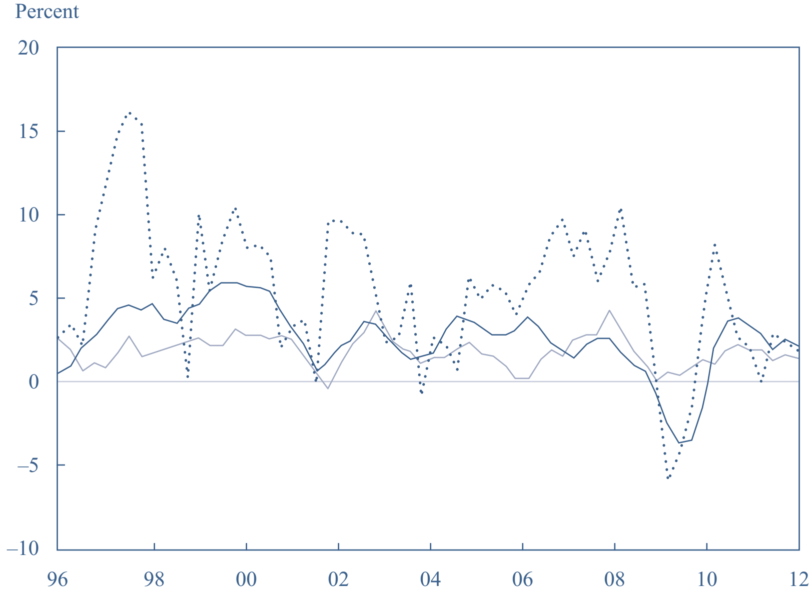
Equity analysts spend the majority of their time focusing on the numerator in Equation 15—that is, forming views about expected earnings and, therefore, about dividends and free cash flow. Given the close relationship between aggregate earnings, or profits, and the business cycle shown in Exhibit 22, an understanding of the business cycle is crucial for earnings projections, particularly in the short term. However, the business cycle will not affect the corporate profits of every company in the same way. The type of product sold or service provided by the company will have an impact on earnings and consequently on equity performance over the business cycle.

Some companies make products or provide services that are relatively insensitive to general economic conditions. Toothpaste might fall into this category. Because the cost of toothpaste usually only represents a small proportion of the overall household budget, people will generally still want to keep their teeth clean even if the economy is in recession, and because they are unlikely to want to clean their teeth more often simply because the economy is booming, the demand for toothpaste will remain fairly stable over the business cycle. Companies and equity sectors that produce such products are referred to as non-cyclical or defensive investments. By contrast, some companies, such as airlines, will produce goods or provide services that are extremely sensitive to the business cycle. In difficult economic conditions, consumers are much more likely to postpone or cancel their vacations or to vacation at home than to reduce their consumption of toothpaste, and businesses are likely to cut back on airline travel. Generally speaking, an annual family vacation will constitute a large proportion of the household budget, and most people do not need a vacation in the same way that they need toothpaste or soap. Businesses may rely on alternatives to expensive travel for meetings, such as video conferencing. By contrast, in good times when real incomes are rising, people are more likely to take more vacations or more expensive ones, and the increase in business activity may necessitate more meetings in new, often distant markets. Economists and investment strategists may view a rise in the earnings of cyclical companies after a period of decline as an indicator of a likely improvement in wider economic growth in the future.

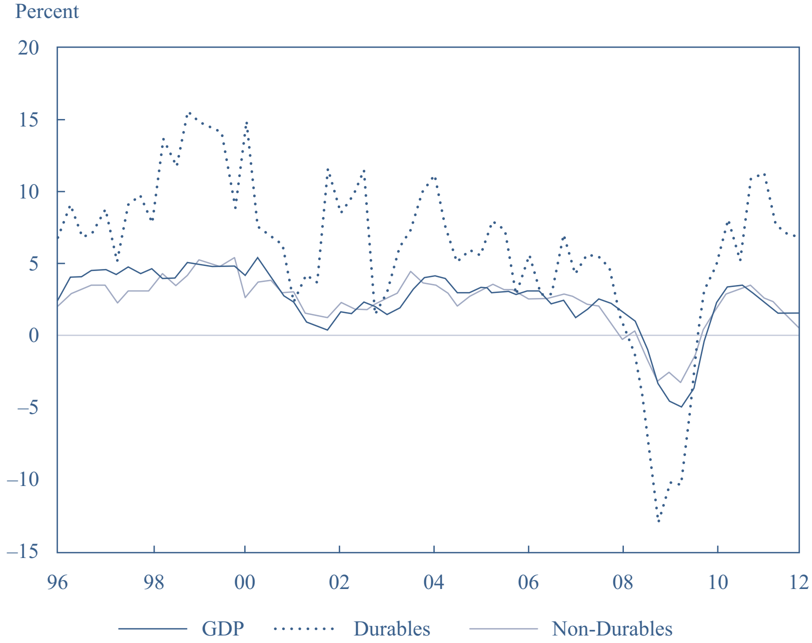
Exhibit 23 shows the annual growth rates (year over year) of real GDP and of the consumption of both durable and non-durable goods for Canada (Panel A) and the United States (Panel B). Both panels of the exhibit show how sensitive durable goods consumption is to the economic cycle. We can expect then that the profits of companies that produce durable as opposed to non-durable goods to be commensurately more volatile too.

Exhibit 23 Year-over-Year Growth Rate of GDP and the Consumption of Durable and Non-Durable Goods, 1996–2012

A. Canada



B. United States



Source: Based on data from Thomson Reuters.

Stock market participants often classify stocks as being cyclical and non-cyclical. Exhibit 24 shows the real earnings growth of the non-cyclical and cyclical goods sectors of both the United States (Panel A) and the United Kingdom (Panel B). The cyclical sectors in this case are represented by companies that produce discretionary consumer goods, whereas the non-cyclical index is represented by an index that includes companies that produce staple (or less discretionary) consumer goods. Panel

A shows the clearest evidence of the greater sensitivity to business conditions of the cyclical sector. Real earnings growth rises and falls dramatically over the business cycle. By contrast, although the real earnings of non-cyclical companies vary across the business cycle, the peaks and troughs are less extreme. The time variation in UK real earnings over the business cycle is also evident from the exhibit. However, the difference between the real earnings growth of the UK's cyclical and non-cyclical sectors is less clear, although the cyclical sector tended to experience more significant troughs in real earnings growth over this period.

Exhibit 24 The Real Earnings Growth of Discretionary and Staple Consumer Goods Companies, 1974–2012

A. United States

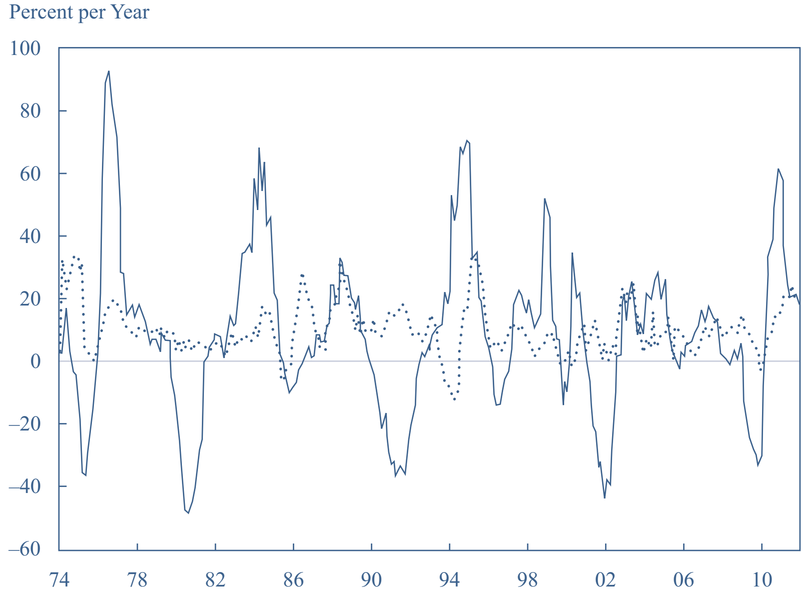
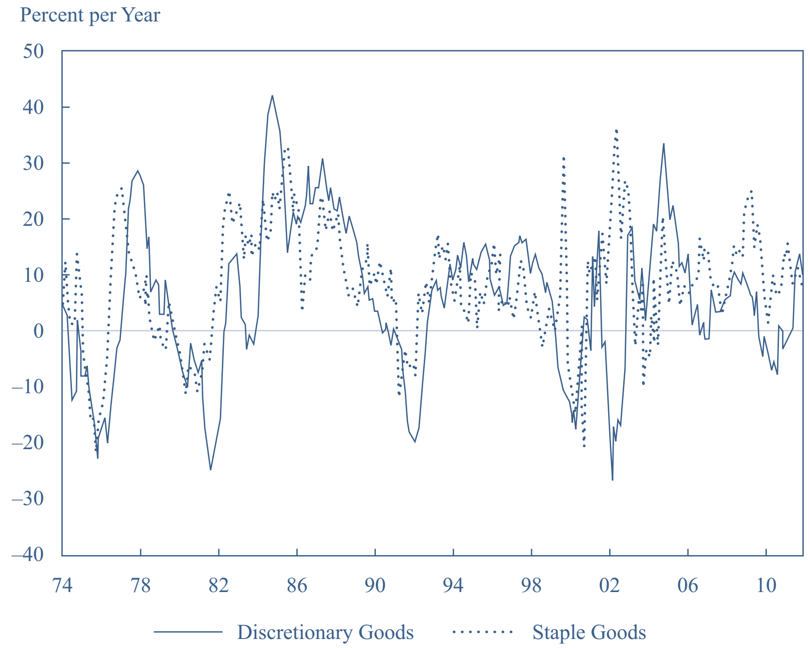


Exhibit 24 (Continued)

B. United Kingdom



Source: Based on data from Thomson Reuters.

There are, of course, other factors that determine the earnings growth of an equity or equity sector: the financial structure of the company, the quality and experience of its management, and the ease with which new entrants can establish themselves to compete away any abnormal profits. However, the relationship between the business cycle and the nature of the type of good or service sold will remain important. Indeed, in a booming economy, even bad managers of companies with poor financial structures can generate or appear to generate profits—for example, WorldCom and Enron. But tougher, recessionary conditions often expose weak companies as demand turns down and financing becomes harder to access.

HOW BIG IS THE EQUITY RISK PREMIUM?

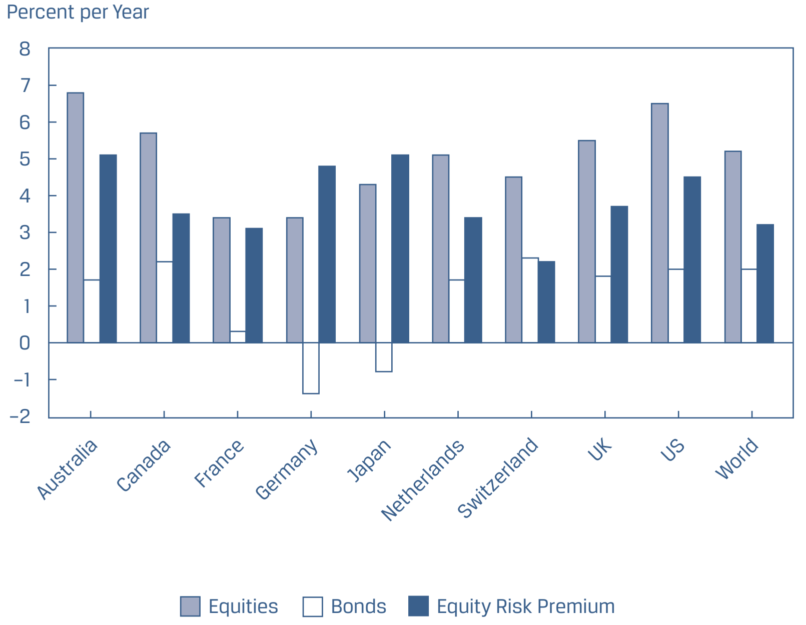
21

i. explain the relationship between the consumption hedging properties of equity and the equity risk premium;

The real earnings of companies are clearly affected by the underlying economy. This relationship is positive in that when the economy turns down, so (normally) do corporate profits. But it is in these bad times that investors need their investments to offset these worsening earnings. Because of the pro-cyclicality of economies and corporate profits (in aggregate), equities are not a good hedge against bad consumption outcomes, which, in turn, means that investors will require a risk premium. But how big should this premium $\left(\lambda_{t,s}^{i}\right)$ be?

It is impossible to quantify the equity risk premium *ex ante*. But we can at least look at its *ex post* value using very long runs of data. Exhibit 25 shows the real annual return on equities and government bonds over the period of 1900–2017 for a range of developed-economy equity markets. Over this very long period, equities in each country have outperformed government bonds. The bars representing the *ex post* equity risk premium range from 2.2% per year in Switzerland to 5.1% per year in Australia. US equities, which constitute the world's largest equity market, have outperformed US Treasuries by 4.5% per year on average over the 117 years under measurement, whereas the global equity market has produced an equity risk premium of 3.2% per year (premium versus bonds). Of course, there is no guarantee that a premium earned in the past will be earned in the future, but this long span of data shows that the *ex ante* equity risk premium for developed-economy equity markets could be somewhere between 3% and 5% per year.

Exhibit 25 *Ex-Post* Real Returns on Equities, Bonds, and Equity Risk Premiums



Source: Based on data from Chapter 1 of Dimson, Marsh, Staunton (2018).

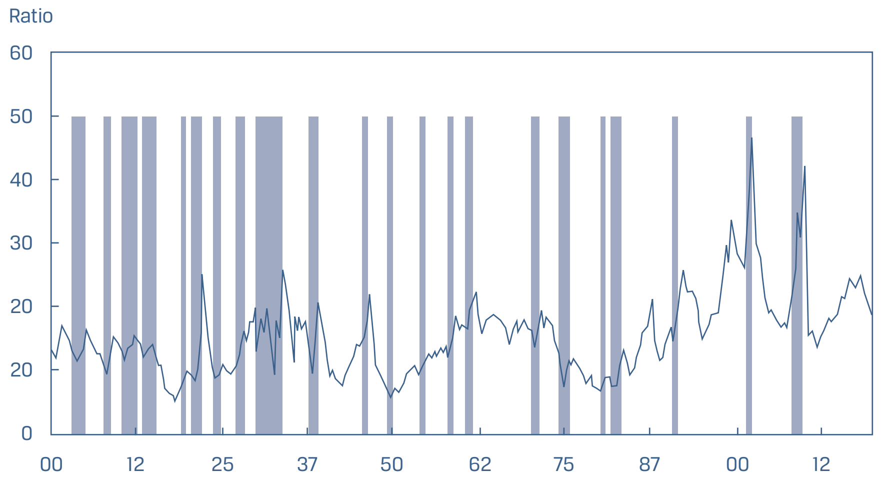
22 VALUATION MULTIPLES

j describe cyclical effects on valuation multiples;

Analysis of company earnings prospects is usually the central focus of equity analysts and strategists. To help compare equities within and among sectors, they will generally monitor valuation multiples, such as the price-to-earnings ratio (P/E) or the price-to-book ratio (P/B). P/E is calculated as the ratio of the current share price to the earnings per share (EPS) generated by the company. This ratio tells investors the price they are paying for the shares as a multiple of the company's earnings per share. Investors use this ratio to compare the valuations attributed to individual equities, sectors, and markets. For instance, if a stock is trading with a low P/E relative to the rest of the market, it implies that investors are not willing to pay a high price for a dollar's worth of the company's earnings. The reason may be that the market believes the prospect of strong earnings growth in the future is low. Alternatively, a share trading with a very high P/E relative to the rest of the market indicates that investors are willing to pay a higher price for each dollar's worth of the company's earnings. They may be willing to do so because they expect this company's earnings to grow rapidly in the future. When the EPS used to estimate the ratio refers to last year's earnings, the P/E is referred to as being a historical or a trailing P/E. However, when the EPS is based on an estimate of future earnings, it is referred to as the leading or forward P/E. If a company's EPS is expected to grow, then its historical P/E will be greater than its forward P/E. However, what constitutes a high or low P/E very much depends on the market, sector, or company in question and, in particular, on the economic backdrop. US P/Es between 1900 and 2018 are shown in Exhibit 26.

Another popular valuation multiple is the price-to-book ratio, which measures the ratio of the company's share price to its net assets or its assets minus liabilities attributed to each share. The P/B tells investors the extent to which the value of their shares is "covered" by the company's net assets. Some of these assets, such as office buildings, are tangible, whereas others, such as patents and copyrights, are intangible. Furthermore, some of the assets are actually on the balance sheet and hence part of book value, whereas others are not. It also indicates the strength of investors' expectations about the company's ability to generate a high return on its net assets, adjusted for risk. The higher the ratio, the greater the expectations for growth but the lower the safety margin if things do not turn out as expected. Again, what constitutes a high or low P/B is determined by the market, sector, and stock in question.

Exhibit 26 US P/Es, 1900–2018



Note: Shaded areas indicate recessions.
Sources: Based on data from Shiller (2000) and www.econ.yale.edu/~shiller/.

One of the problems for equity strategists is to ascertain whether the P/E (or P/B) is high or low. The average trailing US P/E between 1900 and 1990 was 13.5, indicating that investors were willing to pay \$13.5 for a dollar's worth of the previous year's earnings; by the late 1990s and early 2000s, they were willing to pay \$45. The expansion in the US equity market's P/E during the 1990s was a global phenomenon. Equity strategists justified the rise in the price of earnings with many ad hoc explanations—for example, the end of the cold war, better macro-policy that would ensure that major recessions were a thing of the past, and the internet revolution, to name just a few. However, with regard to the basic pricing relationship, shown in Equation 15, the high P/E could be the result of a number of factors, including

a an increase in expectation of future real earnings growth $\left(E_t\left[\widetilde{CF}^{i}_{t+s}\right]\right)$;

b falling real interest rates ($l_{t,s}$), possibly associated with falling volatility in real GDP growth;

c a decrease in inflation expectations ($\theta_{t,s}$);

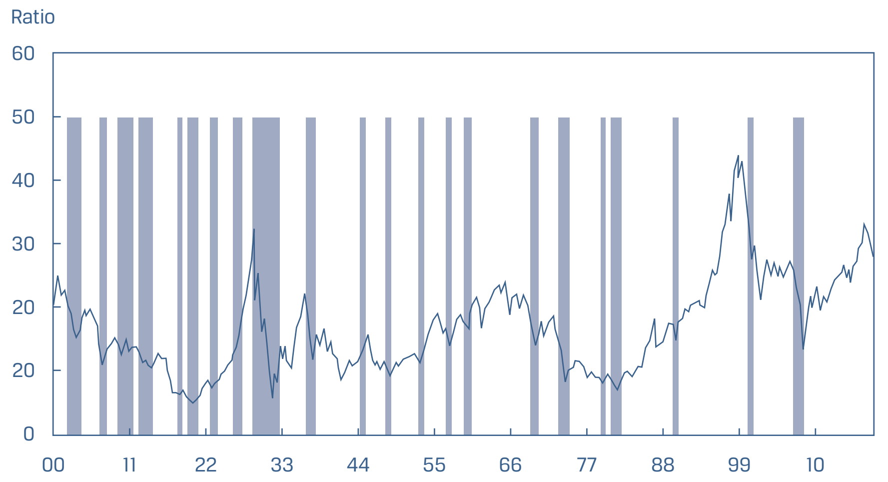
d a decline in uncertainty about future inflation ($\pi_{t,s}$); or

e a decrease in the equity risk premium $\left(\lambda^{i}_{t,s}\right)$.

Other things being equal, any one of these changes or all of them combined could justify higher equity prices (P) relative to current earnings (E) and thus higher equilibrium P/Es. There were some investors, however, who were not convinced that such high P/Es relative to historical levels were justifiable, particularly on the grounds of much higher future earnings growth. The US Federal Reserve Board chairman, Alan Greenspan, alarmed by the rise in P/Es, described the valuation of equity markets in 1996 as essentially the result of "irrational exuberance" on the part of equity investors.

Robert Shiller has proposed an alternative valuation multiple—the real cyclically adjusted P/E (CAPE). The CAPE is derived in the same way as the P/E, but the "P" represents the real (or inflation-adjusted) price of the equity market and the "E" is a 10-year moving average of the market's real (or inflation-adjusted) earnings. Deflating the real equity price by a moving average of real earnings irons out the short-term volatility in this indicator over time. Exhibit 27 shows this ratio for the United States from 1900 to 2018. The very high price that equity investors were willing to pay for equities in 1929 and 1999 and, to a lesser extent, in 1965 is still apparent. It is worth noting that the average real return on US equities in the 10 years following the peaks in the CAPE in 1929, 1965, and 1999 were –0.3% per year, –5.4% per year, and –4.1% per year, respectively. Conversely, the average real return over the 10-year period after the two lowest values of the CAPE in 1921 and 1980 were 12.3% per year and 7.3% per year, respectively.

Exhibit 27 Real US Cyclically Adjusted P/E (CAPE), 1900–2018



Note: Shaded areas indicate recessions.
Source: Based on data from Shiller (2000) and www.econ.yale.edu/~shiller/.

COMMERCIAL REAL ESTATE

23

l. describe the economic factors affecting investment in commercial real estate.

The basic pricing formula can be applied to asset classes besides bonds and equities. To demonstrate how the basic pricing framework presented in Equation 1 can be extended to other asset classes, we will consider commercial real estate.

23.1 Regular Cash Flow from Commercial Real Estate Investments

When investors invest in commercial real estate, the cash flow they hope to receive is derived from the rents paid by the tenants. These rents are normally collected net of ownership costs, such as those related to the upkeep of the building, according to a fixed schedule from the businesses that lease the property from the investors who act as landlords. Although practices vary from country to country, the rental agreement will be reviewed regularly and may be reset. In some countries, rents are subjected to "upward only" restrictions, which means that existing tenants will not see their rents fall, only potentially rise. Rents may also be indexed so that they rise in line with a pre-specified index of (usually) consumer prices.

Thus to a large extent, the rental income can be viewed as being analogous to the coupon income derived from a bond. Because a well-diversified portfolio of commercial property could be expected to generate a stream of rental income for investors, they might view such a portfolio as being similar to a well-diversified portfolio of bonds. The credit quality of a commercial property portfolio will be determined by the credit quality of the underlying tenants, in much the same way that the credit quality of a bond portfolio will be determined by the credit ratings of the bond issuers of the constituent bonds. Generally speaking, the lower the credit quality of the tenants, the less likely they will be to pay their rent on time or at all.

23.1.1 *The Equity Component of an Investment in Commercial Real Estate*

Investors in commercial real estate will receive regular cash flows derived from the rents paid by tenants, but there is another important element to property investment that is less bond-like. When a bond matures, the investor generally receives the face value of the bond along with the final coupon. But when the lease on a property expires, the investors (acting as landlords) will take back possession of the property and will have to decide whether to re-rent it to another tenant, to sell it to another investor, or to redevelop it for a future sale. The determining factor is likely to be the value of the property at the time. Its value may have risen dramatically over time, or it might now be worth much less. The value of the property will arguably be determined by two key factors: the property's location and the state of the underlying economy. If, during the time of the lease, the area in which the property is situated has become more popular, then the property might be sold at a profit or it might be worth redeveloping the property. Similarly, if the lease expires when general economic activity is high and thus there is strong demand for property, then the sale or redevelopment option might be worth pursuing. But if, when the lease expires, the location is deemed to be less desirable or the economy is weak, then redevelopment may not be an option, future rents may have to be lower on the property, and investors may come to the view that the property should be sold, even at a loss.

The potential for profit or loss and the uncertainty related to this profit from redevelopment add an equity-like dimension to investment in commercial real estate. In other words, this potential and uncertainty add either a positive increment to cash flow or a negative one. To this extent, some investors like to think about the cash flow derived from a commercial real estate portfolio as being part bond, part equity.

23.1.2 *Illiquidity and Investment in Commercial Real Estate*

There is a third aspect to investing in commercial real estate that is also crucial: its illiquidity. Anyone who has sold a home knows that it usually takes a great deal of time and effort to put the property up for sale, to find a buyer, and (if a buyer can be found) to finalize the deal. For similar reasons, it can take months and sometimes years to exit from a commercial property investment, and the high transactions costs often discourage investors further from liquidating holdings. By contrast, it is relatively easy in normal market conditions to transform a holding in developed-economy government bonds, investment-grade corporate debt, or publicly traded equities into cash. Generally speaking, most of the asset classes that we have considered so far in this reading are liquid relative to an investment in commercial property.

23.2 The Pricing Formula for Commercial Real Estate

Commercial real estate is a "special" asset class; it can be viewed as being part equity, part bond, and it is usually very illiquid. However, with some minor adaptions of the generic pricing formula in Equation 1, we can still capture all of the salient features of the price of commercial real estate, as follows:

$$P_t^i = \sum_{s=1}^{N} \frac{E_t\left(\widetilde{CF}_{t+s}^i\right)}{\left(1 + l_{t,s} + \theta_{t,s} + \pi_{t,s} + \gamma_{t,s}^i + \kappa_{t,s}^i + \phi_{t,s}^i\right)^s}. \tag{17}$$

The pricing formula shown in Equation 17 acknowledges that the expected cash flow from an investment in commercial real estate, $E_t\left(\widetilde{CF}_{t+s}^i\right)$, will be uncertain because tenants may default on the rental agreement. The quality of this rental income will depend on the quality of the tenants, just as the reliability or quality of the coupons from a corporate bond will be dependent on the credit standing of the corporate bond issuer. Furthermore, the property's value in the future cannot be known with certainty.

But what should the discount rate look like? To understand the construction of the discount rate in Equation 17, consider the following tenants and associated rental/leasing agreements:

1 a developed-economy government tenant that agrees to pay rental income that is indexed to inflation $(1 + l_{t,s})$,

2 a developed-economy government tenant that agrees to pay fixed nominal rental income $(1 + l_{t,s} + \theta_{t,s} + \pi_{t,s})$, and

3 a corporate tenant that agrees to pay a fixed nominal rental income $(1 + l_{t,s} + \theta_{t,s} + \pi_{t,s} + \gamma_{t,s}^i)$.

In each case, the expressions in parentheses represent the composition of the discount rate that would be applied to the cash flows of bonds issued by these entities: (1) is analogous to the purchase of a real default-free government bond, (2) is analogous to the purchase of a nominal default-free government bond, and (3) is analogous to the purchase of a credit risky nominal bond. In each case, though, we need to add a risk premium to take into account the uncertainty relating to the value of the property at the end of the lease. This premium is analogous to the equity risk premium, $\kappa^{i}_{t,s}$.

Finally, we have to take into account the illiquidity of a commercial property investment. Because investors cannot easily convert their property investments into cash, there exists the possibility that they will not be able to liquidate their investment in bad economic times. In other words, other things being equal, illiquidity acts to reduce an asset class's usefulness as a hedge against bad consumption outcomes. Because of this, investors will demand a liquidity risk premium, which we have expressed as $\phi^{i}_{t,s}$ in Equation 17.

The discount rates that investors would apply to an investment in commercial property in each of the three instances previously listed are, therefore,

1. $1 + l_{t,s} + \kappa^{i}_{t,s} + \phi^{i}_{t,s}$,
2. $1 + l_{t,s} + \theta_{t,s} + \pi^{i}_{t,s} + \kappa^{i}_{t,s} + \phi^{i}_{t,s}$, and
3. $1 + l_{t,s} + \theta_{t,s} + \pi^{i}_{t,s} + \gamma^{i}_{t,s} + \kappa^{i}_{t,s} + \phi^{i}_{t,s}$.

The relative sizes of the components listed will vary depending on the length of the lease, the quality of the tenant, and the location of the property.

EXAMPLE 17

A Real Estate Investment Decision

An analyst estimates that the real risk-free rate is 1.25%, average inflation over the next year will be 2.5%, and the premium required by investors for inflation uncertainty is 0.50%. He also observes that the yield on a 10-year senior unsecured bond issued by Supermarket plc is 5.75%. From these figures, he deduces that the credit spread on Supermarket plc's 10-year debt is 1.50%.

The same analyst is asked to review for a client (an investor) the opportunity to buy a site currently occupied by Supermarket plc. Once the investor purchases the property, Supermarket plc will lease it back and pay $500,000 annual rent in arrears to the investor. "Rent in arrears" in this case means that the first annual rental payment is due in 12 months, covering the first year's tenancy, and the second is due in 24 months, and so on. Like the Supermarket bond, the lease on the property has 10 full years to expire. At the end of this period, the property and land will revert to the investor, and the analyst estimates that the resale value of the property after 10 years will be $10 million, net of all transactions costs.

The investor tells the analyst that it normally expects to receive a risk premium of 0.50% on any cash flow from a commercial property investment to compensate it for the uncertainty of the final value of the property and the uncertainty relating to the receipt of rental income, plus a liquidity premium of 1.0% on these cash flows. The investor's required return on the property is thus 7.25% (= 5.75% + 0.5% + 1.0%). If the purchase price of this piece of commercial property is $8.2 million, should the analyst recommend the purchase to the client?

Discount Rate: 7.25%

Payment Due (years)	Cash Flow	Present Value
1	$500,000	$466,200
2	500,000	434,686
3	500,000	405,301

(continued)

	Discount Rate: 7.25%	
Payment Due (years)	**Cash Flow**	**Present Value**
4	500,000	377,903
5	500,000	352,357
6	500,000	328,538
7	500,000	306,330
8	500,000	285,622
9	500,000	266,314
10	10,500,000	5,214,543
Implied property value		$8,437,796

The cash flows in the table, along with their associated present values, demonstrate that at a discount rate of 7.25%, the property would be priced at $8,437,796. Any asking price above this value would imply a return of less than the investor's hurdle rate of 7.25%, whereas any price below this price implies a return above this hurdle rate. On the basis of this information, the analyst should recommend that the client go ahead with the investment.

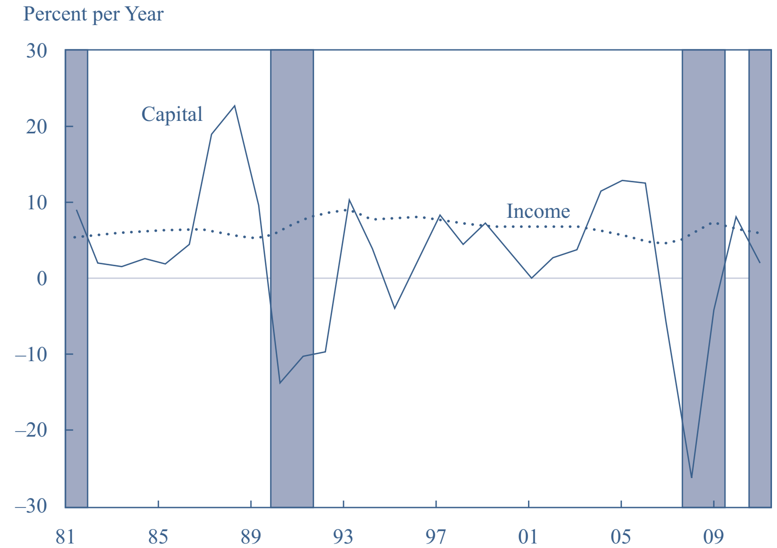
24 COMMERCIAL REAL ESTATE AND THE BUSINESS CYCLE

l. describe the economic factors affecting investment in commercial real estate.

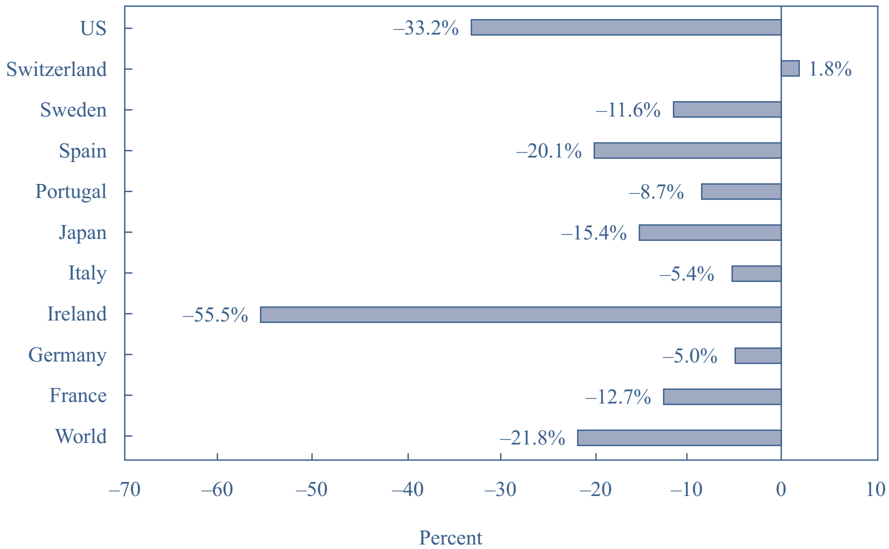
The nature of the cash flows from commercial property and the complex structure of the discount rate will all be influenced by the evolution of the underlying economy. Panel A of Exhibit 28 shows the annual growth rate of UK commercial property income over a 30-year period. It is remarkably stable over this period, averaging 6.5% per year; in other words, UK commercial property rental income has grown by approximately 6.5% annually in nominal terms (or 2.5% in real terms) over the 30-year period. The stability of this income stream, over a number of business cycles, suggests that investors might calculate its present value using a very low discount rate. But as well as showing the annual change in rent from a portfolio of UK commercial property, Panel A also shows the annual percentage change in the capital value of the United Kingdom's commercial property market. Whereas rental income appears to have been relatively stable (in nominal terms) and almost immune to the business cycle, commercial property capital values are much more sensitive to the economic cycle. Between 1990 and 1992, as the UK economy experienced a deep recession, UK commercial property prices fell by a cumulative 30%. Over the course of the United Kingdom's 1990s recession, the capital value of the UK property market fell by 26%. In Panel B of the exhibit, we present the capital value changes in a number of markets around the world between 2008 and 2009. It is clear that the global recession had a significant impact on commercial property prices. For example, in Ireland, one of the developed economies arguably worst hit by the crisis, commercial property prices fell by 55.5%.

Exhibit 28 Commercial Property

A. UK Commercial Property Returns, 1981–2011



B. International Comparison of Commercial Property Value Changes, 2008–2009



Note: Shaded areas in Panel A indicate recessions.
Source: Based on data from Investment Property Databank (www.ipd.com).

Taken together, the two panels in Exhibit 28 show that even though nominal rental income might be relatively stable, the capital values of commercial property are highly sensitive to the economic environment. A recession will generally cause these values to fall, whereas more robust economic conditions will tend to cause commercial property prices to rise, often dramatically. For example, the recovering and then strong global economy between 2003 and 2006 caused world commercial property prices to rise by nearly 20%. Over the same period, they rose by 41% in the United Kingdom and by a staggering 51% in Ireland.

The pro-cyclical nature of commercial property prices means that investors will generally demand a relatively high risk premium in return for investing in this asset class. The reason is that commercial property does not appear to be a very good hedge against bad economic outcomes. In fact, the sharp declines in capital values in recessionary periods resemble the sort of declines that investors in equity experience, although these occurrences are more frequent with equity investment. Thus, the sort of risk premium that investors will demand from their commercial property investments arguably will be closer to that demanded on equities than on default-free government bonds.

Finally, although it is difficult to derive a value for the property risk premium, it is likely to vary over time with economic conditions and to be relatively highly and positively correlated with the risk premiums on corporate bonds and equities.

EXAMPLE 18

Valuation and the Business Cycle

Describe how real estate valuation is distinguished from valuation of public equities.

Solution

Real estate does not trade in public markets (the exception being REITs). Compared with the valuation of public equities, the valuation of real estate should reflect a discount for relative lack of liquidity.

SUMMARY

In this reading, we have sought to explain the fundamental connection between the prices of financial assets and the underlying economy. The connection should be strong because ultimately all financial assets represent a claim on the real economy. Because all financial assets offer a means of deferring consumption, to make the connection tangible we have explored the relationship between these asset prices and the consumption and saving decisions of economic agents.

- At any point in time, the market value of any financial security is simply the sum of discounted values of the cash flows that the security is expected to produce. The timing and magnitude of these expected cash flows will thus be an integral part of the security's market value, as will the discount rate applied to these expected cash flows, which is the sum of a real default-free interest rate, expected inflation, and possibly several risk premiums. Each of these elements will be influenced by the business cycle. It is through these components that the real economy exerts its influence on the market value of financial instruments.
- The average level of real short-term interest rates is positively related to the trend rate of growth of the underlying economy and also to the volatility of economic growth in the economy. Other things being equal, these relationships mean that we should expect to find that the average level of real short-term interest rates is higher in an economy with high and volatile growth and lower in an economy with lower, more stable growth.

- On average, over time, according to the Taylor rule, a central bank's policy rate should comprise the sum of an economy's trend growth plus inflation expectations, which might, in turn, be anchored to an explicit inflation target. This policy rate level is referred to as the neutral rate. Other things being equal, when inflation is above (below) the targeted level, the policy rate should be above (below) the neutral rate, and when the output gap is positive (negative), the policy rate should also be above (below) the neutral rate. The policy rate can thus vary over time with inflation expectations and the economy's output gap.
- Short-term nominal rates will be closely related to a central bank's policy rate of interest and will comprise the real interest rate that is required to balance the requirements of savers and investors plus investors' expectations of inflation over the relevant borrowing or lending period. Short-term nominal interest rates will be positively related to short-term real interest rates and to inflation expectations.
- If bond investors were risk neutral, then the term structure of interest rates would be determined by short-term interest rate expectations. But bond investors are risk averse, which means that they will normally demand a risk premium for investing in even default-free government bonds. This risk premium will generally rise with the maturity of these bonds because longer-dated government bonds tend to be less negatively correlated with consumption and, therefore, represent a less useful consumption hedge for investors. Overall, the shape of the curve will be determined by a combination of short-term interest rates and inflation expectations as well as risk premiums. In turn, these factors will be influenced by the business cycle and policymakers.
- The yield differential between default-free conventional government bonds and index-linked equivalents will be driven by inflation expectations and a risk premium. The risk premium will be largely influenced by investors' uncertainty about future inflation.
- The difference between the yield on a corporate bond and that on a government bond with the same currency denomination and maturity is referred to as the measured credit spread. It is conceptually akin (but not equal) to the risk premium demanded by investors in compensation for the additional credit risk that they bear compared with that embodied in the default-free government bond. It tends to rise in times of economic weakness, as the probability of default rises, and tends to narrow in times of robust economic growth, when defaults are less common.
- The uncertainty about and time variation in future equity cash flows (dividends) is a distinct feature of equity investment, as opposed to corporate bond investment. This feature explains why we would expect the equity premium to be larger than the credit premium. In times of economic weakness or stress, the uncertainty about future dividends will tend to be higher, and we should thus expect the equity risk premium to rise in such an economic environment.
- Given the uncertain nature of the cash flows generated by equities, investors will demand an equity risk premium because the consumption hedging properties of equities are poor. In other words, equities tend not to pay off in bad times. Because in the event of company failure an equity holder will lose all of his or her investment whereas an investor in the company's bonds may recover a significant portion of his or her investment, it would be reasonable to assume that a risk-averse investor would demand a higher premium on an equity holding than on a corporate bond holding. The two premiums will tend to be positively correlated over time and will tend to be influenced by the business cycle in similar ways.

- The P/E tends to rise during periods of economic expansion and to fall during recessions. A "high" P/E could be the result of a number of factors, including the following: falling real interest rates, a decline in the equity risk premium, an increase in the expectation of future real earnings growth, an expectation of lower operating and/or financial risk, or a combination of all of these factors. All of these components will be influenced by the business cycle.
- The market value of an investment in commercial property can be derived in much the same way as the market value of an investment in equity. The cash flows come in the form of rent, which can be enhanced with additional redevelopment values as leases on properties expire. These cash flows are uncertain, and the uncertainty surrounding them will tend to rise when the economy turns down. We might thus expect the risk premium demanded on commercial property investments to rise in these times.
- The pro-cyclical nature of commercial property prices means that investors will generally demand a relatively high risk premium in return for investing in this asset class. The reason is that commercial property is not a very good hedge against bad economic outcomes. In addition, the illiquid nature of property investment means that investors may also demand a liquidity premium for investing in this asset class.

REFERENCES

Altug, Sumru, and Pamela Labadie. 2008. *Asset Pricing for Dynamic Economies*. Cambridge, UK: Cambridge University Press.

Cochrane, John H. 2005. *Asset Pricing*. Princeton, NJ: Princeton University Press.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2018. *Credit Suisse Global Investment Returns Yearbook 2018*. Credit Suisse Research Institute.

Elton, Edward J., Martin J. Gruber, Stephen J. Brown, and William N. Goetzmann. 2014. *Modern Portfolio Theory and Investment Analysis*, 9th Ed. New York: John Wiley & Sons.

Litterman, R., and J.A. Scheinkman. 1991. "Common Factors Affecting Bond Returns." *Journal of Fixed Income*, vol. 1, no. 1 (June): 54–61.

Shiller, Robert J. 2000. *Irrational Exuberance*. Princeton, NJ: Princeton University Press.

Steeley, James M. 1990. "Modelling the Dynamics of the Term Structure of Interest Rates." *Economic and Social Review*, vol. 21, no. 4 (July): 337–361.

Taylor, John B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, vol. 39: 195–214.

PRACTICE PROBLEMS

1 All else equal, which of the following would *most likely* explain the fall in price of a particular company's shares?

A The expected inflation rate falls.

B The company's future cash flows are expected to increase.

C The yield to maturity on real default-free investments rises.

2 The prices of one-period, real default-free government bonds are likely to be *most* sensitive to changes in:

A investors' inflation expectations.

B the expected volatility of economic growth.

C the covariance between investors' inter-temporal rates of substitution and the expected future prices of the bonds.

3 The covariance between a risk-averse investor's inter-temporal rate of substitution and the expected future price of a risky asset is typically:

A negative.

B zero.

C positive.

4 Default-free real interest rates tend to be relatively high in countries with high expected economic growth because investors:

A increase current borrowing.

B have high inter-temporal rates of substitution.

C have high uncertainty about levels of future consumption.

5 Positive output gaps are usually associated with:

A deflation.

B high unemployment.

C economic growth beyond sustainable capacity.

6 All else equal, an investor expects future inflation to increase, but the uncertainty of future inflation to fall. For such an investor the break-even inflation rate:

A is uncertain.

B is expected to fall.

C is expected to rise.

7 The difference between the yield on a zero-coupon, default-free nominal bond and the yield on a zero-coupon, default-free real bond of the same maturity reflects:

A investors' expectations about future inflation only.

B a premium for the uncertainty of future inflation only.

C both, investors' expectations about future inflation and a premium for the uncertainty of future inflation.

8 One interpretation of an upward sloping yield curve is that the returns to short-dated bonds are:

A uncorrelated with bad times.

© 2017 CFA Institute. All rights reserved.

B more positively correlated with bad times than are returns to long-dated bonds.

C more negatively correlated with bad times than are returns to long-dated bonds.

9 An analyst, who measures yield as a combination of interest rates and premiums, observes an upward-sloping, default-free government bond nominal yield curve. Which of the following statements is correct?

A Interest rates must be expected to rise in the future.

B Bond risk premiums must be expected to rise in the future.

C Expectations relating to the future direction of interest rates are indeterminate.

10 During a recession, the slope of the yield curve for default-free government bonds is *most likely* to:

A flatten.

B steepen.

C become inverted.

11 A corporate bond has a remaining maturity of 1 year, has a face value of EUR100, and is currently priced at EUR90.90. The real risk-free rate is 3.25%. Inflation is expected to be 2.0% next year, and the premium required by investors for inflation uncertainty is 0.25%.

The implied credit risk premium embedded in the bond's price is *best* described as:

A equal to (100/90.90) – 1 = 10%.

B 10% reduced by the real risk-free rate and expected inflation.

C 10% reduced by the real risk-free rate, expected inflation, and the premium for inflation uncertainty.

12 A decrease in the prices of AAA-rated corporate bonds during a recession would *most likely* be the result of:

A expectations of higher inflation.

B increases in credit risk premiums.

C increases in short-term, default-free interest rates.

13 During an economic period when spreads between corporate and government bonds are narrowing, and spreads between higher- and lower-rated corporate bond categories are also narrowing, it can be expected that:

A government bonds will outperform corporate bonds.

B lower-rated corporate bonds will outperform higher-rated corporate bonds.

C higher-rated corporate bonds will outperform lower-rated corporate bonds.

14 The sensitivity of a corporate bond's spread to changes in the business cycle is *most likely* to be:

A uncorrelated with the level of cyclicality in the company's business.

B positively correlated with the level of cyclicality in the company's business.

C negatively correlated with the level of cyclicality of the company's business.

15 The category of bonds whose spreads can be expected to widen the *most* during an economic downturn are bonds from the:

A cyclical sector with low credit ratings.

B cyclical sector with high credit ratings.

C non-cyclical sector with low credit ratings.

16 When assessing investment opportunities in equities, investors should:

A assign higher equity risk premiums to non-cyclical companies, relative to cyclical companies.

B forecast lower volatility in the growth rate of earnings for cyclical companies, relative to non-cyclical companies.

C forecast higher growth rates in earnings for cyclical companies coming out of a recession, relative to non-cyclical companies.

17 Risk-averse investors demanding a large equity risk premium are *most likely* expecting their future consumption outcomes and equity returns to be:

A uncorrelated.

B positively correlated.

C negatively correlated.

18 Which of the following financial assets is likely to offer the *most* effective hedge against bad consumption outcomes?

A Equities.

B Short-dated, default-free government bonds.

C Long-dated, default-free government bonds.

19 Other things equal, equilibrium price-to-earnings ratios (P/Es) will *most likely* decrease if:

A real interest rates decrease.

B inflation is expected to increase.

C there is less uncertainty about future inflation.

20 Which of the following statements relating to commercial real estate is correct?

A Rental income from commercial real estate is generally unstable across business cycles.

B Commercial real estate investments generally offer a good hedge against bad consumption outcomes.

C The key difference in the discount rates applied to the cash flows of equity investments and commercial real estate investments relates to liquidity.

21 With regard to the credit risk of the sovereign debt issued by country governments, which of the following statements is correct? The credit risk premium on such debt is:

A zero because governments can print money to settle their debt.

B negligibly small because no country has defaulted on sovereign debt.

C a non-zero and positive quantity that varies depending on a country's creditworthiness.

The following information relates to Questions 22–28

Julie Carlisle is a financial planner at a large wealth management firm. One of her clients, Esteban Blake, just received a sizable inheritance. He invests a portion of the inheritance in an annuity that will immediately increase his income by a substantial amount. He enlists Carlisle's help to invest the remaining amount of the inheritance.

Blake informs Carlisle that he would like some short-term bonds in his portfolio. Carlisle proposes purchasing a one-year domestic government zero-coupon bond. It has a face value of $100 and is currently priced at $96.37. Carlisle estimates the one-year real risk-free rate at 1.15% and expects inflation over the next year to be 2.25%.

In an effort to provide Blake with some exposure to international markets, Carlisle proposes three countries to look for investment opportunities. Selected data on the three countries are presented in Exhibit 1.

Exhibit 1 Selected Macroeconomic Data

	Nominal GDP Growth	Inflation Rate	Volatility of Real GDP Growth	Yield Curve Shape	Trailing 12-Month Equity Index P/E
Country #1	6.5%	4.0%	Low	Flat	16.5
Country #2	5.0%	2.5%	High	Upward slope	17.3
Country #3	3.5%	2.0%	Low	Flat	18.2

In her analysis, Carlisle observes that the spread between the three-year default-free nominal bond and the default-free real zero-coupon bond in Country #3 is 2.0%.

Blake expresses concern that stocks may be currently overvalued in Country 3 given its 20-year historical equity index P/E of 16.0. Carlisle comments,

> I think the equilibrium P/E in Country #3 has increased because of changes in market conditions.

Carlisle predicts that Country #3 will slip into a recession next quarter. She thinks it will be short-lived, lasting only 12 months or so, and considers the impact of such a recession on the performance of the country's stocks and bonds.

Exhibit 2 Three-Year Corporate Bonds from Country #3

Corporate Bond	Moody's Investors Service Rating	Spread*
Bond A	Aaa	1.4%
Bond B	Baa1	3.2%
Bond C	B3	5.3%

*Spread versus three-year sovereign bond

22 Holding all else constant, the change in Blake's income will *most likely* result in:

- **A** an increase in his marginal utility of consumption.
- **B** an increase in his inter-temporal rate of substitution.
- **C** a decrease in his required risk premium for investing in risky assets.

23 The implied premium for inflation uncertainty for the one-year government zero-coupon bond proposed by Carlisle is *closest* to:

- **A** 0.23%.
- **B** 0.37%.
- **C** 1.10%.

24 Based on the data in Exhibit 1, current real short-term interest rates would *most likely* be highest in:

A Country #1.

B Country #2.

C Country #3.

25 The recent change in Country #3's break-even inflation rate suggests that the expected rate of inflation over the next three years is:

A less than 2.0%.

B equal to 2.0%.

C greater than 2.0%.

26 Which of the following changes in market conditions *best* supports Carlisle's comment regarding the equilibrium P/E for Country #3?

A An increase in the equity risk premium

B A decrease in uncertainty about future inflation

C A decrease in expectation of future real earnings growth

27 If Carlisle's prediction about the economy of Country #3 is realized, the yield curve in Country #3 will *most likely*:

A remain flat.

B become upward sloping.

C become downward sloping.

28 Based on Exhibit 2, if Carlisle's prediction for Country #3 is realized, then over the next 12 months:

A Bond A would be expected to outperform Bond C.

B Bond B would be expected to outperform Bond A.

C Bond C would be expected to outperform Bond B.

SOLUTIONS

1 C is correct. According to the fundamental pricing equation, the market value of an asset is affected by economic factors that influence the asset's expected future cash flows, default-free interest rates, expected inflation rates, or the asset's risk premium. From Equation 1, expected cash flows are in the numerator, while expected inflation and the real risk-free rate are in the denominator. Consequently, a rise in the real risk-free rate (the yield to maturity on a default-free instrument) will lead to a fall in the price of a risky asset, such as stock, by increasing the rate at which its cash flows are discounted.

2 B is correct. Only changes in default-free real interest rates will affect the price of real, default-free bonds. The average level of default-free real interest rates is positively related to the volatility of economic growth in the economy; thus, changes in the expected volatility of economic growth would likely lead to changes in default-free real interest rates, which in turn would affect the prices of real, default-free government bonds.

3 A is correct. For risk-averse investors, when the expected future price of the investment is high (low), the marginal utility of future consumption relative to that of current consumption is low (high). Hence, the covariance of the inter-temporal rate of substitution with asset price is expected to be negative for risk-averse investors.

4 A is correct. The average level of default-free real interest rates is positively related to the expected rate of growth of the underlying economy and also to the volatility of economic growth in the economy. During periods of high expected economic growth, investors are less worried about the future and their consumption abilities in the future; that is, their inter-temporal rate of substitution is low, so they borrow more today and save less. Other things being equal, this means that the average level of default-free real interest rates (the reciprocal of the rate of substitution, see Equation 4) should be higher in an economy with high growth and lower in an economy with lower, more stable growth.

5 C is correct. An economy operating with a positive output gap—that is, where the level of actual GDP exceeds potential GDP—is producing beyond its sustainable capacity. Positive output gaps are usually associated with high and/or rising inflation, while high levels of unemployment usually accompany negative output gaps.

6 A is correct. The break-even inflation rate is the difference between the yield on a zero-coupon, default-free nominal bond and on a zero-coupon, default-free real bond of the same maturity. The rate incorporates changing expectations about inflation and changing perceptions about the uncertainty of the future inflation environment. Consequently, if inflation is expected to rise while the uncertainty about future inflation falls (in Equation 10, $\theta_{t,s}$ rises but $\pi_{t,s}$ falls), it is unclear in which direction break-even inflation rates will move.

7 C is correct. The difference between the yield on a zero-coupon, default-free nominal bond and the yield on a zero-coupon, default-free real bond of the same maturity is known as the break-even inflation rate. This break-even inflation rate will incorporate the inflation expectations of investors over the investment horizon of the two bonds, plus a risk premium to compensate investors for uncertainty about future inflation. Break-even inflation rates are not simply the market's best estimate of future inflation over the relevant investment horizon, because break-even inflation rates also include a risk premium to compensate investors for their uncertainty about future inflation.

8 C is correct. One interpretation of an upward-sloping yield curve is that returns to short-dated bonds are more negatively correlated with bad times than are returns to long-dated bonds. This interpretation is based on the notion that investors are willing to pay a premium and accept a lower return for short-dated bonds if they believe that long-dated bonds are not a good hedge against economic "bad times."

9 C is correct. An upward-sloping yield curve may be caused by a combination of expected rate increases and positive bond risk premiums. It may also be a combination of expectations that interest rates will be unchanged in the future coupled with positive bond risk premiums. Lastly, an upward-sloping yield curve may actually be a reflection of expected rate cuts that are more than offset by the existence of positive bond risk premiums. So, expectations relating to the future direction of interest rates are indeterminate.

10 B is correct. During a recession, short rates are often lower because central banks tend to lower their policy rate in these times because the output gap is likely to be negative. However, the impact of such monetary policy on longer-term rates will not be as strong, so long rates may not fall by as much as short rates. The central bank will usually be expected to bring short-term rates back to normal as the recession recedes, and the risk-free rates will increase as economic growth recovers. Thus, the slope of the yield curve will typically steepen during a recession.

11 C is correct. The implied credit risk premium embedded in the bond's price is the yield (10%) less the default-risk-free nominal interest rate, which includes a premium for inflation uncertainty. See Example 15. The credit risk premium can be calculated as 4.51% in this case:

$$\gamma_{t,s}^{i} = \frac{100}{90.90} - (1 + 0.0325 + 0.02 + 0.0025).$$

$$\gamma_{t,s}^{i} = 4.51\%.$$

12 B is correct. During recessions, the risk premium that investors demand on financial assets, particularly those that are not default-free, such as corporate bonds, may rise because investors in general may be less willing and able to take on heightened default risk during such periods. Specifically, the credit risk premium demanded by investors tends to rise in times of economic weakness, when the probability of a corporate default and bankruptcy is highest.

13 B is correct. When spreads are narrowing, investors seem to be less discerning between issues with weak versus strong credit, and the rate of improvement will tend to be greater for those bonds issued by entities with a relatively weaker ability to pay. Thus, during times when corporate bond spreads are narrowing relative to government bonds and the spreads between higher- and lower-rated bond categories are also narrowing, corporate bonds will generally outperform government bonds and lower-rated corporate bonds will tend to outperform higher-rated corporate bonds.

14 B is correct. The sensitivity of a corporate bond's spread to changes in the business cycle and the level of cyclicality tend to be positively correlated. The greater the level of cyclicality, the greater the sensitivity of the bond's spread to changes in the business cycle.

15 A is correct. During an economic downturn, the spreads of corporate bonds can be expected to widen, because the risk premium that investors demand on risky financial assets will increase. When spreads widen, the spreads on bonds issued by corporations with a low credit rating and that are part of the cyclical sector will tend to widen the most.

16 C is correct. During recessions, cyclical companies are likely to experience sharp declines in earnings, more so than non-cyclical companies. In contrast, while coming out of a recession, cyclical companies are likely to generate higher earnings growth relative to non-cyclical companies.

17 B is correct. If investors demand high equity risk premiums, they are likely expecting their future consumption and equity returns to be positively correlated. The positive correlation indicates that equities will exhibit poor hedging properties, because equity returns will be high (i.e., pay off) during "good times" and will be low (i.e., not pay off) during "bad times." In other words, the covariance between risk-averse investors' inter-temporal rates of substitution and the expected future prices of equities is highly negative, resulting in a positive and large equity risk premium. This is the case because in good times, when equity returns are high, the marginal value of consumption is low. Similarly, in bad times, when equity returns are low, the marginal value of consumption is high. Holding all else constant, the larger the magnitude of the negative covariance term, the larger the risk premium.

18 B is correct. The relative certainty about the real payoff from short-dated, default-free government bonds and, therefore, the relative certainty about the amount of consumption that the investor will be able to undertake with the payoff indicate that an investment in such bonds would be a good hedge against bad consumption outcomes.

19 B is correct. Other things being equal, an increase in inflation expectations would result in lower equity prices relative to current earnings. This would result in lower equilibrium P/Es.

20 C is correct. To arrive at an appropriate discount rate to be used to discount the cash flows from a commercial real estate investment, a liquidity premium is added to the discount rate applicable to equity investments. The added liquidity premium provides additional compensation for the risk that the real estate investment may be very illiquid in bad economic times.

21 C is correct. Credit premiums have been an important component of the expected return on bonds issued by countries (sovereign debt). The credit premium varies from country to country depending on how creditworthy investors consider it to be. The fact that countries have both printed money to pay back debt and defaulted on it gives rise to a non-zero credit risk premium.

22 C is correct. The additional annuity payment substantially increases Blake's income and wealth, which decreases his marginal utility of consumption. As a result, the average loss of marginal utility from any risk taking decreases as his wealth increases. Thus, he requires a lower risk premium and is willing to buy more risky assets.

23 B is correct. The pricing equation for a default-free nominal coupon-paying bond is

$$P_t^i = \sum_{s=1}^{N} \frac{CF_{t+s}^i}{\left(1 + l_{t,s} + \theta_{t,s} + \pi_{t,s}\right)^s}.$$

For a one-year bond, the pricing formula reduces to

$$P_t = \frac{CF_{t+1}}{\left(1 + l_{t,1} + \theta_{t,1} + \pi_{t,1}\right)^1}.$$

Thus, the implied premium for inflation uncertainty for the one-year government zero-coupon bond is calculated as

$$\pi_{t,1} = \frac{CF_{t+1}}{P_t} - \left(1 + l_{t,1} + \theta_{t,1}\right)$$

$$= \frac{100}{96.37} - (1 + 0.0115 + 0.0225)$$

$$= 1.0377 - 1.0340$$

$$= 0.0037\text{, or } 0.37\%.$$

24 B is correct. Real short-term interest rates are positively related to both real GDP growth and the volatility of real GDP growth. Country 1 and Country 2 have the highest real GDP growth, as estimated by the difference between nominal GDP growth and average inflation (6.5% – 4.0% = 2.5% and 5.0% – 2.5% = 2.5%, respectively), while Country 3 has the lowest real GDP growth (3.5% – 2.0% = 1.5%). Looking at the volatility of real GDP growth, Country 2 has high real GDP growth volatility, whereas Country 1 and Country 3 have low real GDP growth volatility. Therefore, Country 2 would most likely have the highest real short-term interest rates.

25 A is correct. The difference, or spread, between the yields on the country's three-year default-free nominal bond and on the default-free real zero-coupon bond is 2.0%. This spread is known as the break-even rate of inflation (BEI), which is composed of the expected rate of inflation plus a risk premium for the uncertainty of future inflation. Because this risk premium component is most likely positive, because investors are unlikely to be very confident in their ability to predict inflation accurately, the expected rate of inflation component would be less than 2.0%.

26 B is correct. Stock prices are a function of expected cash flows discounted by inflation expectations, the uncertainty of future inflation, and the equity risk premium, among other factors. Holding all else equal, a decline in the uncertainty of future inflation would result in lower discount rates and higher valuations. This result would support a higher equilibrium P/E, thus justifying Country 3's current trailing P/E being higher than its historical average.

27 B is correct. The yield curve in Country 3 is currently flat (Exhibit 1), and Carlisle predicts a recession. During a recession, short-term rates tend to be lower because central banks tend to lower their policy rate in these times. However, the impact of monetary policy on longer-term rates will not be as strong because the central bank will usually be expected to bring short-term rates back to normal as the recession recedes. Thus, the slope of the yield curve will likely become upward sloping during the recession.

28 A is correct. If Country 3 experiences a recession over the next 12 months, the credit spreads for corporate bonds would be expected to widen as investors sell the low-quality debt of issuers with high default risk and trade up to the higher-quality debt of issuers with low default risk. The issuers with a good credit rating (such as Aaa rated Bond A) tend to outperform those with lower ratings (such as B3 rated Bond C) as the spread between low- and higher-quality issuers widens. As a result, Bond A would be expected to outperform Bond C over the next 12 months.

READING

43

Analysis of Active Portfolio Management

by Roger G. Clarke, PhD, Harindra de Silva, PhD, CFA, and Steven Thorley, PhD, CFA

Roger G. Clarke, PhD (USA). Harindra de Silva, PhD, CFA, is at Analytic Investors, Wells Fargo Asset Management (USA). Steven Thorley, PhD, CFA, is at the Marriott School, BYU (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to:*
☐	**a.** describe how value added by active management is measured;
☐	**b.** calculate and interpret the information ratio (*ex post* and *ex ante*) and contrast it to the Sharpe ratio;
☐	**c.** describe and interpret the fundamental law of active portfolio management, including its component terms—transfer coefficient, information coefficient, breadth, and active risk (aggressiveness);
☐	**d.** explain how the information ratio may be useful in investment manager selection and choosing the level of active portfolio risk;
☐	**e.** compare active management strategies, including market timing and security selection, and evaluate strategy changes in terms of the fundamental law of active management;
☐	**f.** describe the practical strengths and limitations of the fundamental law of active management.

INTRODUCTION

1

The Markowitz (1952) framework of what was originally called modern portfolio theory (MPT) has now become the prominent paradigm for communicating and applying principles of risk and return in portfolio management. Much of the mathematics and terminology of mean–variance portfolio theory was subsequently combined with the notion of informational efficiency by Sharpe (1964) and other financial economists to develop equilibrium models, such as the traditional capital asset pricing model. Separately, the tools of MPT were applied by Treynor and Black (1973) to guide investors in their selection of securities when prices differ from their equilibrium values.

© 2020 CFA Institute. All rights reserved.

The application of portfolio theory to active management was further developed by Grinold (1989) in "The Fundamental Law of Active Management" and by Black and Litterman (1992).

We summarize the principles of active portfolio management using the terminology and mathematics of the fundamental law introduced by Grinold (1989) and further developed by Clarke, de Silva, and Thorley (2002). Active management theory deals with how an investor should construct a portfolio given an assumed competitive advantage or skill in predicting returns. Thus, active management relies on the assumption that financial markets are not perfectly efficient. Although investors might ultimately care about total risk and return, when asset management is delegated to professional investors in institutional settings (e.g., pension funds) the appropriate perspective is risk and return relative to a benchmark portfolio. In addition to the principal–agent problem in delegated asset management, the availability of passively managed portfolios requires a focus on value added above and beyond the alternative of a low-cost index fund.

We assume an understanding of basic portfolio theory, including the mathematics of expected values, variances, and correlation coefficients, as well as some familiarity with the related disciplines of mean–variance optimization and multi-factor risk models. The following sections introduce the mathematics of value added through active portfolio management, including the concepts of active weights, relative returns, and performance attribution systems. The subsequent section compares the well-known Sharpe ratio for measuring the total risk-adjusted value added with the information ratio for measuring relative risk-adjusted value added. This section also makes a distinction between *ex ante*, or expected, risk and return versus *ex post*, or realized, risk and return and explains that the information ratio is the best criterion for evaluating active investors. We then introduce the fundamental law that describes how relative skill, breadth of application, active management aggressiveness, and the constraints in portfolio construction combine to affect value added. The remaining sections provide examples of active portfolio management strategies in both the equity and fixed-income markets, describe some of the practical limitations of the fundamental law, and provide a summary of the concepts and principles.

2 ACTIVE MANAGEMENT AND VALUE ADDED

a describe how value added by active management is measured;

The objective of active management is to add value in the investment process by doing better than a benchmark portfolio. Value added is a relative performance comparison to investing in the benchmark portfolio, often called passive investing. If the investor outperforms the benchmark portfolio, value added is positive. If the investor underperforms the benchmark portfolio, value added is negative. In the latter case, the investor would have been better off during the measurement period by simply holding the benchmark portfolio, particularly net of fees and expenses. Examples of indexes that are used as benchmark portfolios include the MSCI All Country World Index and the Bloomberg Barclays Global Aggregate Bond Index, which represent the performance of global equities and global bonds, respectively.

2.1 Choice of Benchmark

A benchmark or passive portfolio should have a number of qualities to serve as a relevant comparison for active management:

- The benchmark is representative of the assets from which the investor will select.
- Positions in the benchmark portfolio can actually be replicated at low cost.
- Benchmark weights are verifiable *ex ante*, and return data are timely *ex post*.

An available security market index is often used as the benchmark portfolio. The most common market indexes weight the individual assets by their market capitalization. Capitalization weighting has played a prominent role in the development of capital market theory because such indexes are generally self-rebalancing and can be simultaneously held by many investors. Float-adjusted market capitalization-weighted indexes represent an incremental improvement over non-float-adjusted indexes by accounting for the percentage of a security or asset that is not privately held and thus available to the general investing public. One important consequence of using a float-adjusted capitalization-weighted market index as the benchmark is that when all relevant assets are included in the market, the value added from active management becomes a zero-sum game with respect to the market. Because the market portfolio represents the average performance across all investors that own securities before costs, active investors as a group cannot outperform the market (i.e., active management is a zero-sum game). For benchmarks that have a narrower definition than the total market, active management is not a zero-sum game because investors can select assets outside the benchmark.

The return on the benchmark portfolio, R_B, is based on the returns to the individual securities and the weights of each security in the portfolio:

$$R_B = \sum_{i=1}^{N} w_{B,i} R_i \tag{1}$$

where R_i is the return on security i, $w_{B,i}$ is the benchmark weight of security i, and N is the number of securities. Similarly, the return on an actively managed portfolio, R_P, is a function of the weights of the securities, i, held in the portfolio, $w_{P,i}$, and the returns to the individual securities:

$$R_P = \sum_{i=1}^{N} w_{P,i} R_i \tag{2}$$

The benchmark might include securities that are not part of the actively managed portfolio and thus would have a weight of zero by definition or simply be left out of the calculation in Equation 2. Similarly, an investor could include securities in the active portfolio that are not in the benchmark, and those would have a benchmark weight of zero in Equation 1. Please note that for simplicity, the same notation, N, is used in the summation in the expression for the managed portfolio return and the benchmark return, although fewer or more securities may be in the managed portfolio than in the benchmark.

2.2 Measuring Value Added

The value added or "active return" of an actively managed portfolio is typically calculated as the simple difference between the return on that portfolio and the return on the benchmark portfolio,

$R_A = R_P - R_B$,

and can thus be either positive or negative. A risk-adjusted calculation of value added, which we will refer to as the managed portfolio's alpha, incorporates some estimate of the managed portfolio's risk relative to the benchmark, often captured by the portfolio's beta, $\alpha_P = R_P - \beta_P R_B$. Unfortunately, the term *alpha* in practice is often used to refer to active return as well, which implicitly assumes that the beta of the managed portfolio relative to the benchmark is 1.

Equations 1 and 2 can be combined to illustrate the important principle that value added is ultimately driven by the differences in managed portfolio weights and benchmark weights: $\Delta w_i = w_{P,i} - w_{B,i}$. These values are called the active weights of the managed portfolio, and the symbol Δ (Greek letter delta) is used to indicate the difference from the benchmark weights. Combining Equations 1 and 2 and employing this definition for active weights yields the conceptually important result that value added is the sum product of the active weights and asset returns:

$$R_A = \sum_{i=1}^{N} \Delta w_i R_i$$

Given that the sum of the active weights is zero, we can also write the value added as the sum product of active weights and active security returns:

$$R_A = \sum_{i=1}^{N} \Delta w_i R_{Ai} \tag{3}$$

where $R_{Ai} = R_i - R_B$. Equation 3 indicates that positive value added is generated when securities that have returns greater than the benchmark are overweighted and securities that have returns less than the benchmark are underweighted.

Whereas many applications of value added focus on individual securities as the assets, we first illustrate the concept with a simple numerical example of a composite portfolio that has just two assets—a stock portfolio and a bond portfolio. Suppose the benchmark is a 60/40 weighted composite portfolio of stocks and bonds. The investor believes that over the next year stocks will outperform bonds, so the investor holds a portfolio that is weighted 70% stocks and 30% bonds. The managed portfolio is said to be *overweight* stocks by 10 percentage points and *underweight* bonds by 10 percentage points—in other words, an active weight of −10 percentage points on bonds. Assume that *ex post* (i.e., "after the fact"), the return on the stock market turned out to be 14.0% and the return on the bond market turned out to be just 2.0%. In this case, the return on the managed portfolio is 0.70(14.0) + 0.30(2.0) = 10.4% and the return on the benchmark is 0.60(14.0) + 0.40(2.0) = 9.2%.

From these final numbers, one could directly calculate the value added as 10.4 − 9.2 = 1.2%. But using Equation 3, a more informative calculation of value added showing the contributions from each segment is R_A = 0.10(14.0 − 9.2) − 0.10(2.0 − 9.2) = 0.5 + 0.7 = 1.2%. This breakout suggests that a 0.5% return relative to the benchmark was generated by being overweight stocks, and a 0.7% return was generated simultaneously by being underweight bonds—for a total of 1.2%. Of course, the actual returns might have been different—with the stock market return being lower than the bond market return, resulting in negative value added in the managed portfolio. For example, if the stock market had a return of −14.0% instead of +14.0%, the portfolio return and benchmark return would have been −9.2% and −7.6%, respectively. Then the value added from this single overweight/underweight decision would have been R_A = 0.10(−14.0) − 0.10(2.0) = −1.4% − 0.2% = −1.6%.

EXAMPLE 1

Value Added and Country Equity Markets

Consider the MSCI EAFE Index as the benchmark for an actively managed portfolio that includes allocations to individual countries, as given in the following exhibit. The portfolio (both benchmark and managed) weights are for the beginning of 2018. The portfolio manager actively changes country allocations but does not engage in security selection.

Country	Benchmark Weight	Portfolio Weight	2018 Return
United Kingdom	17%	16%	−7.6%
Japan	25%	14%	−9.0%
France	11%	8%	−3.5%
Germany	9%	24%	−15.8%
Other Countries	38%	38%	−0.1%

Source: Data from MSCI.

1 Which countries have the largest overweight and largest underweight in the managed portfolio compared with the benchmark portfolio? What are the active weights for these two countries?

2 Using active weights and total returns, what was the value added of the managed portfolio over the benchmark portfolio in the calendar year 2018?

Solution to 1:

Germany has the largest overweight at 24 – 9 = +15%, and Japan has the largest underweight at 14 – 25 = −11%.

Solution to 2:

The value added is $-0.01(-7.6) - 0.11(-9.0) - 0.03(-3.5) + 0.15(-15.8) = -1.2\%$. Note that the "Other Countries" active weight is zero, so this asset does not contribute anything to the portfolio's active return. The value added can also be calculated using relative returns in Equation 3 with the same net result.

2.3 Decomposition of Value Added

In contrast to the previous simple example, performance attribution systems often attempt to decompose the value added into *multiple* sources. The most common decomposition is between value added due to asset allocation and value added due to security selection. Consider a composite portfolio of stocks and bonds where the asset allocation weights differ from a composite benchmark *and* each asset class is actively managed by selecting individual securities. The total value added is the difference between the actual portfolio return and the benchmark return:

$$R_A = \sum_{j=1}^{M} w_{P,j} R_{P,j} - \sum_{j=1}^{M} w_{B,j} R_{B,j}$$

The first summation has both portfolio weights and the returns on actively managed portfolios, designated by the "P" subscript. The second summation has both benchmark weights and benchmark returns, designated by the "B" subscript. The subscript $j = 1$ to M counts the number of asset classes, leaving the notation subscript $i = 1$ to N for use elsewhere to count the securities within each asset class.

We can rewrite the total value added as the sum of the active asset allocation decisions and the weighted sum of the value added from security selection, $R_{A,j} = R_{P,j} - R_{B,j}$, within each asset class:

$$R_A = \sum_{j=1}^{M} \Delta w_j R_{B,j} + \sum_{j=1}^{M} w_{P,j} R_{A,j} \tag{4}$$

although this formulation arbitrarily assigns an interactive effect to security selection. The performance attribution system in Equation 4 may be easier to conceptualize with just two asset classes, stocks and bonds (in other words, with $M = 2$). Using *stocks* and *bonds* as the subscripts, Equation 4 becomes:

$$R_A = \left(\Delta w_{stocks} R_{B,stocks} + \Delta w_{bonds} R_{B,bonds}\right) + \left(w_{P,stocks} R_{A,stocks} + w_{P,bonds} R_{A,bonds}\right)$$

The first (parenthetical) term is the value added from the asset allocation decision. The second term is the value added from security selection within the stock and bond portfolios. The active weights in the first term refer to differences from the policy portfolio. For example, the long-term policy portfolio might be 60/40 stocks versus bonds, and the investor deviates from this policy portfolio from year to year based on beliefs about the returns to each asset class.

To give a numerical example, consider the fund returns for the calendar year 2018 in the following table.

Fund	Fund Return (%)	Benchmark Return (%)	Value Added (%)
Fidelity Magellan	–5.6	–4.5	–1.1
PIMCO Total Return	–0.3	0.0	–0.3
Portfolio Return	–3.9	–2.7	–1.2

Specifically, the Fidelity Magellan mutual fund had a return of –5.6%, compared with a –4.5% return for its benchmark, the S&P 500 Index. In the same year, the PIMCO Total Return Fund had a return of –0.3%, compared with a 0.0% return for its benchmark, the Bloomberg Barclays US Aggregate Index. Consider an investor who invested in both actively managed funds, with 68% of the total portfolio in Fidelity and 32% in PIMCO. Assume that the investor's policy portfolio (strategic asset allocation) specifies weights of 60% for equities and 40% for bonds.

- As shown in the table, Fidelity Magellan added value of $R_A = R_P - R_B = -5.6\% - (-4.5)\% = -1.1\%$, and PIMCO Total Return added value of $R_A = R_P - R_B = -0.3\% - (0.0\%) = -0.3\%$. These value added numbers represent the skill in security selection within each individual fund.
- Using the actual weights of 68% and 32% in the Fidelity and PIMCO funds, the combined value added from security selection was 0.68(–1.1%) + 0.32(–0.3%) = –0.8%.
- The active asset allocation weights in 2018 were 68% – 60% = +8% for equities and −8% for bonds, so the value added by the active asset allocation decision was 0.08(–4.5%) – 0.08(0.0%) = –0.4%. The total value added by the investor's active asset allocation decision *and* by the mutual funds through security selection was –0.8% – 0.4% = –1.2%. To confirm this total value added, note that the

return on the investor's portfolio was 0.68(−5.6%) + 0.32(−0.3%) = −3.9% and the return on the policy portfolio was 0.60(−4.5%) + 0.40(0.0%) = −2.7%, for a difference of −3.9% − (−2.7) = −1.2%.

Performance attribution systems can be expanded to include several asset classes—for example, stocks, bonds, real estate, and cash (in other words, with M = 4 in Equation 4). For a given asset class, the performance attribution system might also include value added from the selection of industries or sectors relative to the benchmark. For example, an equity portfolio might measure value added from over- and underweighting different industry sectors, as well as individual stock selection within those sectors, and a fixed-income portfolio might decompose value added from the mix of sovereign government bonds versus corporate bonds, as well as individual bond selection.

In summary, deviations from portfolio benchmark weights drive the value added by active portfolio management. If every asset in the managed portfolio is held at its benchmark weight, there would be no value added relative to the benchmark. The total value added can be decomposed into various sources that capture the contribution from different decisions, such as asset allocation and security selection.

3 COMPARING RISK AND RETURN, THE SHARPE RATIO AND THE INFORMATION RATIO

b calculate and interpret the information ratio (*ex post* and *ex ante*) and contrast it to the Sharpe ratio;

The risk–return trade-off of a portfolio can be represented in either *absolute* or *relative* terms. The Sharpe ratio provides an absolute expected (*ex ante*) or realized (*ex post*) reward-to-risk measure. As we have noted, however, value added is a relative return comparison. The information ratio provides a benchmark relative expected (*ex ante*) or realized (*ex post*) reward-to-risk measure.

3.1 The Sharpe Ratio

The Sharpe ratio is used to compare the portfolio return in excess of a riskless rate with the volatility of the portfolio return. The ratio provides a measure of how much the investor is receiving in excess of a riskless rate for assuming the risk of the portfolio. The Sharpe ratio, SR_P, is calculated for any portfolio, either actively managed or a benchmark, using the formula

$$SR_P = \frac{R_P - R_F}{\sigma_P} \quad (5)$$

where R_P is the portfolio return, R_F is the risk-free rate, and σ_P is the standard deviation of the portfolio return. In this context, the standard deviation of the portfolio return is often called either volatility or total risk. The Sharpe ratio can be used as an *ex ante* measure of *expected* return and risk, in which case the general formula in Equation 5 would have the expected portfolio return, $E(R_P)$, minus the risk-free rate in the numerator and a forecast of volatility in the denominator. As subjective forecasts, the expected return and standard deviation of return will likely vary among different investors.

The Sharpe ratio can also be used to measure the *ex post* or *realized* performance of a portfolio over some time period. In that case, when applied to multiple time periods, the numerator in Equation 5 is the difference between the average realized portfolio return, $\overline{R_P}$, and the average risk-free rate, $\overline{R_F}$, and the denominator in Equation 5 is the sample standard deviation. The convention for Sharpe ratios is to annualize both the portfolio average return and the portfolio risk. For example, if the past return data are measured monthly, the average monthly return can be multiplied by 12 and the monthly return volatility can be multiplied by the square root of 12. The logic for multiplying the standard deviation by the *square root* of 12 is that variance (i.e., standard deviation squared), under certain assumptions, increases proportionally with time.

Although this scaling convention is common in practice, multiplying monthly returns by a factor of 12 for averages and the square root of 12 for standard deviations ignores the multiplicative (i.e., compound) nature of returns over time. Simple multiplication factors (e.g., 250 and the square root of 250 for annualizing trading-day returns) are only technically correct if the underlying returns are independent and continuously compounded or logarithmic. Similarly, annualized compound returns for the two values in the numerator of the Sharpe ratio (i.e., the portfolio return and the riskless rate) may be used instead of the annualized difference of arithmetic returns. The various methodologies produce slightly different results but should not be a serious problem as long as comparisons between different portfolios use the same approach.

Exhibit 1 Benchmark Sharpe Ratios for 1994–2018 (based on a risk-free rate of 2.3%)

	MSCI World	S&P 500	Russell 2000	MSCI EAFE	Bloomberg Barclays US Aggregate
Average annual return	7.9%	9.9%	10.3%	6.3%	5.0%
Return standard dev.	14.5%	14.4%	19.1%	15.8%	3.5%
Sharpe ratio	0.38	0.53	0.41	0.25	0.77

Exhibit 1 reports the annualized monthly historical return data (not compounded) in US dollars for several different benchmark portfolios for the 25-year period from 1994 to 2018. Long-term *ex post* Sharpe ratios for equity benchmarks have typically fallen within a range of 0.20–0.60, although over a shorter horizon they will vary over a wider range and can be either negative or positive. The Sharpe ratio for the Bloomberg Barclays US Aggregate fixed-income benchmark in Exhibit 1 is particularly high because of the secular decline in interest rates over this 25-year period that boosted the average return for fixed income. Exhibit 2 reports historical return data and Sharpe ratios from 1994 to 2018 for some well-known actively managed mutual funds over the same period. The Sharpe ratios in both exhibits are based on a risk-free rate of 2.3%, the average annualized US Treasury bill return during this 25-year period. The comparison of Sharpe ratios between funds intentionally uses data from the same measurement period. One should not compare the Sharpe ratio of one fund over one period with that of another fund over a different period.

Exhibit 2 Active Fund Sharpe Ratios for 1994–2018 (based on a risk-free rate of 2.3%)

	Fidelity Magellan	Growth Fund of America	Templeton World	T. Rowe Price Small Cap	JPMorgan Bond
Average annual return	8.5%	11.1%	7.9%	11.6%	5.2%
Return standard dev.	16.5%	15.7%	15.2%	16.7%	3.6%
Sharpe ratio	0.38	0.56	0.37	0.56	0.80

Note: The selection of funds for illustration was made without any intended implication, positive or negative, concerning their performance relative to other possible choices.

An important property is that the Sharpe ratio is unaffected by the addition of cash or leverage in a portfolio. Consider a combined portfolio with a weight of w_P on the actively managed portfolio and a weight of $(1 - w_P)$ on risk-free cash. The return on the combined portfolio is $R_C = w_P R_P + (1 - w_P)R_F$, and the volatility of the combined portfolio is just $\sigma_C = w_P\sigma_P$ because the $(1 - w_P)R_F$ portion is risk free. Applying these two relationships in Equation 5 gives the Sharpe ratio for the combined portfolio as

$$\mathrm{SR}_C = \frac{R_C - R_F}{\sigma_C} = \frac{w_P\left(R_P - R_F\right)}{w_P\sigma_P} = \mathrm{SR}_P$$

which is the same as the Sharpe ratio of the actively managed portfolio. Note that the weight in the combined portfolio, w_P, could be greater than 1, so $(1 - w_P)$ could be negative, indicating that leverage created by *borrowing* risk-free cash and investing in risky assets also does not affect the portfolio's Sharpe ratio.

The proposition that independent of preferences investors should form portfolios using two funds—one of which is the risk-free asset and the other the risky asset portfolio with the highest Sharpe ratio—is known as two-fund separation. On the one hand, if the expected volatility of the risky asset portfolio is higher than the investor prefers, the volatility can be reduced by holding more cash and less of the risky portfolio. On the other hand, if the expected volatility of the risky portfolio is lower than the investor allows, the volatility and expected return can be increased by leverage. For example, suppose an investor believes the performance of the Growth Fund of America shown in Exhibit 2 will repeat going forward but only allows a volatility of 10%. The investor might invest 64% of assets in the Growth Fund of America and 36% in cash to reduce overall portfolio risk. The expected return of the combined portfolio is 0.64(11.1%) + 0.36(2.3%) = 7.9%. The volatility of the combined portfolio is 0.64(15.7%) = 10.0%. The Sharpe ratio of the combined portfolio is (7.9% – 2.3%)/10.0% = 0.56, the same as the 0.56 Sharpe ratio of the Growth Fund of America shown in Exhibit 2.

EXAMPLE 2

Adjusting Risk and Return Using the Sharpe Ratio

Consider an investor choosing between two risky portfolios: a large-cap stock portfolio and a small-cap stock portfolio. Although forecasts about the future are subjective, suppose for simplicity that the investor expects that the future statistics will be those in the following table, with a risk-free rate of 2.3%. The forecasted 0.42 Sharpe ratio of the small-cap portfolio is higher than the 0.40 ratio of the large-cap portfolio, but suppose the investor does not want the high 19.2% volatility associated with the small-cap stocks.

	Large Cap	Small Cap
Expected return	8.2%	10.3%
Expected volatility	14.6%	19.2%
Sharpe ratio	0.40	0.42

1 What percentage of the portfolio would an investor need to hold in cash to reduce the risk of a portfolio invested in the small-cap portfolio and cash to the same risk level as that of the large-cap portfolio?

2 Based on your answer to 1, calculate the Sharpe ratio of the small-cap plus cash portfolio.

3 Compare the expected return of the small-cap plus cash portfolio with the expected return of the large-cap portfolio.

Solution to 1:

We want to reduce the 19.2% volatility to 14.6% by adding cash. The weight of small-cap stocks in the combined portfolio must therefore be 14.6/19.2 = 76%, leaving a 24% weight in risk-free cash. With that amount of cash, the volatility of the combined portfolio will be 0.76(19.2%) = 14.6%, the same as the large-cap portfolio.

Solution to 2:

The Sharpe ratio of the combined portfolio is unaffected by the amount in cash, so it remains 0.42.

Solution to 3:

The expected return of the combined portfolio is 0.76(10.3%) + 0.24(2.3%) = 8.4%, 20 basis points (bps) higher than the 8.2% expected return on the large-cap portfolio, but with the same risk as the large-cap portfolio. To reconfirm, the Sharpe ratio of the combined portfolio is (84% – 2.3%)/14.6% = 0.42, the same as the original 0.42 value.

3.2 The Information Ratio

The simplest definition of the information ratio compares the active return from a portfolio relative to a benchmark with the volatility of the active return, called "active risk" or "benchmark tracking risk." The information ratio can be thought of as a way to measure the consistency of active return, as most investors would prefer a more evenly generated value added (low active risk) rather than a lumpy active return pattern. Like the more formal distinction between active portfolio return and alpha, active risk has a more exact beta-adjusted counterpart, which Grinold and Kahn (1999) called "residual risk." In this discussion, the information ratio is based on the implicit assumption that the beta of the managed portfolio relative to the benchmark is exactly 1.0, although in practice that assumption can be relaxed. For example, Fischer and Wermers (2013) present the information ratio that does not assume beta is 1.

The information ratio tells the investor how much active return has been earned, or is expected to be earned, for incurring the level of active risk. Active return, R_A, is the difference between the managed portfolio return, R_P, and the benchmark portfolio return, R_B. The information ratio of an actively managed portfolio, IR, is calculated by dividing the active return by active risk:

$$\mathrm{IR} = \frac{R_P - R_B}{\sigma\left(R_P - R_B\right)} = \frac{R_A}{\sigma_A} \tag{6}$$

where σ(·) is the standard deviation function—in this case, the standard deviation of the excess return of the portfolio (R_P) over the return of the benchmark (R_B). As with the Sharpe ratio, the typical convention is to annualize both the active return and the active risk. The information ratio can refer to the investor's *ex ante*, or forecasted, active return. Thus, the numerator in Equation 6 would be replaced by the expected returns—that is, $E(R_A) = E(R_P) - E(R_B)$—and the denominator would be the expected active risk. Alternatively, the calculation of an *ex post*, or historical, information ratio would use realized average active returns and the realized sample standard deviation of the active return.

Two investment strategies and associated terminology can help reinforce the conceptual distinction between the Sharpe ratio and the information ratio. First, a "closet index fund" (a fund that advertises itself as being actively managed but is actually close to being an index fund) will have a Sharpe ratio that is close to the benchmark because the excess return and volatility will be similar to the benchmark. However, the closet index fund will have a small amount of active risk, although positive by definition like any volatility estimate. While there may be little active risk, the information ratio of a closet index fund will likely be close to zero or slightly negative if value added cannot overcome the management fees. If one has the actual holdings of the fund, closet indexing is easy to detect on the basis of a measurement called "active share," a measure of how similar a portfolio is to its benchmark. [Cremers and Petajisto (2009) defined active share as half the sum of the absolute values of the active weights.] As a second example, the Sharpe ratio and the information ratio for a market-neutral long–short equity fund (a fund with offsetting long and short positions that has a beta of zero with respect to the market) would be identical if we consider the benchmark to be the riskless rate because the excess return and active return would be the same calculation, as would be total risk and active risk.

Exhibit 3 shows historical information ratios for the mutual funds in Exhibit 2, with the benchmark portfolio for each calculation shown at the bottom of Exhibit 3. The average active return in the first row of Exhibit 3 can be calculated by subtracting the specified benchmark average return in Exhibit 1 from the average fund return in Exhibit 2. The active risk is the annualized standard deviation of the return differences from 1994 to 2018, which cannot be verified with just the summary data in Exhibits 1 and 2.

As shown in Exhibit 3, *ex post* information ratios will be negative if the active return is negative. In fact, under the zero-sum property of active management, the average realized information ratio across investment funds with the same benchmark should be about zero. The realized information ratios in Exhibit 3 are within a range of about −0.30 to +0.30, although the range would be much wider over shorter periods. Of course, *ex ante*, or before the fact, if an investor did not expect the information ratio to be positive, he or she would simply invest in the benchmark. Note that ranking by active risk, a relative measure, does not necessarily equate to ranking by total risk, an absolute measure. For example, the relative risk of Fidelity Magellan in Exhibit 3 is slightly lower than the relative risk of the Growth Fund of America; however, the absolute risk of Fidelity Magellan in Exhibit 2 is slightly higher.

Exhibit 3 Active Fund Information Ratios for 2014–2018

	Fidelity Magellan	Growth Fund of America	Templeton World	T. Rowe Price Small Cap	JPMorgan Bond
Active return	−1.4%	1.2%	0.0%	1.4%	0.2%
Active risk	5.1%	6.2%	5.0%	4.7%	1.0%

(continued)

Exhibit 3 (Continued)

	Fidelity Magellan	Growth Fund of America	Templeton World	T. Rowe Price Small Cap	JPMorgan Bond
Information ratio	−0.27	0.20	0.00	0.29	0.19
Benchmark	S&P 500	S&P 500	MSCI World	Russell 2000	Bloomberg Barclays US Aggregate

Unlike the Sharpe ratio, the information ratio is affected by the addition of cash or the use of leverage. For example, if the investor adds cash to a portfolio of risky assets, the information ratio for the combined portfolio will generally shrink. However, the information ratio of an unconstrained portfolio is unaffected by the aggressiveness of active weights. Specifically, if the active security weights, Δw_i, defined as deviations from the benchmark portfolio weights, are all multiplied by some constant, c, the information ratio of an actively managed portfolio will remain unchanged.

Recall the expression for the active return of a managed portfolio in Equation 3. If each active weight in Equation 3 is multiplied by some constant, c, then the active return on the altered portfolio, R_C, is

$$R_C = \sum_{i=1}^{N} c\Delta w_i R_{Ai} = c\sum_{i=1}^{N} \Delta w_i R_{Ai} = cR_A$$

Similarly, the active risk of the altered portfolio is $c\sigma_A$, so the information ratio of the altered portfolio is

$$\text{IR}_C = \frac{cR_A}{c\sigma_A} = \text{IR}$$

the same as that of the actively managed portfolio with no proportional increase in the active weights. Specifically, if the active weights in a managed portfolio are all doubled, the expected active return (or realized average active return) would be doubled, along with the expected or realized active risk, leaving the information ratio unchanged.

Of course, an outside investor would not be able to adjust the active risk of an existing fund by changing the individual asset active weight positions, but the same objective can be met by taking positions in the benchmark portfolio. For example, if the active risk of a fund is 5.0%, combining that fund in an 80/20 mix with the benchmark portfolio (i.e., a benchmark portfolio weight of 20%) will result in an active risk of the combined portfolio of 0.80(5.0%) = 4.0%, with a proportional reduction in the active return. Similarly, the investor can short sell the benchmark portfolio and use the proceeds to invest in the actively managed fund to increase the active risk and return. Note that in practice, institutional investors might simply reduce the amount they would have otherwise invested in the benchmark portfolio—or, if possible, another actively managed fund—rather than employ an explicit short sell.

4 CONSTRUCTING OPTIMAL PORTFOLIOS

b calculate and interpret the information ratio (*ex post* and *ex ante*) and contrast it to the Sharpe ratio;

An important concept from basic portfolio theory is that with a risk-free asset, the portfolio on the efficient frontier of risky assets that is tangent to a ray extended from the risk-free rate is optimal because it has the highest possible Sharpe ratio. Thus, given the opportunity to adjust absolute risk and return with cash or leverage, the overriding objective is to find the single risky asset portfolio with the maximum Sharpe ratio, whatever the investor's risk aversion. A similarly important property in active management theory is that given the opportunity to adjust active risk and return by investing in both the actively managed and benchmark portfolios, the squared Sharpe ratio of an actively managed portfolio is equal to the squared Sharpe ratio of the benchmark plus the information ratio squared:

$$\mathrm{SR}_P^2 = \mathrm{SR}_B^2 + \mathrm{IR}^2 \tag{7}$$

Equation 7 implies that the active portfolio with the highest (squared) information ratio will also have the highest (squared) Sharpe ratio. (Note that Equation 7 is not practical for comparisons of investment skill involving negative IR because the sign is lost in squaring.) As a consequence, according to mean−variance theory, the expected information ratio is the single best criterion for assessing active performance among various actively managed funds with the same benchmark (Grinold 1989). For any given asset class, an investor should choose the manager with the highest expected skill as measured by the information ratio, because investing with the highest information-ratio manager will produce the highest Sharpe ratio for the investor's portfolio.

The preceding discussion on adjusting active risk raises the issue of determining the *optimal* amount of active risk, without resorting to utility functions that measure risk aversion. For unconstrained portfolios, the level of active risk that leads to the optimal result in Equation 7 is

$$\sigma(R_A) = \frac{\mathrm{IR}}{\mathrm{SR}_B}\sigma_B \tag{8}$$

where σ_B is the standard deviation of the benchmark return. (Note that the right-hand side of the equation should be multiplied by the benchmark beta of the actively managed portfolio if that value is different from 1.) This Sharpe ratio-maximizing level of active risk or "aggressiveness" comes from the general mean–variance optimality condition that the ratio of expected active return to active return variance of the managed portfolio be set equal to the ratio of expected benchmark excess return to benchmark return variance:

$$\frac{E(R_A)}{\sigma_A^2} = \frac{E(R_B - R_F)}{\sigma_B^2}$$

For example, if the actively managed portfolio has an information ratio of 0.30 and active risk of 8.0% and the benchmark portfolio has an expected excess return of 6.4% and total risk of 16.0% resulting in a Sharpe ratio of 0.40, then according to Equation 8, the optimal amount of aggressiveness in the actively managed portfolio is (0.30/0.40)16.0% = 12.0%. If the actively managed portfolio is constructed with this amount of active risk, the Sharpe ratio will be $(0.40^2 + 0.30^2)^{1/2} = 0.50$, as shown in Equation 7. To verify this Sharpe ratio, note that the more aggressively managed portfolio in this example has an expected active return of (0.30)12.0% = 3.6% over the benchmark, or a total expected excess return of 6.4% + 3.6% = 10.0%. By definition, the total risk of the actively managed portfolio is the sum of the benchmark return variance and active return variance,

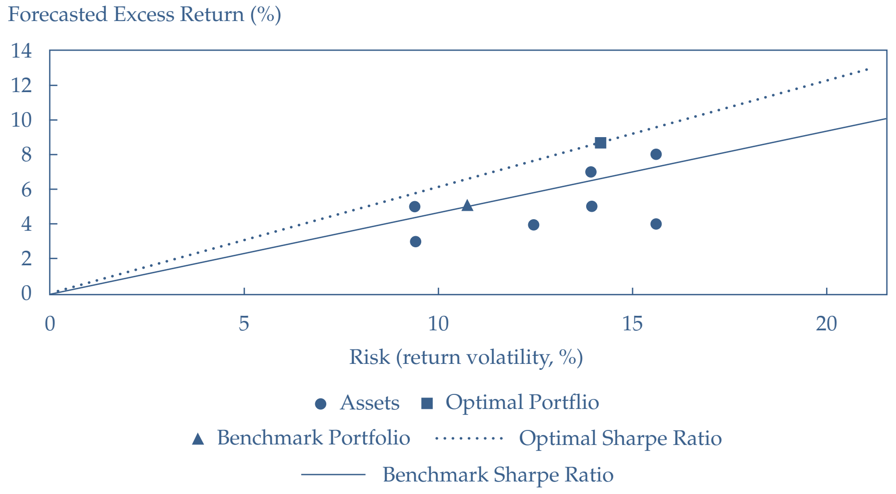
$$\sigma_P^2 = \sigma_B^2 + \sigma_A^2$$

At the optimal active risk of 12.0%, the total portfolio risk is $(16.0^2 + 12.0^2)^{1/2} = 20.0\%$, verifying the maximum possible Sharpe ratio of 10.0/20.0 = 0.50.

The initial actively managed portfolio has active risk of only 8.0%, whereas the optimal amount required under the assumed information ratio needed to maximize the Sharpe ratio is 12.0%. The actively managed portfolio would thus need to be managed more aggressively to increase the active risk while preserving the same information ratio; alternatively, the investor could short the benchmark and use the proceeds to increase the amount invested in the actively managed fund. The proportion required to be invested in the actively managed fund would be 12.0/8.0 = 1.5 times; shorting the benchmark by 0.5 times would fund the increase.

For readers familiar with risk–return charts in basic portfolio theory, Exhibit 4 will help illustrate these concepts.

Exhibit 4 Portfolio Risk and Return



Several individual risky assets are plotted in Exhibit 4 in terms of their forecasted return in excess of the risk-free rate ("excess return") on the vertical axis and risk on the horizontal axis. The values for the individual assets are based on subjective assessments supplied by the investor. The theory described here explains how to optimally employ those expectations assuming they are based on reasonable judgment. Using the benchmark portfolio weights (not shown), the risks and expected returns of the individual assets combine into the benchmark portfolio risk and expected return shown in Exhibit 4. Because the expected returns plotted along the vertical axis are in excess of the risk-free rate, the slope of a line that emanates from the origin (zero risk and zero excess return) is the Sharpe ratio of the benchmark portfolio. Specifically, the Sharpe ratio of the benchmark portfolio (i.e., slope of the dark line) in Exhibit 4 is the expected excess return of 5.0% divided by return volatility of 10.8%, or 5.0%/10.8% = 0.46.

Because of diversification, the Sharpe ratio of the benchmark portfolio is higher than those of most of the individual assets; however, the benchmark portfolio does not have the highest possible Sharpe ratio of all portfolios that can be constructed from these assets. In fact, the optimal portfolio (i.e., mean–variance efficient frontier portfolio with the highest possible Sharpe ratio) shown in Exhibit 4 has an expected excess return of 8.7% and return volatility of 14.2%, resulting in a Sharpe ratio of 8.7%/14.2% = 0.61 (i.e., slope of the dotted line). This higher Sharpe ratio could be retained in the portfolio while adjusting the level of risk through the use of cash or leverage. For example, the risk of the optimal portfolio could be reduced along the dotted line to the benchmark portfolio risk of 10.8% with an expected excess return of 0.61(10.8%) = 6.6%, compared with the benchmark expected excess return of 5.0%.

Exhibit 5 Portfolio Active Risk and Return

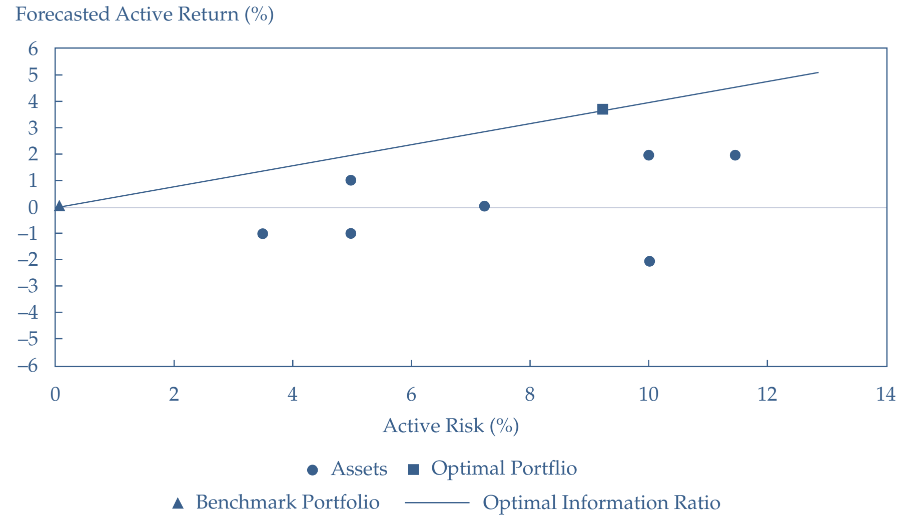


Exhibit 5 plots the same individual assets, the benchmark portfolio, and the optimal portfolio from Exhibit 4 in terms of their expected *active* return on the vertical axis and *active* risk on the horizontal axis. By definition, the benchmark portfolio is plotted at the origin in Exhibit 5 with zero active return and zero active risk. The individual assets have both positive and negative expected active returns compared with the benchmark portfolio return, whereas the optimal portfolio has a positive active return of 3.8% and active risk of 9.4%. The information ratio of the optimal portfolio is therefore 3.8%/9.4% = 0.40, the slope of the dark line in Exhibit 5. The information ratio of the optimal portfolio is higher than that of any of the individual assets; in fact, it can be shown to be the square root of the sum of the squared values of the individual assets' information ratios, similar to Equation 7, including those assets with negative information ratios. The asset weights required for the construction of this optimal portfolio are the subject of the next section, but here we note they might be negative for negative-IR assets—that is, short sells of individual assets may be required.

Although the information ratio will remain constant at 0.40, various levels of aggressiveness can be applied to the actively managed portfolio in Exhibit 5 by scaling the optimal active weights or, alternatively, taking a position in the benchmark portfolio—leading to portfolios that plot along the dark line. But to construct the *optimal* actively managed portfolio given the assumed information ratio, the active risk must be adjusted to a level of (0.40/0.46)10.8% = 9.4% in accordance with Equation 8. Specifically, this level of aggressiveness is required to construct the optimal portfolio in Exhibit 4, and according to Equation 7, the Sharpe ratio of this optimal portfolio is $(0.46^2 + 0.40^2)^{1/2} = 0.61$.

As we will see later, optimal levels of active risk in equity management practice are typically lower than those shown in this numerical example because the underlying portfolios are constrained to be long only, leading to information ratios that are substantially lower. As the information ratio gets close to zero, either because of constraints or because the manager is judged to be less skilled, the optimal amount of active risk in Equation 8 goes to zero (i.e., the optimal portfolio becomes the passive benchmark portfolio).

EXAMPLE 3

Expected Value Added Based on the Information Ratio

Suppose that the historical performance of the Fidelity Magellan and Growth Fund of America mutual funds from Exhibits 2 and 3 are indicative of the future performance of hypothetical funds, "Fund I" and "Fund II." In addition, suppose that the historical performance of the S&P 500 benchmark portfolio shown in Exhibit 1 is indicative of expected returns and risk going forward, as shown in the following excerpts. We use historical values in this problem for convenience, but in practice the forecasted, or expected, values for both the benchmark portfolio and the active funds would be subjectively determined by the investor.

Excerpted from Exhibits 1 and 2 (based on a risk-free rate of 2.3%)

	S&P 500	Fidelity Magellan (Fund I)	Growth Fund of America (Fund II)
Average annual return	9.9%	8.5%	11.1%
Return standard dev.	14.4%	16.5%	15.7%
Sharpe ratio	0.53	0.38	0.56

Excerpted from Exhibit 3

	Fidelity Magellan (Fund I)	Growth Fund of America (Fund II)
Active return	–1.4%	1.2%
Active risk	5.1%	6.2%
Information ratio	–0.27	0.20
Benchmark	S&P 500	S&P 500

1 State which of the two actively managed funds, Fund I or Fund II, would be better to combine with the passive benchmark portfolio and why.

2 Calculate the possible improvement over the S&P 500 Sharpe ratio from the optimal deployment of Fund II, which has an expected information ratio of 0.20.

3 Fund I comes with an active (i.e., benchmark relative) risk of 5.1%, but the investor wants to adjust the active risk to 5.4%. Describe how that adjustment would be made. (No calculations are required; give a qualitative description.)

4 Fund II comes with an active risk of 6.2%. Determine the weight of the benchmark portfolio required to create a combined portfolio with the highest possible expected Sharpe ratio.

Solution to 1:

Fund II is better, as measured by the combined Sharpe ratio, because Fund II has the higher expected information ratio: 0.20 compared with −0.27 in Fund I.

Solution to 2:

Properly combined with the S&P 500 benchmark portfolio, Fund II has the potential to increase the expected Sharpe ratio from 0.53 for the passive benchmark portfolio to an expected Sharpe ratio of $(0.53^2 + 0.20^2)^{1/2} = 0.57$.

Solution to 3:

To increase the active risk of Fund I, the investor would need to be more aggressive in managing the portfolio, take a short (i.e., negative) position in the benchmark, or, more simply, invest less than he or she otherwise would have in the benchmark or another actively managed fund.

Solution to 4:

According to Equation 8, the optimal amount of active risk is (0.20/0.53)14.4% = 5.4%. A positive position in the benchmark is needed to adjust the active weight down from 6.2%. Specifically, the benchmark portfolio weight needed to adjust the active risk in Fund II is 1 − 5.4%/6.2% = 13%.

Note that at the 5.4% optimal level of active risk, Fund II has an expected active return of 0.20(5.4%) = 1.1%, a total expected excess return of 7.6% + 1.1% = 8.7%, and a total risk of $(14.4^2 + 5.4^2)^{1/2} = 15.4\%$. The result is an expected Sharpe ratio of 8.7/15.4 = 0.57, the same as the value calculated for Question 2.

In summary, the information ratio is active return over active risk (in contrast to the excess return-to-risk measure known as the Sharpe ratio). Information ratios help investors focus on the relative valued added by active management. The information ratio is unaffected by the aggressiveness of the active weights (i.e., deviations from benchmark weights) in the managed portfolio because both the active return and the active risk increase proportionally. The potential improvement in an active portfolio's expected Sharpe ratio compared with the benchmark's Sharpe ratio is a function of the squared information ratio. Thus, the expected information ratio becomes the single best criterion for constructing an actively managed portfolio, and the *ex post* information ratio is the best criterion for evaluating the past performance of various actively managed funds.

5 ACTIVE SECURITY RETURNS AND THE BASIC FUNDAMENTAL LAW OF ACTIVE MANAGEMENT

c describe and interpret the fundamental law of active portfolio management, including its component terms—transfer coefficient, information coefficient, breadth, and active risk (aggressiveness);

d explain how the information ratio may be useful in investment manager selection and choosing the level of active portfolio risk;

The fundamental law is a framework for thinking about the potential value added through active portfolio management. The framework can be used to size individual asset active weights, estimate the expected value added of an active management strategy, or measure the realized value added after the fact; however, the most common use is the description and evaluation of active management strategies. The law itself is a mathematical relationship that relates the expected information ratio of an actively managed portfolio to a few key parameters.

5.1 Active Security Returns

On the basis of the prior section, we assume that the investor is concerned about maximizing the managed portfolio's active return subject to a limit on active risk (also called "benchmark tracking risk"). To this end, the investor uses forecasts for each security of the active return, R_{Ai}, or thus the benchmark relative return,

$$R_{Ai} = R_i - R_B, \tag{9}$$

for the N individual assets that might be included in the portfolio. Our notation for the investor's forecasts of the active security returns is μ_i (Greek letter mu). The term μ_i can be thought of as the security's expected active return, $\mu_i = E(R_{Ai})$, referring to the investor's subjective expectation, in contrast to an expectation based on a formal equilibrium model.

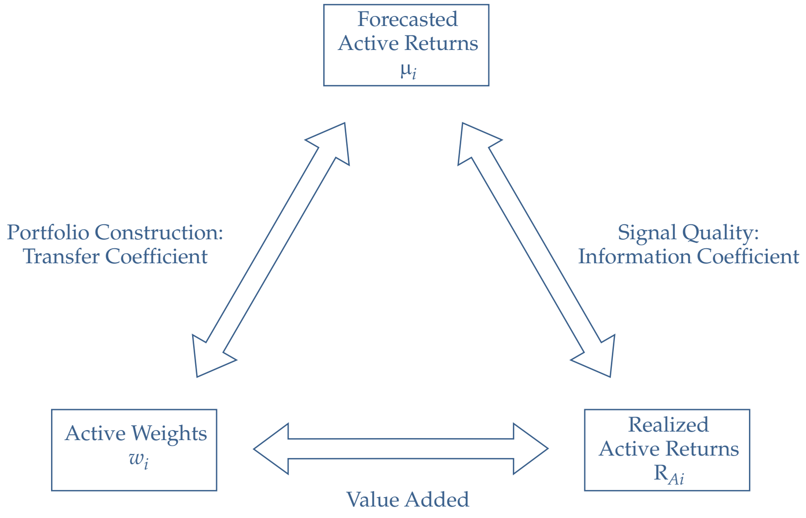
Although we focus on the simple definition of active security return in Equation 9, there are several possible choices depending on the assumed risk model (i.e., statistical model of returns) and the desired trade-off between a conceptual treatment and more complex but implementable formulas. For example, Equation 9 can be modified to define the active security return as the residual return in a single-factor model, $R_{Ai} = R_i - \beta_i R_B$, where β_i is the sensitivity of the security return to the benchmark return. Although this expression may appear to be related to the CAPM, the benchmark return may or may not be the market return. Moreover, the fundamental law does not require the empirical validity of the CAPM, the multi-factor APT (arbitrage pricing theory), or any other equilibrium theory of required returns. The individual security active return can also be defined as the residual return in a multi-factor model:

$$R_{Ai} = R_i - \sum_{j=1}^{K} \beta_{j,i} R_j$$

with K market-wide factor returns, R_j, and security sensitivities, $\beta_{j,i}$, to those factors.

Exhibit 6 provides a conceptual diagram in which to think about the various parameters in the fundamental law of active management. At the three corners of the triangle are the sets of forecasted active returns, μ_i, active portfolio weights, Δw_i, and realized active returns, R_{Ai}. The base of the triangle reflects the realized value added through active management, defined as the difference between the realized returns on the actively managed portfolio and the benchmark portfolio. Value added is the sum of the products of active weights and active returns for the i = 1 to N securities in the portfolio, as shown in Equation 3. The value of this sum is ultimately a function of the correlation coefficient between the active weights, Δw_i, and realized active returns, R_{Ai}.

Exhibit 6 The Correlation Triangle



To understand the role of the correlation coefficient, consider the following algebraic expansion of Equation 3 that uses COV, STD (σ), and COR (ρ) to designate the covariance, standard deviation, and correlation coefficient functions, respectively:

$$\begin{aligned} R_A &= \sum_{i=1}^{N} \Delta w_i R_{Ai} \\ &= \rho\left(\Delta w_i, R_{Ai}\right) N \\ &= \rho\left(\Delta w_i, R_{Ai}\right)\sigma\left(\Delta w_i\right)\sigma\left(R_{Ai}\right) N \end{aligned}$$

The exact equalities in this expansion depend on the fact that the cross-sectional means of active weights and active returns are zero. Specifically, the population covariance between two variables, X and Y, is calculated as $\mathrm{COV}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$, or simply $\mathrm{COV}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} X_i Y_i$.

Similarly, the population variance for a single variable is $\mathrm{VAR}(X) = \frac{1}{N}\sum_{i=1}^{N} X_i^2$ if the mean is zero.

In Exhibit 6, the arrows in the legs of the triangle represent the correlation between the quantities at the corners of the triangle to which the arrows point. While the arrow at the base of the triangle reflects value added, a better understanding of the sources and limitations of value added can be obtained by examining the correlations on the two vertical legs. First, there is little hope of adding value if the investor's forecasts of active returns do not correspond at least loosely to the realized active returns. Signal quality is measured by the correlation between the forecasted active returns, μ_i, at the top of the triangle, and the realized active returns, R_{Ai}, at the right corner, commonly called the information coefficient (IC). Investors with higher IC, or ability to forecast returns, will add more value over time but only to the extent that those forecasts are exploited in the construction of the managed portfolio. The correlation between any set of active weights—Δw_i, in the left corner, and forecasted active returns, μ_i, at the top of the triangle—measures the degree to which the investor's forecasts are translated into active weights, called the transfer coefficient (TC).

The mathematics of the fundamental law were introduced by Grinold (1989) and further developed by Clarke, de Silva, and Thorley (2002). The mean–variance-optimal active security weights for uncorrelated active returns, subject to a limit on active portfolio risk, are given by

$$\Delta w_i^* = \frac{\mu_i}{\sigma_i^2} \frac{\sigma_A}{\sqrt{\sum_{i=1}^{N} \frac{\mu_i^2}{\sigma_i^2}}}$$

where σ_A represents active portfolio risk and σ_i is the forecasted volatility of the active return on security *i*. This formula for active weights makes intuitive sense. The deviation (positive or negative) from the benchmark weight for security *i* are higher for larger values of the forecasted active return, μ_i, but are reduced by forecasted volatility, σ_i. In addition, the active weights are scaled by the active risk of the portfolio, σ_A, so that the desire for more active portfolio risk requires larger individual active weights.

In addition to employing mean–variance optimization, proofs of the fundamental law generally assume that active return forecasts are scaled prior to optimization using the Grinold (1994) rule:

$$\mu_i = \text{IC}\sigma_i S_i, \tag{10}$$

where IC is the expected information coefficient and S_i represents a set of standardized forecasts of expected returns across securities, sometimes called "scores." Scores with a cross-sectional variance of 1 are used in Equation 10 to ensure that the scaling process using the multipliers σ_i (separate values for individual securities) and IC (one value for all securities) result in expected active returns of the correct magnitude. Specifically, if the assumed IC value is low, then the cross-sectional variation of the expected active returns in Equation 10 will be low. However, the exact process for calculating expected active returns may be more involved than the simple rule indicated in Equation 10 depending on how the investor's views on individual asset returns are originally formulated.

Using the Grinold rule shown in Equation 10, the mean–variance optimal active weights are

$$\Delta w_i^* = \frac{\mu_i}{\sigma_i^2} \frac{\sigma_A}{\text{IC}\sqrt{\text{BR}}} \tag{11}$$

where IC stands for information coefficient and BR, which has replaced the symbol *N*, stands for breadth. We discuss these two key fundamental law parameters next.

As previously stated, IC is the *ex ante* (i.e., anticipated) cross-sectional correlation between the *N* forecasted active returns, μ_i, and the *N* realized active returns, R_{Ai}. To be more accurate, IC is the *ex ante risk-weighted* correlation

$$\text{IC} = \rho\left(\frac{R_{Ai}}{\sigma_i}, \frac{\mu_i}{\sigma_i}\right) \tag{12}$$

where $\rho(\cdot)$ indicates correlation. As a correlation coefficient, IC can take on values anywhere from −1.00 to +1.00, although small positive values less than 0.20 are often the norm. The *ex ante*, or anticipated, IC must be positive or the investor would not pursue active management but simply invest in the passive benchmark. Later, we will discuss the realized, or *ex post*, information coefficient in terms of measuring active management performance after the fact—where the realized information coefficient might be either positive or negative, leading to positive or negative value added.

The other important fundamental law parameter in Equation 11 is BR, or breadth, conceptually equal to the number of independent decisions made per year by the investor in constructing the portfolio. The simplest case for the calculation of breadth is a single-factor risk model, where the only source of correlation between the securities is the common market factor and decisions about the active return for any given security are independent from one year to the next. In this case, breadth is equal to the number of securities: Each active return is independent from the other active return forecasts for that period and independent from the forecast for that security in subsequent periods.

However, most risk models will incorporate other factors—for example, economic sectors or industries. If the risk model includes the assumption that all the securities within a given industry are positively correlated, then part of the forecast that the active returns for securities in that industry will be higher or lower is based on just one perspective the investor has about the industry. In this case, breadth is intuitively lower than the number of securities. Alternatively, breadth can be higher than the number of securities if factors in the risk model suggest that their active returns are negatively correlated. For these more complicated cases, breadth will be a non-integer number, as noted in Clarke, de Silva, and Thorley (2006).

Similarly, if some aspect of a security is fairly constant over time and the investor makes decisions about expected active return based on that characteristic, then breadth over time is lower. Alternatively, if the investor makes quarterly or monthly forecasts about a security that are truly independent over time, then breadth can be as high as the number of securities times the number of rebalancing periods per year.

EXAMPLE 4

Scaling Active Return Forecasts and Sizing Active Weights

Consider the simple case of four individual securities whose active returns are assumed to be uncorrelated with each other and have active return volatilities of 25.0% and 50.0%. After some analysis, an active investor believes the first two securities will outperform the other two over the next year and thus assigns scores of +1 and −1 to the first and second groups, respectively. The scenario is depicted in the following exhibit:

Security	Score	Volatility
#1	1.0	25.0%
#2	1.0	50.0%
#3	−1.0	25.0%
#4	−1.0	50.0%

1 Assume that the anticipated accuracy of the investor's ranking of securities is measured by an information coefficient of IC = 0.20. What are the forecasted active returns for each of the four securities using the scaling rule $\mu_i = \text{IC}\sigma_i S_i$?

2 Given the assumptions that the four securities' active returns are uncorrelated with each other and forecasts are independent from year to year, what is the breadth of the investor's forecasts?

3 Suppose the investor wants to maximize the expected active return of the portfolio subject to an active risk constraint of 9.0%. Calculate the active weights that should be assigned to each of these securities using the formula $\Delta w_i^* = \frac{\mu_i}{\sigma_i^2}\frac{\sigma_A}{\text{IC}\sqrt{\text{BR}}}$.

Solution to 1:

The forecasted active return to Security #1 is 0.20(25.0%)(1.0) = 5.0%. Similar calculations for the other three securities are shown in the following exhibit.

Security	Score	Active Return Volatility	Expected Active Return
#1	1.0	25.0%	5.0%
#2	1.0	50.0%	10.0%
#3	−1.0	25.0%	−5.0%
#4	−1.0	50.0%	−10.0%

Solution to 2:

If the active returns are uncorrelated with each other and the forecasts are independent from year to year, then the investor has made four separate decisions and breadth is BR = 4, the number of securities.

Solution to 3:

The size of the active weight for Security #1 is $\Delta w_i^* = \dfrac{0.05}{0.25^2}\dfrac{0.09}{0.20\sqrt{4}} = 18\%$.

Similar calculations for the other four securities are shown in the following exhibit.

Security	Expected Active Return	Active Return Volatility	Active Weight
#1	5.0%	25.0%	18.0%
#2	10.0%	50.0%	9.0%
#3	−5.0%	25.0%	−18.0%
#4	−10.0%	50.0%	−9.0%

5.2 The Basic Fundamental Law

On the basis of Equation 3, the anticipated value added for an actively managed portfolio, or expected active portfolio return, is the sum product of active security weights and forecasted active security returns:

$$\mathrm{E}(R_A) = \sum_{i=1}^{N} \Delta w_i \mu_i$$

Using the optimal active weights in Equation 11 and forecasted active security returns in Equation 10, the expected active portfolio return is

$$\mathrm{E}(R_A)^* = \mathrm{IC}\sqrt{\mathrm{BR}}\sigma_A \tag{13}$$

where the * indicates that the actively managed portfolio is constructed from *optimal* active security weights, Δw_i^*. Remember that the algebra for this result assumes that breadth is the number of securities: BR = N. A more general proof where breadth is different from the number of securities is provided in Clarke, de Silva, and Thorley (2006).

The basic fundamental law of active management in Equation 13 states that the optimal expected active return, $\mathrm{E}(R_A)^*$, is the product of three key parameters: the assumed information coefficient, IC, the square root of breadth, BR, and portfolio active risk, σ_A. Using Equation 13, we can also express the information ratio of the unconstrained optimal portfolio, $\mathrm{E}(R_A)^*/\sigma_A$, as the product of just two terms: $\mathrm{IR}^* = \mathrm{IC}\sqrt{\mathrm{BR}}$.

EXAMPLE 5

The Basic Fundamental Law

Consider the simple case of four individual securities whose active returns are uncorrelated with each other and forecasts are independent from year to year. The active return forecasts, active risks, and the active weights for each security are shown in the following exhibit.

Security	Expected Active Return	Active Return Volatility	Active Weight
#1	5.0%	25.0%	18%
#2	10.0%	50.0%	9%
#3	−5.0%	25.0%	−18%
#4	−10.0%	50.0%	−9%

1 Suppose that the benchmark portfolio for these four securities is equally weighted (i.e., $w_{B,i}$ = 25% for each security) and that the forecasted return on the benchmark portfolio is 10.0%. What are the portfolio weights and the total expected returns for each of the four securities?

2 Calculate the forecasted total return and active return of the managed portfolio.

3 Calculate the active risk of the managed portfolio.

4 Verify the basic fundamental law of active management using the expected active return and active risk of the managed portfolio. The individual security active return forecasts and active weights were sized using an information coefficient of IC = 0.20, breadth of BR = 4, and active risk of σ_A = 9.0%.

Solution to 1:

The portfolio weight for Security #1 is the benchmark weight plus the active weight, 25% + 18% = 43%. The total expected return for Security #1 is the expected benchmark return plus the expected active return, 10.0% + 5.0% = 15.0%. Similar calculations for the other three securities are shown in the following exhibit.

Security	Total Weight	Total Return Forecast
#1	43%	15.0%
#2	34%	20.0%
#3	7%	5.0%
#4	16%	0.0%
	100%	

Solution to 2:

The forecasted total return of the portfolio is the sum of portfolio weights times total returns for each security: 0.43(15.0) + 0.34(20.0) + 0.07(5.0) + 0.16(0.0) = 13.6%. The expected active return of the portfolio is the managed portfolio return minus the benchmark return: 13.6 − 10.0 = 3.6%. Alternatively, the calculation is the sum of active weights times active returns for each security: 0.18(5.0%) + 0.09(10.0%) − 0.18(−5.0%) − 0.09(−10.0%) = 3.6%.

Solution to 3:

The active risk of the managed portfolio is the square root of the sum of active weights squared times the active volatility squared for each security, which gives $[0.18^2 \times 25.0^2 + 0.09^2 \times 50.0^2 + (-0.18)^2 \times 25.0^2 + (-0.09)^2 \times 50.0^2]^{1/2} = 9.0\%$.

Solution to 4:

The basic fundamental law states that the expected active portfolio return is $\mathrm{IC}\sqrt{\mathrm{BR}}\sigma_A = 0.20 \times 4^{1/2} \times 9.0 = 3.6\%$, which is consistent with the calculation in the Solution to 2. Alternatively, the information ratio of 3.6/9.0 = 0.40 confirms the basic fundamental law that $\mathrm{IR}^* = \mathrm{IC}\sqrt{\mathrm{BR}} = 0.20 \times 4^{1/2} = 0.40$.

6 THE FULL FUNDAMENTAL LAW, *EX* POST PERFORMANCE MEASUREMENT

c describe and interpret the fundamental law of active portfolio management, including its component terms—transfer coefficient, information coefficient, breadth, and active risk (aggressiveness);

d explain how the information ratio may be useful in investment manager selection and choosing the level of active portfolio risk;

Although we were able to derive an analytic (i.e., formula-based) solution for the set of unconstrained optimal active weights in Equation 11, a number of practical or strategic constraints are often imposed in practice. For example, if the unconstrained active weight of a particular security is negative and large, that might lead to a negative absolute weight or short sell of the security. Many investors are constrained to be long only, either by regulation or by preference because of the extra complexity and costs of short selling. For quantitatively oriented investors, optimal solutions for active weights under long-only constraints, limits on turnover, ESG screens, or other constraints generally require the use of a numerical optimizer. Alternatively, one can use the fundamental law framework to better analyze the active weights that are subjectively determined by less quantitative techniques.

Let Δw_i (without an *) represent the *actual* active security weights for a constrained portfolio—in contrast to the optimal active weights, Δw_i^*, specified in Equation 11. As explained previously, the transfer coefficient, TC, is essentially the cross-sectional correlation between the forecasted active security returns and actual active weights. To be more precise, for a single-factor risk model, TC is the following *risk-weighted* correlation:

$$\mathrm{TC} = \rho(\mu_i/\sigma_i, \Delta w_i \sigma_i).$$

Based on the correspondence between optimal active weights and forecasted active returns in Equation 11, the transfer coefficient can also be expressed as the risk-weighted correlation between the optimal active weights and the actual active weights, $\mathrm{TC} = \rho\left(\Delta w_i^* \sigma_i, \Delta w_i \sigma_i\right)$.

As a correlation coefficient, TC can take on values anywhere from −1.00 to +1.00, although TC values are typically positive and range from about 0.20 to 0.90. A low TC results from the formal or informal constraints imposed on the structure of the portfolio. In fact, at TC = 0.00, there would be no correspondence between the active return forecasts and active weights taken and thus no expectation of value added from active management. In contrast, TC = 1.00 (no binding constraints) represents a perfect correspondence between active weights taken and forecasted active returns, allowing the full expected value added to be reflected in the portfolio structure. The portfolio TC could even conceivably be negative if relative weights are negatively correlated with current expected returns because the portfolio needs rebalancing.

Including the impact of the transfer coefficient, the expanded fundamental law is expressed in the following equation:

$$E(R_A) = (TC)(IC)\sqrt{BR}\sigma_A \tag{14}$$

where an * is not used because the managed portfolio is constructed from *constrained* active security weights, Δw_i. The expanded fundamental law of active management shown in Equation 14, which we will refer to simply as the *fundamental law* henceforward, states that the expected active return, $E(R_A)$, is the product of four key parameters: the transfer coefficient, TC, the assumed information coefficient, IC, the square root of breadth, BR, and portfolio active risk, σ_A. Using Equation 14, we can also express the portfolio's information ratio, $E(R_A)/\sigma_A$, as the product of just three terms: $IR = (TC)(IC)\sqrt{BR}$.

The fundamental law as stated in Equation 14, although more practical than Equation 13, is still based on a simple risk model for the individual securities. Specifically, the equations in this section are based on the simplifying assumption of a single-index model; so, the active security returns are residual returns and are uncorrelated with each other. If we go even further in terms of simplicity and assume that the individual securities all have the same residual volatility, then the correlation formulas for IC and TC do not need to be risk weighted. Alternatively, we could move in the direction of more complexity by using the single-factor risk model with factor sensitivity: $R_{Ai} = R_i - \beta_i R_B$. In quantitative portfolio management practice, even more sophisticated multi-factor risk models are used with correspondingly more complex fundamental law parameter values, although the basic form of Equation 14 is preserved.

EXAMPLE 6

The Expanded Fundamental Law

Consider the simple case of four individual securities whose active returns are uncorrelated with each other and forecasts are independent from year to year. The securities have a range of active return forecasts, risks, optimal active weights, and actual active weights as given in the following exhibit. The optimal active weights are based on a formula for maximizing the active return of a managed portfolio for a given level of active risk. The actual active weights are the result of a numerical optimizer with a number of constraints, in addition to the active risk constraint of 9.0%.

Security	Expected Active Return	Active Return Volatility	Optimal Active Weight	Actual Active Weight
#1	5.0%	25.0%	18%	6%
#2	10.0%	50.0%	9%	4%
#3	−5.0%	25.0%	−18%	7%
#4	−10.0%	50.0%	−9%	−17%

1 Calculate the transfer coefficient (TC) as the risk-weighted correlation coefficient between the four active return forecasts and the four actual active weights. Compare this number with the transfer coefficient for the optimal active weights.

2 The forecasted active return of the optimal portfolio is the sum of the active weights times active returns for each security: 0.18(5.0) + 0.09(10.0) + (–0.18)(–5.0) + (–0.09)(–10.0) = 3.6%. The active risk of the optimal portfolio is the square root of the sum of active weights squared times the active volatility squared for each security: $[(0.18)^2(25.0)^2 + (0.09)^2 (50.0)^2 + (-0.18)^2 (25.0)^2 + (-0.09)^2 (50.0)^2]^{1/2}$ = 9.0%. Calculate the forecasted active return and active risk of the managed portfolio using the *actual* rather than unconstrained optimal active weights.

3 Verify the expanded fundamental law of active management using the active portfolio return, active portfolio risk, and transfer coefficient calculations in Parts 1 and 2. The individual active return forecasts and optimal active weights were sized using an information coefficient of IC = 0.20 and breadth of BR = 4.

Solution to 1:

The transfer coefficient is the correlation between the risk-weighted expected active returns and actual active weights: TC = $\rho(\Delta w_i\sigma_i,\mu_i/\sigma_i)$, where ρ denotes correlation (you can use the Microsoft Excel function CORREL) with four pairs of numbers. The risk-weighted values for Security #1 are $\Delta w_1\sigma_1$ = 0.06(25.0) = 1.5% and μ_1/σ_1 = 5.0/25.0 = 20.0%. The correlation coefficient across all four securities (calculated in Excel) is TC = 0.58. The transfer coefficient for the optimal active weights is by definition 1.0 but can be verified by the calculated correlation coefficient. The risk-weighted values for Security #1 are then $\Delta w_1\sigma_1$ = 0.18(25.0) = 4.5% and μ_1/σ_1 = 5.0/25.0 = 20.0%.

Solution to 2:

The forecasted active return of the managed portfolio is 0.06(5.0) + 0.04(10.0) + 0.07(–5.0) + (–0.17)(–10.0) = 2.1%. The active risk of the managed portfolio is the square root of the sum of actual active weights squared times the active volatility squared for each security, $[(0.06)^2(25.0)^2 + (0.04)^2 (50.0)^2 + (0.07)^2 (25.0)^2 + (-0.17)^2 (50.0)^2]^{1/2}$ = 9.0%, as specified by the active risk constraint.

Solution to 3:

The expanded fundamental law states that the expected portfolio active return will be $E(R_A) = (TC)(IC)\sqrt{BR}\sigma_A = 0.58 \times 0.20 \times (4)^{1/2} \times 9.0 = 2.1\%$, consistent with the direct calculation in the Solution to 2.

We close this sub-section by noting that the transfer coefficient, TC, also comes into play when calculating the optimal amount of active risk for an actively managed portfolio with constraints. Specifically, with constraints and using notation consistent with expressions in the fundamental law, Equation 8 becomes

$$\sigma_A = TC\frac{IR^*}{SR_B}\sigma_B$$

where IR* is the information ratio of an otherwise unconstrained portfolio. Employing this optimal level of aggressiveness leads to a maximum possible value of the constrained portfolio's squared Sharpe ratio:

$$SR_P^2 = SR_B^2 + (TC)^2(IR^*)^2$$

As noted previously, the active risk of an actively managed fund can be adjusted to its optimal level while preserving the information ratio by adding long or short positions in the benchmark portfolio. For further insight, note that with a transfer coefficient of 0.00, the optimal amount of active risk calculated is zero. In other words, the investor should just invest in the benchmark portfolio.

We now illustrate the impact of the transfer coefficient with the following example. If the actively managed portfolio has a transfer coefficient of 0.50 and an unconstrained information ratio of 0.30 and the benchmark portfolio has a Sharpe ratio of 0.40 and risk of 16.0%, then the optimal amount of aggressiveness in the actively managed portfolio is 0.50(0.30/0.40)16.0 = 6.0%. If the actively managed portfolio is constructed with this amount of active risk, the Sharpe ratio will be $(0.40^2 + 0.50^2 \times 0.30^2)^{1/2} = 0.43$. If the constrained portfolio has an active risk of 8.0%, the active risk can be lowered to the optimal level of 6.0% by mixing 1 – 6.0/8.0 = 25% in the benchmark and 75.0% in the actively managed fund.

6.1 *Ex Post* Performance Measurement

Most of the fundamental law perspectives discussed up to this point relate to the expected value added through active portfolio management. Actual performance in any given period will vary from its expected value in a range determined by the benchmark tracking risk. We now turn our attention to examining actual performance, the *ex post* analysis of the realized value added.

The key determinant of the sign and magnitude of the realized value added in Equation 3 is the degree to which the portfolio has positive active weights on securities that realize positive relative returns and negative active weights on securities that realize negative relative returns. In other words, actual performance is measured by the relationship between relative weights and realized relative returns. Knowing how actual returns match up with realized returns (the *realized* information coefficient, IC_R) allows the investor to examine what realized return to expect given the transfer coefficient. Specifically, expected value added conditional on the realized information coefficient, IC_R, is

$$\mathrm{E}\left(R_A \mid \mathrm{IC}_R\right) = \left(\mathrm{TC}\right)\left(\mathrm{IC}_R\right)\sqrt{\mathrm{BR}}\,\sigma_A \quad (15)$$

Equation 15 is similar to the fundamental law, shown in Equation 14, but in Equation 15 the realized information coefficient, IC_R, replaces the *expected* information coefficient, IC.

We can represent any difference between the actual active return of the portfolio and the conditional expected active return with a noise term:

$$R_A = \mathrm{E}(R_A \mid \mathrm{IC}_R) + \text{Noise}. \quad (16)$$

Equation 16 states that the realized value added of an actively managed portfolio can be divided into two parts. The first part comes from the expected value added given the realized skill of the investor that period. The second part represents any noise that results from constraints that impinge on the optimal portfolio structure.

Equation 15 also leads to an *ex post* (i.e., realized) decomposition of the portfolio's active return variance into two parts: variation due to the realized information coefficient and variation due to constraint-induced noise. Clarke, de Silva, and Thorley (2005) showed that the two parts of the realized variance are proportional to TC^2 and $1 - TC^2$. For example, with a TC value of, say, 0.60, only TC^2 = 36% of the realized variation in performance is attributed to variation in the realized information coefficient, and $1 - TC^2$ = 64% comes from constraint-induced noise. Low-TC investors will frequently experience periods when the forecasting process succeeds but actual performance is poor or when actual performance is good even though the return-forecasting process fails.

EXAMPLE 7

Ex Post Performance

Consider an active management strategy that includes BR = 100 investment decisions (e.g., 100 individual stocks, whose active returns are uncorrelated, and annual rebalancing), an expected information coefficient of IC = 0.05, a transfer coefficient of TC = 0.80, and annualized active risk of σ_A = 4.0%. Thus, the expected value added according to the fundamental law is

$$E(R_A) = (TC)(IC)\sqrt{BR}\sigma_A = 0.80 \times 0.05 \times \sqrt{100} \times 4.0\% = 1.6\%$$

1. Suppose that the *realized* information coefficient in a given period is −0.10, instead of the expected value of IC = 0.05. In the absence of constraint-induced noise, what would be the value added that period?
2. Suppose that the actual return on the active portfolio was −2.6%. Given the −0.10 realized information coefficient, how much of the forecasted active return was offset by the noise component?
3. What percentage of the performance variance (i.e., tracking risk squared) in this strategy over time is attributed to variation in the realized information coefficient (i.e., forecasting success), and what percentage of performance variance is attributed to constraint-induced noise?

Solution to 1:

The value added, without including constraint-induced noise (which has an expected value of zero) is

$$E(R_A \mid IC_R) = (TC)(IC_R)\sqrt{BR}\sigma_A = 0.80 \times (-0.10) \times \sqrt{100} \times 4.0\% = -3.2\%$$

In other words, conditional on the actual information coefficient, the investor should expect an active return that is negative because the realized information coefficient is negative.

Solution to 2:

The noise portion of the active return is the difference between the actual active return and the forecasted active return: −2.6 − (−3.2) = 0.6%. In other words, the noise component helped offset the negative value added from poor return forecasting. Of course, the constraint-induced noise component could just as easily have gone the other way, exacerbating the negative value added. Note that the negative realized active return of −2.6% is well within the range associated with the tracking error (active risk) of 4.0% per period.

Solution to 3:

Given the transfer coefficient of TC = 0.80, TC^2 = 64% of the variation in performance over time is attributed to the success of the forecasting process, leaving 36% due to constraint-induced noise.

APPLICATIONS OF THE FUNDAMENTAL LAW AND GLOBAL EQUITY STRATEGY

7

d explain how the information ratio may be useful in investment manager selection and choosing the level of active portfolio risk;

e compare active management strategies, including market timing and security selection, and evaluate strategy changes in terms of the fundamental law of active management;

In this section, we discuss three specific applications of active portfolio management: one application to a global equity strategy with different sets of active return forecasts and constraints and two applications to US fixed income. These applications will further illustrate how the fundamental law is used to evaluate active portfolio strategies, including security selection and market timing.

7.1 Global Equity Strategy

In our first example, we show how the fundamental law can be used to calculate the expected active return for an actively managed portfolio benchmarked to the MSCI All Country World Index (ACWI). This global equity example focuses on the cross-sectional characteristics of the fundamental law, whereas the US fixed-income examples that follow also include time-series implications of the law. The investable assets in this example are the individual MSCI market indexes—including the 21 EAFE (Europe, Australasia, and Far East) markets, the United States, Canada, and the Emerging Markets Index—for a total of 24 assets. "Now" is the beginning of the calendar year 2019. For purposes of illustration, we will assume that the future will be like the past in terms of active risk, and thus we will base our estimates on the US dollar return to the MSCI market indexes from 2009 to 2018. In practice, managerial judgment or a commercial model would be used to forecast risk. The various rankings of the markets' forecasted active returns for the calendar year 2019 are hypothetical.

The *ex ante* expected active risk of each asset is equal to the annualized historical standard deviation of beta-adjusted differences between the individual market return and the ACWI return, as shown in the third column of Exhibit 7. For example, the active risk of the United Kingdom is 6.4% and the active risk of Japan is 9.1%. Note that the risk estimates are for active returns (i.e., the difference between the individual asset and benchmark returns). The total risk of each market would be higher based on the estimated risk of the benchmark and the benchmark beta.

Exhibit 7 Long–Short Global Equity Fund for 2019 (risk statistics based on MSCI returns from 2009 to 2018)

Market	Score	Active Return Volatility	Expected Active Return	Active Weight	ACWI Benchmark Weight	Portfolio Weight
United Kingdom	2.0	6.4%	1.3%	15.7%	5.2%	20.9%
Japan	0.0	9.1%	0.0%	–2.0%	7.6%	5.7%
France	–2.0	8.5%	–1.7%	–12.0%	3.4%	–8.5%
Germany	0.0	8.4%	0.0%	2.8%	2.7%	5.6%
Switzerland	2.0	7.8%	1.6%	9.7%	2.7%	12.3%
Australia	0.0	11.1%	0.0%	–1.0%	2.1%	1.1%
Spain	–2.0	16.0%	–3.2%	–5.3%	1.0%	–4.3%

(continued)

Exhibit 7 (Continued)

Market	Score	Active Return Volatility	Expected Active Return	Active Weight	ACWI Benchmark Weight	Portfolio Weight
Sweden	0.0	10.1%	0.0%	−1.2%	0.8%	−0.4%
Hong Kong SAR	1.0	12.0%	1.2%	2.3%	1.2%	3.5%
Netherlands	0.0	7.9%	0.0%	1.8%	1.1%	2.9%
Italy	−1.0	15.1%	−1.5%	−0.8%	0.7%	−0.1%
Singapore	0.0	11.8%	0.0%	−1.8%	0.4%	−1.3%
Belgium	1.0	10.2%	1.0%	4.4%	0.3%	4.7%
Denmark	0.0	12.3%	0.0%	−1.2%	0.5%	−0.7%
Finland	−1.0	13.7%	−1.4%	−2.5%	0.3%	−2.2%
Norway	0.0	13.7%	0.0%	0.1%	0.2%	0.3%
Israel	1.0	15.3%	1.5%	1.4%	0.2%	1.6%
Ireland	0.0	14.9%	0.0%	0.3%	0.2%	0.5%
Austria	−1.0	14.6%	−1.5%	−1.6%	0.1%	−1.6%
Portugal	0.0	15.2%	0.0%	0.9%	0.0%	0.9%
New Zealand	1.0	14.2%	1.4%	1.8%	0.1%	1.8%
United States	0.0	3.8%	0.0%	−5.3%	54.3%	48.9%
Canada	−1.0	9.2%	−0.9%	−6.7%	3.0%	−3.7%
Emerging	0.0	9.0%	0.0%	0.1%	11.9%	12.0%
Total	0.0			0.0%	100.0%	100.0%

Transfer Coefficient	Information Coefficient	Breadth
0.995	0.099	24.5

Active Return	Active Risk	Information Ratio
0.98%	2.00%	0.49

The 24 individual assets in Exhibit 7 are listed approximately by size in the EAFE benchmark, followed by the United States, Canada, and the emerging markets. For example, the United Kingdom has a benchmark weight of 5.2% and Canada has a benchmark weight of 3.0%. Scores representing an active investor's forecasts of the relative performance of each asset during 2019 are assigned to each market. The scores are one of five numerical values that represent a managerial forecast of strong outperformance (2.0), weak outperformance (1.0), neutral performance (0.0), weak underperformance (−1.0), and strong underperformance (−2.0). The number of scores in each of these five categories is based on the requirement that the scores sum to zero and have a cross-sectional standard deviation of 1.

The active return forecasts in the fourth column of Exhibit 7 are based on the Grinold rule in Equation 10 of "IC times volatility times score," where IC is the *ex ante* information coefficient that measures the assumed accuracy of the investor's relative rankings, as illustrated by the right leg of the correlation triangle in Exhibit 6. In this example, we use an assumed information coefficient of 0.10; thus, forecasted and realized active security returns are expected to have a cross-sectional correlation coefficient of 0.10. For example, the active return forecast for the United Kingdom, which has a score of 2.0, is 0.10(6.4)(2.0) = 1.3%. Alternatively, the active return forecast for Japan is 0.0% because the score is 0. As explained later, the information coefficient used in fundamental law accounting will be adjusted down to 0.095, as shown at the bottom of Exhibit 7, to account for the assignment of scores in this particular example.

The active weights for each market are based on the active return forecast and a numerical optimizer (i.e., Excel Solver) with the objective to maximize the expected active return of the portfolio, subject to a 2.00% constraint on active risk. Note that while the active weights for each market are generally correlated with the forecasted active returns in Exhibit 7, they are not perfectly proportional for two reasons. First, the optimizer also takes into account the estimated correlations between each market's active return, based on the MSCI monthly return data from 2009 to 2018. Exhibit 8 reports the estimated active return correlations for the eight largest EAFE countries; the full correlation matrix is not reported to conserve space. For example, the correlation coefficient between the United Kingdom (GB) and Japan (JP) is fairly low at −0.02, while the correlation coefficient between France (FR) and Germany (DE) is higher at 0.30. Note that these correlation coefficients are for *active* returns (i.e., the differences between the individual market and ACWI benchmark returns). The correlations for *total* market returns would all be positive and much higher—for example, values that range from 0.4 to 0.9.

Exhibit 8 Active Return Correlation Coefficients for Eight Countries (based on MSCI returns from 2009 to 2018)

Country	GB	JP	FR	DE	CH	AU	ES	SE	
GB	1.000	−0.02	0.21	0.08	0.13	0.00	0.18	0.14	…
JP	−0.02	1.000	−0.08	−0.03	0.04	−0.08	0.01	−0.07	…
FR	0.21	−0.08	1.000	0.30	0.16	−0.03	0.34	0.15	…
DE	0.08	−0.03	0.30	1.000	0.10	−0.07	0.19	0.15	…
CH	0.13	0.04	0.16	0.10	1.000	0.06	0.11	0.10	…
AU	0.00	−0.08	−0.03	−0.07	0.06	1.000	0.01	0.06	…
ES	0.18	0.01	0.34	0.19	0.11	0.01	1.000	0.11	…
SE	0.14	−0.07	0.15	0.15	0.10	0.06	0.11	1.000	…
	…	…	…	…	…	…	…	…	…

The second reason that the active weights in Exhibit 7 are not perfectly proportional to the forecasted active returns is that the active weights are constrained by the optimizer to sum to zero. For example, the highest active weight in Exhibit 7 is for the United Kingdom, at 15.7%, and the lowest active weight is for France, at −12.0%. These active weights are added to the benchmark weights to give the total portfolio weights in the last column of Exhibit 7. For example, the total weight for the United States is 48.9%, even though the active weight is −5.3%, because the US benchmark weight in the ACWI is 54.3%. In fact, the optimization in Exhibit 7 is for a relatively unconstrained long–short portfolio where the sum of the positive total weights is about 120% and the sum of the negative total weights is about −20%, what might be called a "120/20 long–short" strategy in practice.

Because the optimization is basically unconstrained, the transfer coefficient or risk-weighted correlation between active return forecasts and active weights shown at the bottom of Exhibit 7 is 0.995, almost perfect. The transfer coefficient in this example takes into account all the risk statistics (i.e., forecasted active volatilities *and* forecasted correlations), but it is not exactly 1.0 because of the budget constraint that the active weights sum to zero. Alternatively, if the sum of active weights were allowed to be non-zero, effectively allowing for risk-free cash or leverage in the equity portfolio to meet the budget constraint, the transfer coefficient would be exactly 1.0. The breadth of the strategy shown at the bottom of Exhibit 7 is 24.5, slightly higher than the number of individual assets, 24.0, because the risk model includes active return correlation coefficients that are different from zero. If all the off-diagonal correlations in the extended table in Exhibit 8 were exactly zero, then breadth would be exactly 24.0, instead of 24.5.

The fundamental law in Equation 14 states that the expected active return on the portfolio is $E(R_A) = (TC)(IC)\sqrt{BR}\sigma_A = 0.995 \times 0.099 \times (24.5)^{1/2} \times 2.00 = 0.98\%$. Alternatively, the expected active return of 0.98%, shown at the bottom of Exhibit 7, is calculated as the sum of the active weights times active returns. Thus, the accuracy of the fundamental law is quite high. The fundamental law is often expressed in terms of the information ratio, or forecasted active return over active risk. Using this framework, the validation of the fundamental law is $IR = (TC)(IC)\sqrt{BR} = 0.995 \times 0.099 \times (24.5)^{1/2} = 0.49$, equal to actual forecasted active return divided by active risk, 0.98/2.00 = 0.490. Because the information ratio in this relatively unconstrained portfolio is unaffected by the aggressiveness of the strategy, we would get the same IR value if the active risk were allowed to be higher. For example, if the active risk specified to the optimizer were increased to 3.00%, the forecasted active return would increase to 1.47%, an information ratio of, again, 1.47/3.00 = 0.49.

Exhibit 9 continues examining the global equity strategy but uses a slightly different assignment of scores than Exhibit 7 to illustrate how this change affects the values in the fundamental law. Specifically, the scores for Germany (DE) and the United Kingdom (UK) have been switched as well as the scores for Switzerland (CH) and Australia (AU). While the breadth in Exhibit 9 is unchanged at 24.5, the information coefficient has increased slightly to 0.105, compared with 0.099 in Exhibit 7. Even though the assumed IC used to create the expected active returns in Exhibit 9 is still 0.10, the IC used in fundamental law accounting has increased because the new assignment of scores represents a slightly more ambitious forecast. For example, the active (i.e., benchmark relative) returns for France and Germany in Exhibit 9 are now forecasted to go strongly in opposite directions, even though they are positively correlated according to the risk model in Exhibit 8. Given the increase in IC and slight change in TC, the fundamental law calculation for Exhibit 9 is now $IR = (TC)(IC)\sqrt{BR} = 0.997 \times 0.105 \times (24.5)^{1/2} = 0.532$, equal to the actual value of 0.52.

Exhibit 9 Long–Short Global Equity Fund with Different Scores for 2019 (risk statistics based on MSCI returns from 2009 to 2018)

Market	Score	Active Return Volatility	Expected Active Return	Active Weight	ACWI Benchmark Weight	Portfolio Weight
United Kingdom	0.0	6.4%	0.0%	1.7%	5.2%	6.9%
Japan	0.0	9.1%	0.0%	−0.3%	7.6%	7.3%
France	−2.0	8.5%	−1.7%	−11.4%	3.4%	−8.0%
Germany	2.0	8.4%	1.7%	14.5%	2.7%	17.2%
Switzerland	0.0	7.8%	0.0%	−2.0%	2.7%	0.6%
Australia	2.0	11.1%	2.2%	7.7%	2.1%	9.9%
Spain	−2.0	16.0%	−3.2%	−4.9%	1.0%	−4.0%
Sweden	0.0	10.1%	0.0%	−1.2%	0.8%	−0.4%

Exhibit 9 (Continued)

Market	Score	Active Return Volatility	Expected Active Return	Active Weight	ACWI Benchmark Weight	Portfolio Weight
Hong Kong SAR	1.0	12.0%	1.2%	2.8%	1.2%	4.0%
Netherlands	0.0	7.9%	0.0%	3.0%	1.1%	4.1%
Italy	−1.0	15.1%	−1.5%	−0.9%	0.7%	−0.2%
Singapore	0.0	11.8%	0.0%	−1.4%	0.4%	−1.0%
Belgium	1.0	10.2%	1.0%	5.8%	0.3%	6.1%
Denmark	0.0	12.3%	0.0%	−0.7%	0.5%	−0.1%
Finland	−1.0	13.7%	−1.4%	−3.2%	0.3%	−2.9%
Norway	0.0	13.7%	0.0%	1.3%	0.2%	1.5%
Israel	1.0	15.3%	1.5%	1.5%	0.2%	1.7%
Ireland	0.0	14.9%	0.0%	0.2%	0.2%	0.4%
Austria	−1.0	14.6%	−1.5%	−1.8%	0.1%	−1.7%
Portugal	0.0	15.2%	0.0%	1.4%	0.0%	1.4%
New Zealand	1.0	14.2%	1.4%	1.4%	0.1%	1.4%
United States	0.0	3.8%	0.0%	−5.0%	54.3%	49.2%
Canada	−1.0	9.2%	−0.9%	−5.3%	3.0%	−2.3%
Emerging	0.0	9.0%	0.0%	−2.9%	11.9%	9.0%
Total	0.0			0.0%	100.0%	100.0%

Transfer Coefficient	Information Coefficient	Breadth
0.997	0.105	24.5

Active Return	Active Risk	Information Ratio
1.04%	2.00%	0.52

We now apply constraints to the global equity strategy to focus on the transfer coefficient. Specifically, Exhibit 10 shows two *constrained* portfolio optimizations using the same score assignments and thus active return forecasts as in Exhibit 7. The first optimization, shown on the left-hand side of Exhibit 10, has two constraints. First, the portfolio is constrained to be long only (i.e., a negative active weight for any given market cannot be bigger than the benchmark weight). For example, France has an active weight of −3.4%, bounded by the benchmark weight of 3.4%, so that the total weight for France in the managed portfolio is zero. Second, the portfolio weights are constrained to not be more than 10.0% over or under the benchmark weight (i.e., the absolute value of any given market active weight cannot be greater than 10.0%). For example, the active weights for the United Kingdom and Switzerland are limited to 10.0% and the active weight for the United States is limited to −10.0%.

Exhibit 10 Constrained Global Equity Funds for 2019 (risk statistics based on MSCI returns from 2009 to 2018)

Market	Active Weight	ACWI Benchmark Weight	Portfolio Weight	Active Weight	ACWI Benchmark Weight	Portfolio Weight
United Kingdom	10.0%	5.2%	15.2%	8.6%	5.2%	15.2%
Japan	−6.6%	7.6%	1.0%	−7.8%	7.76%	0.0%
France	−3.4%	3.4%	0.0%	−3.7%	3.4%	0.0%
Germany	−2.7%	2.7%	0.0%	−3.5%	2.7%	0.0%
Switzerland	10.0%	2.7%	12.7%	10.0%	2.7%	12.7%
Australia	−2.1%	2.1%	0.0%	−2.8%	2.1%	0.0%
Spain	−1.0%	1.0%	0.0%	−1.2%	1.0%	0.0%
Sweden	−0.8%	0.8%	0.0%	−1.2%	0.8%	0.0%
Hong Kong SAR	5.9%	1.2%	7.1%	9.5%	1.2%	5.7%
Netherlands	−1.1%	1.1%	0.0%	−1.0%	1.1%	0.0%
Italy	−0.7%	0.7%	0.0%	−0.8%	0.7%	0.0%
Singapore	−0.4%	0.4%	0.0%	−0.5%	0.4%	0.0%
Belgium	6.1%	0.3%	6.4%	−0.4%	0.3%	0.0%
Denmark	−0.5%	0.5%	0.0%	−0.4%	0.5%	0.0%
Finland	−0.3%	0.3%	0.0%	−0.3%	0.3%	0.0%
Norway	−0.2%	0.2%	0.0%	−0.3%	0.2%	0.0%
Israel	4.2%	0.2%	4.4%	10.0%	0.2%	10.2%
Ireland	−0.2%	0.2%	0.0%	−0.1%	0.2%	0.0%
Austria	−0.1%	0.1%	0.0%	−0.1%	0.1%	0.0%
Portugal	0.0%	0.0%	0.0%	−0.1%	0.0%	0.0%
New Zealand	4.7%	0.1%	4.8%	10.0%	0.1%	10.1%
United States	−10.0%	54.3%	44.3%	−10.0%	54.3%	44.3%
Canada	−3.0%	3.0%	0.0%	−3.7%	3.0%	0.0%
Emerging	−7.7%	11.9%	4.2%	−10.0%	11.9%	1.9%
Total	0.0%	100.0%	100.0%	0.0%	100.0%	100.0%

	Transfer Coefficient	Information Coefficient	Breadth	Transfer Coefficient	Information Coefficient	Breadth
	0.694	0.099	24.5	0.567	0.099	24.5

	Active Return	Active Risk	Information Ratio	Active Return	Active Risk	Information Ratio
	0.68%	2.00%	0.34	0.76%	2.74%	0.28

The long-only and maximum over- or underweight constraints substantially reduce the transfer of active return forecasts into active weights, as shown by the transfer coefficient of 0.694 at the bottom of the left side of Exhibit 10, compared with 0.995 for the same scores and active return forecasts in Exhibit 7. The impact of this transfer coefficient on expected active return according to the fundamental law is $E(R_A) = (TC)(IC)\sqrt{BR}\sigma_A = 0.694 \times 0.099 \times (24.5)^{1/2} \times 2.00 = 0.68\%$, compared with 0.98% for the unconstrained portfolio in Exhibit 7. Similarly, the impact of this transfer coefficient measured by the information ratio is $\text{IR} = (\text{TC})(\text{IC})\sqrt{\text{BR}} = 0.694 \times 0.099 \times (24.5)^{1/2} = 0.34$, compared with 0.49 for the unconstrained portfolio. In other words, the expected active return and information ratio are reduced by almost a third because of the constraints imposed in portfolio construction.

As previously mentioned, an increase in the allowed active risk from 2.00% to 3.00% in the unconstrained portfolio in Exhibit 7 proportionally increases the active return, leaving the information ratio at about 0.49. However, an increase in allowed active risk to 3.00% does *not* preserve the information ratio of the constrained portfolio, as shown by the optimization on the right-hand side of Exhibit 10. Specifically, the higher active risk leads to more variation in unconstrained active weights, as shown in Equation 11; thus, the constraints become more binding. For example, the active weight for New Zealand, which is 4.7% on the left-hand side of Exhibit 10, is capped at the maximum possible value of 10.0% on the right-hand side of Exhibit 10. The result is a further reduction in the transfer coefficient from 0.694 to 0.567, leading to a reduction in the information ratio to $\text{IR} = (\text{TC})(\text{IC})\sqrt{\text{BR}} = 0.567 \times 0.099 \times (24.5)^{1/2} = 0.28$, compared with 0.34 at the lower active portfolio risk of 2.0%.

The key concept is that although an unconstrained IR is invariant to the level of active risk, as shown by the dark line in Exhibit 5, the IR for a *constrained* portfolio generally decreases with the aggressiveness of the strategy. Specifically, the dark line in Exhibit 5 for a constrained portfolio would curve downward from left to right in accordance with an increasingly lower transfer coefficient. Thus, the constraints that are imposed on the portfolio should inform the decision of how aggressively to apply an active management strategy.

EXAMPLE 8

Compare and Contrast Active Management Strategies

Consider two active management strategies: individual stock selection, with a benchmark composed of 100 securities, and industrial sector selection, with a benchmark of nine sectors. The active security returns are defined as residuals in a risk model and thus are essentially uncorrelated, and forecasts are independent from year to year. Suppose the individual stock investor is expected to exhibit skill as measured by an information coefficient of 0.05, while the industrial sector investor has a higher information coefficient of 0.15.

1. Conceptually, what is the breadth (i.e., number of independent decisions per year) of each active management strategy?
2. Calculate the expected information ratio for each strategy under the assumption that each investor's forecasts can be implemented without constraints, such as the long-only constraint or a limit on turnover each year.
3. Suppose the aggressiveness of each active management strategy is established by a portfolio active risk target of 3.0% per year. What is the expected active return to each strategy?
4. Under the more realistic assumption that the individual security selection strategy is constrained to be long only and has turnover limits, the transfer coefficient has a value of 0.60. Calculate the constrained information ratio and expected active return of the security selection strategy.
5. Suppose the aggressiveness of the constrained individual security selection strategy is increased to a portfolio active risk target of 4.0% per year. Conceptually, what is likely to happen to the information ratio, and why?

Solution to 1:

Given that the active asset returns in each strategy are uncorrelated and forecasts are independent from year to year, the breadth of the security selection strategy is BR = 100 and of the sector selection strategy is BR = 9.

Solution to 2:

The expected information ratio of the unconstrained security selection strategy is calculated as $\text{IR} = (\text{IC})\sqrt{\text{BR}} = 0.05 \times \sqrt{100} = 0.50$, while the information ratio of the industrial sector selection strategy is $\text{IR} = (\text{IC})\sqrt{\text{BR}} = 0.15 \times \sqrt{9} = 0.45$.

Solution to 3:

The expected active return to the unconstrained security selection strategy is 0.50(3.0) = 1.50%, while the expected active return of the industrial sector selection strategy is 0.45(3.0) = 1.35%.

Solution to 4:

The information ratio of the constrained security selection strategy is $\text{IR} = (\text{TC})(\text{IC})\sqrt{\text{BR}} = 0.60 \times 0.05 \times \sqrt{100} = 0.30$, rather than 0.50, and the expected active return is 0.30(3.0) = 0.90%, rather than 1.50%.

Solution to 5:

A more aggressive implementation of the constrained security selection strategy will likely result in larger deviations of constrained weights from unconstrained weights and thus a lower transfer coefficient. For example, the transfer coefficient might drop from 0.60 to 0.50, leading to an information ratio of only $\text{IR} = (\text{TC})(\text{IC})\sqrt{\text{BR}} = 0.50 \times 0.05 \times \sqrt{100} = 0.25$. Thus, instead of a proportional increase in the expected active return associated with an increase in the active portfolio risk from 3.0% to 4.0%, the expected active return would only increase from 0.9% to 0.25(4.0) = 1.0%.

8 FIXED-INCOME STRATEGIES

d explain how the information ratio may be useful in investment manager selection and choosing the level of active portfolio risk;

e compare active management strategies, including market timing and security selection, and evaluate strategy changes in terms of the fundamental law of active management;

Two additional examples of the fundamental law in practice are based on the Bloomberg Barclays US fixed-income index returns. Consider first an active management strategy of over- and underweighting credit exposure once a quarter using corporate investment-grade and high-yield bond portfolios as assets. Let the benchmark portfolio be composed of 70% investment-grade bonds and 30% high-yield bonds. Each quarter, the active investor makes a single dichotomous decision either to overweight the investment-grade asset (and thus underweight the high-yield asset) or to overweight the high-yield asset (and thus underweight the investment-grade asset). In addition to switching to a fixed-income example, we are also now moving into a time-series application of the fundamental law instead of the purely cross-sectional application.

For example, consider two bond portfolios, an investment-grade portfolio and a high-yield portfolio. The quarterly return volatility of the IG (investment-grade) asset is 2.84%, and the quarterly return volatility of the HY (high-yield) asset is 4.64%, with an estimated correlation between the two of 0.575. The *active* risk of this decision is the volatility of the differential returns between the two bond portfolios, $[(2.84)^2 - 2(2.84)(4.64)(0.575) + (4.64)^2]^{1/2} = 3.80\%$. In effect, the active investor assigns a "score" of either +1.0 or –1.0 on credit exposure each quarter, with an *annualized* active risk of $3.80 \times (4)^{1/2} = 7.60\%$. Suppose the fixed-income investor expects to call the market correctly 55% of the time (i.e., 11 out of 20 quarters). If the investor makes the correct decision 55% of the time and an incorrect decision 45% of the time, then the time-series information coefficient is 0.55 – 0.45 = 0.10.

If a time series of T predicted dichotomous (i.e., plus or minus 1.0) scores, $S_{P,t}$, and a time series of T realized dichotomous scores, $S_{R,t}$, both have zero means, then the time-series covariance between the two is $\mathrm{COV}(S_P,S_R)=\frac{1}{T}\sum_{t=1}^{T}S_{P,t}S_{R,t}$. The product of the two scores at time period t is 1.0 if the scores have the same sign (i.e., the decision is correct) and –1.0 if the scores have different signs (i.e., the decision is incorrect). Because the scores have unit variances, the correlation coefficient is equal to the covariance. Thus, the time-series correlation is equal to the number of correct decisions minus the number of incorrect decisions all over total decisions, or, in other words, the percentage correct minus the percentage incorrect.

Without a limit on active risk, the expected active return can be calculated using a simple probability-weighted average: 0.55(3.80) + 0.45(–3.80) = 38 bps per quarter. But to illustrate the fundamental law, we use the Grinold rule in Equation 10 of "alpha equals IC times volatility times score": 0.10(3.80)(1.0) = 38 bps.

The investor decides to limit the annual active risk to 2.00% and thus sets the active weight (i.e., deviation from the 70/30 benchmark weights) at 2.00/7.60 = 26.3%. Under the assumption that active returns are uncorrelated over time, the breadth of this strategy is 4.0, the four quarterly rebalancing decisions made each year. Thus, in quarters when the investor believes credit risk will pay off, the managed portfolio is invested 70.0% – 26.3% = 43.7% in investment-grade bonds and 30% + 26.3% = 56.3% in high-yield bonds. Alternatively, in quarters where the investor believes credit risk will not pay off, the active portfolio has 70.0% + 26.3% = 96.3% in investment-grade bonds and only 30% – 26.3% = 3.7% in high-yield bonds. According to the simple form of the fundamental law, the expected annualized active return to this strategy is $\mathrm{E}(R_A) = (\mathrm{IC})\sqrt{\mathrm{BR}}\sigma_A = 0.10 \times (4.0)^{1/2} \times 2.00 = 40$ bps a year, or 10 bps per quarter. Alternatively, given the active weight of 26.3% motivated by the desire to limit active risk, the expected quarterly return can be calculated more directly as 0.263 × 38 = 10 bps. Given the small breadth of this strategy, the annual information ratio is only $\mathrm{IR} = (\mathrm{IC})\sqrt{\mathrm{BR}} = 0.10 \times (4.0)^{1/2} = 0.20$.

The key concept in this illustration is that the breadth of the strategy is only 4, meaning four active management decisions per year. The same small-breadth problem also applies to quarterly tactical asset allocation decisions in a simple strategy that switches between equity and cash. There are so few opportunities to make an active decision in these "market-timing" strategies that the investor's accuracy as measured by the information coefficient must be quite high to achieve even a modest information ratio. A full description of the breadth calculation requires relatively complex matrix formulas that take into account the correlations between security returns. However, one "rule of thumb" is that breadth is approximately BR = $N/[1 + (N - 1)\rho]$, where N is the number of securities and ρ is the average correlation between the active security returns. In this fixed-income example, $\rho = 0.0$, so breadth is BR = 4.0.

A natural question is whether the expected information ratio can be increased by switching more frequently—say, monthly. Although it is somewhat more complicated to show, the basic answer is yes—*if* the information coefficient of 0.10 can be maintained and *if* the credit exposure decisions in this example are truly independent over time. For example, making monthly decisions that do not change during the quarter (i.e., signals of +1.0, +1.0, and +1.0 in January, February, and March) will *not* increase the information ratio of 0.20. However, if the monthly signals are truly uncorrelated with each other, then the information ratio in this example would be $\text{IR} = (\text{IC})\sqrt{\text{BR}} = 0.10 \times \sqrt{12} = 0.35$. Although somewhat implausible, if an investor made daily decisions (250 trading days a year) that were truly independent and were *still* correct 55% of the time, the expected information ratio could potentially increase to $\text{IR} = (\text{IC})\sqrt{\text{BR}} = 0.10 \times \sqrt{250} = 1.58$.

The high 1.58 information ratio indicates that the investor could earn an expected active return of 3.16% with active risk of only 2.00%. With such a high information ratio, the investor might be inclined to increase the aggressiveness of the credit risk strategy—for example, doubling to an expected return of 2 × 3.16% = 6.32% and active risk of 2 × 2.00% = 4.00%. Besides the issue of transaction costs, this more aggressive strategy would likely bump up against various constraints. For example, at the higher 4.00% active risk, the required active weights would be plus and minus 4.00/7.60 = 52.6%. In other words, a tilt against credit risk would require a total portfolio weight of 70% + 52.6% = 122.6% in investment-grade bonds funded by a −22.6% *short* position in high-yield bonds.

The essential logic of this example is not confined to a dichotomous decision; the same general perspectives would hold if the single credit risk signal were continuous—for example, numbers like −0.57 or 1.32. Then under the more aggressively applied active risk target of 4.0%, a signal of −0.57 would require an active weight of −0.57(4.0)/7.6 = −30.0%. With a benchmark portfolio of 70.0% investment-grade and 30.0% high-yield bonds, this active weight translates into a 100% position in investment-grade bonds and no position in high-yield bonds. Alternatively, for a positive credit risk signal of 1.32, the required active weight would be 1.32(4.00)/7.60 = 69.5% (i.e., 100% in high-yield bonds and almost no position in investment-grade bonds). In other words, for this more aggressive strategy under a long-only constraint, the transfer coefficient would be less than 1 and the expanded fundamental law, $\text{IR} = (\text{TC})(\text{IC})\sqrt{\text{BR}}$, would come into play. Under a normal distribution for scores, the transfer coefficient of this strategy is 0.62, so the expected information ratio is only $\text{IR} = 0.62 \times 0.10 \times \sqrt{250} = 0.98$, not 1.58. For an active risk of 4.00%, the expected active return is thus only 0.98 × 4.00% = 3.92%, not 6.32%. Please note that the transfer coefficient in this example is based on the calculation $\Phi(1.32) - \Phi(-0.57) = 0.62$, where $\Phi(S)$ is the cumulative standard normal distribution function. Given long-only limits on positions, the actual active risk of the constrained portfolio would be lower than 4.0%. In other words, the actual active weights (determined by a numerical optimizer) would need to be larger than the simple formula (S)4.0/7.6 to get back up to an actual active risk of 4.0%.

For our second fixed-income example, consider an active management strategy using the five US Treasury bond portfolios in Exhibit 11 as the individual assets. Let the neutral benchmark be an equally weighted composite portfolio of the five, or 20% invested in each asset, but with annual rebalancing. In other words, we are now moving back into a purely cross-sectional application of the fundamental law.

Exhibit 11 Bloomberg Barclays US Treasury Bond Average Returns and Risk (return statistics from 2009 to 2018)

	Treas. 0–1	Treas. 1–3	Treas. 3–7	Treas. 7–10	Treas. 10–20
Avg. Ret.	0.40%	0.90%	2.21%	3.15%	3.89%
Volatility	0.17%	0.85%	3.20%	5.86%	7.95%

Exhibit 12 shows the volatility of the historical return differences between each asset and the equally weighted benchmark. Note that while the absolute volatility of each asset return goes up with maturity in Exhibit 11, the *active* volatility with respect to the benchmark is highest for the assets with the shortest maturity, at 3.45%, and the longest maturity, at 4.57%. Exhibit 12 also shows the estimated active (i.e., benchmark relative) return correlation matrix, which has both positive and negative values, in contrast to the absolute return correlation matrix (not shown), which would only have large positive values. For example, the correlation between the 0–1-year T-bond active return and the 1–3-year T-bond active return in Exhibit 12 is *positive* 0.49, showing that these shorter-maturity active returns tend to move together. However, the correlation between the 0–1-year active return and the 7–10-year active return is *negative*, at −0.49, showing that these two diverse maturity active returns tend to move apart.

Exhibit 12 US Treasury Bond Estimated Active Return Risk and Correlations (return statistics from 2009 to 2018)

	Treas. 0–1	Treas. 1–3	Treas. 3–7	Treas. 7–10	Treas. 10–20
Active Vol.	3.45%	2.85%	1.05%	2.40%	4.57%

Active Corr.	**Treas. 0–1**	**Treas. 1–3**	**Treas. 3–7**	**Treas. 7–10**	**Treas. 10–20**
Treas. 0–1	1.000	0.49	0.21	−0.49	−0.47
Treas. 1–3	0.49	1.000	0.26	−0.49	−0.49
Treas. 3–7	0.21	0.26	1.000	−0.19	−0.33
Treas. 7–10	−0.49	−0.49	−0.19	1.000	0.46
Treas. 10–20	−0.47	−0.49	−0.33	0.46	1.000

The breadth associated with the risk estimates in Exhibit 12 is 9.4, even though there are only 5 assets. The breadth is different from the number of assets because the off-diagonal values in the correlation matrix are substantially different from zero. Exhibit 13 shows the fundamental law calculations for two sets of scores given an active portfolio risk target of 1.0% per year. The first set of scores has positive values for the shorter-maturity bonds and negative scores for the longer-maturity bonds. The associated active returns are calculated using the Grinold rule in Equation 10 and an assumed information coefficient of 0.20; for example, the active return for 10–20-year T-bonds is 0.20 × 4.57% × −1.76 = −1.61%. The active weights in Exhibit 13 are calculated by an optimizer given the constraint on active risk of 1.00%. For example, the active weight for the 10–20-year T-bonds is −19.1%, shown in the upper half of Exhibit 13. Given the benchmark weights of 20% for each asset, this results in a total weight of only 20 − 19.1 = 0.9% in the managed portfolio.

Although the information coefficient used to scale the active returns was 0.20, the first set of scores in Exhibit 13 does not represent a very ambitious forecast, so the information coefficient used in the fundamental law calculation is 0.12. The intuition for the large downward adjustment in the information coefficient is that the positive scores for the shorter-maturity bonds and the negative scores for the longer-maturity bonds are all based on essentially one active decision that interest rates will rise. Specifically, the expected active (i.e., benchmark relative) return for the managed fixed-income portfolio is $E(R_A) = (IC)\sqrt{BR}\sigma_A = 0.12 \times (9.4)^{1/2} \times 1.00 = 37$ bps a year.

Exhibit 13 Signals and Weights for a Fixed-Income Portfolio with Breadth of 9.4 and Active Risk of 1.00% (return statistics from 2009 to 2018)

	Treas. 0–1	Treas. 1–3	Treas. 3–7	Treas. 7–10	Treas. 10–20	IC	Active Ret.
Score	0.63	0.67	0.92	−0.46	−1.76	0.12	0.37%
Active Ret.	0.43%	0.38%	0.19%	−0.22%	−1.61%		
Active Wgt.	−1.6%	−2.1%	15.4%	7.4%	−19.1%		
Total Wgt.	18.4%	17.9%	35.4%	27.4%	0.9%		
Score	−0.22	1.20	0.23	0.57	−1.77	0.18	0.55%
Active Ret.	−0.15%	0.68%	0.05%	0.27%	−1.62%		
Active Wgt.	−11.3%	17.0%	−12.8%	24.3%	−17.2%		
Total Wgt.	8.7%	37.0%	7.2%	44.3%	2.8%		

In contrast, the second set of scores in Exhibit 13 is a more ambitious set of active forecasts that specify a modification in the shape of the yield curve. As a result, the information coefficient is 0.18, not much lower than the 0.20 value used to scale the active returns, and the expected active return for the portfolio using the fundamental law is $E(R_A) = (IC)\sqrt{BR}\sigma_A = 0.18 \times (9.4)^{1/2} \times 1.00 = 55$ bps a year. The fundamental law in terms of the expected information ratio for the second set of scores in Exhibit 13 is $IR = (IC)\sqrt{BR} = 0.18 \times (9.4)^{1/2} = 0.55$, alternatively calculated as the expected active return over active risk, 55/100 = 0.55.

At this relatively high information ratio, the investor may be inclined to increase the active risk to, say, 2.00% instead of 1.00%. However, given that the longest-maturity asset has a total weight that is approaching zero (i.e., 2.8%, as shown in the lower right-hand corner of Exhibit 13), such a strategy would likely require shorting; if short sells are not allowed, the transfer coefficient would likely end up being less than 1.00.

EXAMPLE 9

Breadth and Rebalancing in Active Management Strategies

Consider an active portfolio management strategy that involves decisions on overweighting or underweighting four individual assets. For example, the assets might be ETFs for four country equity markets or four different fixed-income

ETFs. The active returns to Assets #1 and #2 are positively correlated, as are the active returns to Assets #3 and #4. However, the assumed risk model for active returns has no other non-zero correlations. The correlation structure in this risk model is shown in the following 4-by-4 correlation matrix, and the breadth calculation is BR = 3.2. For simplicity, we will assume that the portfolio management decisions are dichotomous; thus, each year the investor forecasts two of the assets to outperform the benchmark and the other two assets to underperform.

Correlations	#1	#2	#3	#4
#1	1.00	0.25	0.00	0.00
#2	0.25	1.00	0.00	0.00
#3	0.00	0.00	1.00	0.25
#4	0.00	0.00	0.25	1.00

1. Conceptually speaking (i.e., exact numbers are not necessary), why is the breadth less than the number of assets for this strategy?
2. Suppose the investor predicts that Assets #1 and #2 will outperform and that Assets #3 and #4 will underperform. Conceptually speaking (i.e., exact numbers are not necessary), how will these scores affect the information coefficient in the fundamental law compared with a prediction that Assets #1 and #3 will outperform and Assets #2 and #4 will underperform?
3. Suppose the active investor rebalances monthly instead of just once a year. Explain how this would affect the information ratio of this strategy, clearly stating your assumptions.

Solution to 1:

According to the risk model, the active returns to Assets #1 and #2 tend to move together, with a correlation coefficient of 0.25, as do the active returns for Assets #3 and #4. As a result, the 3.2 breadth of this strategy is lower than the number of assets, $N = 4$.

Solution to 2:

According to the risk model, the active returns to Assets #1 and #2 tend to move together, so a forecast that both will outperform is not as ambitious as a forecast that one will outperform while the other underperforms. As a result, the information coefficient will be adjusted downward by more under the first set of forecasts than under the second set of forecasts.

Solution to 3:

Rebalancing monthly instead of annually could increase the breadth by a factor of 12 but only if the active management decisions for each asset are truly uncorrelated over time. For example, the breadth could increase to as much as $12 \times 3.2 = 38.4$. However, to increase the information ratio, one would have to assume that the information coefficient remains at the same level and that there are no constraints to fully implementing the active management decisions (i.e., a transfer coefficient of 1.00). For example, turnover constraints might limit the degree to which the monthly active management decisions could be fully implemented into new active positions, resulting in a lower transfer coefficient.

In summary, these examples illustrate how the information coefficient, IC, measures the strength of the return-forecasting process, or signal. The information coefficient is the correlation between the forecasted and realized security active returns and is anticipated to be positive or active management is not justified. Breadth, BR, measures the number of independent decisions made by the investor each year and is equal to the number of securities if the active returns are cross-sectionally uncorrelated. Similarly, breadth increases with the number of rebalancing periods but only if the active returns are uncorrelated over time.

Like the information coefficient, the transfer coefficient, TC, is a simple multiplicative factor in the fundamental law. It measures the extent to which constraints reduce the expected value added of the investor's forecasting ability. In the absence of constraints, the transfer coefficient is approximately 1.00, resulting in the basic form of the fundamental law. However, in practice, investors often work under constraints that result in TC values between 0.20 and 0.80. The lower transfer coefficient suggests that average performance in practice is only a fraction (20%–80%) of what would otherwise be predicted by the basic form of the fundamental law.

9 PRACTICAL LIMITATIONS

f describe the practical strengths and limitations of the fundamental law of active management.

The limitations of the fundamental law include both practical considerations, such as ignoring transaction costs and taxes, and more conceptual issues, such as dynamic implementation over time. In this section, we focus on two limitations: the *ex ante* measurement of skill using the information coefficient and assumptions of independence in forecasts across assets and over time. The fundamental law extends the mean–variance-optimization approach to relative performance and hence has many of the same limitations of mean–variance optimization. In our discussion, we do not deal with the shortcomings of mean–variance optimization in general (e.g., assumptions of normality in return distributions or the degree of risk aversion) or the technical problems associated with the estimation and use of a risk model (e.g., the correct set of risk factors, nonlinearities, and non-stationary returns). The fundamental law takes as given that mean–variance optimization to balance risk and return against a benchmark is the correct objective function and that the investor has a way to adequately model risk.

9.1 *Ex Ante* Measurement of Skill

A core element of the fundamental law is the information coefficient, generally defined as the correlation between the portfolio investor's forecasts and actual outcomes. Active investors assume that the financial market they are trading in is not perfectly efficient in terms of public information and that they have some differential skill in competing with other active investors; otherwise, active management is generally not justified. Behaviorally, one might argue that investors tend to overestimate their own skills as embedded in the assumed IC, but even if that bias did not exist, questions about assessing an accurate level of skill remain. Furthermore, forecasting ability probably differs among different asset segments and varies over time.

For example, Qian and Hua (2004) expanded the basic form of the fundamental law by including the uncertainty about the level of skill, or the reality that the realized information coefficient can vary over time. Specifically, they showed that realized active portfolio risk, σ_A, is a product of both the benchmark tracking risk predicted by the risk model, denoted σ_{RM}, and the additional risk induced by the uncertainty of the information coefficient, denoted σ_{IC}:

$$\sigma_A = \sigma_{IC}\sqrt{N}\sigma_{RM} \tag{17}$$

Their insight about "strategy risk" is derived under the simplifying assumptions that portfolio positions are unconstrained, TC = 1.00, and that breadth is the number of securities, BR = N, but can be expanded to include both refinements. In other words, they suggest that a more accurate representation of the basic fundamental law using the expression in Equation 17 is

$$\mathrm{E}(R_A) = \frac{\mathrm{IC}}{\sigma_{IC}}\sigma_A \tag{18}$$

The key impact of accounting for the uncertainty of skill is that actual information ratios are substantially lower than predicted by an objective application of the original form of the fundamental law. Specifically, security (i.e., individual stock) selection strategies can be analytically and empirically confirmed to be 45%–91% of original estimates using the fundamental law. Like the refinement for implementation issues associated with constraints as measured by the transfer coefficient, strategy risk reduces expected and average realized information ratios. The higher the uncertainty about forecasting ability, the smaller the likely expected value added.

9.2 Independence of Investment Decisions

As we have discussed, the number of individual assets, N, is not an adequate measure of strategy breadth, BR, when the active returns between individual assets are correlated, as defined by the risk model, and forecasts are not independent from period to period. Specifically, decisions to overweight all the stocks in a given industry or all the countries in a given region because they are responding to similar influences cannot be counted as completely independent decisions, so breadth in these contexts is lower than the number of assets. Similarly, when fundamental law concepts are applied to hedging strategies using derivatives or other forms of arbitrage, breadth can increase well beyond the number of securities.

For example, arbitrage of just two securities—say, a country equity market ETF traded on two different exchanges—can have extremely high breadth (i.e., the expected active return on the strategy is large compared with the active risk). To illustrate, Clarke, de Silva, and Thorley (2006) showed that a practical measure of breadth is

$$\mathrm{BR} = \frac{N}{1 + (N - 1)\rho} \tag{19}$$

where ρ is the same correlation coefficient in all the off-diagonal elements of the risk model. For just two securities, $N = 2$, and a correlation coefficient associated with near-arbitrage opportunities, $\rho = -0.8$, breadth could be BR = 2/[1 − (2 − 1)0.8] = 10.0 so that information ratios are quite high for even modest values of IC or forecasting skill.

Another example of the limitation of the fundamental law due to the lack of decision independence is the active management of fixed-income portfolios. Most descriptions of the fundamental law are based on individual stock selection strategies where the risk of equity securities is decomposed into systematic and idiosyncratic factors by a risk model. Once the systematic risk factors are removed, the active asset returns (defined as the idiosyncratic returns) are essentially independent, so breadth can be more easily determined. In contrast, almost all bonds represent some form of duration risk, as well as credit risk and optionality, so returns are highly correlated in more subtle ways. In addition, the implicit assumption of normality in the realized return distribution of bonds with default risk and embedded options is clearly unwarranted.

The limitation of independent decisions within the fundamental law also affects time-series implementation. In particular, increasing the rebalancing frequency may increase the realized information ratio but only to the extent that sequential active return forecasts are independent from period to period. Refinements on the concept of breadth—for example, Buckle (2004)—have improved the cross-sectional operationalization of the fundamental law, but more work is needed to provide conceptually useful modifications of the fundamental law in a multi-period, multi-asset setting.

In summary, the fundamental law is a useful conceptual framework in many active management applications and can even produce operational measurements of the essential elements of an active management strategy. But an understanding of the limitations of the law is warranted—particularly the issues of uncertainty in the level of assumed skill and the measurement of breadth in the face of time-dependent rebalancing policies and multi-period optimization.

EXAMPLE 10

Limitations of the Fundamental Law

Consider an active portfolio management strategy of selecting individual stocks in the S&P 500 on a monthly basis. The investor does a quick calculation of the fundamental law based on an information coefficient of IC = 0.05 and BR = 12 × 500 = 6,000, giving an astounding information ratio of IR = 3.87. In other words, at an active portfolio risk of 3.0%, the expected active return would be 3.87(3.0) = 11.6%.

Provide at least two different explanations of *why* the information ratio in this example could be too high based on practical limitations of the fundamental law.

Solution:

Potential answers include the following:

1. Cross-sectional dependence: The active returns on the 500 stocks in the S&P 500 are probably correlated, so the number of independent monthly decisions is lower than 500. For example, the investor could be forecasting outperformance of all the stocks in a given industrial sector and underperformance of all the stocks in another sector.
2. Time-series dependence: The decisions on any particular stock may be correlated from month to month. For example, the forecasting process might be based on the earnings yield (reported EPS over price), which changes slowly over time. A stock that is forecasted to outperform in one month is likely to retain the outperformance forecast for several months in a row.
3. Uncertainty: Although an information coefficient of 0.05 appears to be modest, the basic form of the fundamental law does not account for uncertainty in the information coefficient or the likelihood that the information coefficient changes over time and could be different for different sets of stocks.
4. Constraints: An answer that involves accounting for such constraints as long only or turnover limits using a transfer coefficient is a weaker answer because the impact of constraints and the transfer coefficient is a well-known refinement of the fundamental law, even though it does not appear to be used in this example.

SUMMARY

We have covered a number of key concepts and principles associated with active portfolio management. Active management is based on the mathematics and principles of risk and return from basic mean–variance portfolio theory but with a focus on value added compared with a benchmark portfolio. Critical concepts include the following:

- Value added is defined as the difference between the return on the managed portfolio and the return on a passive benchmark portfolio. This difference in returns might be positive or negative after the fact but would be expected to be positive before the fact or active management would not be justified.
- Value added is related to active weights in the portfolio, defined as differences between the various asset weights in the managed portfolio and their weights in the benchmark portfolio. Individual assets can be overweighted (have positive active weights) or underweighted (have negative active weights), but the complete set of active weights sums to zero.
- Positive value added is generated when positive-active-weight assets have larger returns than negative-active-weight assets. By defining individual asset active returns as the difference between the asset total return and the benchmark return, value added is shown to be positive if and only if end-of-period realized active asset returns are positively correlated with the active asset weights established at the beginning of the period.
- Value added can come from a variety of active portfolio management decisions, including security selection, asset class allocation, and even further decompositions into economic sector weightings and geographic or country weights.
- The Sharpe ratio measures reward per unit of risk in absolute returns, whereas the information ratio measures reward per unit of risk in benchmark relative returns. Either ratio can be applied *ex ante* to expected returns or *ex post* to realized returns. The information ratio is a key criterion on which to evaluate actively managed portfolios.
- Higher information ratio portfolios can be used to create higher Sharpe ratio portfolios. The optimal amount of active management that maximizes a portfolio's Sharpe ratio is positively related to the assumed forecasting accuracy or *ex ante* information coefficient of the active strategy.
- The active risk of an actively managed strategy can be adjusted to its desired level by combining it with a position in the benchmark. Furthermore, once an investor has identified the maximum Sharpe ratio portfolio, the total volatility of a portfolio can be adjusted to its desired level by combining it with cash (two-fund separation concept).
- The fundamental law of active portfolio management began as a conceptual framework for evaluating the potential value added of various investment strategies, but it has also emerged as an operational system for measuring the essential components of those active strategies.
- Although the fundamental law provides a framework for analyzing investment strategies, the essential inputs of forecasted asset returns and risks still require judgment in formulating the expected returns.
- The fundamental law separates the expected value added, or portfolio return relative to the benchmark return, into the basic elements of the strategy:
  - *skill* as measured by the information coefficient,
  - *structuring* of the portfolio as measured by the transfer coefficient,

- *breadth* of the strategy measured by the number of independent decisions per year, and
- *aggressiveness* measured by the benchmark tracking risk.

The last three of these four elements may be beyond the control of the investor if they are specified by investment policy or constrained by regulation.

- The fundamental law has been applied in settings that include the selection of country equity markets in a global equity fund and the timing of credit and duration exposures in a fixed-income fund.
- The fundamental law of active management has limitations, including uncertainty about the *ex ante* information coefficient and the conceptual definition of breadth as the number of independent decisions by the investor.

REFERENCES

Black, Fischer, and Robert Litterman. 1992. "Global Portfolio Optimization." *Financial Analysts Journal* 48 (5): 28–43.

Buckle, David. 2004. "How to Calculate Breadth: An Evolution of the Fundamental Law of Active Portfolio Management." *Journal of Asset Management* 4 (6): 393–405.

Clarke, Roger, Harindra de Silva, and Steven Thorley. 2002. "Portfolio Constraints and the Fundamental Law of Active Management." *Financial Analysts Journal* 58 (5): 48–66.

Clarke, Roger, Harindra de Silva, and Steven Thorley. 2005. "Performance Attribution and the Fundamental Law." *Financial Analysts Journal* 61 (5): 70–83.

Clarke, Roger, Harindra de Silva, and Steven Thorley. 2006. "The Fundamental Law of Active Portfolio Management." *Journal of Investment Management* 4 (3): 54–72.

Cremers, K.J. Martijn, and Antti Petajisto. 2009. "How Active Is Your Fund Manager?" *Review of Financial Studies* 22 (9): 3329–65.

Elton, Edward, and Martin Gruber. 1973. "Estimating the Dependence Structure of Share Prices." *Journal of Finance* 28 (5): 1203–32.

Fischer, Bernd, and Russell Wermers. 2013. *Performance Evaluation and Attribution of Security Portfolios*. Oxford, UK: Elsevier Inc.

Grinold, Richard C. 1989. "The Fundamental Law of Active Management." *Journal of Portfolio Management* 15 (3): 30–37.

Grinold, Richard C. 1994. "Alpha Is Volatility Times IC Times Score, or Real Alphas Don't Get Eaten." *Journal of Portfolio Management* 20 (4): 9–16.

Grinold, Richard C., and Ronald N. Kahn. 1999. *Active Portfolio Management: A Quantitative Approach for Providing Superior Returns and Controlling Risk*. 2nd ed. New York: McGraw-Hill.

Markowitz, Harry M. 1952. "Portfolio Selection." *Journal of Finance* 7 (1): 77–91.

Qian, Edward, and Ronald Hua. 2004. "Active Risk and Information Ratio." *Journal of Investment Management* 2 (3): 20–34.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance* 19 (3): 425–42.

Treynor, J., and Fischer Black. 1973. "How to Use Security Analysis to Improve Portfolio Selection." *Journal of Business* 46:66–86.

PRACTICE PROBLEMS

1 Wei Liu makes two statements about active portfolio management:

Statement 1 The "active return" of an actively managed portfolio is the difference between the portfolio's return and the return on the benchmark portfolio, and it is equal to the managed portfolio's alpha.

Statement 2 The active weights are the differences in the managed portfolio's weights and the benchmark's weights.

Are Liu's statements correct?

A Only Statement 1 is correct.

B Only Statement 2 is correct.

C Both statements are correct.

2 The benchmark weights and returns for each of the five stocks in the Capitol Index are given in the following table. The Tukol Fund uses the Capitol Index as its benchmark, and the fund's portfolio weights are also shown in the table.

Stock	Portfolio Weight (%)	Benchmark Weight (%)	20X2 Return (%)
1	30	24	14
2	30	20	15
3	20	20	12
4	10	18	8
5	10	18	10

What is the value added (active return) for the Tukol Fund?

A 0.00%

B 0.90%

C 1.92%

3 Consider the following asset class returns for calendar year 20X2:

Asset Class	Portfolio Weight (%)	Benchmark Weight (%)	Portfolio Return (%)	Benchmark Return (%)
Domestic equities	55	40	10	8
International equities	20	30	10	9
Bonds	25	30	5	6

What is the value added (or active return) for the managed portfolio?

A 0.25%

B 0.35%

C 1.05%

4 Gertrude Fischer mentions two properties of the Sharpe ratio and the information ratio that she says are very useful.

Property 1 The Sharpe ratio is unaffected by the addition of cash or leverage in a portfolio.

© 2016 CFA Institute. All rights reserved.

Property 2 The information ratio for an unconstrained portfolio is unaffected by the aggressiveness of the active weights.

Are Fischer's two properties correct?

A Yes.

B No. Only Property 1 is correct.

C No. Only Property 2 is correct.

The following information relates to Questions 5 and 6

	S&P 500	Indigo Fund
Expected annual return	9.0%	10.5%
Return standard deviation	18.0%	25.0%
Sharpe ratio	0.333	0.30
Active return		1.2%
Active risk		8.0%
Information ratio		0.15

Note: Data are based on a risk-free rate of 2.3%.

5 What is the maximum Sharpe ratio that a manager can achieve by combining the S&P 500 benchmark portfolio and the Indigo Fund?

A 0.333

B 0.365

C 0.448

6 Which of the following pairs of weights would be used to achieve the highest Sharpe ratio and optimal amount of active risk through combining the Indigo Fund and benchmark portfolio, respectively?

A 1.014 on Indigo and −0.014 on the benchmark

B 1.450 on Indigo and −0.450 on the benchmark

C 1.500 on Indigo and −0.500 on the benchmark

7 The benchmark portfolio is the S&P 500. Which of the following three portfolios can be combined with the benchmark portfolio to produce the highest combined Sharpe ratio?

	S&P 500	Portfolio A	Portfolio B	Portfolio C
Expected annual return	9.0%	10.0%	9.5%	9.0%
Return standard deviation	18.0%	20.0%	20.0%	18.0%
Sharpe ratio	0.333	0.350	0.325	0.333

	S&P 500	Portfolio A	Portfolio B	Portfolio C
Active return	0	1.0%	0.5%	0
Active risk	0	10.0%	3.0%	2.0%

Note: Data are based on a risk-free rate of 2.3%.

A Portfolio A

B Portfolio B

C Portfolio C

8 Based on the fundamental law of active management, if a portfolio manager has an information ratio of 0.75, an information coefficient of 0.1819, and a transfer coefficient of 1.0, how many securities are in the portfolio manager's fund, making the assumption that the active returns are uncorrelated.

A About 2

B About 4

C About 17

9 Two analysts make the following statements about the transfer coefficient in the expanded fundamental law of active management:

Analyst One says, "The transfer coefficient measures how well the realized returns correlate with the anticipated returns, adjusted for risk."

Analyst Two says, "The transfer coefficient measures how well the realized returns correlate with the active weights, adjusted for risk."

Which, if either, analyst is correct?

A Only Analyst One is correct.

B Only Analyst Two is correct.

C Neither analyst is correct.

10 The expanded fundamental law of active management is stated as follows:

$$\mathrm{E}(R_A) = (\mathrm{TC})(\mathrm{IC})\sqrt{\mathrm{BR}}\,\sigma_A$$

Which component on the righthand side represents the extent to which the portfolio manager's expectations are realized? The

A transfer coefficient, TC.

B information coefficient, IC.

C breadth, BR.

11 An analyst is given the following information about a portfolio and its benchmark. In particular, the analyst is concerned that the portfolio is a closet index fund. The T-bill return chosen to represent the risk-free rate is 0.50%.

	Benchmark	Portfolio
Return	8.75%	8.90%
Risk	17.50%	17.60%
Active return	0.00%	0.15%
Active risk	0.00%	0.79%
Sharpe ratio	0.4714	0.4773
Information ratio	N/A	0.1896

Which of the following three statements *does not* justify your belief that the portfolio is a closet index?

I. The Sharpe ratio of the portfolio is close to the Sharpe ratio of the benchmark.

II. The information ratio of the portfolio is relatively small.

III. The active risk of the portfolio is very low.

A Statement I

B Statement II

C Statement III

12 You are considering three managers for a small-cap growth mandate. After careful analysis, you produce the following forward-looking expectations about the managers' active risk and active return:

	Manager A	Manager B	Manager C
Active return	0.7%	0.6%	1.2%
Active risk	3.2%	3.1%	6.3%

If you intend to rely on the information ratio to make your decision, which manager should you choose?

A Manager A

B Manager B

C Manager C

13 You have a portfolio 100% allocated to a manager with an *ex post*, active risk at 8.0%. You choose to allocate a 75% position to the active manager and 25% to the benchmark to bring the portfolio back to your target active risk of 6.0%. If the manager's information ratio is 0.50, what happens to the information ratio of the portfolio after the reallocation?

A The information ratio increases because the lower active risk reduces the denominator of the ratio.

B The information ratio remains unchanged because allocations between the active portfolio and the benchmark don't affect the information ratio.

C The information ratio decreases because allocating some of the portfolio to the benchmark means that the external manager generates less active return.

The following information relates to Questions 14 and 15

You are analyzing three investment managers for a new mandate. The following table provides the managers' ex-ante active return expectations and portfolio weights. The last two columns include the risk and the *ex post*, realized active returns for the four stocks. Use the following data for the following two questions:

	Manager 1		Manager 2		Manager 3			Realized
	Δw	$E(R_A)$	Δw	$E(R_A)$	Δw	$E(R_A)$	Risk	R_A
Security 1	−0.125	0.03	0.2	0.04	−0.05	0.025	0.17	0.06
Security 2	0.025	0.04	0	0.01	0.05	0.015	0.10	0.07

	Manager 1		Manager 2		Manager 3			Realized
	Δw	$E(R_A)$	Δw	$E(R_A)$	Δw	$E(R_A)$	Risk	R_A
Security 3	0.075	0.05	−0.1	0	0.05	0.005	0.12	0.04
Security 4	0.025	0.06	−0.1	0.02	−0.05	0.015	0.25	0.02

14 Suppose all three managers claim to be good at forecasting returns. According to the expanded fundamental law of active management, which manager is the best at efficiently building portfolios by anticipating future returns?

A Manager 1

B Manager 2

C Manager 3

15 Suppose all three managers claim to be efficient in portfolio construction. According to the expanded fundamental law of active management, which manager is the best at building portfolios to make full use of their ability to correctly anticipate returns?

A Manager 1

B Manager 2

C Manager 3

---

16 Manager 1 has an information coefficient of 0.15, a transfer coefficient of 1.0, and invests in 50 securities. Manager 2 has a different strategy, investing in more securities; however, he is subject to investment constraints that reduce his transfer coefficient. Manager 2 has an information coefficient of 0.10, a transfer coefficient of 0.8, and invests in 100 securities. The investment selections of each manager are independent decisions. If both managers target an active risk of 5.0%, which manager will have the greater expected active return?

A Manager 1

B Manager 2

C Both managers will have the same active return.

17 Nick Young is concerned that Goudon Partners, one of his money managers, overestimates its expected active return because Goudon overstates its strategy breadth. Young makes two notes about his concern:

Note 1 Although Goudon claims that the number of independent asset decisions is high because it uses 200 stocks, many of these stocks cluster in industries where the same general analysis applies to several stocks.

Note 2 Goudon claims that each stock is independent and evaluated each month, or 12 times per year. These analyses are not independent because some of their strategies, such as favoring a particular industry or favoring value stocks, persist beyond one month. For example, a strategy of favoring low-P/E stocks will persist for several months and the investment decisions are not independent.

If his judgments are correct, are Young's notes about the overstatement of breadth correct?

A Only Note 1 is correct.

B Only Note 2 is correct.

C Both Notes 1 and 2 are correct.

18 Caramel Associates uses the fundamental law to estimate its expected active returns. Two things have changed. First, Caramel will lower its estimate of the information coefficient because they felt their prior estimates reflected overconfidence. Second, their major clients have relaxed several constraints on their portfolios—including social screens, prohibitions on short selling, and constraints on turnover. Which of these changes will increase the expected active return?

A Only the lower information coefficient.

B Only the relaxation of several portfolio constraints.

C Both the lower information coefficient and the relaxation of portfolio constraints.

The following information relates to Questions 19–25

James Frazee is chief investment officer at H&F Capital Investors. Frazee hires a third-party adviser to develop a custom benchmark for three actively managed balanced funds he oversees: Fund X, Fund Y, and Fund Z. (Balanced funds are funds invested in equities and bonds.) The benchmark needs to be composed of 60% global equities and 40% global bonds. The third-party adviser submits the proposed benchmark to Frazee, who rejects the benchmark based on the following concerns:

Concern 1: Many securities he wants to purchase are not included in the benchmark portfolio.

Concern 2: One position in the benchmark portfolio will be somewhat costly to replicate.

Concern 3: The benchmark portfolio is a float-adjusted, capitalization-weighted portfolio.

After the third-party adviser makes adjustments to the benchmark to alleviate Frazee's concerns, Frazee accepts the benchmark portfolio. He then asks his research staff to develop risk and expected return forecasts for Funds X, Y, and Z as well as for the benchmark. The forecasts are presented in Exhibit 1.

Exhibit 1 Forecasted Portfolio Statistics for Funds X, Y, and Z and the Benchmark

	Fund X	Fund Y	Fund Z	Benchmark
Portfolio weights:				
Global equities (%)	60.0	65.0	68.0	60.0
Global bonds (%)	40.0	35.0	32.0	40.0
Expected return (%)	10.0	11.6	13.2	9.4
Expected volatility (%)	17.1	18.7	22.2	16.3
Active risk (%)	5.2	9.2	15.1	N/A
Sharpe ratio (SR)	0.45	0.50	0.49	0.44

Note: Data are based on a risk-free rate of 2.3%.

Frazee decides to add a fourth offering to his group of funds, Fund W, which will use the same benchmark as in Exhibit 1. Frazee estimates Fund W's information ratio to be 0.35. He is considering adding the following constraint to his portfolio construction model: Fund W would now have maximum over- and underweight constraints of 7% on single-country positions.

Frazee conducts a search to hire a manager for the global equity portion of Fund W and identifies three candidates. He asks the candidates to prepare risk and return forecasts relative to Fund W's benchmark based on their investment strategy, with the only constraint being no short selling. Each candidate develops independent annual forecasts with active return projections that are uncorrelated and constructs a portfolio made up of stocks that are diverse both geographically and across economic sectors. Selected data for the three candidates' portfolios are presented in Exhibit 2.

Exhibit 2 Forecasted Portfolio Data for Equity Portion of Fund W

	Candidate A	Candidate B	Candidate C
Rebalancing	Annually	Annually	Annually
Number of securities	100	64	36
Information ratio (IR)	0.582	0.746	0.723
Transfer coefficient (TC)	0.832	0.777	0.548
Information coefficient*	0.07	0.12	0.22

* Information coefficient based on previously managed funds.

Frazee asks Candidate C to re-evaluate portfolio data given the following changes:

Change 1: Fix the number of securities to 50.

Change 2: Rebalance on a semi-annual basis.

Change 3: Add maximum over- or underweight constraints on sector weightings.

19 Which of Frazee's concerns *best* justifies his decision to reject the proposed benchmark?

- A Concern 1
- B Concern 2
- C Concern 3

20 Based on Exhibit 1, the expected active return from asset allocation for Fund X is:

- A negative.
- B zero.
- C positive.

21 Based on Exhibit 1, which fund is expected to produce the greatest consistency of active return?

- A Fund X
- B Fund Y
- C Fund Z

22 Based on Exhibit 1, combining Fund W with a fund that replicates the benchmark would produce a Sharpe ratio *closest* to:

A 0.44.
B 0.56.
C 0.89.

23 If Frazee added the assumption he is considering in Fund W's portfolio construction, it would *most likely* result in:
A a decrease in the optimal aggressiveness of the active strategy.
B the information ratio becoming invariant to the level of active risk.
C an increase in the transfer of active return forecasts into active weights.

24 Based on the data presented in Exhibit 2, the candidate with the greatest skill at achieving active returns appears to be:
A Candidate A.
B Candidate B.
C Candidate C.

25 Which proposed change to Fund W would *most likely* decrease Candidate C's information ratio?
A Change 1
B Change 2
C Change 3

The following information relates to Questions 26–29

John Martinez is assessing the performance of the actively managed diversified asset portfolio. The diversified asset portfolio is invested in equities, bonds, and real estate, and allocations to these asset classes and to the holdings within them are unconstrained.

Selected return and financial data for the portfolio for 2019 are presented in Exhibit 1.

Exhibit 1 Diversified Asset Portfolio 2019 Portfolio Performance

	Sub-Portfolio Return (%)	Benchmark Return (%)	Portfolio Allocation (%)	Strategic Asset Allocation (%)
Equities sub-portfolio	36.9	31.6	63	60
Bond sub-portfolio	−2.4	−2.6	28	35
Real estate sub-portfolio	33.4	28.3	9	5

Martinez uses several risk-adjusted return metrics to assess the performance of the diversified asset portfolio, including the information ratio and the Sharpe ratio. Selected risk, return, and statistical data for the portfolio are presented in Exhibit 2.

Exhibit 2 Diversified Asset Portfolio Data, 2000-2019

	Transfer Coefficient (TC)	Information Coefficient (IC)	Breadth (BR)
Equities sub-portfolio	0.90	0.091	21
Bond sub-portfolio	0.79	0.087	23
Real estate sub-portfolio	0.86	0.093	19

Martinez has recently hired Kenneth Singh to help him evaluate portfolios. Martinez asks Singh about the possible effects on the portfolio's information ratio if cash were added to the diversified asset portfolio or if the aggressiveness of the portfolio's active weights were increased. Singh responds with two statements:

Statement 1 Adding cash to the portfolio would change the portfolio's information ratio.

Statement 2 Increasing the aggressiveness of active weights would not change the portfolio's information ratio.

26 Based on Exhibit 1, the value added to the diversified asset portfolio attributable to the security selection decision in 2019 was *closest* to:

A 2.3%.

B 3.9%.

C 6.1%.

27 Based on Exhibit 1, the value added of the diversified asset portfolio attributable to the asset allocation decision in 2019 was *closest* to:

A 2.3%.

B 3.9%.

C 6.1%.

28 Based on data in Exhibit 2 and using the information ratio as the criterion for evaluating performance, which sub-portfolio had the best performance in the period 2000–2019?

A The bond sub-portfolio.

B The equities sub-portfolio.

C The real estate sub-portfolio.

29 Which of Singh's statements regarding the information ratio is correct?

A Only Statement 1

B Only Statement 2

C Both Statement 1 and Statement 2

SOLUTIONS

1 B is correct. Although the first part of Statement 1 is correct (active return, or value added, equals the difference between the managed portfolio return and the benchmark return), active return is not the same as alpha. In other words, $R_A = R_P - R_B$, while $\alpha_P = R_P - \beta_P \times R_B$. Statement 2 correctly defines active weights.

2 B is correct. The portfolio active return is equal to the portfolio return minus the benchmark return:

$$R_A = R_P - R_B.$$

The portfolio return is $R_P = \sum_{i=1}^{n} w_{P,i} R_i$;

$$R_P = 0.30(14\%) + 0.30(15\%) + 0.20(12\%) + 0.10(8\%) + 0.10(10\%) = 12.9\%.$$

The benchmark return is $R_B = \sum_{i=1}^{n} w_{B,i} R_i$;

$$R_B = 0.24(14\%) + 0.20(15\%) + 0.20(12\%) + 0.18(8\%) + 0.18(10\%) = 12.0\%.$$

Thus, the active return is

$$R_A = R_P - R_B = 12.9\% - 12.0\% = 0.9\%.$$

Note that this same correct answer can be obtained in two other equivalent ways. The active weights are the differences between the portfolio and benchmark weights, or $\Delta w_i = w_{P,i} - w_{B,i}$. Computing the active weights from the table provided, the active return is

$$\begin{aligned} R_A &= \sum_{i=1}^{N} \Delta w_i R_i \\ &= 0.06(14\%) + 0.10(15\%) + 0(12\%) - 0.08(8\%) - 0.08(10\%) \\ &= 0.9\%. \end{aligned}$$

Finally, we could express the active security returns as their differences from the benchmark return, or $R_{Ai} = R_i - R_B$. Computing the active security returns from the table provided, the portfolio active return is the sum product of the active weights and the active security returns:

$$\begin{aligned} R_A &= \sum_{i=1}^{N} \Delta w_i R_{Ai} \\ &= 0.06(2\%) + 0.10(3\%) + 0(0\%) - 0.08(-4\%) - 0.08(-2\%) \\ &= 0.9\%. \end{aligned}$$

3 C is correct. The active return is equal to the portfolio return minus the benchmark return:

$$R_A = R_P - R_B = \sum_{j=1}^{M} w_{P,j} R_{P,j} - \sum_{j=1}^{M} w_{B,j} R_{B,j}$$

The portfolio return is $R_P = \sum_{i=1}^{n} w_{P,i} R_i = 0.55(10\%) + 0.20(10\%) + 0.25(5\%) =$ 8.75%.

The benchmark return is $R_B = \sum_{i=1}^{n} w_{B,i} R_i$ = 0.40(8%) + 0.30(9%) + 0.30(6%) = 7.70%.

Thus, $R_A = R_P - R_B$ = 8.75% − 7.70% = 1.05%.

4 A is correct. Both properties are correct. For Property 1, if w_P is the weight of an actively managed portfolio and $(1 - w_P)$ is the weight on risk-free cash, changing w_P does not change the Sharpe ratio, as can be seen in this equation:

$$\text{SR}_C = \frac{R_C - R_F}{\sigma_C} = \frac{w_P\left(R_P - R_F\right)}{w_P \sigma_P} = \text{SR}_P$$

For Property 2, the information ratio of an unconstrained portfolio is unaffected by multiplying the active security weights, Δw_i, by a constant.

5 B is correct. The highest squared Sharpe ratio of an actively managed portfolio is

$$\text{SR}_P^2 = \text{SR}_B^2 + \text{IR}^2 = 0.333^2 + 0.15^2 = 0.1334$$

The highest Sharpe ratio is $\text{SR}_P = \sqrt{0.1334} = 0.365$.

6 A is correct. The optimal amount of active risk is

$$\sigma_A = \frac{\text{IR}}{\text{SR}_B}\sigma_B = \frac{0.15}{0.333}18.0\% = 8.11\%$$

The weight on the active portfolio (Indigo) would be 8.11%/8.0% = 1.014, and the weight on the benchmark portfolio would be 1 − 1.014 = −0.014.

We can demonstrate that these weights achieve the maximum Sharpe ratio (of 0.365). Note that 8.11% is the optimal level of active risk and that Indigo has an expected active return of 1.014(1.2%) = 1.217% over the benchmark and a total excess return of 6.0% + 1.217% = 7.217%. The portfolio total risk is

$$\sigma_P{}^2 = \sigma_B{}^2 + \sigma_A{}^2 = 18.0^2 + 8.111^2 = 389.788.$$

Taking the square root, σ_P = 19.743, and the optimal Sharpe ratio is indeed 7.217/19.743 = 0.365.

7 B is correct. The optimal active portfolio is the portfolio with the highest information ratio, the ratio of active return to active risk. The IRs for the three active portfolios are as follows:

IR_A = 1.0/10.0 = 0.10

IR_B = 0. 5/3.0 = 0.167

IR_C = 0/2.0 = 0.00

Portfolio B has the highest IR and is the best active portfolio; it is therefore the best portfolio to combine with the benchmark.

8 C is correct. Using the equation IR* = IC × $\sqrt{\text{BR}}$ and assuming that breadth can be interpreted as number of securities in the portfolio, solving for breadth in the equation yields $\left(\frac{0.75}{0.1819}\right)^2$ = 17.000.

9 C is correct. The transfer coefficient measures how well the anticipated (*ex ante*), risk-adjusted returns correlate with the risk-adjusted active weights. This is also expressed in the equation for the transfer coefficient: TC = $\rho(\mu_i/\sigma_i, \Delta w_i \sigma_i)$.

10 B is correct. The IC measures an investment manager's ability to forecast returns.

11 B is correct. A closet index will have a very low active risk and will also have a Sharpe ratio very close to the benchmark. Therefore, Statements I and III are consistent with a closet index portfolio. A closet index's information ratio can be indeterminate (because the active risk is so low) and is often negative due to management fees.

12 A is correct. Manager A has the highest information ratio. The information ratio is defined as $IR = \dfrac{\text{active return}}{\text{active risk}}$. The managers in this example have the following information ratios:

	Manager A	Manager B	Manager C
Information ratio	0.7/3.2 = 0.219	0.6/3.1 = 0.194	1.2/6.3 = 0.190

13 B is correct. The information ratio is unaffected by rebalancing the active portfolio and the benchmark portfolio. In this case, the active return and active risk are both reduced by 25% and the information ratio will be unchanged.

14 C is correct. The proper statistic to calculate is the information coefficient, and it is defined as follows:

$$IC = \rho\left(\frac{R_{Ai}}{\sigma_i}, \frac{\mu_i}{\sigma_i}\right)$$

A manager is a good forecaster if his or her *ex ante*, active return expectations (forecasts) are highly correlated with the realized active returns. The information coefficient requires that these forecasts and realized returns be risk-weighted. When this is done for the three managers, the risk-weighted forecasts and realized returns are:

	Risk-Weighted Forecasts, μ_i/σ_i			R_{Ai}/σ_i
	Manager 1	**Manager 2**	**Manager 3**	**Realized**
Security 1	0.176	0.235	0.147	0.353
Security 2	0.400	0.100	0.150	0.700
Security 3	0.417	0.000	0.042	0.333
Security 4	0.240	0.080	0.060	0.080

The ICs are found by calculating the correlations between each manager's forecasts and the realized risk-weighted returns. The three managers have the following ICs:

	Manager 1	Manager 2	Manager 3
Information coefficient	0.5335	0.0966	0.6769

Manager 3 has the highest IC.

15 B is correct. The proper statistic to calculate is the transfer coefficient, and it is defined as follows:

$$TC = \rho(\mu_i/\sigma_i, \Delta w_i \sigma_i)$$

The TC is the cross-sectional correlation between the forecasted active security returns and the actual active weights, adjusted for risk.

	Risk-Weighted Forecasts, μ_i/σ_i			Risk-Adjusted Weights, $\Delta w_i\sigma_i$		
	Manager 1	Manager 2	Manager 3	Manager 1	Manager 2	Manager 3
Security 1	0.1765	0.2353	0.1471	–0.0213	0.0340	–0.0085
Security 2	0.4000	0.1000	0.1500	0.0025	0.0000	0.0050
Security 3	0.4167	0.0000	0.0417	0.0090	–0.0120	0.0060
Security 4	0.2400	0.0800	0.0600	0.0063	–0.0250	–0.0125

The three managers have the following TCs:

	Manager 1	Manager 2	Manager 3
Transfer coefficient	0.7267	0.8504	–0.0020

Manager 2 has the highest TC.

16 A is correct. Manager 1's IR = TC × IC × $\sqrt{\text{BR}}$ = 1.0 × 0.15 × $\sqrt{50}$ = 1.06. Manager 2's IR = 0.8 × 0.10 × $\sqrt{100}$ = 0.80. Manager 1's active return is 1.06(5.0) = 5.3%, and Manager 2's expected active return is 0.80(5.0) = 4.0%. Manager 1 has the greater expected active return.

17 C is correct. If the decisions about each of the 200 stocks are not independent, and if the decisions about a stock from one month to the next are not independent, then Goudon Partners is overstating its estimates of its breadth and its expected active returns.

18 B is correct. Although the relaxation of portfolio constraints will increase the transfer coefficient (and expected active returns), the lower information coefficient reduces the information ratio and the expected active return.

19 A is correct. Because the benchmark does not contain many assets that Frazee wants to invest in, the benchmark may not be representative of his investment approach. Concern 2, as stated, is less important because it does not imply that the cost of replicating the benchmark is a serious concern. Finally, Concern 3 actually states a generally positive feature of the benchmark.

20 B is correct. Active return from asset allocation is derived from differences between the benchmark weight and the portfolio weight across asset classes. For Fund X, the expected active return from asset allocation is calculated as

$$\sum_{j=1}^{M} \Delta w_j R_{B,j} = (60-60)R_{B,e} + (40-40)R_{B,b} = 0$$

where Δw_j is the difference in the active portfolio and the benchmark asset weights, $R_{B,e}$ is the benchmark's return from global equities, and $R_{B,b}$ is the benchmark's return from global bonds.

Because Fund X has the same asset weights as the benchmark across the two asset classes (60% global equities, 40% global bonds), the expected active return from asset allocation is zero.

21 C is correct. The IR measures the consistency of active return. The IR is calculated for the three funds as follows:

$$\text{IR} = \frac{R_P - R_B}{\sigma(P_P - R_B)} = \frac{R_A}{\sigma_A}$$

IR for Fund X = (10.0 – 9.4)/5.2 = 0.6/5.2 = 0.12.

IR for Fund Y = (11.6 – 9.4)/9.2 = 2.2/9.2 = 0.24.

IR for Fund Z = (13.2 − 9.4)/15.1 = 3.8/15.1 = 0.25.

Fund Z has the largest IR and thus is expected to produce the greatest consistency of active return.

22 B is correct. Given the IR for Fund W of 0.35 and the benchmark's SR of 0.44, the combination of the benchmark portfolio and Fund W would produce an SR of 0.56, calculated as follows:

$$SR_P^2 = SR_B^2 + IR^2$$

$$SR_P = (0.44^2 + 0.35^2)^{0.5} = 0.56.$$

23 A is correct. The new assumption adds constraints to Fund W. The IR for a constrained portfolio generally decreases with the aggressiveness of the strategy because portfolio constraints reduce the transfer of active return forecasts into active weights. Furthermore, the optimal active risk is given by the following formula:

$$\sigma_A = TC\frac{IR}{SR_B}\sigma_B$$

The addition of portfolio constraints reduces the TC, thus also reducing the optimal active risk.

So, having maximum over- and underweight constraints on single-country positions decreases the optimal aggressiveness of the active management strategy.

24 B is correct. The IR measures the consistency of active return generation. A higher ratio generally indicates better managerial skill at achieving active returns on a risk-adjusted basis. The IR for Candidate B (0.746) is higher than the IR for Candidate A (0.582) and Candidate C (0.723).

Thus, Candidate B appears to have the greatest skill, as indicated by the highest IR of 0.746.

25 C is correct. The IR is calculated as $IR = (TC)(IC)\sqrt{BR}$, where BR is breadth. Change 3, establishing new constraints of caps on the over- and underweight of sectors, reduces the correlation of optimal active weights with the actual active weights, which results in a decreased TC and thus a decrease in the IR. Change 1 (increasing portfolio size from 36 to 50) and Change 2 (increasing the frequency of rebalancing from annually to semi-annually) would both likely have the effect of increasing the BR of the portfolio, which would increase the IR.

26 B is correct. Based on the differences in returns for the portfolio and benchmark in Exhibit 1, the value added by each asset class within the portfolio is shown in the following table:

	Sub-Portfolio Return (%)	Benchmark Return (%)	Value Added (%)	Portfolio Allocation (%)
Equities sub-portfolio	36.9	31.6	5.3	63
Bond sub-portfolio	−2.4	−2.6	0.2	28
Real estate sub-portfolio	33.4	28.3	5.1	9

The value added from security selection is calculated as the sum of the actual portfolio weights multiplied by each sub-portfolio's value added measure. Thus, the value added from security selection is calculated as: Value added from security selection = 0.63(5.3%) + 0.28(0.2%) + 0.09(5.1%) = 3.9%.

A is incorrect. It represents the value added from asset allocation (2.3%).

C is incorrect. It represents the total value added (2.3% + 3.9% = 6.1%, with rounding).

27 A is correct. The value added from asset allocation is calculated as the sum of the differences in the weights between the strategic (benchmark) allocation and the actual sub-portfolio allocation multiplied by each sub-portfolio's benchmark return.

	Benchmark Return (%)	Actual Asset Allocation (%)	Strategic Asset Allocation (%)	Actual – Strategic Asset Allocation (%)
Equities sub-portfolio	31.6	63	60	+3
Bond sub-portfolio	−2.6	28	35	−7
Real estate sub-portfolio	28.3	9	5	+4

Thus, the value added by the active asset allocation decision is calculated as

Value added from asset allocation decision = 0.03(31.6%) − 0.07(−2.6%) + 0.04(28.3%) = 2.3%.

B is incorrect. It is the value added from security selection.

C is incorrect. It is the total value added.

28 B is correct. The information ratio for a portfolio can be expressed as follows:

$$IR = (TC)(IC)\sqrt{BR}$$

The information ratios for the three sub-portfolios are calculated as follows:

	Information Ratio
Equities sub-portfolio	$0.90 \times 0.091 \times (21)^{0.5} = 0.38$
Bond sub-portfolio	$0.79 \times 0.087 \times (23)^{0.5} = 0.33$
Real estate sub-portfolio	$0.86 \times 0.093 \times (19)^{0.5} = 0.35$

Based on the information ratio, the equities sub-portfolio outperformed the real estate sub-portfolio. The information ratio for the equities sub-portfolio of 0.38 was higher than the information ratio for the real estate sub-portfolio of 0.35 and the bond sub-portfolio of 0.33.

29 C is correct. The information ratio for a portfolio of risky assets will generally shrink if cash is added to the portfolio. Because the diversified asset portfolio is an unconstrained portfolio, its information ratio would be unaffected by an increase in the aggressiveness of active weights.

READING

44

Trading Costs and Electronic Markets

by Larry Harris, PhD, CFA

Larry Harris, PhD, CFA, is at the USC Marshall School of Business (USA).

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** explain the components of execution costs, including explicit and implicit costs;
☐	**b.** calculate and interpret effective spreads and VWAP transaction cost estimates;
☐	**c.** describe the implementation shortfall approach to transaction cost measurement;
☐	**d.** describe factors driving the development of electronic trading systems;
☐	**e.** describe market fragmentation;
☐	**f.** identify and contrast the types of electronic traders;
☐	**g.** describe characteristics and uses of electronic trading systems;
☐	**h.** describe comparative advantages of low-latency traders;
☐	**i.** describe the risks associated with electronic trading and how regulators mitigate them;
☐	**j.** describe abusive trading practices that real-time surveillance of markets may detect.

COSTS OF TRADING

1

a explain the components of execution costs, including explicit and implicit costs;

This reading draws from *Trading and Electronic Markets: What Investment Professionals Need to Know,* by Larry Harris, PhD, CFA, Research Foundation of CFA Institute. © 2015 CFA Institute. All rights reserved.

© 2019 CFA Institute. All rights reserved.

Securities research, portfolio management, and securities trading support the investment process. Of the three, trading is often the least understood and least appreciated function. Among the questions addressed in this reading are the following:

- What are explicit and implicit trading costs, and how are they measured?
- How is a limit order book interpreted?
- How have trading strategies adapted to market fragmentation?
- What types of electronic traders can be distinguished?

This reading is organized as follows: Sections 1–2 discuss the direct and indirect costs of trading.[1] Section 3 discusses developments in electronic trading and the effects they had on transaction costs and market fragmentation. Section 4 identifies the most important types of electronic traders. Section 5 describes electronic trading facilities and some important ways traders use them. Sections 6–7 discuss risks posed by electronic trading and how regulators control them. The final section summarizes the reading.

1.1 Costs of Trading

Understanding the costs of trading is critical for ensuring optimal execution and transaction cost management for portfolios. Because trading costs are a significant source of investment performance slippage, investment sponsors and their investment managers pay close attention to trading processes.

The costs of trading include fixed costs and variable costs. For buy-side institutions, fixed trading costs include the costs of employing buy-side traders, the costs of equipping them with proper trading tools (electronic systems and data), and the costs of office space (trading rooms or corners). Small buy-side institutions often avoid these costs by not employing buy-side traders. Their portfolio managers submit their orders directly to their brokers. Variable transaction costs arise from trading activity and consist of explicit and implicit costs.

Explicit costs are the direct costs of trading, such as broker commission costs, transaction taxes, stamp duties, and fees paid to exchanges. They are costs for which a trader could receive a receipt.

Implicit costs, by contrast, are indirect costs caused by the market impact of trading. Buyers often must raise prices to encourage sellers to trade with them, and sellers often must lower prices to encourage buyers. The price concessions that impatient traders make to complete their trades are called the market impacts of their trades. For small orders, market impact often is limited to buying at bid prices and selling at lower ask prices. Small market orders generally have small market impact because these orders often are immediately filled by traders willing to trade at quoted bid and offer prices, or even better prices. Larger orders have greater market impact when traders must move the market to fill their orders. In these cases, traders must accept larger price concessions (less attractive prices) to execute their orders in entirety. Although no receipt can be given for implicit costs, they are real nonetheless.

Implicit costs result from the following issues:

- The **bid–ask spread** is the ask price (the price at which a trader will sell a specified quantity of a security) minus the bid price (the price at which a trader will buy a specified quantity of a security). Traders who want to trade quickly buy at higher prices and sell at lower prices than those willing to wait for others to trade with them.

1 CFA Institute would like to thank Ananth Madhavan, PhD, at BlackRock (USA) for his contribution to this section, which includes material first written by him.

- **Market impact** (or price impact) is the effect of the trade on transaction prices. Traders who want to fill large orders often must move prices to encourage others to trade with them.
- **Delay costs** (also called slippage) arise from the inability to complete the desired trade immediately. Traders fail to profit when they fill their orders after prices move as they expect.
- **Opportunity costs** (or unrealized profit/loss) arise from the failure to execute a trade promptly. Traders fail to profit when their orders fail to trade and price move as they expect.

1.1.1 *Dealer Quotes*

Dealers provide liquidity to other traders when they allow traders to buy and sell when those traders want to trade. Those traders may be the clients known to the dealers, or they may be unknown traders whose orders exchanges assign to standing dealer orders and quotes.

Unlike brokers, dealers trade for their accounts when filling their customers' orders. When dealers buy or sell, they increase or reduce their inventories. Dealers profit by selling at ask prices that are higher than the bid prices at which they buy. If buying interest is greater than selling interest, dealers raise their ask prices to discourage buyers and raise their bid prices to encourage sellers. Likewise, if selling interest is greater than buying interest, dealers lower their ask prices to encourage buyers and lower their bid prices to discourage sellers.

Dealers help markets function well by being continuously available to take the other side of a trade when other traders want to trade. Dealers thus make markets more continuous. They are especially important in markets for infrequently traded securities in which buyers and sellers rarely are present at the same time. For example, most bond markets are overwhelmingly dealer markets because most bonds rarely trade. If an investor wants to sell a rarely traded bond, the investor might have a long wait before another investor interested in buying that bond arrives. Instead, a dealer generally will buy the bond and then try to market it to potential buyers. Practitioners say that dealers "make market" when they offer to trade.

1.1.2 *Bid–Ask Spreads and Order Books*

The prices at which dealers will buy or sell specified quantities of a security are, respectively, their **bid prices** and **ask prices**. (Ask prices are also known as offer prices.) The excess of the ask price over the bid price is the dealer's **bid–ask spread**.

When several dealers offer bid prices, the **best bid** is the offer to buy with the highest bid price. The best bid is also known as the **inside bid**. The **best ask**, also known as the **best offer** or **inside ask**, is the offer to sell with the lowest ask price.

The spread between the best bid price and the best ask price in a market is the market bid–ask spread, which is also known as the **inside spread**. It will be smaller (tighter or narrower) than the individual dealer spreads if the dealer with the highest bid price is not also the dealer with the lowest ask price.

For example, suppose that a portfolio manager gives the firm's trading desk an order to buy 1,000 shares of Economical Chemical Systems, Inc. (ECSI). Three dealers (coded A, B, and C) make a market in those shares. When the trader views the market in ECSI at 10:22 a.m. on his computer screen, the three dealers have put in the following limit orders to trade at an exchange market:

- Dealer A: *bid*: 98.85 for 600 shares; *ask*: 100.51 for 1,000 shares
- Dealer B: *bid*: 98.84 for 500 shares; *ask*: 100.55 for 500 shares
- Dealer C: *bid*: 98.82 for 700 shares; *ask*: 100.49 for 200 shares

The bid–ask spreads of Dealers A, B, and C are, respectively,

- 100.51 – 98.85 = 1.66
- 100.55 – 98.84 = 1.71
- 100.49 – 98.82 = 1.67

The best bid price, 98.85 by Dealer A, is lower than the best ask price, 100.49 by Dealer C. The market spread is thus 100.49 – 98.85 = 1.64, which is lower than any of the dealers' spreads.

The trader might see the quote information organized on his screen as shown in Exhibit 1. In this display, called a **limit order book**, the bids and asks are separately ordered from best to worst with the best at the top. The trader also notes that the **midquote price** (halfway between the market bid and ask prices) is (100.49 + 98.85)/2 = 99.67.

Exhibit 1 The Limit Order Book for Economical Chemical Systems, Inc.

Bids				Asks			
Dealer	Time Entered	Price	Size	Dealer	Time Entered	Price	Size
A	10:21 a.m.	98.85	600	C	10:21 a.m.	100.49	200
B	10:21 a.m.	98.84	500	A	10:21 a.m.	100.51	1,000
C	10:19 a.m.	98.82	700	B	10:19 a.m.	100.55	500

Note: The bids are ordered from highest to lowest, while the asks are ordered from lowest to highest. These orderings are from best bid or ask to worst bid or ask.

If the trader on the firm's trading desk submits a market buy order for 1,000 shares, the trader would purchase 200 shares from Dealer C at 100.49 per share and 800 shares from Dealer A at 100.51 per share.

Note that filling the second part of the order cost the trader 0.02 per share more than the first part because Dealer C's ask size was insufficient to fill the entire order. Large orders have price impact when they move down the book as they fill. The price impact of an order depends on its size and the available liquidity.

If this market were not an exchange market, the trader might choose to direct the buy order to a specific dealer—for example, to Dealer A. The trader may do so for many reasons. The trader may believe that Dealer A more likely will honor her quote than would Dealer C. Alternatively, the trader may believe that Dealer A more likely will settle the trade than Dealer C. Such considerations are especially important in markets for which no clearinghouse guarantees that all trades will settle—for example, most currency markets. Institutions active in such markets may screen counterparties on credit criteria. Finally, the trader might fear that Dealer A will cancel her quote when she (or a computer managing her quote) sees that a trade took place at 100.49. Sending the order first to Dealer A thus could produce a better average price.

1.1.3 *Implicit Transaction Cost Estimates*

Investment managers and traders measure transaction costs so that they can better predict the cost of filling orders and so that they can better manage the brokers and dealers who fill their orders. Buyers, of course, want to trade at low prices, while sellers want to trade at high prices. Expensive trades are purchases arranged at high prices or sales arranged at low prices.

To estimate transaction costs, analysts compare trade prices to a benchmark price. Commonly used price benchmarks include the midquote price at the time of the trade, the midquote price at the time of the order submission, and a volume-weighted average price around the time of the trade. These three benchmarks, respectively, correspond to the effective spread, implementation shortfall, and VWAP methods of transaction cost estimation.

2 EFFECTIVE SPREADS AND VOLUME-WEIGHTED COST ESTIMATES

b calculate and interpret effective spreads and VWAP transaction cost estimates;

c describe the implementation shortfall approach to transaction cost measurement;

The market spread is a measure of trade execution costs. It is how much traders would lose per quantity traded if they simultaneously submitted buy and sell market orders that respectively execute at the ask and bid prices. The loss is the cost of trading, because this strategy otherwise accomplishes nothing. Given that two trades generated the cost, the cost per trade is one half of the quoted spread.

The prices that traders receive when trading often differ from quoted prices. Smaller orders sometimes fill at better prices; larger orders often fill at worse prices. Standing orders offering liquidity fill at same-side prices (buy at bid, sell at ask), if they fill at all.

The effective spread provides a more general estimate of the cost of trading. It uses the midquote price (the average, or midpoint, of the bid and the ask prices at the time the order was entered) as the benchmark price:

$$\text{Effective spread transaction cost estimate} =$$

$$\text{Trade size} \times \begin{cases} \text{Trade price} - \left(\dfrac{\text{Bid} + \text{Ask}}{2}\right) & \text{for buy orders} \\ \left(\dfrac{\text{Bid} + \text{Ask}}{2}\right) - \text{Trade price} & \text{for sell orders} \end{cases}$$

For a buy order filled at the ask, the estimated implicit cost of trading is half the bid–ask spread, because Ask − [(Bid + Ask)/2] = [(Ask − Bid)/2]. Multiplying this midquote price benchmark transaction cost estimate by 2 produces a statistic called the **effective spread**. It is the spread that traders would have observed if the quoted ask (for a purchase) or the bid (for a sale) were equal to the trade price.

The effective spread is a sensible estimate of transaction costs when orders are filled in single trades. If an order fills at a price better than the quoted price (e.g., a buy order fills at a price below the ask price), the order is said to receive **price improvement** and the spread is effectively lower. Price improvement occurs when trade execution prices are better than quoted prices. An order that fills at a price outside the quoted spread has an effective spread that is larger than the quoted spread. Such results occur when trade execution prices are worse than quoted prices.

The effective spread is a poor estimate of transaction costs when traders split large orders into many parts to fill over time. Such orders often move the market and cause bid and ask prices to rise or fall. The impact of the order on market prices, called **market impact**, makes trading expensive—especially for the last parts to fill—but the effective spread will not fully identify this cost if it is computed separately for each trade.

For example, suppose that a buy order for 10,000 shares fills in two trades. The prices and sizes of these trades and the best bids and offers in the market when the trades occurred appear in the following table:

Trade	Trade Price	Trade Size	Prevailing Bid	Prevailing Offer
#1	10.21	4,000	10.19	10.21
#2	10.22	6,000	10.20	10.22

For this buy order, the effective spread transaction cost per share is 0.01, or [(10.21– 10.19)/2] and [(10.22 – 10.20)/2], for both trades (the effective spreads are both 0.02). Thus, the total transaction cost estimate measured using the midquote price benchmark is 100 = 0.01 × 10,000. This estimate is problematic because it reflects the higher price of the second trade, which was likely caused by the market impact of the trader's first trade.

Effective spreads also do not measure **delay costs** (also called slippage) that arise from the inability to complete the desired trade immediately because of its size in relation to the available market liquidity. Delay costs also arise when portfolio managers or their traders fail to create and route orders quickly to the markets where they will fill most quickly. Analysts often measure delay costs on the portion of the order carried over from one day to the next. Delay is costly when price moves away from an order (up for a buy order, down for a sell order), often because information leaks into the market before or during the execution of the order.

When delays in execution cause a portion of the order to go unfilled, the associated cost is called **opportunity cost**. For example, suppose a futures trader places an order to buy 10 contracts with a limit price of 99.00, good for one day, when the market quote is 99.01 to 99.04. The order does not execute, and the contract closes at 99.80. If the order could have been filled at 99.04, the difference (99.80 – 99.04 = 0.76) reflects the opportunity cost per contract. By trading more aggressively, the trader might have avoided these costs. Opportunity costs are difficult to measure. In the example, the one-day time frame is arbitrary, and the assumption that the order could fill at 99.04 may be suspect. The estimate usually is sensitive to the time frame chosen for measurement and to assumptions about the prices at which orders could trade.

2.1 Implementation Shortfall

The implementation shortfall method of measuring trading costs addresses the problems associated with the effective spread method. Implementation shortfall is also attractive because it views trading from an investment management perspective and measures the total cost of implementing an investment decision by capturing all explicit and implicit costs. The implementation shortfall method includes the market impact costs and delay costs as well as opportunity costs, which are often significant for large orders.

Implementation shortfall compares the values of the actual portfolio with that of a paper portfolio constructed on the assumption that trades could be arranged at the prices that prevailed when the decision to trade is made. The prevailing price—also called the decision price, the arrival price, or the strike price—is generally taken to be the midquote price at the time of the trade decision. The excess of the paper value over the actual value is the **implementation shortfall**. The coverage of implementation shortfall is continued at Level III.

2.2 VWAP Transaction Cost Estimates

Volume-weighted average price (VWAP) is one of the most widely used benchmark prices that analysts use to estimate transaction costs. Analysts typically compute the VWAP using all trades that occurred from the start of the order until the order was completed, a measure that is often referred to as "interval VWAP." The VWAP is the sum of the total dollar value of the benchmark trades divided by the total quantity of the trades. The VWAP transaction cost estimate formula is as follows:

$$\text{VWAP transaction cost estimate} =$$

$$\text{Trade size} \times \begin{cases} \text{Trade VWAP} - \text{VWAP benchmark} & \text{for buy orders} \\ \text{VWAP benchmark} - \text{Trade VWAP} & \text{for sell orders} \end{cases}$$

The VWAP transaction cost estimate is popular in part because it is easy to interpret. It answers this question: Did you get a better or worse average price than all traders trading when you were trading?

Interpreting VWAP transaction cost estimates is problematic when the trades being evaluated are a substantial fraction of all trades in the VWAP benchmark, or, more generally, when the trades took place at the same rate as other trades in the market. In both cases, the Trade VWAP and the VWAP benchmark will be nearly equal, which would suggest that the evaluated trades were not costly. But this conclusion would be misleading if the trade had substantial price impact. For example, if a large trader were the only buyer for a given trading period (or interval), the VWAP transaction cost estimate would be zero regardless of the market impact.

This bias toward zero helps explain why the measure is so popular. Investment managers like to show their investment sponsors transaction cost estimates that suggest that trading is not expensive.

EXAMPLE 1

Transaction Cost Analyses for an Illiquid Stock

Arapahoe Tanager, portfolio manager of a Canadian small-cap equity mutual fund, and his firm's chief trader, Lief Schrader, are reviewing the execution of a ticket to sell 12,000 shares of Alpha Company, limit C$9.95. The order was traded over the day.

Schrader split the ticket into three orders that executed that day as follows:

A A market order to sell 2,000 shares executed at a price of C$10.15. Upon order submission, the market was C$10.12 bid for 3,000 shares, 2,000 shares offered at C$10.24.

B A market order to sell 3,000 shares executed at a price of C$10.11. Upon order submission, the market was C$10.11 bid for 3,000 shares, 2,000 shares offered at C$10.22.

C Toward the end of the trading day, Schrader submitted an order to sell the remaining 7,000 shares, limit C$9.95. The order executed in part, with 5,000 shares trading at an average price of C$10.01. Upon order submission, the market was C$10.05 bid for 3,000 shares, 2,000 shares offered at C$10.19. This order exceeded the quoted bid size and "walked down" the limit order book (i.e., after the market bid was filled, the order continued to sell at lower prices). After the market closed, Schrader allowed the order to cancel. Tanager did want to sell the 2,000 unfilled shares on the next trading day.

Only two other trades in Alpha Company occurred on this day: 2,000 shares at C$10.20 and 1,000 shares at C$10.15. The last trade price of the day was C$9.95; it was C$9.50 on the following day.

1 For each of the three fund trades, compute the quoted spread. Also, compute the average quoted spreads prevailing at the times of each trade.
2 For each of the three fund trades, compute the effective spread (use the average fill price for the third trade). Also, compute the average effective spread.
3 Explain the relative magnitudes of quoted and effective spreads for each of the three fund trades.
4 Calculate the VWAP for all 13,000 Alpha Company shares that traded that day and for the 10,000 shares sold by the mutual fund. Compute the VWAP transaction cost estimate for the 10,000 shares sold.

Solution to 1:

The quoted spread is the difference between the ask and bid prices. For the first order, the quoted spread is C$10.24 – C$10.12 = C$0.12. Similarly, the quoted spreads for the second and third orders are C$0.11 and C$0.14, respectively. The average quoted spread is (C$0.12 + C$0.11 + C$0.14)/3 = C$0.1233.

Solution to 2:

The effective spread for a sell order is 2 × (Midpoint of the market at the time of order entry – Trade price). For the first order, the midpoint of the market at the time of order entry is (C$10.12 + C$10.24)/2 = C$10.18, so that the effective spread is 2 × (C$10.18 – C$10.15) = C$0.06.

The effective spread for the second order is 2 × [(C$10.11 + C$10.22)/2 – C$10.11] = C$0.11.

The effective spread for the third order is 2 × [(C$10.05 + C$10.19)/2 – C$10.01] = C$0.22.

The average effective spread is (C$0.06 + C$0.11 + C$0.22)/3 = C$0.13.

Solution to 3:

The first trade received price improvement because the shares sold at a price above the bid price. Therefore, the effective spread is less than the quoted spread. No price improvement occurred for the second trade because the shares sold at the bid price. Also, the second trade had no price impact beyond trading at the bid; the entire order traded at the quoted bid. Accordingly, the effective and quoted spreads are equal. The effective spread for the third trade is greater than the quoted spread because the large order size, which was greater than the bid size, caused the order to walk down the limit order book. The average sale price was less than the bid so that the effective spread was higher than the quoted spread.

Solution to 4:

The VWAP for the day is the total dollar volume divided by the total number of shares traded. The dollar volume is 2,000 shares × C$10.15 + 3,000 shares × C$10.11 + 5,000 shares × C$10.01 + 2,000 shares × C$10.20 + 1,000 shares at C$10.15 = C$131,230. Dividing this by the 13,000-share total volume gives a VWAP of C$10.0946. A similar calculation using only the sales sold by the mutual fund gives a trade VWAP of C$10.0680. The VWAP transaction cost estimate for the sale is the difference multiplied by the 10,000 shares sold: C$266.15 = 10,000 shares × (C$10.0946 – C$10.0680) [differences due to rounding].

DEVELOPMENT OF ELECTRONIC MARKETS

3

d describe factors driving the development of electronic trading systems;

e describe market fragmentation;

The application of new information technologies to trading processes produced radical changes in how investment managers trade. Automated trading systems and trading strategies replaced manual processes. New electronic exchanges, alternative trading systems, electronic traders, and securities dramatically changed trading in most markets. The resulting efficiencies generally improved market quality, but electronic trading also produced new regulatory concerns. High levels of fragmentation and electronification now characterize most global trading markets.

3.1 Electronic Trading

Trading at organized exchanges now depends critically on automated electronic systems used both by exchanges and by their trader clients. The exchanges use electronic systems to arrange trades by matching orders submitted by buyers with those submitted by sellers. Traders use electronic systems to generate the orders that the exchanges process. The most important electronic traders are dealers, arbitrageurs, and buy-side institutional traders who use algorithmic trading tools provided by their brokers to fill their large orders.

The two types of systems are co-dependent: Traders need high-speed order processing and communication systems to implement their electronic trading strategies, and the exchanges need electronic exchange systems to process the vast numbers of orders that these electronic traders produce. The adoption of electronic exchange systems led to huge growth in automated order creation and submission systems.

The widespread use of electronic trading systems significantly decreased trading costs for buy-side traders. Costs fell as exchanges obtained greater cost efficiencies from using electronic matching systems instead of floor-based, manual trading systems. These technologies also decreased costs and increased efficiencies for the dealers and arbitrageurs, who provide much of the liquidity offered at exchanges. Competition forced them to pass along many of the benefits of their new technologies to buy-side traders in the form of narrower spreads quoted for larger sizes. New electronic buy-side order management systems also decreased buy-side trading costs by allowing a smaller number of buy-side traders to process more orders and to process them more efficiently than manual traders.

3.2 Advantages of Electronic Trading Systems

Compared with floor-based trading systems, electronic order-matching systems enjoy many advantages:

- Most obviously, electronic systems are cheap to operate once built. Operating in server rooms, they require less physical space than trading floors. Also, in contrast to floor-based trading systems, electronic trading systems do not require exchange officials to record and report prices.
- Electronic exchange systems do exactly what they are programmed to do. When properly programmed, they precisely enforce the exchange's trading order precedence and pricing rules without error or exception.
- Electronic exchange systems can also keep perfect audit trails so that forensic investigators can determine the exact sequence and timing of events that may interest them.

- Electronic exchange systems that support hidden orders keep those orders perfectly hidden. Unlike floor brokers, they never inadvertently or fraudulently reveal their clients' hidden orders to others.
- In contrast to floor-based brokers and exchange officials, electronic order-matching systems can operate, for the most part, on a continuous, "around-the-clock" basis.
- Finally, electronic exchanges can operate when bad weather or other events would likely prevent workers from convening on a floor.

These efficiencies led to great growth. Electronic trading systems have largely displaced floor-based trading systems in all instruments for which order-driven markets are viable. Order-driven markets—markets in which orders submitted by traders are arranged based on a rules-based, order-matching system run by an exchange, a broker, or an alternative trading system (ATS)—are now organized by most exchanges and electronic communication networks (ECNs).

Additionally, computers have come to dominate the implementation of many trading strategies because they are so efficient and so unlike human traders:

- Computers have infinite attention spans and a very wide attention scope. They can continuously watch and respond to information from many instruments and many markets simultaneously and essentially forever.
- Their responses are extraordinarily fast.
- Computers are perfectly disciplined and do only what they are instructed (programmed) to do.
- Computers do not forget any information that their programmers want to save.

3.3 Electronification of Bond Markets

The electronic market structures of equity, futures, and options markets have attracted tremendous attention throughout the world. Much less attention has been given to the market structures of corporate and municipal bond markets, most of which, from the customer's point of view, have changed little since the late 19th century. Despite the efforts of many creative developers of electronic bond trading systems, most public investors in these markets still trade largely over the counter with dealers. The potential for electronic trading systems in these markets—and the attendant growth in electronic trading strategies—is quite large. Such systems undoubtedly will reflect the fact that bond issues—especially municipal bonds—vastly outnumber stock issues. Accordingly, except for the most actively traded bonds, limit order book trading systems will not be successful because buyers and sellers rarely will be present at the same time.

However, systems can be built that would allow public investors to trade with each other when both sides are present in the market. These systems would provide order display facilities, where public investors and proprietary traders could post limit orders so that all traders could see them. Like marketable orders, limit orders seek to obtain the best price immediately available; additionally, they instruct not to accept a price higher than a specified limit price when buying or a price lower than a specified limit price when selling. If these facilities also had automatic execution mechanisms and regulations or legal decisions to prevent dealers from trading through displayed orders when arranging their trades, bond transaction costs would drop substantially and bond trading would become much more active. Many such electronic bond order-matching systems already exist, but they primarily serve dealers and not public investors. Recent empirical research suggests that public investors would greatly benefit if their brokers provided them with direct access to these systems as they presently do in the equity markets. Instead, most broker/dealers commonly interpose themselves.

3.4 Market Fragmentation

Markets for many asset classes have become increasingly fragmented throughout the world because venues trading the same instruments have proliferated and trading in any given instrument now occurs in multiple venues. Available liquidity for an instrument on any one exchange now often represents just a small fraction of the aggregate liquidity for that instrument. **Market fragmentation**—trading the same instrument in multiple venues—increases the potential for price and liquidity disparities across venues because buyers and sellers often are not in the same venues at the same time.

For example, in the United States, order flow in exchange-listed equities is now divided among 11 exchanges, 40 alternative trading systems, and numerous dealers. In the late 20th century, however, trading mainly occurred on three primary exchanges, a few minor regional exchanges, and in the offices of some large institutional broker/dealers. Alternative trading systems (ATSs), also known as electronic communication networks (ECNs) or multilateral trading facilities (MTFs), are increasingly important trading venues. They function like exchanges but do not exercise regulatory authority over their subscribers except concerning the conduct of their trading in their trading systems.

With increasing market fragmentation, traders filling large orders now adapt their trading strategies to search for liquidity across multiple venues and across time to control the market impacts of their trades. Electronic algorithmic trading techniques, such as liquidity aggregation and smart order routing, help traders manage the challenges and opportunities presented by fragmentation. Liquidity aggregators create "super books" that present liquidity across markets for a given instrument. These tools offer global views of market depth (available liquidity) for each instrument regardless of which trading venue offers the liquidity. For example, the best bid, or highest price a buyer is willing to pay, for a Eurodollar future may be on the Chicago Mercantile Exchange (CME) and the second best on ELX Markets, a fully electronic futures exchange. Smart order-routing algorithms send orders to the markets that display the best-quoted prices and sizes.

3.5 Effects on Transaction Costs

Numerous studies show that transaction costs declined with the growth of electronic trading over time. Some studies also show that at a given point in time, lower transaction costs are found in those markets with the greatest intensity of electronic trading. These time-series and cross-sectional results are not surprising. They result from the greater cost efficiencies associated with electronic trading.

With the growth of electronic trading, bid–ask spreads decreased substantially. These decreases lowered transaction costs for retail traders and institutions trading small orders.

Overall transaction costs also decreased for large orders, many of which are now broken into smaller parts for execution. A study of the execution costs of tens of thousands of equity orders for US stocks involving tens of millions of dollars of principal value shows that the implementation shortfall cost of filling those orders dropped with the growth of electronic trading. This evidence suggests that any profits obtained by parasitic traders from front running orders are smaller than the cost savings obtained by buy-side traders from trading in electronic markets using algorithms.

4 TYPES OF ELECTRONIC TRADERS

f identify and contrast the types of electronic traders;

The proliferation of electronic exchange trading systems has led to the adoption of electronic trading by proprietary traders, buy-side traders, and the electronic brokers that serve them. Proprietary traders include dealers, arbitrageurs, and various types of front runners—all of whom are profit-motivated traders. In contrast, buy-side traders trade to fill orders for investment and risk managers who use the markets to establish positions from which they derive various utilitarian and profit-motivated benefits. Electronic brokers serve both types of traders.

Electronic traders differ in how they send orders to markets. Those proprietary traders who are registered as broker/dealers usually send their orders directly to exchanges. Those who are not broker/dealers must send their orders to brokers, who then forward them to exchanges. These brokers are said to provide sponsored access to their proprietary electronic trader clients. Brokers who provide sponsored access have very fast electronic order processing systems that allow them to forward orders to exchanges as quickly as possible while still undertaking the regulatory functions necessary to protect the markets and themselves from various financial and operational risks associated with brokering orders for proprietary electronic traders.

Electronic trading strategies are most profitable or effective when they can act on new information quickly. Accordingly, proprietary traders and electronic brokers build automated trading systems that are extremely fast. These systems often can receive information of interest to the trader, process it, and place a trading instruction at an exchange in less than a few milliseconds—and sometimes much faster.

The events that interest electronic traders include:

- trade reports and quote changes in the securities or contracts that they trade;
- similar data for instruments that are correlated with the securities or contracts that they trade;
- indexes that summarize these data across markets and for various instrument classes;
- changes in limit order books; and
- news releases from companies, governments, and other producers and aggregators of information.

Electronic traders typically receive information about these events via high-speed electronic data feeds. Not all electronic traders analyze all these different information sources, but many do.

Electronic proprietary traders include high-frequency traders and low-latency traders. High-frequency and low-latency (i.e., extremely fast) traders must often trade very quickly in response to new information to be profitable. They are distinguished by how often they trade.

High-frequency traders (HFTs) generally complete round trips composed of a purchase followed by a sale (or a sale followed by a purchase) within a minute and often as quickly as a few milliseconds. During a day, they may trade in and out of an actively traded security or contract more than a thousand times—but usually only in small sizes.

Low-latency traders include news traders who trade on electronic news feeds and certain parasitic traders. Parasitic traders are speculators who base their predictions about future prices on information they obtain about orders that other traders intend, or will soon intend, to fill. Parasitic traders include front runners, who trade in front of traders who demand liquidity, and quote matchers, who trade in front of traders who supply liquidity. When trying to open or close positions, low-latency traders often need to send or cancel orders very quickly in response to new information. In contrast to HFTs, low-latency traders may hold their positions for as long as a day and sometimes longer.

The distinction between HFTs and low-latency traders is relatively new. Many commentators do not make any distinction, calling all electronic traders who need to trade quickly HFTs.

4.1 The Major Types of Electronic Traders

Electronic news traders subscribe to high-speed electronic news feeds that report news releases made by corporations, governments, and other aggregators of information. They then quickly analyze these releases to determine whether the information they contain will move the markets and, if so, in which direction. They trade on this information by sending marketable orders—instructions to fill the order at the best available price—to wherever they expect they may be filled. News traders profit when they can execute against stale orders—orders that do not yet reflect the new information.

For example, stock prices usually rise when a company announces earnings of 25 pence a share when the consensus forecast is only 10 pence. Electronic news traders who receive the initial press release will use their computers to parse the text of the release to find the earnings number. The computers then will compare that number with the consensus forecast, which they have stored in their memory rather than on disk to reduce access time. If the 15 pence difference is sufficiently large, news traders may send one or more marketable buy orders to exchanges for execution. News traders must be very quick to ensure that they get to the market before others do. If they are too late, the price may have changed already or liquidity suppliers may have canceled their quotes.

Some news traders also process news releases that do not contain quantitative data. Using natural language-processing techniques, they try to identify the importance of the information for market valuations. For example, a report stating that "our main pesticide plant shut down because of the accidental release of poisonous chemicals" might be marked as having strong negative implications for values. Electronic news traders would sell on this information. If they are correct, the market will drop as other, slower traders read, interpret, and act on the information. If they are wrong, the market will not react to the information. In that case, news traders will reverse their position and lose the transaction costs associated with their round-trip trades. (Note that these transaction costs could be high if many news traders made the same wrong inference.) Because round-trip transaction costs usually are lower than the profits that electronic news traders can occasionally make when significant news arrives, news traders often may trade with the expectation of being right only occasionally.

Electronic dealers, like all dealers, make markets by placing bids (prices at which they are willing to buy) and offers (prices at which they are willing to sell) with the expectation that they can profit from round trips at favorable net spreads. Those who trade at the highest frequencies tend to be very wary. On the first indication that prices may move against their inventory positions (i.e., price decreases if they are long or own the asset; price increases if they are short or sold an asset they do not own), they immediately take liquidity by executing on the opposite side to reduce their exposure. They generally will not hold large inventory positions in actively traded stocks. As soon as they reach their inventory limit on one side of the market or the other, they cease bidding or offering on that side. Electronic dealers often monitor electronic news feeds. They may immediately cancel all their orders in any security mentioned in a news report. If the news is material, they do not want to offer liquidity to news traders to whom they would lose. If the news is immaterial, they merely lose whatever opportunity to trade may have come their way while out of the market.

Electronic dealers, like all other dealers, also keep track of scheduled news releases. They cancel their orders just before releases to avoid offering liquidity to traders who can act faster than they can. They also may try to reduce their inventories before a scheduled release to avoid holding a risky position.

Electronic arbitrageurs look across markets for arbitrage opportunities in which they can buy an undervalued instrument and sell a similar overvalued one. The combination of these two positions is called an arbitrage portfolio, and the positions are called legs. Electronic arbitrageurs try to construct their arbitrage portfolios at minimum cost and risk.

Electronic front runners are low-latency traders who use artificial intelligence methods to identify when large traders, or many small traders, are trying to fill orders on the same side of the market. They will purchase when they believe that an imbalance of buy orders over sell orders will push the market up and sell when they believe the opposite. Their order anticipation strategies try to identify predictable patterns in order submission. They may search for patterns in order submissions, trades, or the relations between trades and other events.

In most jurisdictions, dealers and brokers cannot legally front run orders that their clients have submitted. These orders include large orders that they know their clients are breaking up to fill in small pieces. But dealers and brokers can study records of their clients' past orders to identify patterns in their behavior that would allow them to predict orders not yet submitted.

Some front runners also look for patterns in executed trades. For example, suppose that a trader sees that trades of a given size have been occurring at the offer every 10 minutes for an hour. If the trader has seen this pattern of trading before, the trader may suspect that the activity will continue. If so, the trader may buy on the assumption that a trader is in the market filling a large buy order by breaking it into smaller pieces.

Buy-side traders, and the brokers who provide them with algorithms to manage large orders, are aware of the efforts that electronic traders make to detect and front run their orders. Accordingly, they randomize their strategies to make them more difficult to detect. They submit orders at random times instead of at regular intervals, and they submit various sizes instead of the same size. Although these techniques make detection more difficult, hiding large, liquidity-demanding trades is always challenging because sophisticated traders can ultimately identify them by the inevitable relation between prices and volumes that they create. Electronic front runners look for these patterns, often using very advanced, automated data-mining tools.

Finally, some front runners examine the relation between trades and other events to predict future trades. Traders who identify these events quickly may be able to profit by buying ahead of retail or institutional traders. Because many traders initiate trades in response to common stimuli or in response to predictable situations, traders who can identify patterns in the relations between trades and events may profit from trading ahead. When the time between the stimulus and the response is short, electronic traders have a clear advantage.

Electronic quote matchers try to exploit the option values of standing orders. Standing orders are limit orders waiting to be filled. Options to trade are valuable to quote matchers because they allow them to take positions with potentially limited losses. Quote matchers buy when they believe they can rely on standing buy orders to get out of their positions, and they sell when they can do the same with standing sell orders. Traders say that quote matchers lean on these orders. If prices then move in the quote matchers' favor, they profit for as long as they stay in the security or contract. But if the quote matchers conclude that prices are moving against them, they immediately try to exit by trading with the standing orders and thereby limiting their losses.

For example, a fast quote matcher may buy when a slow trader is bidding at 20. If the price subsequently rises, the quote matcher will profit. If the quote matcher believes that the price will fall, the quote matcher will sell the position to the buyer at 20 and thereby limit his losses. The main risk of the quote-matching strategy is that the standing order may be unavailable when the quote matcher needs it. Standing orders disappear when filled by another trader or when canceled.

Most large buy-side traders use electronic order management systems (OMSs) to manage their trading. These systems keep track of the orders that their portfolio managers want to be filled, which orders have been sent out to be filled, and which fills have been obtained. Buy-side OMSs generally allow the buy-side trader to route orders to brokers for further handling, along with instructions for how the orders should be handled. These entities may include exchanges, brokers, dealers, and various alternative trading systems. The OMSs typically have dashboards that allow the buy-side trader to see summaries of all activity of interest so that the trader can better manage the trading process. Finally, the OMSs help the buy-side traders report and confirm the trades to all interested parties.

Buy-side traders often employ electronic brokers to arrange their trades. In addition to supporting standard order instructions, such as limit or market orders, these brokers often provide a full suite of advanced orders, trading tactics, and algorithms. The broker's electronic trading system generally manages these advanced orders, tactics, and algorithms, but in some cases, exchange computers may perform these functions.

ELECTRONIC TRADING SYSTEM: CHARACTERISTICS AND USES

5

g describe characteristics and uses of electronic trading systems;

h describe comparative advantages of low-latency traders;

Traders value speed because it allows them to act before other traders can act. This section identifies the three situations where speed is valuable, how exchanges and traders build and use fast trading systems, and some select examples of how electronic trading changed trading strategies.

5.1 Why Speed Matters

Electronic traders must be fast to trade effectively, regardless of whether they are proprietary traders or buy-side traders. Electronic traders have three needs for speed:

1 **Taking**. Electronic traders sometimes want to take a trading opportunity before others do. A new trading opportunity may attract many traders, and an existing trading opportunity may attract many traders when market events cause it to become more valuable (e.g., a standing limit order to sell becomes much more attractive when the prices of correlated securities rise). Often only the first trader to reach the attractive opportunity will benefit. Thus, electronic traders must be fast so they can beat other traders to attractive trading opportunities.

2 **Making**. Market events often create attractive opportunities to offer liquidity. For example, at most exchanges when prices rise, the first traders to place bids at improved prices acquire time precedence at those prices that may allow them

to trade sooner or at better prices than they otherwise would be able to trade. Therefore, electronic traders must be fast so they can acquire priority when they want it and before other traders do.

3 **Canceling**. Frequently, traders must quickly cancel orders they no longer want to fill, often because market events have increased the option values of those orders. For example, if traders have limit buy orders standing at the best bid and large trades take place at other exchanges at the same price, these traders may reasonably conclude that prices may drop and that they may obtain better executions at a lower price. They must cancel their orders as quickly as possible to reduce the probability that they will trade.

Note that electronic traders do not simply need to be fast to trade effectively: They must be faster than their competitors. Little inherent value comes from being fast; the value lies in being faster. The reason electronic trading systems have such low latencies (i.e., are extremely fast) is because electronic traders have been trying for years to be faster than their competitors.

Electronic order-handling systems used by exchanges also have grown faster as exchanges compete for order flows from electronic traders. Electronic traders often will not send orders to exchanges where they cannot quickly cancel them, especially if other exchanges have faster trading systems. Accordingly, exchanges with slow order-handling systems have lost market share.

Latency is the elapsed time between the occurrence of an event and a subsequent action that depends on that event. For example, the event might be a trade at one exchange, and the action might be the receipt by another exchange of an instruction to cancel a standing order that a trader has sent upon learning of the trade. Electronic traders measure these latencies in milliseconds or microseconds (millionths of a second).

The latency of a linear multi-step process is the sum of the latencies of each step in the process. The submission of an order instruction by a trader in response to an event consists of three major steps, each of which involves many smaller steps beyond the scope of this discussion:

1 The trader must learn that the event took place.

2 The trader must respond to the new information with a new order instruction.

3 The trader must send, and the exchange must receive, the new instruction.

Traders must use very fast communication systems to minimize the latencies associated with steps 1 and 3 (communicating in and out), and they must use very fast computer systems to minimize the latency associated with step 2 (responding).

5.2 Fast Communications

Electronic traders and brokers use several strategies to minimize their communication times. These strategies involve minimizing communication distances and maximizing line speeds. Note that the relevant measure of communication distance is the total of two distances that signals must travel. The first distance is from where the event is reported (often an exchange but sometimes another type of news source) to the computer that will process the information. The second distance is from the computer to the exchange trading system where the trader wants to deliver an order instruction.

Electronic traders and brokers locate their computers as close as possible to the exchanges at which they trade to minimize latencies resulting from physics: No message can travel faster than the speed of light. At 300,000 kilometers (186,000 miles) per second in a vacuum, light travels 300 kilometers in a millisecond. Although the speed of light is incredibly fast, a fast computer with a clock speed of 5 GHz (billion cycles per second) can do 5 million operations in a millisecond—which often is more than required to receive information, process it, and send out an order instruction in response.

Communication latencies are particularly important when messages must travel significant distances. For example, the great circle (shortest) distances between Chicago and New York and between New York and London are, respectively, 1,146 kilometers and 5,576 kilometers. Thus, round-trip communications between these two pairs of cities have minimum latencies of approximately 8 and 37 milliseconds simply because of the speed of light. (The actual minimum latencies are longer because the speed of light in standard optical fiber is 31% slower than the speed of light in a vacuum.) Such delays illustrate that no electronic trader located at any significant distance from where information is created or must be delivered can effectively compete with traders who have minimized these combined distances.

Many exchanges allow electronic traders to place their servers in the rooms where the exchange servers operate, a practice called collocation. Exchanges charge substantial fees for collocation space and related services, such as air conditioning and power. Note that even within collocation centers, concerns about fairness dictate that the communication lines connecting proprietary servers to exchange servers all be of the same length for all customers buying the same class of collocation service.

Electronic traders and brokers also use the fastest communication technologies they can obtain to collect and transmit information when any distance separates the places where information events occur from the places where they act on those events. To that end, they use the fastest and most direct communication lines that are available. For example, they prefer line-of-sight microwave channels to fiber-optic and copper channels because of the differences in speed of electromagnetic wave propagation through these materials. (Microwaves travel through air at just slightly below the speed of light, whereas signals travel through fiber-optic channels and copper wires only two-thirds as quickly.) They also ensure that their communications pass through the fewest electronic routers and switches possible because passage through each of these devices adds its latency to the total latency of the line.

Finally, electronic traders and brokers subscribe to special high-speed data feeds directly from exchanges and other data vendors. The vendors charge premium prices for these services, which are delivered over very high-speed communication lines. Some exchanges provide multiple classes of data services that vary by speed to price-discriminate among their clients.

5.3 Fast Computations

Once electronic traders receive information about an event of interest, they must decide whether to act on that information and how. Those traders who can make decisions faster than their competitors will trade more profitably. Electronic traders minimize the latencies associated with their decision making by using several strategies.

First and most obviously, they use very fast computers. They overclock their processors (i.e., run them faster than the processor designers intended) and use liquid cooling systems to keep them from melting. They store all information in fast memory to avoid the latencies associated with physical disk drives, which cannot deliver information while their heads are seeking the right track and can only deliver information as fast as their disks spin once the right track is found. They sometimes use specialized processors designed to solve their specific trading problems quickly, and they may even use processors etched on gallium arsenide rather than silicon.

Electronic traders also must run very efficient software. They often use simple and specialized operating systems to avoid the overhead associated with supporting operating system functions they do not use. Remarkably, many electronic trading systems run under variants of the original MS-DOS operating system because of its simplicity.

Electronic traders optimize their computer code for speed. They often write important functions that they repeatedly use in assembler language to ensure that they run quickly. (Code written in high-level languages, such as C++, tends to be slower because their compilers are designed to handle all types of code, not just code written to solve trading problems.) And they avoid using such languages as Python because they are interpreter languages that compile (create executable machine code) as they run, rather than compiling only once when first written.

Some electronic trading problems change so frequently that speed of coding is more important than speed of execution. For example, some problems depend on ever-changing sets of conditions or exceptions that present or constrain profit opportunities. For such problems, traders use high-level languages (e.g., Python), because they can code faster and more accurately in these languages than in lower-level languages, such as C++. If they expect that the software will remain useful, they may later recode their routines in other languages to make them run faster.

Some electronic traders also reduce latency by creating contingency tables that contain prearranged action plans. For example, suppose that a bid rises in a market in which electronic traders are active. In response to the increased bid, traders may want to raise their bids or offers. The decision to do so may depend on their inventory positions and perhaps on many other factors as well. To decide what to do following an increased bid may require substantial analyses, which take time. Traders can reduce their decision latencies by doing these analyses before the bid increases instead of afterward. Seeing the increased bid, they can respond by simply looking up the optimal response in a contingency table stored in memory. To be most useful, the contingency tables must be kept up to date and must include responses for most-likely events. In this example, traders presumably would also have precomputed responses for a decrease in the bid, among many other contingencies.

EXAMPLE 2

Latency

Explain why low-latency is important to electronic traders.

Solution:

Electronic traders need a comparative speed advantage to 1) take advantage of market opportunities before others do, 2) receive time precedence that would allow them to trade sooner when offering liquidity to others, and 3) ensure order cancellation when they no longer want to fill the order. To gain a comparative advantage relative to others, electronic traders try to minimize latency—the time between an event occurring and a subsequent action, typically the submission of an order instruction, based upon that event. To minimize latency, electronic traders invest in very fast communication systems and very fast computer systems.

5.4 Advanced Orders, Tactics, and Algorithms

Buy-side traders often use electronic brokers and their systems for advanced orders, trading tactics, and algorithms provided by their electronic brokers to search for liquidity.

Advanced order types. Advanced orders generally are limit orders with limit prices that change as market conditions change. An example would be a pegged limit order for which the trader would like to maintain a bid or an offer at a specified distance relative to some benchmark. Suppose that a trader wants to peg a limit buy order two ticks below the current ask. A broker who supports this instruction may forward it to an exchange that supports the instruction if the probability of the order's filling at that exchange is favorable compared with other exchanges. When the ask rises or falls, the exchange system will immediately cancel the order and replace it with a new limit order to keep the order at two ticks below the current ask. If the exchange does not support this instruction, the broker's computer will manage the order by submitting a limit order priced two ticks below the current ask and adjusting it as necessary to maintain the peg when the market moves. Effective management of a pegged limit order requires an electronic trading system with very low latency. If the order is not adjusted quickly enough, it risks being executed at an unfavorable price (in this example, if prices drop) or being resubmitted after other orders have been placed at the new price so the probability of execution at that price will be lower (if prices rise). Traders sometimes call pegged limit orders floating limit orders.

Trading tactics. A trading tactic is a plan for executing a simple function that generally involves the submission of multiple orders. Note that the distinction between advanced orders and tactics can be arbitrary, and not all traders will use the same language to describe various trading functions. An example of a trading tactic is an instruction to sweep through every market at a given price to find hidden trading opportunities.

Suppose that the best exposed bid among all trading venues is 20.00 and the best exposed offer is 20.02. Because many trading systems permit traders to hide their orders, hidden buyers or sellers may be willing to trade at the 20.01 midpoint. Depending on the exchange, at least three types of orders could permit a trade at the midpoint. First, among exchanges that permit hidden orders, one or more exchanges may be holding a hidden limit order at 20.01. Second, among exchanges that permit discretionary limit orders, one or more exchanges may be holding a discretionary limit order that can be filled at the midpoint. For example, suppose that an exchange is holding a limit order to buy at 19.99 with 0.02 discretion. This order can be filled at 20.01 if a suitable sell limit order arrives at that price. Finally, among exchanges and dark pools that permit midspread orders, one or more exchanges or dark pools may be holding such an order. Dark pools are trading venues that do not publish their liquidity and are only available to selected clients. A midspread order is a limit order that is pegged to the midpoint of the quoted bid–ask spread.

To find such hidden liquidity, an electronic trading system may submit an immediate or cancel (IOC) order priced at 20.01 to the exchange that the trader expects will most likely have hidden liquidity on the needed side of the market. If such liquidity exists, the order will execute up to the minimum of the sizes of the two orders. If not, the exchange will immediately cancel the order and report the cancellation. If the order has any remaining unfilled size, the electronic trading system will search for liquidity at another exchange. This process will continue until the order is filled or until the trader decides that further search is probably futile. This sweeping tactic is most effective when the electronic trading system managing it has very low latency. A slow system may lose an opportunity to trade if someone else takes it first. Also, a slow system that obtains one or more partial fills may lose opportunities to trade at other exchanges if the proprietary electronic trading systems managing the standing orders that provide those opportunities cancel their standing orders when they suspect someone is sweeping the market, as they might if they see trade reports inside the quoted spread.

An example of another trading tactic is placing a limit order at some price with the hope that it will fill at that price. If the order does not fill after some time period (which might be random or based on information), the electronic trading system will cancel the order and resubmit it with an improved price (i.e., a higher price for a buy order or a lower price for a sell order). The process is repeated until the order fills.

Algorithms. Algorithms ("algos" for short) are programmed strategies for filling orders. Algorithms may use combinations or sequences of simple orders, advanced orders, or multiple orders to achieve their objectives. Buy-side traders use algorithms, often provided by brokers, extensively to trade small orders and to reduce the price impacts of large trades. For example, many algorithms break up large orders and submit the pieces to various markets over time. Breaking up orders makes it difficult for other traders to infer that a trader is trying to fill a large order. The algorithms typically submit the orders at random times, in random sizes, and sometimes to randomly selected exchanges to hide their common origin.

The rates at which algorithms try to fill large orders may depend on market volumes or on elapsed time. For example, VWAP algorithms attempt to obtain a volume-weighted average fill price that is close to (or better than) the volume-weighted average price (VWAP) of all trades arranged within a prespecified time interval. To minimize the variation between the actual average fill price and the VWAP over the interval, these algorithms try to participate in an equal fraction of all trading volume throughout the interval. To do so, they forecast volumes based on the historical volume profile and on current volumes. The algorithm trades more during periods of historically high volume (e.g., around market open and close) and when the market has been more active than normal. It trades less during periods of relatively low volume. In practice, the execution rate will vary because volumes will differ from expectations. Buy-side traders use VWAP algorithms when spreading the order over time and when obtaining the average market price within an interval is acceptable to them or their portfolio managers.

Many algorithms use floating limit orders with the hope of obtaining cheap executions. If they fail to fill after some time period, they may switch to more-aggressively priced orders or to marketable orders to ensure that they fill. Large traders who use algorithms to manage their orders are especially concerned about hiding their intentions from front runners. Many electronic traders use artificial intelligence systems to detect when large traders are present in the market. In particular, they look for patterns that large traders may leave. For example, a poorly designed algorithm may submit orders exactly at the same millisecond within a second whenever it submits an order. A clever trader who is aware of this regularity may detect when a large trader is in the market and, equally important, when the trader has completed filling his order. To avoid these problems, algorithm designers often randomize order submission times and sizes to avoid producing patterns that might give them away. They also sometimes try to hide their orders among other orders so that front runners cannot easily identify their intentions.

Developing good algorithms requires extensive research into the origins of transaction costs. Algorithm authors must understand transaction costs well so that they can design algorithms that will trade effectively. To that end, algorithm providers build and estimate models of the costs of trading orders of various sizes, models of the impact trades of a given size or frequency will have on prices, and models of the probabilities that limit orders will fill under a variety of conditions. They must also predict volumes accurately. The most effective algorithms are based on the best research and implemented on the fastest and most capable electronic systems.

Good algorithms generally obtain low-cost executions by knowing when and where to offer liquidity via limit orders, when to use market orders, and how to most effectively keep the market from being aware of their efforts. They reduce the price impacts of large trades and greatly reduce the costs of managing many small trades.

EXAMPLE 3

Use of Electronic Brokers

You have recently been hired recently as a junior buy-side analyst. Part of your training (on-boarding) has been to sit with the trading desk to learn how the desk trades through its electronic brokers. In a meeting with your manager, she asks you to explain the use of electronic brokers for advanced orders, trading tactics, and algorithmic trading tools that your electronic brokers provide. What would you say?

Solution:

The use of electronic brokers and their systems is valuable for such advanced order types as pegged or floating limit orders, whose limit prices change as market conditions change. Traders use these order types to supply liquidity at a specified distance from the market. These orders require continuous real-time evaluation to determine if an order cancellation or replacement is needed as market conditions change. The use of electronic brokers relieves the need for the trader to continuously monitor the market to cancel and resubmit orders when prices change. An electronic broker is also valuable for orders placed a few ticks outside the best market that will be among the last orders to supply liquidity to a large trader, hopefully at a good price.

Electronic brokers also allow their clients to access order execution tactics (presented as another complex order type) that involve multiple submissions that may "sweep" through markets to uncover hidden liquidity. These tactics allow traders to submit multiple orders with a single instruction.

Finally, electronic brokers also provide algorithmic trading tools. Algorithms are automated (programmed trading strategies for combinations of simple and single, advanced, or multiple orders and various trading tactics) to fill small orders efficiently based on various criteria. They often break up large orders into smaller pieces to minimize the market impact of filling the order. They may route the orders to multiple venues at the same time or to the same venue at various times. For example, VWAP algorithms attempt to fill orders at the volume-weighted average price (or better) of all trades over a specified interval. The systems running algorithms that place standing limit orders must be very fast to cancel orders in trading. In these cases, low latency is critical to ensure order cancellation before unfavorable executions occur. Fast systems also help ensure that traders are first to respond when market conditions change and to maintain time precedence.

5.5 Select Examples of How Electronic Trading Changed Trading Strategies

The growth in electronic trading systems changed how traders interact with the market. Proprietary traders, buy-side traders, and brokers adapted their trading strategies to use new electronic tools and facilities. Select characteristics of electronic trading are described below.

Hidden orders. Hidden orders are very common in electronic markets. Hidden orders are orders that are exposed (or shown) only to the brokers or exchanges who receive them. Traders—especially large traders—submit them when they do not want to reveal the existence of the trading options that their standing orders provide to the markets. Traders concerned about quote matchers can protect themselves to some extent by submitting hidden limit orders. Note that hidden limit orders are the electronic equivalent of giving orders to floor brokers to fill with the understanding that the floor brokers may expose the orders only if they can arrange trades. Such orders work better at electronic exchanges than at floor-based exchanges because computers never inadvertently or intentionally display these orders improperly. In electronic markets, the most common type of order by far is the immediate or cancel (IOC) limit order. Traders use these orders to discover hidden orders that may stand in the spread between a market's quoted bid and ask prices. Because they cancel immediately if they do not find liquidity, these orders are also hidden and thus do not reveal trade intentions.

Some electronic traders try to discover hidden orders by pinging the market: They submit a small IOC limit order for only a few shares at the price at which they are looking for hidden orders. If the pinging order trades, they know that a hidden order is present at that price; however, they do not know the full size of the order (which they can discover only by trading with it). Traders then may use this information to adjust their trading strategies.

All traders who subscribe to a complete trade feed that includes odd-lot transactions (substandard transaction sizes) can see the results of a ping that discovers liquidity. At almost all exchanges, however, only the pinger will know on which side of the market the hidden liquidity lies. Nonetheless, the information produced by someone else's successful ping can be useful to various traders. It indicates that someone in the market is concerned enough about liquidity conditions that pinging is worthwhile and that hidden liquidity is available on one side of the market.

Leapfrog. When bid–ask spreads are wide, dealers often are willing to trade at better prices than they quote. They quote wide spreads because they hope to trade at more favorable prices. When another trader quotes a better price, dealers often immediately quote an even better price. For example, if the market is 20 bid, offered at 28, and a buy-side trader bids at 21, a dealer might instantly bid at 22. (The improved price might also come from a quote matcher.) This behavior frustrates buy-side traders, who then must quote a better price to maintain order precedence. If the spread is sufficiently wide, a game of leapfrog may ensue as the dealer jumps ahead again.

Flickering quotes. Electronic markets often have flickering quotes, which are exposed limit orders that electronic traders submit and then cancel shortly thereafter, often within a second. Electronic dealers and algorithmic buy-side traders submit and repeatedly cancel and resubmit their orders when they do not want their orders to stand in the market; rather, they want other traders to see that they are willing to trade at the displayed price. Traders who wish to trade with a flickering quote can place a hidden limit order at the price where the quote is flickering. If the flickering order returns, it will hit their hidden limit order, and then they will trade with it.

Electronic arbitrage. Electronic arbitrageurs use electronic trading systems to implement three types of arbitrage trading strategies:

1 **Take liquidity on both sides**. The costliest and least risky arbitrage trading strategy involves using marketable orders to fill both legs, or positions (i.e., buying an undervalued instrument and selling a similar overvalued instrument), of the arbitrage portfolio. This strategy is profitable only if the arbitrage spread is sufficiently large, but competition among arbitrageurs ensures that such large arbitrage spreads are quite rare. Arbitrageurs can seldom simultaneously take liquidity in two markets for identical instruments and make a profit. To effectively execute this strategy, arbitrageurs must use very fast trading systems so that they can lock in the arbitrage spread before prices in one or both markets change.

2 **Offer liquidity on one side**. In this strategy, arbitrageurs offer liquidity in one or both markets in which they trade. When they obtain a fill in one market, they immediately take liquidity in the other market to complete the construction of their arbitrage portfolio. This strategy produces lower-cost executions, but it is a bit riskier than the first strategy.

   For example, suppose that Markets A and B are both quoting 20 bid, offered at 21 for the same instrument. An arbitrageur may place a bid at 19 in Market A with the hope that a large seller will come along who takes all liquidity at 20 (i.e., fills all bids at 20) in Market A and then proceeds to fill the arbitrageur's order at 19. If so, the arbitrageur will immediately try to sell to the 20 bid in Market B. If the arbitrageur is quick enough, he may be able to fill his order before the bidder at 20 in Market B cancels that bid and before any other trader—particularly the large trader—takes it. If successful, the arbitrageur realizes a profit of 1. Of course, the arbitrageur will immediately cancel his 19 bid in Market A if the 20 bid in Market B disappears.

3 **Offer liquidity on both sides.** The final arbitrage strategy involves offering liquidity in both markets. In this strategy, after the first order to execute fills, the arbitrageur continues to offer liquidity to complete the second trade. This strategy is the riskiest strategy because arbitrageurs are exposed to substantial price risk when one leg is filled and the other is not. Moreover, if prices are moving because well-informed traders are on the same side in both markets—as they might be if the well-informed traders possess information about common risk factors—the leg providing liquidity to the informed traders will fill quickly, whereas the other leg probably will not fill.

   Arbitrageurs using this strategy trade much like dealers—switching from offering (supplying) liquidity to taking (demanding) liquidity when they believe that offering liquidity may be too risky. They may also often cancel and resubmit their orders when market conditions change. Thus, they are most effective when they use fast trading systems.

   When the arbitrage spread reverts, as the arbitrageurs expect, the arbitrageurs will reverse their trades, often using the same strategy they used to acquire their arbitrage portfolios. Of course, if the spread never reverts, arbitrageurs will lose regardless of how they trade. They will lose less, however, if they can trade their arbitrage portfolio by offering liquidity in one or both legs.

Machine learning. Machine learning, also known as data mining, uses advanced statistical methods to characterize data structures, particularly relations among variables. These methods include neural nets, genetic algorithms, classifiers, and other methods designed to explain variables of interest using sparse data or data for which the number of potential explanatory variables far exceeds the number of observations.

Machine-learning methods produce models based on observed empirical regularities rather than on theoretical principles identified by analysts. These methods can be powerful when stable processes generate vast amounts of data, such as occurs in active financial markets.

Many trading problems are ideally suited for machine-learning analyses because the problems repeat regularly and often. For such problems, machine-based learning systems can be extraordinarily powerful.

However, these systems are often useless—or worse—when trading becomes extraordinary (e.g., when volatilities shoot up). Machine-learning systems frequently do not produce useful information during volatility episodes because these episodes have few precedents from which the machines can learn. Thus, traders often instruct their electronic trading systems to stop trading—and sometimes to close out their positions—whenever they recognize that they are entering uncharted territory. Many traders shut down when volatility spikes, both because high-volatility episodes are uncommon and thus not well understood and because even if such episodes were well understood, they represent periods of exceptionally high risk.

6 ELECTRONIC TRADING RISKS

i. describe the risks associated with electronic trading and how regulators mitigate them;

The advent of electronic trading affected securities markets in many ways. Investors now benefit from greater trade process efficiencies and reduced transaction costs, but electronic trading also creates new systemic risks for market participants.

6.1 The HFT Arms Race

The competition among high-frequency traders (HFTs) has created an "arms race" in which each trader tries to be faster than the next. Consequently, the state-of-the-art, high-frequency trading technologies necessary to compete successfully are now very expensive, making entry quite costly. These costs form barriers to entry that can create natural monopolies. Although substantial evidence suggests that electronic trading benefits the markets, these benefits may erode if only a few HFTs survive and can exploit their unique positions. Already, many HFTs are quitting the markets because they cannot compete effectively.

More generally, many commentators have observed that most of the costly technologies that high-frequency traders acquire do little to promote better or more-liquid markets. HFTs primarily incur these costs so they can beat their competitors. The utilitarian traders who demand liquidity ultimately pay these costs. Concerns about the costs of the HFT arms race have led to calls for changes in market structure that would diminish the advantages of being faster. Some commentators suggest that markets be slowed by running call markets once a second or more often instead of trading continuously. Others suggest that the order processing be delayed by random intervals to reduce the benefits of being fast and thus the incentives to invest in speed.

6.2 Systemic Risks of Electronic Trading

Electronic trading created new systemic risks that concern regulators and practitioners. A systemic risk is a risk that some failure will hurt more than just the entity responsible for the failure. Systemic risks are particularly problematic when the responsible entity is not required or is unable to compensate others for the costs its failure imposes on them. When people do not bear the full costs of their behaviors, they tend not to be as careful in avoiding damaging behaviors as they otherwise would be.

Systemic risks associated with fast trading may be caused by electronic exchange trading system failures or excessive orders submitted by electronic traders. Electronic exchange trading system failures occur when programmers make mistakes, exchange servers have insufficient capacity to handle traffic, or computer hardware or communication lines fail.

The 18 May 2012 Facebook IPO at NASDAQ is an example of a trading system failure caused by a programming error that unexpectedly high demands on capacity revealed. In this case, two software processes locked into an infinite loop as they took turns responding to each other.

Examples of systemic risks caused by excessive orders submitted by electronic traders include the following:

- *Runaway algorithms* produce streams of unintended orders that result from programming mistakes. The problems sometimes occur when programmers do not anticipate some contingency. The Knight Capital trading failure on 1 August 2012 may be the most extreme example of a runaway algorithm incident. Owing to a software programming mistake, Knight sent millions of orders to the markets over a 45-minute period when it intended only to fill 212 orders, some of which normally might have been broken up but none of which would have generated so many orders. These orders produced 4 million executions involving 397 stocks. Knight lost $400 million in the incident.
- *Fat finger errors* occur when a manual trader submits a larger order than intended. They are called fat finger errors because they sometimes occur when a trader hits the wrong key or hits a key more often than intended. These types of errors are not unique to electronic trading systems, but their consequences are often greater in electronic systems because of the speed at which they operate and because clerks often catch these errors in manual trading systems before they cause problems.
- *Overlarge orders* demand more liquidity than the market can provide. In these events, a trader—often inexperienced—will try to execute a marketable order that is too large for the market to handle without severely disrupting prices in the time given to fill the order. The 6 May 2010 Flash Crash occurred as a result of such an order. The crash was triggered when a large institutional trader tried to sell $4.1 billion in E-mini S&P 500 futures contracts using an algorithm over a short period. The algorithm was designed to participate in a fixed fraction of the market volume. When the initial trades depressed S&P 500 futures prices, trading volumes increased substantially as arbitrageurs and others started to trade. The increase in trading volumes caused the algorithm to increase the rate of its order submissions, which exacerbated the problem. The market reverted to its former levels after the Chicago Mercantile Exchange briefly halted trading in the E-mini S&P 500 futures contract, and the large order eventually was filled.
- *Malevolent order streams* are created deliberately to disrupt the markets. The perpetrators may be market manipulators; aggrieved employees, such as traders or software engineers; or terrorists. Traders conducting denial-of-service attacks designed to overwhelm their competitors' electronic trading systems with excessive quotes also may create malevolent order streams.

The solutions to the systemic risk problems associated with electronic trading systems are multifold:

- Most obviously, traders must test software thoroughly before using it in live trading. Exchanges often conduct mock trading sessions to allow developers to test their software.
- Rigorous market access controls must ensure that only those orders coming from approved sources enter electronic order-matching systems.
- Rigorous access controls on software developers must ensure that only authorized developers can change software. Best practice mandates that these controls also include the requirement that all software be read, understood, and vouched for by at least one developer besides its author.

- The electronic traders who generate orders and the electronic exchanges that receive orders must surveil their order flow in real time to ensure that it conforms to preset parameters that characterize its expected volume, size, and other characteristics. When the order flow is different than expected, automatic controls must shut it off immediately.
- Brokers must surveil all client orders that clients introduce into electronic trading systems to ensure that their clients' trading is appropriate. Brokers must not allow their clients to enter orders directly into exchange trading systems—a process called sponsored naked access—because it would allow clients to avoid broker oversight.
- Some exchanges have adopted price limits and trade halts to stop trading when prices move too quickly. These rules stop trading when excess demands for liquidity occur. They also prevent the extreme price changes that can occur in electronic markets when market orders arrive and no liquidity is present. Most brokers now automatically convert market orders into marketable limit orders to ensure that they do not trade at unreasonable prices.

Historical Event: The Flash Crash

The 6 May 2010 Flash Crash was the most notable market structure event in recent memory. During the crash, which started at about 2:42 p.m. ET, the E-mini S&P 500 futures contract dropped approximately 5% in 5 minutes and then recovered nearly fully in the next 10 minutes. The price volatility spilled from the equity futures market into the stock market, where some stocks traded down more than 99% or up more than 1,000%. In the immediate aftermath of the crash, regulators decided that more than 20,000 trades in more than 300 securities that occurred more than 60% away from earlier prices would be broken (canceled).

This extraordinary event raised many concerns about security market structure—in particular, how the adoption of electronic trading may have increased potential systemic risks. This subsection describes the events that led up to the crash, what happened during the crash, and the regulatory responses to the crash.

The Event and Its Causes

On Thursday, 6 May 2010, the stock market traded down throughout the day at an accelerating rate. By 2:30 p.m., it had lost about 4% from its previous close. Contemporaneous commentators attributed the fall to concerns about Greek sovereign debt and the implications of a Greek default for other markets. During the day, many traders who had been providing liquidity to the market were accumulating substantial long positions as people demanded to sell. As the day wore on, their willingness to continue to accumulate additional inventory decreased. Moreover, day traders, who do not normally carry inventory overnight, also were considering how and when they would sell their losing positions.

Presumably, in response to the European concerns and perhaps other concerns, portfolio managers at Waddell & Reed Financial Inc. (W&R) decided to reduce US equity exposure in their $27 billion Asset Strategy Fund by selling 75,000 June 2010 E-mini S&P 500 futures contracts with a nominal value of approximately $4.1 billion. They gave this order to their buy-side trader, who proceeded to fill it using an algorithm that split the order into small pieces for execution. Although the order was the largest single order submitted to the E-mini futures market that year, it was not without precedent. Two earlier orders in the previous year were of similar size or larger, one of which had been submitted by W&R. Those orders had been filled in more stable markets and over longer periods of time than W&R's 6 May order. The order started to execute at 2:32 p.m.

W&R's head trader, who normally would have handled such a large order, was out of the office that day. Instead, a less-senior trader in his office handled the order.

The trader set parameters on the algorithm to target an execution rate of 9% of the trading volume calculated over the previous minute without regard to price or time. This trading strategy was more aggressive than the one W&R had used to fill its large order from the previous year. The trader probably set an aggressive rate because he feared that the firm would obtain a worse execution if prices continued to fall. The more aggressive strategy contributed to the crash.

When the initial trades depressed S&P 500 futures prices, trading volumes increased substantially as arbitrageurs and others started to trade, many of them trading with each other as they normally did. The arbitrageurs bought the futures and sold equities and equity ETFs (exchange-traded funds), such as the SPDR S&P 500 Trust (ticker SPY). Some arbitrageurs also sold call option contracts and bought put option contracts. The increase in trading volumes caused the algorithm to increase the rate of its order submissions as it tried to keep up with its mandate to participate in 9% of the market volume. The increasing order submission rate exacerbated the problem.

Initially, high-frequency traders and other liquidity suppliers in the E-mini futures markets supplied liquidity to W&R's order and accumulated long positions. Between 2:41 p.m. and 2:44 p.m., these short-term traders sold these positions as the algorithm continued to pump more orders into the market. During this 4-minute period, the E-mini dropped 3%. By the end of this period, buy-side depth (total size of standing buy orders) in the E-mini contract dropped to only 1% of the average depth observed earlier in the day. The E-mini contract then dropped 1.7% in the next 15 seconds.

The arbitrage trades caused the equity markets to drop. In many securities— especially the ETFs—falling prices triggered stock loss market orders, which further depressed prices. The levered ETFs were particularly affected because their high volatilities make them popular with technical traders and retail traders, many of whom routinely place stop orders to protect their positions.

As the prices changed quickly, many traders who were providing liquidity in the futures and equity markets dropped out because they were unwilling to trade in the face of such extreme volatility. Many also had already accumulated large inventory positions from earlier in the day and did not want to buy more. Interestingly, researchers later discovered that the largest and most active high-frequency trading firms did not withdraw. Nonetheless, limit order books thinned out—especially on the buy side—as traders canceled standing orders and as sellers filled those buy orders still standing.

In some stocks, all standing buy orders were exhausted and trading stopped. In other stocks, all buy orders except those placed with a limit price of only a cent or two were exhausted. In these stocks, exchange trading systems blindly filled market sell orders at extraordinarily low prices. In a few other stocks, the withdrawal of liquidity suppliers from the market also removed essentially all liquidity from the sell side of the market. Some stocks then traded at prices as high as $100,000 when market buy orders were filled against sell orders placed at extraordinarily high prices.

The slide stopped at 2:45:28 p.m. when a Chicago Mercantile Exchange trading rule called Stop Logic Functionality caused the exchange's computers to halt trading briefly in the E-mini S&P 500 futures contract and to clear the limit order book of all standing limit orders. The rule is triggered when it becomes apparent that pending order executions would cause prices to jump too far. The futures contract dropped about 5% from when the algorithm started to trade at 2:32 p.m. to the market halt at 2:45 p.m. The algorithm sold about 35,000 contracts during this period.

When trading resumed 5 seconds later, the buy-side algorithm continued to trade, but many liquidity suppliers were now willing to provide liquidity. Prices rose quickly in orderly markets.

The episode largely ended when the big W&R order completed filling at around 2:51 p.m., about 20 minutes after it started. However, the market remained quite volatile during the remainder of the day as traders adjusted their positions and responded to the extreme volatility.

Following the crash, regulators broke all trades that had occurred more than 60% away from the previous close.

Implications for Traders

The Flash Crash provided three important lessons for observant traders:

- First, market orders are incompatible with electronic order-matching systems that do not curb trading when prices move too quickly. Had traders priced all their orders, no trades would have taken place at unreasonably high or low prices. Following the crash, many retail brokers adopted a policy of converting all customer market orders into marketable limit orders with limit prices set about 10% above the current ask for buy orders and 10% below the current bid for sell orders.
- Second, institutional traders using algorithms must be careful not to demand more liquidity than orderly markets can provide. Most buy-side investors probably immediately recognized that W&R lost a substantial amount of its clients' money owing to the extraordinarily high transaction costs associated with the trade. To obtain a crude estimate of this loss, assume that the algorithm traded all \$4.1 billion of its order at a uniform rate throughout the 5% price reversal. The average market impact of the trade would have been 2.5%, which implies total transaction costs of about \$100 million, or 0.37% of the \$27 billion in assets of the W&R Asset Strategy Fund. Such significant losses attract attention. Within a week, many algorithm writers probably coded limits into their algorithms to help prevent them from being used irresponsibly.
- Finally, algorithm writers and the traders who use algorithms must pay much more attention to the dangers of using algorithms that can create destructive feedback loops. They particularly must understand how algorithms respond to market conditions that they may create themselves.

Regulatory Responses

Following the Flash Crash, regulators adopted new rules to prevent a similar crash from happening again. They placed curbs that halt trades in a stock for 5 minutes if prices move up or down by more than 10% for large stocks and 20% for smaller stocks. This rule ensures that prices cannot move too quickly, but it does not prevent traders from behaving foolishly. Had it been in effect during the Flash Crash, the rule would have stopped trades from occurring at ridiculously low or high prices, but it would not have stopped the W&R trader from submitting an unrealistically aggressive order.

Regulators also adopted rules to establish when and which trades will be broken in the event of another extreme price change. Such rules should help ensure that liquidity suppliers who are afraid that their trades may be broken do not withdraw from the market prematurely.

EXAMPLE 4

Electronic Trading and Transaction Costs

Describe the impact of electronic trading on transaction costs.

Solution:

Growth in electronic trading has resulted in greater trade process efficiencies and reduced transaction costs for investors. Electronic systems are much cheaper to operate than floor-based systems (requiring less physical space and fewer exchange personnel). These systems can operate on a close-to-continuous basis at far greater scale and scope and at much faster speeds than humans. Process efficiencies from electronic trading have led to significant decreases in bid–ask spreads, which have lowered transaction costs for investors.

DETECTING ABUSIVE TRADING PRACTICES

7

j describe abusive trading practices that real-time surveillance of markets may detect.

Regulators around the world recognize that real-time market monitoring and surveillance systems allow faster responses to potential crises and market abuses with the potential for rapid intervention to prevent or minimize damages. Many trading venues have long used real-time surveillance technologies, but their use is not consistent across all markets. The goal of real-time market surveillance is to detect potential market abuse while it is happening. Real-time surveillance often can detect the following damaging behaviors:

Front running. Front running involves buying in front of anticipated purchases and selling in front of anticipated sales. In most jurisdictions, front running is illegal if the front runners acquire their information about orders improperly—for example, by a tip from a broker handling a large order.

Some traders use electronic artificial intelligence systems to identify when traders are filling large orders over time by breaking them up into small pieces. When these traders suspect that buyers or sellers are working large orders, they will trade ahead on the same side with the hope of benefiting when the large traders move prices as they fill their orders. This front-running strategy is legal if the information on which it is based is properly obtained— for example, by watching a market data feed.

Front running increases transaction costs for the traders whose orders are front run because the front runners take liquidity that the front-run traders otherwise would have taken for themselves.

Market manipulation. In general, market manipulation consists of any trading strategy whose purpose is to produce misleading or false market prices, quotes, or fundamental information to profit from distorting the normal operation of markets. Market manipulators are parasitic traders who attempt to fool or force others into making disadvantageous trades. Many market manipulation strategies exist—including bluffing, squeezing, cornering, and gunning.

In most jurisdictions, market manipulation strategies are illegal. Enforcement is often difficult, however, because the exact infractions can be hard to define and because prosecutors generally must prove scienter (a legal term meaning intent or knowledge of wrongdoing), which can be difficult when defendants suggest alternative explanations for their behavior.

Market manipulation strategies usually involve one or more of the following improper market activities:

- *Trading for market impact* involves trading to raise or lower prices deliberately. A market manipulator often is willing to incur substantial transaction costs to raise or lower the price of a security to influence other traders' perceptions of value.
- R*umormongering* is the dissemination of false information about fundamental values or about other traders' trading intentions to alter investors' value assessments. Financial analysts must be careful to ensure that they base their analyses on valid information and not on false information designed to fool them into making poor decisions. Note that although rumormongering is illegal in most jurisdictions, simply reporting one side of an issue is not illegal. Financial analysts, therefore, must also be careful to ensure that they base their analyses on balanced information and not on information that is true but selectively presented to them with the purpose of distorting their analyses.

- *Wash trading* consists of trades arranged among commonly controlled accounts to create the impression of market activity at a particular price. The purpose of wash trading is to fool investors into believing that a market is more liquid than it truly is and to thereby increase investors' confidence both in their ability to exit positions without substantial cost and in their assessments of security values. Manipulators also can achieve these purposes by falsely reporting trades that never occurred, which is essentially what happens when they arrange trades among commonly controlled accounts.
- *Spoofing*, also known as *layering*, is a trading practice in which traders place exposed standing limit orders to convey an impression to other traders that the market is more liquid than it is or to suggest to other traders that the security is under- or overvalued. For example, suppose that a spoofer wants to buy stock cheaply or quickly. The spoofer might place a hidden buy order in the market. The spoofer then places one or more exposed sell limit orders in the market to convey the impression that prices may soon fall. Seeing the spoofing sell orders, one or more traders may conclude that values may be lower than market prices suggest. On that basis, they may sell into the spoofer's buy order, enabling the spoofer to obtain a quick and possibly cheaper purchase than the spoofer otherwise would have obtained had the spoofer not placed the spoofing sell orders. Of course, immediately following the execution of the buy order, the spoofer will cancel the sell orders.

  Spoofing is risky because the spoofing orders that spoofers submit might execute before their intended orders execute. Spoofers can manage this risk by keeping track of the orders in the limit order book ahead of their spoofing orders. If these orders fill before the spoofers' intended orders fill, spoofers will cancel their spoofing orders to prevent them from executing. To effectively manage these processes, spoofers use electronic systems to monitor trading and to ensure that they can quickly cancel their orders as soon as they no longer want them to stand.

Market manipulators often use these improper market activities singly or in combination when they try to fool or force other traders into trades that will ultimately prove to be disadvantageous to them. Market manipulation strategies include:

- **Bluffing**. Bluffing involves submitting orders and arranging trades to influence other traders' perceptions of value. Bluffers often prey on momentum traders, who buy when prices are rising and sell when prices are falling. For example, consider typical "pump-and-dump" schemes in which bluffers buy stock to raise its price and thereby encourage momentum traders to buy. The bluffers then sell the stock to the momentum traders at higher prices. To further the scheme, bluffers may engage in such activities as rumormongering or wash trading. Note also that bluffers may time their purchases to immediately follow the release of valid positive information about the security and thereby fool traders into overvaluing the material significance of the new information.

  In a pump-and-dump manipulation, the bluffer tries to raise prices. Similar manipulations can occur on the short side, though they are less common. In such manipulations, manipulators take short positions and then try to repurchase shares at lower prices. These manipulations are often called "short and distorts."

  To avoid falling into these traps, financial analysts must ensure that they base their analyses on independent assessments of value. Their analyses must have a proper foundation as required by Standard V(A): Diligence and Reasonable Basis, of the CFA Institute Code of Ethics and Standards of Professional Conduct.

- **Gunning the market**. Gunning the market is a strategy used by market manipulators to force traders to do disadvantageous trades. A manipulator generally guns the market by selling quickly to push prices down with the hope of triggering stop-loss sell orders. A stop-loss (or stop) sell order becomes valid for execution once the specified stop price condition is met by a trade occurring at or below the stop price. For example, suppose that a market manipulator believes that traders have placed many stop-loss sell orders at 50. These sell orders would become valid upon a trade occurring at 50 or below. The manipulator may sell aggressively to push prices down from 51 to 50 and thereby trigger the stop-loss sell orders. The manipulator then may be able to profit by repurchasing at lower prices.
- **Squeezing and cornering**. Squeezing, cornering, and gunning the market are all schemes that market manipulators use to force traders to do disadvantageous trades. In a squeeze or corner, the manipulator obtains control over resources necessary to settle trading contracts. The manipulator then unexpectedly withdraws those resources from the market, which causes traders to default on their contracts, some of which the manipulator may hold. The manipulator profits by providing the resources at high prices or by closing the contracts at exceptionally high prices.

  For example, in short squeezes, manipulators obtain control of a substantial fraction of all available lendable stock shares or bonds. If the securities are overvalued, as they might be if the manipulators are also engaging in a pump and dump, many speculators may be short selling the securities by unknowingly borrowing them from the manipulators. The manipulators then will recall the security loans. If the short sellers ("shorts") cannot borrow the securities from others, they will be forced to buy securities in the market to cover their stock loans. Their purchases will raise prices and allow the manipulators to sell their securities at overvalued prices. Manipulators also may profit by raising the rates they charge to lend their securities. To avoid being caught in a short squeeze, short sellers must be sure that the market for lendable securities has many participants and is not concentrated in the hands of one or more entities acting in concert.

  In commodity market corners, manipulators buy many futures contracts while simultaneously buying in the spot markets much of the deliverable supply of the commodity. When the contract approaches expiration, the manipulators then demand delivery from the shorts, most of whom will not own the deliverable commodity. The shorts then must buy the deliverable supply from the manipulators at exceptionally high prices. Alternatively, they may repurchase their contracts from the manipulators, again at very high prices.

  Corners can occur in commodity markets because most participants in commodity futures contracts do not demand to receive or make delivery when the contract expires. Instead, they close their positions by arranging offsetting trades in the futures market, either because they are simultaneously accepting or making delivery elsewhere or because they are rolling their positions into future contract months. Accordingly, most short sellers neither expect nor intend to make delivery. When forced to make delivery, they are caught short.

  Corners are illegal in most jurisdictions, and they always violate the rules of the exchanges on which futures contracts trade. In general, long holders cannot demand delivery if they do not have a valid business reason for doing so. However, enforcement is complicated by the fact that manipulators may offer plausible reasons for requesting unexpected deliveries. Note also that sometimes, unexpected supply shortages coupled with unexpected legitimate demands for delivery can result in inadvertent short squeezes. Thus, short sellers who do not intend to make delivery should try to close their positions early to ensure that they are not caught in an intentional corner or an inadvertent squeeze.

SUMMARY

This reading explains the implicit and explicit costs of trading as well as widely used methods for estimating transaction costs. The reading also describes developments in electronic trading, the main types of electronic traders, their needs for speed and ways in which they trade. Electronic trading benefits investors through lower transaction costs and greater efficiencies but also introduces systemic risks and the need to closely monitor markets for abusive trading practices. Appropriate market governance and regulatory policies will help reduce the likelihood of events such as the 2010 Flash Crash. The reading's main points include:

- Dealers provide liquidity to buyers and sellers when they take the other side of a trade if no other willing traders are present.
- The bid–ask spread is the difference between the bid and the ask prices. The effective spread is two times the difference between the trade price and the midquote price before the trade occurred. The effective spread is a poor estimate of actual transaction costs when large orders have been filled in many parts over time or when small orders receive price improvement.
- Transaction costs include explicit costs and implicit costs. Explicit costs are the direct costs of trading. They include broker commissions, transaction taxes, stamp duties, and exchange fees. Implicit costs include indirect costs, such as the impact of the trade on the price received. The bid–ask spread, market impact, delay, and unfilled trades all contribute to implicit trading costs.
- The implementation shortfall method measures the total cost of implementing an investment decision by capturing all explicit and implicit trading costs. It includes the market impact costs, delay costs, as well as opportunity costs.
- The VWAP method of estimating transaction costs compares average fill prices to average market prices during a period surrounding the trade. It tends to produce lower transaction cost estimates than does implementation shortfall because it often does not measure the market impact of an order well.
- Markets have become increasingly fragmented as venues trading the same instruments have proliferated. Trading in any given instrument now occurs in multiple venues.
- The advantages of electronic trading systems include cost and operational efficiencies, lack of human bias, extraordinarily fast speed, and infinite span and scope of attention.
- Latency is the elapsed time between the occurrence of an event and a subsequent action that depends on that event. Traders use fast communication systems and fast computer systems to minimize latency to execute their strategies faster than others.
- Hidden orders, quote leapfrogging, flickering quotes, and the use of machine learning to support trading strategies commonly are found in electronic markets.
- Traders commonly use advanced order types, trading tactics, and algorithms in electronic markets.

- Electronic trading has benefited investors through greater trade process efficiencies and reduced transaction costs. At the same time, electronic trading has increased systemic risks.
- Examples of systemic risks posed by electronic traders include: runaway algorithms that produce streams of unintended orders caused by programming mistakes, fat finger errors that occur when a manual trader submits a larger order than intended, overlarge orders that demand more liquidity than the market can provide, and malevolent order streams created deliberately to disrupt the markets.
- Real-time surveillance of markets often can detect order front running and various market manipulation strategies.
- Market manipulators use such improper activities as trading for market impact, rumormongering, wash trading, and spoofing to further their schemes.
- Market manipulation strategies include bluffing, squeezing, cornering, and gunning.

PRACTICE PROBLEMS

The following information relates to Questions 1–10

Brian Johnson is a senior manager at Star Asset Management (SAMN), a large asset management firm in the United States. Tim Martin has just earned his advanced degree in statistics and was hired to support the trading team at SAMN. Martin meets with Johnson to undergo a training relating to SAMN's trading activities.

Johnson begins the training with a review of the limit order book for Light Systems, Inc., which is presented in Exhibit 1. Three dealers make market for the shares of Light Systems. Based on these prices, SAMN's trading desk executes a market sell order for 1,100 shares of Light Systems.

Exhibit 1 Limit Order Book for Light Systems, Inc.

Bid				Ask			
Dealer	Time Entered	Price	Size	Dealer	Time Entered	Price	Size
B	10.10 a.m.	$17.15	900	C	10.11 a.m.	$17.19	1,200
C	10.11 a.m.	$17.14	1,500	B	10.10 a.m.	$17.20	800
A	10.11 a.m.	$17.12	1,100	A	10.12 a.m.	$17.22	1,100

Johnson then discusses a market buy order for 5,000 shares of an illiquid stock. The order was filled in three trades, and details about the three trades are presented in Exhibit 2.

Exhibit 2 Buy Trade Order Details

Trade #	Time	Trade Price	Trade Size	Bid Price	Ask Price
1	9.45 a.m.	$25.20	1,200	$25.17	$25.20
2	9.55 a.m.	$25.22	1,300	$25.19	$25.22
3	11.30 a.m.	$25.27	2,500	$25.22	$25.26

Johnson explains to Martin that the number of venues trading the same instruments has proliferated in recent years, and trading in any given instrument has now been distributed across these multiple venues. As a result, the available liquidity on any one of those exchanges represents just a small portion of the aggregate liquidity for that security. As a result, SAMN has had to adapt its trading strategies, particularly for large trades.

Johnson asks Martin about his views on how the introduction of electronic trading might have impacted SAMN. Martin tells Johnson:

© 2019 CFA Institute. All rights reserved.

Statement 1 Once built, electronic trading systems are more efficient and cheaper to operate than floor-based trading systems.

Statement 2 Electronic trading systems have attracted a lot of new buy-side traders, and the increased competition has resulted in narrower bid–ask spreads.

Statement 3 The introduction of electronic markets has had a much greater impact on the trading of corporate and municipal bonds than on the trading of equities.

Johnson tells Martin that communication speed is SAMN's current highest priority. All of SAMN's competitors have increased their communication speeds in recent months, and Johnson says management wants SAMN to be faster than its competitors. SAMN's trading desk is located in a residential area far from downtown where the exchanges it works with are located. SAMN's trading team is relatively large with experienced investment professionals, and the firm recently invested in fast computers with the latest algorithms.

At the end of the training, Johnson gives Martin his first assignment. The assignment is for Martin to use the vast amount of data that SAMN has collected to design a machine learning (ML) model using advanced statistical methods to characterize data structures and relations. Then he has to build a trading algorithm based on the same model. Since electronic trading has added systemic risk to the market, Johnson asks Martin to suggest ways to minimize the systemic risk introduced by his algorithm. Martin offers two suggestions:

Suggestion 1 Perform extensive testing of the algorithm before its launch.

Suggestion 2 Impose mandatory trading halts if prices change outside a threshold range.

A month into the job, Johnson sends Martin to an investment conference focused on abusive trading practices. Based on what he learned at the conference, Martin recommends to Johnson that SAMN incorporate a new rule that news be validated before a trade triggered by news is executed.

1 Based on Exhibit 1, the inside bid–ask spread for the limit order book for Light Systems is *closest* to:

A $0.04.

B $0.07.

C $0.10.

2 Based on Exhibit 1, the total amount that SAMN will receive, on a per share basis, for executing the market sell order is *closest* to:

A $17.14.

B $17.15.

C $17.22.

3 Based on Exhibit 2, the market impact relating to Trade 2, on a per share basis, is *closest* to:

A $0.02.

B $0.03.

C $0.07.

4 Based on Exhibit 2, the average effective spread of the three trades is *closest to*:

A $0.0333.

B $0.0367.

C $0.0400.

5 The reason for SAMN having to adapt its trading strategies is a result of:
   A latency.
   B market fragmentation.
   C high frequency trading.

6 Which of Martin's statements relating to the introduction of electronic markets is correct?
   A Statement 1
   B Statement 2
   C Statement 3

7 Which of the following changes should SAMN make to address its key priority?
   A Hire more investment professionals
   B Upgrade to more complex operating systems
   C Move the trading desk physically closer to the exchanges it works with

8 The model that Martin is tasked with designing will likely be *most* effective:
   A for testing new markets.
   B in a well-understood market environment.
   C during periods of higher than normal market volatility.

9 Which of Martin's suggestions will *most likely* be effective in limiting the systemic risk introduced by his algorithm?
   A Only Suggestion 1
   B Only Suggestion 2
   C Both Suggestion 1 and Suggestion 2

10 Which market manipulation strategy is *most likely* the target of the new rule suggested by Martin?
   A Rumormongering
   B Gunning the market
   C Trading for market impact

The following information relates to Questions 11–16

Michael Bloomfield is a trader at 2Fast Trading, a proprietary trading company that uses machine learning and algorithms to execute trades. He works with Amy Riley, a junior trader at the company. Bloomfield and Riley meet to review the company's trading systems and several trades in Bloomfield's trading account.

They discuss the increasing impact of market fragmentation on available liquidity for the company's trading strategies. Riley makes the following comments regarding market fragmentation:

Comment 1 Liquidity aggregation and smart order routing help traders manage the challenges and opportunities presented by fragmentation.

Comment 2 With increasing market fragmentation, traders who fill large orders now search for liquidity across multiple venues and across time to control market impact.

Bloomfield tells Riley that he noticed trades of 500 shares of BYYP stock were executed every 20 minutes for an hour. Bloomfield saw the same pattern of trading in the stock during the previous trading day. He instructs Riley to submit an order to purchase BYYP shares on the assumption that a trader seeks liquidity and is executing a large buy order by breaking it into pieces. The prices of these trades and the best bids and offers in the market when the BYYP trades occurred are presented in Exhibit 1.

Exhibit 1 BYYP Trade Details

Trade	Trade Price	Prevailing Bid	Prevailing Offer
1	41.50	41.45	41.50
2	41.75	41.73	41.75

Bloomfield shifts the conversation to AXZ Corp. Bloomfield notes that AXZ's bid–ask spread is narrow, even though AXZ's share price has been experiencing a period of high volatility. After extensive research, Bloomfield will purchase AXZ shares using a trading strategy that does not include standing orders.

Bloomfield then assesses the risks that 2Fast's electronic trading strategies introduce into the market. He is concerned that these risks may bring on more regulation. Bloomfield claims that the risks can be reduced by changing the structure of the market, and those structural changes can maintain 2Fast's primary competitive advantage, which is trading faster than competitors.

Bloomfield mentions that a regulatory body is investigating a competitor's trading practices. The investigation involves a tip that the competitor is manipulating markets by submitting orders and arranging trades to influence other traders' perceptions of value. Specifically, regulators were informed that the competitor has been buying stock to raise its price, thereby encouraging momentum traders to buy, and then selling the stock to them at higher prices. The regulator confirmed that the competitor did not use standing limit orders or commonly controlled accounts for the trades under investigation.

11 Which of Riley's comments related to market fragmentation is accurate?

- **A** Only Comment 1
- **B** Only Comment 2
- **C** Both Comment 1 and Comment 2

12 Bloomfield's strategy to purchase BYYP shares is *best* classified as electronic:

- **A** arbitrage.
- **B** front running.
- **C** quote matching.

13 Based on Exhibit 1, the average effective spread of the BYYP trades is *closest* to:

- **A** \$0.018.
- **B** \$0.035.
- **C** \$0.070.

14 Bloomfield's trading strategy for the purchase of AXZ shares *most likely* includes the use of:

- **A** flickering quotes.
- **B** machine learning.
- **C** leapfrogging quotes.

15 Which structural change for the market associated with electronic trading systems is *most* consistent with Bloomfield's claim?

A Delaying order processing by random intervals

B Exchanges using trade halts when prices move too quickly

C Slowing markets by running call markets once a second or more often instead of trading continuously

16 The competitor company's trading is *best* described as:

A bluffing.

B spoofing.

C wash trading.

SOLUTIONS

1 A is correct. The inside bid–ask spread, or market bid–ask spread, is the difference between the highest bid price and the lowest ask price. The highest bid price for Light Systems is \$17.15, and the lowest ask price is \$17.19. Therefore, the inside bid–ask spread = \$17.19 – \$17.15 = \$0.04.

2 B is correct. SAMN's trading desk executes a market sell order for 1,100 shares. Based on the limit order book, the trader would first sell 900 shares at \$17.15 (highest bid, Dealer B) and then sell the remaining 200 shares at \$17.14 (second highest bid, Dealer C). Therefore, the approximate price per share received by SAMN for selling the 1,100 shares is equal to [(900 × \$17.15) + (200 × \$17.14)] / 1,100 = \$17.1482 per share (\$17.15 rounded).

3 A is correct. Market impact, or price impact, is the effect of a trade on transaction prices. After the first trade (Trade 1) was executed at \$25.20, Trade 2 was executed at \$25.22, which is \$0.02 per share higher than the trade price of Trade 1. So, the execution of Trade 1 led to a price impact of \$0.02 per share on Trade 2.

4 C is correct. The effective bid–ask spread for buy orders is calculated as:

Effective bid–ask spread (buy order) = 2 × {Trade price – [(Ask price + Bid price) / 2)]} or
= 2 × (Trade price – Midpoint of the market at the time an order is entered).

So, the effective bid–ask spreads for the three buy trades are calculated as:

Effective spread of Trade 1 = 2 × {\$25.20 – [(\$25.20 + \$25.17)/2]} = \$0.0300.

Effective spread of Trade 2 = 2 × {\$25.22 – [(\$25.22 + 25.19)/2]} = \$0.0300.

Effective spread of Trade 3 = 2 × {\$25.27 – [(\$25.26 + \$25.22)/2]} = \$0.0600.

The resulting average effective spread is then calculated as:

Average effective spread = (Effective spread of Trade 1 + Effective spread of Trade 2 + Effective spread of Trade 3)/3.

Average effective spread = (\$0.0300 + \$0.0300 + \$0.0600)/3 = \$0.0400.

5 B is correct. According to Johnson, markets have become increasingly fragmented as the number of venues trading the same instruments has proliferated and trading in any given instrument has been split (or fragmented) across these multiple venues. As a result, the available liquidity on any one exchange represents just a small portion of the aggregate liquidity for that instrument. This phenomenon is known as market fragmentation and creates the potential for price and liquidity disparities across venues. As a result, SAMN has had to adapt its trading strategies to this fragmented liquidity to avoid intensifying the market impact of a large trade.

6 A is correct. Once built, electronic systems are indeed cheaper to operate than floor-based trading systems. They require less physical space than do trading floors, and in contrast to floor-based trading systems, they do not require exchange officials to record and report prices. Furthermore, the widespread use of electronic trading systems significantly decreased trading costs for buy-side traders. Costs fell as exchanges obtained greater cost efficiencies from using

electronic matching systems instead of floor-based manual trading systems. These technologies also decreased costs and increased efficiencies for the dealers and arbitrageurs who provide much of the liquidity offered at exchanges. Competition forced them to pass along much of the benefits of their new technologies to buy-side traders in the form of narrower spreads quoted for larger sizes. New electronic buy-side order management systems also decreased buy-side trading costs by allowing a smaller number of buy-side traders to process more orders and to process them more efficiently than manual traders.

While electronic trading has had a significant effect on equity markets, it has not had as much of an effect on the markets for corporate and municipal bonds. The market structures of corporate and municipal bond markets have hardly changed since the late 19th century. Despite the efforts of many creative developers of electronic bond trading systems, most public investors in these markets still trade largely over the counter with dealers.

7 C is correct. The speed required by electronic traders is affected by fast communication and fast computations. The shorter the distance between the trader and the exchange, the faster the communication. Many exchanges allow electronic traders to place their servers in the rooms where the exchange servers operate, a practice called collocation.

8 B is correct. Many trading problems are ideally suited for machine learning analyses because the problems repeat regularly and often. For such problems, machine-based learning systems can be extraordinarily powerful. However, these systems are often useless—or worse—when trading becomes extraordinary, as when volatilities shoot up. Machine learning systems frequently do not produce useful information during volatility episodes because they have few precedents from which the machines can learn. Thus, traders often instruct their electronic trading systems to stop trading—and sometimes to close out their positions—whenever they recognize that they are entering uncharted territory. Many traders shut down when volatility spikes—both because high-volatility episodes are uncommon and thus not well understood and because even if such episodes were well understood, they represent periods of exceptionally high risk.

9 C is correct. Both suggestions will likely be effective in minimizing the systemic risk introduced by electronic trading. First, exhaustive testing of the algorithm prior to its launch can minimize risk relating to programming errors, which could result in an extreme market reaction that could trigger an even more extreme market reaction. Second, imposing mandatory trade halts in case of large price changes (outside a given threshold) would limit potential undesired results and help minimize systemic risk.

10 A is correct. Rumormongering is the dissemination of false information about fundamental values or about other traders' trading intentions in an attempt to alter investors' value assessments. Martin's suggested news validation rule would reduce the likelihood that SAMN would be adversely affected by this market manipulation strategy.

11 C is correct. Both of Riley's comments are correct. Electronic algorithmic trading techniques, such as liquidity aggregation and smart order routing, help traders manage the challenges and opportunities presented by fragmentation. Liquidity aggregators create "super books" that present liquidity across markets for a given instrument. These tools offer global views of market depth (available liquidity) for each instrument regardless of the trading venue that offers the liquidity. Smart order-routing algorithms send orders to the markets that display the best quoted prices and sizes. Additionally, with increasing market

fragmentation, traders filling large orders adapt their trading strategies to search for liquidity across multiple venues and across time to control the market impacts of their trades.

12 B is correct. Bloomfield noticed a pattern of trading in BYYP and decided to front run shares on the assumption that a trader is in the market filling a large buy order by breaking it into pieces. Electronic front runners trade in front of traders who demand liquidity. They identify when large traders or many small traders are trying to fill orders on the same side of the market. The order anticipation strategies of electronic front runners try to identify predictable patterns in order submission. They may search for patterns in order submissions, trades, or the relations between trades and other events.

A is incorrect because electronic arbitrageurs look across markets for arbitrage opportunities in which they can buy an undervalued instrument and sell a similar overvalued one. His decision to purchase BYYP shares is based on the pattern of trading that Bloomfield observed.

C is incorrect because quote matchers trade in front of traders who supply (not demand) liquidity. Bloomfield decides to purchase BYYP shares on the assumption that a trader is in the market seeking (not supplying) liquidity, which is consistent with front running (not quote matching). Quote matchers trade in front of traders who supply liquidity and try to exploit the option values of standing orders. Quote matchers buy when they believe they can rely on standing buy orders to get out of their positions, and they sell when they can do the same with standing sell orders.

13 B is correct. The effective spread is calculated as follows:

$$\text{Effective spread} = 2 \times \left(\text{Trade price} - \text{Midpoint of market at time of order entry}\right)$$

$$\text{Effective spread of Trade 1} = 2 \times \left(\$41.50 - \$41.475\right) = \$0.05$$

$$\text{Effective spread of Trade 2} = 2 \times \left(\$41.75 - \$41.74\right) = \$0.02$$

$$\text{Average Effective Spread} = \left(\$0.05 + \$0.02\right) / 2 = \$0.035$$

14 A is correct. Flickering quotes are exposed limit orders that electronic traders submit and then cancel shortly thereafter, often within a second. Electronic dealers and algorithmic buy-side traders submit and repeatedly cancel and resubmit their orders when they do not want their orders to stand in the market; rather, they want other traders to see that they are willing to trade at the displayed price. Bloomfield does not want his orders to stand in the market; using flickering quotes to purchase AXZ shares would satisfy that objective.

B is incorrect because AXZ shares are currently in a period of high volatility, so Bloomfield would not likely use machine learning to execute his trades. Machine-learning systems frequently do not produce useful information during volatility episodes because these episodes have few precedents from which the machines can learn. Machine-learning methods produce models based on observed empirical regularities rather than on theoretical principles identified by analysts. Many traders shut down when volatility spikes, both because high-volatility episodes are uncommon and thus not well understood and because even if such episodes were well understood, they represent periods of exceptionally high risk.

C is incorrect because market participants use leapfrogging quotes when spreads are wide (not narrow), and Bloomfield noted that the bid–ask spread for AXZ shares is narrow. When bid–ask spreads are wide, dealers often are willing to trade at better prices than they quote. They quote wide spreads because they hope to trade at more favorable prices. When another trader quotes a better price, dealers often immediately quote an even better price. If the spread is sufficiently wide, a game of leapfrog may ensue as the dealer jumps ahead again.

15 B is correct. To reduce the systemic risks associated with fast trading, some exchanges have adopted trade halts when prices move too quickly. These rules stop trading when excess demand for liquidity occurs. They also prevent the extreme price changes that can occur in electronic markets when market orders arrive and no liquidity is present. 2Fast Trading's competitive advantage will be maintained despite exchange trading halts because the company will be free to trade faster than its competitors once trading resumes. Therefore, exchanges using trade halts to stop trading is the risk reduction strategy that most likely maintains 2Fast Trading's competitive advantage and is consistent with Bloomfield's claim that risks can be reduced by changing the structure of the market.

A is incorrect because delaying order processing by random intervals reduces the benefits of high-frequency traders being faster than their competitors and investing in speed. Therefore, delaying order processing by random order intervals does not maintain 2Fast Trading's primary competitive advantage, which is trading faster than competitors, because that advantage will be reduced.

C is incorrect because slowing markets by running call markets once a second or more often instead of trading continuously diminishes the benefits of high-frequency traders being faster than their competitors and investing with speed. Therefore, slowing markets once a second or more often instead of trading continuously does not maintain 2Fast Trading's primary competitive advantage, which is trading faster than competitors, because that advantage will be reduced.

16 A is correct. Bluffing involves submitting orders and arranging trades to influence other traders' perceptions of value. Bluffers often prey on momentum traders, who buy when prices are rising and sell when prices are falling. Similarly, Bloomfield mentioned that regulators were informed that 2Fast's competitor has been submitting orders and arranging trades to influence other traders' perceptions of value; regulators were informed the competitor has been buying stock to raise its price, thereby encouraging momentum traders to buy, and then selling the stock to them at higher prices.

B is incorrect because the competitor did not use standing limit orders—those orders that are used in a spoofing strategy—for the trades the regulator is investigating. Spoofing is a trading practice in which traders place exposed standing limit orders to convey an impression to other traders that the market is more liquid than it is or to suggest to other traders that the security is under- or overvalued.

C is incorrect because the competitor did not use commonly controlled accounts—those accounts that are used in a wash trading strategy—for the trades that regulators are investigating. Wash trading consists of trades arranged among commonly controlled accounts to create the impression of market activity at a particular price. The purpose of wash trading is to fool investors into believing that a market is more liquid than it truly is and to thereby increase investors' confidence both in their ability to exit positions without substantial cost and in their assessments of security values.

Ethical and Professional Standards

STUDY SESSIONS

Study Session 17	Ethical and Professional Standards

TOPIC LEVEL LEARNING OUTCOME

The candidate should be able to demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct (Code and Standards), identify violations of the Code and Standards, and recommend appropriate corrective measures.

The topic of ethics is one of fundamental importance to the investment profession. Acting responsibly with high levels of integrity builds trust, upon which the investment profession is built. Behaving ethically, and in the best interest of clients, is critical to the long-term success of the investment profession and those choosing a career within it.

The Code and Standards form the ethical foundation for the CFA Institute self-regulatory program. The *Standards of Practice Handbook* provides practical guidance in the interpretation and implementation of the Code and Standards. Together these seek to hold CFA Institute members[1] and CFA Program candidates to the highest levels of professional behavior.

1 Eligibility and requirements for becoming a member of CFA Institute vary by jurisdiction. Please consult www.cfainstitute.org for further details.

© 2021 CFA Institute. All rights reserved.

ETHICAL AND PROFESSIONAL STANDARDS

STUDY SESSION

17

Ethical and Professional Standards

The readings in this study session address the CFA Institute Code of Ethics and Standards of Professional Conduct (Code and Standards). The Code and Standards provide guidance to help identify and resolve ethical conflicts present in everyday activities in the investment profession. "Guidance" in the *Standards of Practice Handbook* addresses the practical application of the Code and Standards by reviewing the purpose and scope of each Standard, presenting recommended procedures for compliance, and providing examples of the Standard in practice. The study session ends with a number of case applications.

READING ASSIGNMENTS

Reading 45	Code of Ethics and Standards of Professional Conduct *Standards of Practice Handbook*, Eleventh Edition
Reading 46	Guidance for Standards I–VII *Standards of Practice Handbook*, Eleventh Edition
Reading 47	Application of the Code and Standards: Level II

© 2021 CFA Institute. All rights reserved.

READING

45

Code of Ethics and Standards of Professional Conduct

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** describe the six components of the Code of Ethics and the seven Standards of Professional Conduct;
☐	**b.** explain the ethical responsibilities required of CFA Institute members and candidates in the CFA Program by the Code and Standards.

PREFACE

The *Standards of Practice Handbook* (*Handbook*) provides guidance to the people who grapple with real ethical dilemmas in the investment profession on a daily basis; the *Handbook* addresses the professional intersection where theory meets practice and where the concept of ethical behavior crosses from the abstract to the concrete. The *Handbook* is intended for a diverse and global audience: CFA Institute members navigating ambiguous ethical situations; supervisors and direct/indirect reports determining the nature of their responsibilities to each other, to existing and potential clients, and to the broader financial markets; and candidates preparing for the Chartered Financial Analyst (CFA) examinations.

Recent events in the global financial markets have tested the ethical mettle of financial market participants, including CFA Institute members. The standards taught in the CFA Program and by which CFA Institute members and candidates must abide represent timeless ethical principles and professional conduct for all market conditions. Through adherence to these standards, which continue to serve as the model for ethical behavior in the investment professional globally, each market participant does his or her part to improve the integrity and efficient operations of the financial markets.

The *Handbook* provides guidance in understanding the interconnectedness of the aspirational and practical principles and provisions of the Code of Ethics and Standards of Professional Conduct (Code and Standards). The Code contains high-level aspirational ethical principles that drive members and candidates to create a positive and reputable investment profession. The Standards contain practical ethical

This reading is a verbatim reprint of Standards of Practice Handbook, *Eleventh Edition, updated March 2018, p. i to 12 included (www.cfainstitute.org/en/ethics/codes/standards-practice-handbook).*

© 2014 CFA Institute. All rights reserved.

principles of conduct that members and candidates must follow to achieve the broader industry expectations. However, applying the principles individually may not capture the complexity of ethical requirements related to the investment industry. The Code and Standards should be viewed and interpreted as an interwoven tapestry of ethical requirements. Through members' and candidates' adherence to these principles as a whole, the integrity of and trust in the capital markets are improved.

Evolution of the CFA Institute Code of Ethics and Standards of Professional Conduct

Generally, changes to the Code and Standards over the years have been minor. CFA Institute has revised the language of the Code and Standards and occasionally added a new standard to address a prominent issue of the day. For instance, in 1992, CFA Institute added the standard addressing performance presentation to the existing list of standards.

Major changes came in 2005 with the ninth edition of the *Handbook*. CFA Institute adopted new standards, revised some existing standards, and reorganized the standards. The revisions were intended to clarify the requirements of the Code and Standards and effectively convey to its global membership what constitutes "best practice" in a number of areas relating to the investment profession.

The Code and Standards must be regularly reviewed and updated if they are to remain effective and continue to represent the highest ethical standards in the global investment industry. CFA Institute strongly believes that revisions of the Code and Standards are not undertaken for cosmetic purposes but to add value by addressing legitimate concerns and improving comprehension.

Changes to the Code and Standards have far-reaching implications for the CFA Institute membership, the CFA Program, and the investment industry as a whole. CFA Institute members and candidates are *required* to adhere to the Code and Standards. In addition, the Code and Standards are increasingly being adopted, in whole or in part, by firms and regulatory authorities. Their relevance goes well beyond CFA Institute members and candidates.

Standards of Practice Handbook

The periodic revisions of the Code and Standards have come in conjunction with updates of the *Standards of Practice Handbook*. The *Handbook* is the fundamental element of the ethics education effort of CFA Institute and the primary resource for guidance in interpreting and implementing the Code and Standards. The *Handbook* seeks to educate members and candidates on how to apply the Code and Standards to their professional lives and thereby benefit their clients, employers, and the investing public in general. The *Handbook* explains the purpose of the Code and Standards and how they apply in a variety of situations. The sections discuss and amplify each standard and suggest procedures to prevent violations.

Examples in the "Application of the Standard" sections are meant to illustrate how the standard applies to hypothetical but factual situations. The names contained in the examples are fictional and are not meant to refer to any actual person or entity. Unless otherwise stated (e.g., one or more people specifically identified), individuals in each example are CFA Institute members and holders of the CFA designation. Because factual circumstances vary so widely and often involve gray areas, the explanatory material and examples are not intended to be all inclusive. Many examples set forth in the application sections involve standards that have legal counterparts; ***members***

are strongly urged to discuss with their supervisors and legal and compliance departments the content of the Code and Standards and the members' general obligations under the Code and Standards.

CFA Institute recognizes that the presence of any set of ethical standards may create a false sense of security unless the documents are fully understood, enforced, and made a meaningful part of everyday professional activities. The *Handbook* is intended to provide a useful frame of reference that suggests ethical professional behavior in the investment decision-making process. This book cannot cover every contingency or circumstance, however, and it does not attempt to do so. The development and interpretation of the Code and Standards are evolving processes; the Code and Standards will be subject to continuing refinement.

Summary of Changes in the Eleventh Edition

The comprehensive review of the Code and Standards in 2005 resulted in principle requirements that remain applicable today. The review carried out for the eleventh edition focused on market practices that have evolved since the tenth edition. Along with updates to the guidance and examples within the *Handbook*, the eleventh edition includes an update to the Code of Ethics that embraces the members' role of maintaining the social contract between the industry and investors. Additionally, there are three changes to the Standards of Professional Conduct, which recognize the importance of proper supervision, clear communications with clients, and the expanding educational programs of CFA Institute.

Inclusion of Updated CFA Institute Mission

The CFA Institute Board of Governors approved an updated mission for the organization that is included in the Preamble to the Code and Standards. The new mission conveys the organization's conviction in the investment industry's role in the betterment of society at large.

> ***Mission:***
>
> To lead the investment profession globally by promoting the highest standards of ethics, education, and professional excellence for the ultimate benefit of society.

Updated Code of Ethics Principle

One of the bullets in the Code of Ethics was updated to reflect the role that the capital markets have in the greater society. As members work to promote and maintain the integrity of the markets, their actions should also help maintain the social contract with investors.

> ***Old:***
>
> Promote the integrity of and uphold the rules governing capital markets.
>
> ***New:***
>
> Promote the integrity and viability of the global capital markets for the ultimate benefit of society.

New Standard Regarding Responsibilities of Supervisors [IV(C)]

The standard for members and candidates with supervision or authority over others within their firms was updated to bring about improvements in preventing illegal and unethical actions from occurring. The prior version of Standard IV(C) focused on the detection and prevention of violations. The updated version stresses broader compliance expectations, which include the detection and prevention aspects of the original version.

> ***Old:***
>
> Members and Candidates must make reasonable efforts to detect and prevent violations of applicable laws, rules, regulations, and the Code and Standards by anyone subject to their supervision or authority.
>
> ***New:***
>
> Members and Candidates must make reasonable efforts to ensure that anyone subject to their supervision or authority complies with applicable laws, rules, regulations, and the Code and Standards.

Additional Requirement under the Standard for Communication with Clients and Prospective Clients [V(B)]

Given the constant development of new and exotic financial instruments and strategies, the standard regarding communicating with clients now includes an implicit requirement to discuss the risks and limitations of recommendations being made to clients. The new principle and related guidance take into account the fact that levels of disclosure will differ between products and services. Members and candidates, along with their firms, must determine the specific disclosures their clients should receive while ensuring appropriate transparency of the individual firms' investment processes.

> ***Addition:***
>
> Disclose to clients and prospective clients significant limitations and risks associated with the investment process.

Modification to Standard VII(A)

Since this standard was developed, CFA Institute has launched additional educational programs. The updated standard not only maintains the integrity of the CFA Program but also expands the same ethical considerations when members or candidates participate in such programs as the CIPM Program and the CFA Institute Investment Foundations certificate program. Whether participating as a member assisting with the curriculum or an examination or as a sitting candidate within a program, we expect them to engage in these programs as they would participate in the CFA Program.

> ***Old:***
>
> Conduct as Members and Candidates in the CFA Program
>
> Members and Candidates must not engage in any conduct that compromises the reputation or integrity of CFA Institute or the CFA designation or the integrity, validity, or security of the CFA examinations.

> ***New:***
>
> Conduct as Participants in CFA Institute Programs
>
> Members and Candidates must not engage in any conduct that compromises the reputation or integrity of CFA Institute or the CFA designation or the integrity, validity, or security of CFA Institute programs.

General Guidance and Example Revision

The guidance and examples were updated to reflect practices and scenarios applicable to today's investment industry. Two concepts that appear frequently in the updates in this edition relate to the increased use of social media for business communications and the use of and reliance on the output of quantitative models. The use of social media platforms has increased significantly since the publication of the tenth edition. And although financial modeling is not new to the industry, this update reflects upon actions that are viewed as possible contributing factors to the financial crises of the past decade.

CFA Institute Professional Conduct Program

All CFA Institute members and candidates enrolled in the CFA Program are required to comply with the Code and Standards. The CFA Institute Board of Governors maintains oversight and responsibility for the Professional Conduct Program (PCP), which, in conjunction with the Disciplinary Review Committee (DRC), is responsible for enforcement of the Code and Standards. The DRC is a volunteer committee of CFA charterholders who serve on panels to review conduct and partner with Professional Conduct staff to establish and review professional conduct policies. The CFA Institute Bylaws and Rules of Procedure for Professional Conduct (Rules of Procedure) form the basic structure for enforcing the Code and Standards. The Professional Conduct division is also responsible for enforcing testing policies of other CFA Institute education programs as well as the professional conduct of Certificate in Investment Performance Measurement (CIPM) certificants.

Professional Conduct inquiries come from a number of sources. First, members and candidates must self-disclose on the annual Professional Conduct Statement all matters that question their professional conduct, such as involvement in civil litigation or a criminal investigation or being the subject of a written complaint. Second, written complaints received by Professional Conduct staff can bring about an investigation. Third, CFA Institute staff may become aware of questionable conduct by a member or candidate through the media, regulatory notices, or another public source. Fourth, candidate conduct is monitored by proctors who complete reports on candidates suspected to have violated testing rules on exam day. Lastly, CFA Institute may also conduct analyses of scores and exam materials after the exam, as well as monitor online and social media to detect disclosure of confidential exam information.

When an inquiry is initiated, the Professional Conduct staff conducts an investigation that may include requesting a written explanation from the member or candidate; interviewing the member or candidate, complaining parties, and third parties; and collecting documents and records relevant to the investigation. Upon reviewing the material obtained during the investigation, the Professional Conduct staff may conclude the inquiry with no disciplinary sanction, issue a cautionary letter, or continue proceedings to discipline the member or candidate. If the Professional Conduct staff believes a violation of the Code and Standards or testing policies has occurred, the member or candidate has the opportunity to reject or accept any charges and the proposed sanctions.

If the member or candidate does not accept the charges and proposed sanction, the matter is referred to a panel composed of DRC members. Panels review materials and presentations from Professional Conduct staff and from the member or candidate. The panel's task is to determine whether a violation of the Code and Standards or testing policies occurred and, if so, what sanction should be imposed.

Sanctions imposed by CFA Institute may have significant consequences; they include public censure, suspension of membership and use of the CFA designation, and revocation of the CFA charter. Candidates enrolled in the CFA Program who have violated the Code and Standards or testing policies may be suspended or prohibited from further participation in the CFA Program.

Adoption of the Code and Standards

The Code and Standards apply to individual members of CFA Institute and candidates in the CFA Program. CFA Institute does encourage firms to adopt the Code and Standards, however, as part of their code of ethics. Those who claim compliance should fully understand the requirements of each of the principles of the Code and Standards.

Once a party—nonmember or firm—ensures its code of ethics meets the principles of the Code and Standards, that party should make the following statement whenever claiming compliance:

> "[Insert name of party] claims compliance with the CFA Institute Code of Ethics and Standards of Professional Conduct. This claim has not been verified by CFA Institute."

CFA Institute welcomes public acknowledgement, when appropriate, that firms are complying with the CFA Institute Code of Ethics and Standards of Professional Conduct and encourages firms to notify us of the adoption plans. For firms that would like to distribute the Code and Standards to clients and potential clients, attractive one-page copies of the Code and Standards, including translations, are available on the CFA Institute website (www.cfainstitute.org).

CFA Institute has also published the Asset Manager Code of Professional Conduct, which is designed, in part, to help asset managers comply with the regulations mandating codes of ethics for investment advisers. Whereas the Code and Standards are aimed at individual investment professionals who are members of CFA Institute or candidates in the CFA Program, the Asset Manager Code was drafted specifically for firms. The Asset Manager Code provides specific, practical guidelines for asset managers in six areas: loyalty to clients, the investment process, trading, compliance, performance evaluation, and disclosure. The Asset Manager Code and the appropriate steps to acknowledge adoption or compliance can be found on the CFA Institute website (www.cfainstitute.org).

Acknowledgments

CFA Institute is a not-for-profit organization that is heavily dependent on the expertise and intellectual contributions of member volunteers. Members devote their time because they share a mutual interest in the organization's mission to promote and achieve ethical practice in the investment profession. CFA Institute owes much to the volunteers' abundant generosity and energy in extending ethical integrity.

The CFA Institute Standards of Practice Council (SPC), a group consisting of CFA charterholder volunteers from many different countries, is charged with maintaining and interpreting the Code and Standards and ensuring that they are effective. The SPC draws its membership from a broad spectrum of organizations in the securities field, including brokers, investment advisers, banks, and insurance companies. In most instances, the SPC members have important supervisory responsibilities within their firms.

The SPC continually evaluates the Code and Standards, as well as the guidance in the *Handbook*, to ensure that they are

- representative of high standards of professional conduct,
- relevant to the changing nature of the investment profession,
- globally applicable,
- sufficiently comprehensive, practical, and specific,
- enforceable, and
- testable for the CFA Program.

The SPC has spent countless hours reviewing and discussing revisions to the Code and Standards and updates to the guidance that make up the eleventh edition of the *Handbook*. Following is a list of the current and former members of the SPC who generously donated their time and energy to this effort.

James E. Hollis III, CFA, Chair
Rik Albrecht, CFA
Terence E. Burns, CFA
Laura Dagan, CFA
Samuel B. Jones, Jr., CFA
Ulrike Kaiser-Boeing, CFA
Jinliang (Jack) Li, CFA
Christopher C. Loop, CFA,
James M. Meeth, CFA
Guy G. Rutherfurd, Jr., CFA
Edouard Senechal, CFA
Wenliang (Richard) Wang, CFA
Peng Lian Wee, CFA

ETHICS AND THE INVESTMENT INDUSTRY

Society ultimately benefits from efficient markets where capital can freely flow to the most productive or innovative destination. Well-functioning capital markets efficiently match those needing capital with those seeking to invest their assets in revenue-generating ventures. In order for capital markets to be efficient, investors must be able to trust that the markets are fair and transparent and offer them the opportunity to be rewarded for the risk they choose to take. Laws, regulations, and enforcement play a vital role but are insufficient alone to guarantee fair and transparent markets. The markets depend on an ethical foundation to guide participants' judgment and behavior. CFA Institute maintains and promotes the Code of Ethics and Standards of Professional Conduct in order to create a culture of ethics for the ultimate benefit of society.

Why Ethics Matters

Ethics can be defined as a set of moral principles or rules of conduct that provide guidance for our behavior when it affects others. Widely acknowledged fundamental ethical principles include honesty, fairness, diligence, and care and respect for others. Ethical conduct follows those principles and balances self-interest with both the direct and the indirect consequences of that behavior for other people.

Not only does unethical behavior by individuals have serious personal consequences—ranging from job loss and reputational damage to fines and even jail—but unethical conduct from market participants, investment professionals, and those who service investors can damage investor trust and thereby impair the sustainability of the global capital markets as a whole. Unfortunately, there seems to be an unending parade of stories bringing to light accounting frauds and manipulations, Ponzi schemes, insider-trading scandals, and other misdeeds. Not surprisingly, this has led to erosion in public confidence in investment professionals. Empirical evidence from numerous surveys documents the low standing in the eyes of the investing public of banks and financial services firms—the very institutions that are entrusted with the economic well-being and retirement security of society.

Governments and regulators have historically tried to combat misconduct in the industry through regulatory reform, with various levels of success. Global capital markets are highly regulated to protect investors and other market participants. However, compliance with regulation alone is insufficient to fully earn investor trust. Individuals and firms must develop a "culture of integrity" that permeates all levels of operations and promotes the ethical principles of stewardship of investor assets and working in the best interests of clients, above and beyond strict compliance with the law. A strong ethical culture that helps honest, ethical people engage in ethical behavior will foster the trust of investors, lead to robust global capital markets, and ultimately benefit society. That is why ethics matters.

Ethics, Society, and the Capital Markets

CFA Institute recently added the concept "for the ultimate benefit of society" to its mission. The premise is that we want to live in a socially, politically, and financially stable society that fosters individual well-being and welfare of the public. A key ingredient for this goal is global capital markets that facilitate the efficient allocation of resources so that the available capital finds its way to places where it most benefits that society. These investments are then used to produce goods and services, to fund innovation and jobs, and to promote improvements in standards of living. Indeed, such a function serves the interests of the society. Efficient capital markets, in turn, provide a host of benefits to those providing the investment capital. Investors are provided the opportunity to transfer and transform risk because the capital markets serve as an information exchange, create investment products, provide liquidity, and limit transaction costs.

However, a well-functioning and efficient capital market system is dependent on trust of the participants. If investors believe that capital market participants—investment professionals and firms—cannot be trusted with their financial assets or that the capital markets are unfair such that only insiders can be successful, they will be unlikely to invest or, at the very least, will require a higher risk premium. Decreased investment capital can reduce innovation and job creation and hurt the economy and society as a whole. Reduced trust in capital markets can also result in a less vibrant, if not smaller, investment industry.

Ethics for a global investment industry should be universal and ultimately support trust and integrity above acceptable local or regional customs and culture. Universal ethics for a global industry strongly supports the efficiency, values, and mission of the industry as a whole. Different countries may be at different stages of development in establishing standards of practice, but the end goal must be to achieve rules, regulations, and standards that support and promote fundamental ethical principles on a global basis.

Capital Market Sustainability and the Actions of One

Individuals and firms also have to look at the indirect impacts of their actions on the broader investment community. The increasingly interconnected nature of global finance brings to the fore an added consideration of market sustainability that was, perhaps, less appreciated in years past. In addition to committing to the highest levels of ethical behavior, today's investment professionals and their employers should consider the long-term health of the market as a whole.

As recent events have demonstrated, apparently isolated and unrelated decisions, however innocuous when considered on an individual basis, in aggregate can precipitate a market crisis. In an interconnected global economy and marketplace, each participant should strive to be aware of how his or her actions or the products he or she distributes may have an impact on capital market participants in other regions or countries.

Investment professionals should consider how their investment decision-making processes affect the global financial markets in the broader context of how they apply their ethical and professional obligations. Those in positions of authority have a special responsibility to consider the broader context of market sustainability in their development and approval of corporate policies, particularly those involving risk management and product development. In addition, corporate compensation strategies should not encourage otherwise ethically sound individuals to engage in unethical or questionable conduct for financial gain. Ethics, sustainability, and properly functioning capital markets are components of the same concept of protecting the best interests of all. To always place the interests of clients ahead of both investment professionals' own interests and those of their employer remains a key ethos.

The Relationship between Ethics and Regulations

Some equate ethical behavior with legal behavior: If you are following the law, you must be acting appropriately. Ethical principles, like laws and regulations, prescribe appropriate constraints on our natural tendency to pursue self-interest that could harm the interests of others. Laws and regulations often attempt to guide people toward ethical behavior, but they do not cover all unethical behavior. Ethical behavior is often distinguished from legal conduct by describing legal behavior as what is required and ethical behavior as conduct that is morally correct. Ethical principles go beyond that which is legally sufficient and encompass what is the right thing to do.

Given many regulators' lack of sufficient resources to enforce well-conceived rules and regulations, relying on a regulatory framework to lead the charge in establishing ethical behavior has its challenges. Therefore, reliance on compliance with laws and regulation alone is insufficient to ensure ethical behavior of investment professionals or to create a truly ethical culture in the industry.

The recent past has shown us that some individuals will succeed at circumventing the regulatory rules for their personal gain. Only the application of strong ethical principles, at both the individual level and the firm level, will limit abuses. Knowing the rules or regulations to apply in a particular situation, although important, may not be sufficient to ensure ethical conduct. Individuals must be able both to recognize areas that are prone to ethical pitfalls and to identify and process those circumstances and influences that can impair ethical judgment.

Applying an Ethical Framework

Laws, regulations, professional standards, and codes of ethics can guide ethical behavior, but individual judgment is a critical ingredient in making principled choices and engaging in appropriate conduct. When faced with an ethical dilemma, individuals must have a well-developed set of principles; otherwise, their thought processes can lead to, at best, equivocation and indecision and, at worst, fraudulent conduct and destruction of the public trust. Establishing an ethical framework for an internal thought process prior to deciding to act is a crucial step in engaging in ethical conduct.

Most investment professionals are used to making decisions from a business (profit/loss) outlook. But given the importance of ethical behavior in carrying out professional responsibilities, it is critical to also analyze decisions and potential conduct from an ethical perspective. Utilizing a framework for ethical decision making will help investment professionals effectively examine their conduct in the context of conflicting interests common to their professional obligations (e.g., researching and gathering information, developing investment recommendations, and managing money for others). Such a framework will allow investment professionals to analyze their conduct in a way that meets high standards of ethical behavior.

An ethical decision-making framework can come in many forms but should provide investment professionals with a tool for following the principles of the firm's code of ethics. Through analyzing the particular circumstances of each decision, investment professionals are able to determine the best course of action to fulfill their responsibilities in an ethical manner.

Commitment to Ethics by Firms

A firm's code of ethics risks becoming a largely ignored, dusty compilation if it is not truly integrated into the fabric of the business. The ability to relate an ethical decision-making framework to a firm's code of ethics allows investment professionals to bring the aspirations and principles of the code of ethics to life—transforming it from a compliance exercise to something that is at the heart of a firm's culture.

An investment professional's natural desire to "do the right thing" must be reinforced by building a culture of integrity in the workplace. Development, maintenance, and demonstration of a strong culture of integrity within the firm by senior management may be the single most important factor in promoting ethical behavior among the firm's employees. Adopting a code that clearly lays out the ethical principles that guide the thought processes and conduct the firm expects from its employees is a critical first step. But a code of ethics, while necessary, is insufficient.

Simply nurturing an inclination to do right is no match for the multitude of daily decisions that investment managers make. We need to exercise ethical decision-making skills to develop the muscle memory necessary for fundamentally ethical people to make good decisions despite the reality of agent conflicts. Just as coaching and practice transform our natural ability to run across a field into the technique and endurance required to run a race, teaching, reinforcing, and practicing ethical decision-making skills prepare us to confront the hard issues effectively. It is good for business, individuals, firms, the industry, and the markets, as well as society as a whole, to engage in the investment management profession in a highly ethical manner.

Ethical Commitment of CFA Institute

An important goal of CFA Institute is to ensure that the organization and its members and candidates develop, promote, and follow the highest ethical standards in the investment industry. The CFA Institute Code of Ethics (Code) and Standards of Professional Conduct (Standards) are the foundation supporting the organization's quest to uphold the industry's highest standards of individual and corporate practice and to help serve the greater good. The Code is a set of principles that define the overarching conduct CFA Institute expects from its members and CFA Program candidates. The Code works in tandem with the Standards, which outline professional conduct that constitutes fair and ethical business practices.

For more than 50 years, CFA Institute members and candidates have been required to abide by the organization's Code and Standards. Periodically, CFA Institute has revised and updated its Code and Standards to ensure that they remain relevant to the changing nature of the investment profession and representative of the highest standard of professional conduct. Within this *Handbook*, CFA Institute addresses ethical principles for the profession, including individual professionalism; responsibilities to capital markets, clients, and employers; ethics involved in investment analysis, recommendations, and actions; and possible conflicts of interest. Although the investment world has become a far more complex place since the first publication of the *Standard of Practice Handbook*, distinguishing right from wrong remains the paramount principle of the Code and Standards.

New challenges will continually arise for members and candidates in applying the Code and Standards because many decisions are not unambiguously right or wrong. The dilemma exists because the choice between right and wrong is not always clear. Even well-intentioned investment professionals can find themselves in circumstances that may tempt them to cut corners. Situational influences can overpower the best of intentions.

CFA Institute has made a significant commitment to providing members and candidates with the resources to extend and deepen their understanding of how to appropriately apply the principles of the Code and Standards. The product offerings from CFA Institute offer a wealth of material. Through publications, conferences, webcasts, and podcasts, the ethical challenges of investment professionals are brought to light. Archived issues of these items are available on the CFA Institute website (www.cfainstitute.org).

By reviewing these resources and discussing with their peers, market participants can further enhance their abilities to apply an effective ethical decision-making framework. In time, this should help restore some of the trust recently lost by investors.

Markets function to an important extent on trust. Recent events have shown the fragility of this foundation and the devastating consequences that can ensue when it is fundamentally questioned. Investment professionals should remain mindful of the long-term health of financial markets and incorporate this concern for the market's sustainability in their investment decision making. CFA Institute and the Standards of Practice Council hope this edition of the *Handbook* will assist and guide investment professionals in meeting the ethical demands of the highly interconnected global capital markets for the ultimate benefit of society.

CFA INSTITUTE CODE OF ETHICS AND STANDARDS OF PROFESSIONAL CONDUCT

Preamble

The CFA Institute Code of Ethics and Standards of Professional Conduct are fundamental to the values of CFA Institute and essential to achieving its mission to lead the investment profession globally by promoting the highest standards of ethics, education, and professional excellence for the ultimate benefit of society. High ethical standards are critical to maintaining the public's trust in financial markets and in the investment profession. Since their creation in the 1960s, the Code and Standards have promoted the integrity of CFA Institute members and served as a model for measuring the ethics of investment professionals globally, regardless of job function, cultural differences, or local laws and regulations. All CFA Institute members (including holders of the Chartered Financial Analyst [CFA] designation) and CFA candidates have the personal responsibility to embrace and uphold the provisions of the Code and Standards and are encouraged to notify their employer of this responsibility. Violations may result in disciplinary sanctions by CFA Institute. Sanctions can include revocation of membership, revocation of candidacy in the CFA Program, and revocation of the right to use the CFA designation.

The Code of Ethics

Members of CFA Institute (including CFA charterholders) and candidates for the CFA designation ("Members and Candidates") must:

- Act with integrity, competence, diligence, and respect and in an ethical manner with the public, clients, prospective clients, employers, employees, colleagues in the investment profession, and other participants in the global capital markets.
- Place the integrity of the investment profession and the interests of clients above their own personal interests.
- Use reasonable care and exercise independent professional judgment when conducting investment analysis, making investment recommendations, taking investment actions, and engaging in other professional activities.
- Practice and encourage others to practice in a professional and ethical manner that will reflect credit on themselves and the profession.
- Promote the integrity and viability of the global capital markets for the ultimate benefit of society.
- Maintain and improve their professional competence and strive to maintain and improve the competence of other investment professionals.

Standards of Professional Conduct

I. PROFESSIONALISM

A Knowledge of the Law

Members and Candidates must understand and comply with all applicable laws, rules, and regulations (including the CFA Institute Code of Ethics and Standards of Professional Conduct) of any government, regulatory organization, licensing agency, or professional association governing their professional activities. In the event of conflict, Members and Candidates must comply with the more strict law, rule, or regulation. Members and Candidates must not knowingly participate or assist in and must dissociate from any violation of such laws, rules, or regulations.

B Independence and Objectivity

Members and Candidates must use reasonable care and judgment to achieve and maintain independence and objectivity in their professional activities. Members and Candidates must not offer, solicit, or accept any gift, benefit, compensation, or consideration that reasonably could be expected to compromise their own or another's independence and objectivity.

C Misrepresentation

Members and Candidates must not knowingly make any misrepresentations relating to investment analysis, recommendations, actions, or other professional activities.

D Misconduct

Members and Candidates must not engage in any professional conduct involving dishonesty, fraud, or deceit or commit any act that reflects adversely on their professional reputation, integrity, or competence.

II. INTEGRITY OF CAPITAL MARKETS

A Material Nonpublic Information

Members and Candidates who possess material nonpublic information that could affect the value of an investment must not act or cause others to act on the information.

B Market Manipulation

Members and Candidates must not engage in practices that distort prices or artificially inflate trading volume with the intent to mislead market participants.

III. DUTIES TO CLIENTS

A Loyalty, Prudence, and Care

Members and Candidates have a duty of loyalty to their clients and must act with reasonable care and exercise prudent judgment. Members and Candidates must act for the benefit of their clients and place their clients' interests before their employer's or their own interests.

B Fair Dealing

Members and Candidates must deal fairly and objectively with all clients when providing investment analysis, making investment recommendations, taking investment action, or engaging in other professional activities.

C Suitability

1 When Members and Candidates are in an advisory relationship with a client, they must:

a Make a reasonable inquiry into a client's or prospective client's investment experience, risk and return objectives, and financial constraints prior to making any investment recommendation or taking investment action and must reassess and update this information regularly.

b Determine that an investment is suitable to the client's financial situation and consistent with the client's written objectives, mandates, and constraints before making an investment recommendation or taking investment action.

c Judge the suitability of investments in the context of the client's total portfolio.

2 When Members and Candidates are responsible for managing a portfolio to a specific mandate, strategy, or style, they must make only investment recommendations or take only investment actions that are consistent with the stated objectives and constraints of the portfolio.

D Performance Presentation

When communicating investment performance information, Members and Candidates must make reasonable efforts to ensure that it is fair, accurate, and complete.

E Preservation of Confidentiality

Members and Candidates must keep information about current, former, and prospective clients confidential unless:

1 The information concerns illegal activities on the part of the client or prospective client,

2 Disclosure is required by law, or

3 The client or prospective client permits disclosure of the information.

IV. DUTIES TO EMPLOYERS

A Loyalty

In matters related to their employment, Members and Candidates must act for the benefit of their employer and not deprive their employer of the advantage of their skills and abilities, divulge confidential information, or otherwise cause harm to their employer.

B Additional Compensation Arrangements

Members and Candidates must not accept gifts, benefits, compensation, or consideration that competes with or might reasonably be expected to create a conflict of interest with their employer's interest unless they obtain written consent from all parties involved.

C Responsibilities of Supervisors

Members and Candidates must make reasonable efforts to ensure that anyone subject to their supervision or authority complies with applicable laws, rules, regulations, and the Code and Standards.

V. INVESTMENT ANALYSIS, RECOMMENDATIONS, AND ACTIONS

A Diligence and Reasonable Basis

Members and Candidates must:

1 Exercise diligence, independence, and thoroughness in analyzing investments, making investment recommendations, and taking investment actions.

2 Have a reasonable and adequate basis, supported by appropriate research and investigation, for any investment analysis, recommendation, or action.

B Communication with Clients and Prospective Clients

Members and Candidates must:

1 Disclose to clients and prospective clients the basic format and general principles of the investment processes they use to analyze investments, select securities, and construct portfolios and must promptly disclose any changes that might materially affect those processes.

2 Disclose to clients and prospective clients significant limitations and risks associated with the investment process.

3 Use reasonable judgment in identifying which factors are important to their investment analyses, recommendations, or actions and include those factors in communications with clients and prospective clients.

4 Distinguish between fact and opinion in the presentation of investment analysis and recommendations.

C Record Retention

Members and Candidates must develop and maintain appropriate records to support their investment analyses, recommendations, actions, and other investment-related communications with clients and prospective clients.

VI. CONFLICTS OF INTEREST

A Disclosure of Conflicts

Members and Candidates must make full and fair disclosure of all matters that could reasonably be expected to impair their independence and objectivity or interfere with respective duties to their clients, prospective clients, and employer. Members and Candidates must ensure that such disclosures are prominent, are delivered in plain language, and communicate the relevant information effectively.

B Priority of Transactions

Investment transactions for clients and employers must have priority over investment transactions in which a Member or Candidate is the beneficial owner.

C Referral Fees

Members and Candidates must disclose to their employer, clients, and prospective clients, as appropriate, any compensation, consideration, or benefit received from or paid to others for the recommendation of products or services.

VII. RESPONSIBILITIES AS A CFA INSTITUTE MEMBER OR CFA CANDIDATE

A Conduct as Participants in CFA Institute Programs

Members and Candidates must not engage in any conduct that compromises the reputation or integrity of CFA Institute or the CFA designation or the integrity, validity, or security of CFA Institute programs.

B Reference to CFA Institute, the CFA Designation, and the CFA Program

When referring to CFA Institute, CFA Institute membership, the CFA designation, or candidacy in the CFA Program, Members and Candidates must not misrepresent or exaggerate the meaning or implications of membership in CFA Institute, holding the CFA designation, or candidacy in the CFA Program.

READING

46

Guidance for Standards I–VII

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;
☐	**b.** recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

STANDARD I(A): PROFESSIONALISM - KNOWLEDGE OF THE LAW

Standard I(A) Knowledge of the Law

Members and Candidates must understand and comply with all applicable laws, rules, and regulations (including the CFA Institute Code of Ethics and Standards of Professional Conduct) of any government, regulatory organization, licensing agency, or professional association governing their professional activities. In the event of conflict, Members and Candidates must comply with the more strict law, rule, or regulation. Members and Candidates must not knowingly participate or assist in and must dissociate from any violation of such laws, rules, or regulations.

Guidance

Highlights:

- *Relationship between the Code and Standards and Applicable Law*

This reading is a verbatim reprint of *Standards of Practice Handbook*, Eleventh Edition, updated March 2018, p. 13 to 174 included (www.cfainstitute.org/en/ethics/codes/standards-practice-handbook).
© 2014 CFA Institute. All rights reserved.

- *Participation in or Association with Violations by Others*
- *Investment Products and Applicable Laws*

Members and candidates must understand the applicable laws and regulations of the countries and jurisdictions where they engage in professional activities. These activities may include, but are not limited to, trading of securities or other financial instruments, providing investment advice, conducting research, or performing other investment services. On the basis of their reasonable and good faith understanding, members and candidates must comply with the laws and regulations that directly govern their professional activities and resulting outcomes and that protect the interests of the clients.

When questions arise, members and candidates should know their firm's policies and procedures for accessing compliance guidance. This standard does not require members and candidates to become experts, however, in compliance. Additionally, members and candidates are not required to have detailed knowledge of or be experts on all the laws that could potentially govern their activities.

During times of changing regulations, members and candidates must remain vigilant in maintaining their knowledge of the requirements for their professional activities. New financial products and processes, along with uncovered ethical missteps, create an environment for recurring and potentially wide-ranging regulatory changes. Members and candidates are also continually provided improved and enhanced methods of communicating with both clients and potential clients, such as mobile applications and web-based social networking platforms. As new local, regional, and global requirements are updated to address these and other changes, members, candidates, and their firms must adjust their procedures and practices to remain in compliance.

Relationship between the Code and Standards and Applicable Law

Some members or candidates may live, work, or provide investment services to clients living in a country that has no law or regulation governing a particular action or that has laws or regulations that differ from the requirements of the Code and Standards. When applicable law and the Code and Standards require different conduct, members and candidates must follow the more strict of the applicable law or the Code and Standards.

"Applicable law" is the law that governs the member's or candidate's conduct. Which law applies will depend on the particular facts and circumstances of each case. The "more strict" law or regulation is the law or regulation that imposes greater restrictions on the action of the member or candidate or calls for the member or candidate to exert a greater degree of action that protects the interests of investors. For example, applicable law or regulation may not require members and candidates to disclose referral fees received from or paid to others for the recommendation of investment products or services. Because the Code and Standards impose this obligation, however, members and candidates must disclose the existence of such fees.

Members and candidates must adhere to the following principles:

- Members and candidates must comply with applicable laws or regulations related to their professional activities.
- Members and candidates must not engage in conduct that constitutes a violation of the Code and Standards, even though it may otherwise be legal.
- In the absence of any applicable law or regulation or when the Code and Standards impose a higher degree of responsibility than applicable laws and regulations, members and candidates must adhere to the Code and Standards. Applications of these principles are outlined in Exhibit 1.

The applicable laws governing the responsibilities of a member or candidate should be viewed as the minimal threshold of acceptable actions. When members and candidates take actions that exceed the minimal requirements, they further support the conduct required of Standard I(A).

CFA Institute members are obligated to abide by the CFA Institute Articles of Incorporation, Bylaws, Code of Ethics, Standards of Professional Conduct, Rules of Procedure, Membership Agreement, and other applicable rules promulgated by CFA Institute, all as amended periodically. CFA candidates who are not members must also abide by these documents (except for the Membership Agreement) as well as rules and regulations related to the administration of the CFA examination, the Candidate Responsibility Statement, and the Candidate Pledge.

Participation in or Association with Violations by Others

Members and candidates are responsible for violations in which they *knowingly* participate or assist. Although members and candidates are presumed to have knowledge of all applicable laws, rules, and regulations, CFA Institute acknowledges that members may not recognize violations if they are not aware of all the facts giving rise to the violations. Standard I(A) applies when members and candidates know or should know that their conduct may contribute to a violation of applicable laws, rules, or regulations or the Code and Standards.

If a member or candidate has reasonable grounds to believe that imminent or ongoing client or employer activities are illegal or unethical, the member or candidate must dissociate, or separate, from the activity. In extreme cases, dissociation may require a member or candidate to leave his or her employment. Members and candidates may take the following intermediate steps to dissociate from ethical violations of others when direct discussions with the person or persons committing the violation are unsuccessful. The first step should be to attempt to stop the behavior by bringing it to the attention of the employer through a supervisor or the firm's compliance department. If this attempt is unsuccessful, then members and candidates have a responsibility to step away and dissociate from the activity. Dissociation practices will differ on the basis of the member's or candidate's role in the investment industry. It may include removing one's name from written reports or recommendations, asking for a different assignment, or refusing to accept a new client or continue to advise a current client. Inaction combined with continuing association with those involved in illegal or unethical conduct may be construed as participation or assistance in the illegal or unethical conduct.

CFA Institute strongly encourages members and candidates to report potential violations of the Code and Standards committed by fellow members and candidates. Although a failure to report is less likely to be construed as a violation than a failure to dissociate from unethical conduct, the impact of inactivity on the integrity of capital markets can be significant. Although the Code and Standards do not compel members and candidates to report violations to their governmental or regulatory organizations unless such disclosure is mandatory under applicable law (voluntary reporting is often referred to as whistleblowing), such disclosure may be prudent under certain circumstances. Members and candidates should consult their legal and compliance advisers for guidance.

Additionally, CFA Institute encourages members, nonmembers, clients, and the investing public to report violations of the Code and Standards by CFA Institute members or CFA candidates by submitting a complaint in writing to the CFA Institute Professional Conduct Program via e-mail (pcprogram@cfainstitute.org) or the CFA Institute website (www.cfainstitute.org).

Investment Products and Applicable Laws

Members and candidates involved in creating or maintaining investment services or investment products or packages of securities and/or derivatives should be mindful of where these products or packages will be sold as well as their places of origination. The applicable laws and regulations of the countries or regions of origination and expected sale should be understood by those responsible for the supervision of the services or creation and maintenance of the products or packages. Members or candidates should make reasonable efforts to review whether associated firms that are distributing products or services developed by their employing firm also abide by the laws and regulations of the countries and regions of distribution. Members and candidates should undertake the necessary due diligence when transacting cross-border business to understand the multiple applicable laws and regulations in order to protect the reputation of their firm and themselves.

Given the complexity that can arise with business transactions in today's market, there may be some uncertainty surrounding which laws or regulations are considered applicable when activities are being conducted in multiple jurisdictions. Members and candidates should seek the appropriate guidance, potentially including the firm's compliance or legal departments and legal counsel outside the organization, to gain a reasonable understanding of their responsibilities and how to implement them appropriately.

Exhibit 1 Global Application of the Code and Standards

Members and candidates who practice in multiple jurisdictions may be subject to varied securities laws and regulations. If applicable law is stricter than the requirements of the Code and Standards, members and candidates must adhere to applicable law; otherwise, they must adhere to the Code and Standards. The following chart provides illustrations involving a member who may be subject to the securities laws and regulations of three different types of countries:

NS: country with no securities laws or regulations

LS: country with ***less*** strict securities laws and regulations than the Code and Standards

MS: country with ***more*** strict securities laws and regulations than the Code and Standards

Applicable Law	Duties	Explanation
Member resides in NS country, does business in LS country; LS law applies.	Member must adhere to the Code and Standards.	Because applicable law is less strict than the Code and Standards, the member must adhere to the Code and Standards.
Member resides in NS country, does business in MS country; MS law applies.	Member must adhere to the law of MS country.	Because applicable law is stricter than the Code and Standards, member must adhere to the more strict applicable law.
Member resides in LS country, does business in NS country; LS law applies.	Member must adhere to the Code and Standards.	Because applicable law is less strict than the Code and Standards, member must adhere to the Code and Standards.

Exhibit 1 (Continued)

Applicable Law	Duties	Explanation
Member resides in LS country, does business in MS country; MS law applies.	Member must adhere to the law of MS country.	Because applicable law is stricter than the Code and Standards, member must adhere to the more strict applicable law.
Member resides in LS country, does business in NS country; LS law applies, but it states that law of locality where business is conducted governs.	Member must adhere to the Code and Standards.	Because applicable law states that the law of the locality where the business is conducted governs and there is no local law, the member must adhere to the Code and Standards.
Member resides in LS country, does business in MS country; LS law applies, but it states that law of locality where business is conducted governs.	Member must adhere to the law of MS country.	Because applicable law of the locality where the business is conducted governs and local law is stricter than the Code and Standards, member must adhere to the more strict applicable law.
Member resides in MS country, does business in LS country; MS law applies.	Member must adhere to the law of MS country.	Because applicable law is stricter than the Code and Standards, member must adhere to the more strict applicable law.
Member resides in MS country, does business in LS country; MS law applies, but it states that law of locality where business is conducted governs.	Member must adhere to the Code and Standards.	Because applicable law states that the law of the locality where the business is conducted governs and local law is less strict than the Code and Standards, member must adhere to the Code and Standards.
Member resides in MS country, does business in LS country with a client who is a citizen of LS country; MS law applies, but it states that the law of the client’s home country governs.	Member must adhere to the Code and Standards.	Because applicable law states that the law of the client’s home country governs (which is less strict than the Code and Standards), member must adhere to the Code and Standards.
Member resides in MS country, does business in LS country with a client who is a citizen of MS country; MS law applies, but it states that the law of the client’s home country governs.	Member must adhere to the law of MS country.	Because applicable law states that the law of the client’s home country governs and the law of the client’s home country is stricter than the Code and Standards, the member must adhere to the more strict applicable law.

STANDARD I(A): RECOMMENDED PROCEDURES

Members and Candidates

Suggested methods by which members and candidates can acquire and maintain understanding of applicable laws, rules, and regulations include the following:

- *Stay informed*: Members and candidates should establish or encourage their employers to establish a procedure by which employees are regularly informed about changes in applicable laws, rules, regulations, and case law. In many instances, the employer's compliance department or legal counsel can provide such information in the form of memorandums distributed to employees in the organization. Also, participation in an internal or external continuing education program is a practical method of staying current.
- *Review procedures*: Members and candidates should review, or encourage their employers to review, the firm's written compliance procedures on a regular basis to ensure that the procedures reflect current law and provide adequate guidance to employees about what is permissible conduct under the law and/or the Code and Standards. Recommended compliance procedures for specific items of the Code and Standards are discussed in this *Handbook* in the "Guidance" sections associated with each standard.
- *Maintain current files*: Members and candidates should maintain or encourage their employers to maintain readily accessible current reference copies of applicable statutes, rules, regulations, and important cases.

Distribution Area Laws

Members and candidates should make reasonable efforts to understand the applicable laws—both country and regional—for the countries and regions where their investment products are developed and are most likely to be distributed to clients.

Legal Counsel

When in doubt about the appropriate action to undertake, it is recommended that a member or candidate seek the advice of compliance personnel or legal counsel concerning legal requirements. If a potential violation is being committed by a fellow employee, it may also be prudent for the member or candidate to seek the advice of the firm's compliance department or legal counsel.

Dissociation

When dissociating from an activity that violates the Code and Standards, members and candidates should document the violation and urge their firms to attempt to persuade the perpetrator(s) to cease such conduct. To dissociate from the conduct, a member or candidate may have to resign his or her employment.

Firms

The formality and complexity of compliance procedures for firms depend on the nature and size of the organization and the nature of its investment operations. Members and candidates should encourage their firms to consider the following policies and procedures to support the principles of Standard I(A):

- *Develop and/or adopt a code of ethics*: The ethical culture of an organization starts at the top. Members and candidates should encourage their supervisors or managers to adopt a code of ethics. Adhering to a code of ethics facilitates solutions when people face ethical dilemmas and can prevent the need for employees to resort to a "whistleblowing" solution publicly alleging concealed misconduct. CFA Institute has published the *Asset Manager Code of Professional Conduct*, which firms may adopt or use as the basis for their codes (visit www.cfainstitute.org).
- *Provide information on applicable laws*: Pertinent information that highlights applicable laws and regulations might be distributed to employees or made available in a central location. Information sources might include primary information developed by the relevant government, governmental agencies, regulatory organizations, licensing agencies, and professional associations (e.g., from their websites); law firm memorandums or newsletters; and association memorandums or publications (e.g., *CFA Institute Magazine*).
- *Establish procedures for reporting violations*: Firms might provide written protocols for reporting suspected violations of laws, regulations, or company policies.

STANDARD I(A): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Notification of Known Violations):

Michael Allen works for a brokerage firm and is responsible for an underwriting of securities. A company official gives Allen information indicating that the financial statements Allen filed with the regulator overstate the issuer's earnings. Allen seeks the advice of the brokerage firm's general counsel, who states that it would be difficult for the regulator to prove that Allen has been involved in any wrongdoing.

> *Comment*: Although it is recommended that members and candidates seek the advice of legal counsel, the reliance on such advice does not absolve a member or candidate from the requirement to comply with the law or regulation. Allen should report this situation to his supervisor, seek an independent legal opinion, and determine whether the regulator should be notified of the error.

Example 2 (Dissociating from a Violation):

Lawrence Brown's employer, an investment banking firm, is the principal underwriter for an issue of convertible debentures by the Courtney Company. Brown discovers that the Courtney Company has concealed severe third-quarter losses in its foreign operations. The preliminary prospectus has already been distributed.

> *Comment*: Knowing that the preliminary prospectus is misleading, Brown should report his findings to the appropriate supervisory persons in his firm. If the matter is not remedied and Brown's employer does not dissociate from the underwriting, Brown should sever all his connections with the underwriting. Brown should also seek legal advice to determine whether additional reporting or other action should be taken.

Example 3 (Dissociating from a Violation):

Kamisha Washington's firm advertises its past performance record by showing the 10-year return of a composite of its client accounts. Washington discovers, however, that the composite omits the performance of accounts that have left the firm during the 10-year period, whereas the description of the composite indicates the inclusion of all firm accounts. This omission has led to an inflated performance figure. Washington is asked to use promotional material that includes the erroneous performance number when soliciting business for the firm.

> *Comment*: Misrepresenting performance is a violation of the Code and Standards. Although she did not calculate the performance herself, Washington would be assisting in violating Standard I(A) if she were to use the inflated performance number when soliciting clients. She must dissociate herself from the activity. If discussing the misleading number with the person responsible is not an option for correcting the problem, she can bring the situation to the attention of her supervisor or the compliance department at her firm. If her firm is unwilling to recalculate performance, she must refrain from using the misleading promotional material and should notify the firm of her reasons. If the firm insists that she use the material, she should consider whether her obligation to dissociate from the activity requires her to seek other employment.

Example 4 (Following the Highest Requirements):

James Collins is an investment analyst for a major Wall Street brokerage firm. He works in a developing country with a rapidly modernizing economy and a growing capital market. Local securities laws are minimal—in form and content—and include no punitive prohibitions against insider trading.

> *Comment*: Collins must abide by the requirements of the Code and Standards, which might be more strict than the rules of the developing country. He should be aware of the risks that a small market and the absence of a fairly regulated flow of information to the market represent to his ability to obtain information and make timely judgments. He should include this factor in formulating his advice to clients. In handling material nonpublic information that accidentally comes into his possession, he must follow Standard II(A)–Material Nonpublic Information.

Example 5 (Following the Highest Requirements):

Laura Jameson works for a multinational investment adviser based in the United States. Jameson lives and works as a registered investment adviser in the tiny, but wealthy, island nation of Karramba. Karramba's securities laws state that no investment adviser registered and working in that country can participate in initial public offerings (IPOs) for the adviser's personal account. Jameson, believing that, as a US citizen working for a US-based company, she should comply only with US law, has ignored this Karrambian law. In addition, Jameson believes that as a charterholder, as long as she adheres to the Code and Standards requirement that she disclose her participation in any IPO to her employer and clients when such ownership creates a conflict of interest, she is meeting the highest ethical requirements.

> *Comment*: Jameson is in violation of Standard I(A). As a registered investment adviser in Karramba, Jameson is prevented by Karrambian securities law from participating in IPOs regardless of the law of her home country. In addition, because the law of the country where she is working is stricter than the Code and Standards, she must follow the stricter requirements of the local law rather than the requirements of the Code and Standards.

Example 6 (Laws and Regulations Based on Religious Tenets):

Amanda Janney is employed as a fixed-income portfolio manager for a large international firm. She is on a team within her firm that is responsible for creating and managing a fixed-income hedge fund to be sold throughout the firm's distribution centers to high-net-worth clients. Her firm receives expressions of interest from potential clients from the Middle East who are seeking investments that comply with Islamic law. The marketing and promotional materials for the fixed-income hedge fund do not specify whether or not the fund is a suitable investment for an investor seeking compliance with Islamic law. Because the fund is being distributed globally, Janney is concerned about the reputation of the fund and the firm and believes disclosure of whether or not the fund complies with Islamic law could help minimize potential mistakes with placing this investment.

> *Comment*: As the financial market continues to become globalized, members and candidates will need to be aware of the differences between cultural and religious laws and requirements as well as the different governmental laws and regulations. Janney and the firm could be proactive in their efforts to acknowledge areas where the new fund may not be suitable for clients.

Example 7 (Reporting Potential Unethical Actions):

Krista Blume is a junior portfolio manager for high-net-worth portfolios at a large global investment manager. She observes a number of new portfolios and relationships coming from a country in Europe where the firm did not have previous business and is told that a broker in that country is responsible for this new business. At a meeting on allocation of research resources to third-party research firms, Blume notes that this broker has been added to the list and is allocated payments for research. However, she knows the portfolios do not invest in securities in the broker's country, and she has not seen any research come from this broker. Blume asks her supervisor about the name being on the list and is told that someone in marketing is receiving the research and that the name being on the list is OK. She believes that what may be going on is that the broker is being paid for new business through the inappropriate research payments, and she wishes to dissociate from the misconduct.

Comment: Blume should follow the firm's policies and procedures for reporting potential unethical activity, which may include discussions with her supervisor or someone in a designated compliance department. She should communicate her concerns appropriately while advocating for disclosure between the new broker relationship and the research payments.

Example 8 (Failure to Maintain Knowledge of the Law):

Colleen White is excited to use new technology to communicate with clients and potential clients. She recently began posting investment information, including performance reports and investment opinions and recommendations, to her Facebook page. In addition, she sends out brief announcements, opinions, and thoughts via her Twitter account (for example, "Prospects for future growth of XYZ company look good! #makingmoney4U"). Prior to White's use of these social media platforms, the local regulator had issued new requirements and guidance governing online electronic communication. White's communications appear to conflict with the recent regulatory announcements.

Comment: White is in violation of Standard I(A) because her communications do not comply with the existing guidance and regulation governing use of social media. White must be aware of the evolving legal requirements pertaining to new and dynamic areas of the financial services industry that are applicable to her. She should seek guidance from appropriate, knowledgeable, and reliable sources, such as her firm's compliance department, external service providers, or outside counsel, unless she diligently follows legal and regulatory trends affecting her professional responsibilities.

STANDARD I(B): PROFESSIONALISM - INDEPENDENCE AND OBJECTIVITY

Members and Candidates must use reasonable care and judgment to achieve and maintain independence and objectivity in their professional activities. Members and Candidates must not offer, solicit, or accept any gift, benefit, compensation, or consideration that reasonably could be expected to compromise their own or another's independence and objectivity.

Guidance

Highlights:

- *Buy-Side Clients*
- *Fund Manager and Custodial Relationships*
- *Investment Banking Relationships*
- *Performance Measurement and Attribution*
- *Public Companies*

- *Credit Rating Agency Opinions*
- *Influence during the Manager Selection/Procurement Process*
- *Issuer-Paid Research*
- *Travel Funding*

Standard I(B) states the responsibility of CFA Institute members and candidates in the CFA Program to maintain independence and objectivity so that their clients will have the benefit of their work and opinions unaffected by any potential conflict of interest or other circumstance adversely affecting their judgment. Every member and candidate should endeavor to avoid situations that could cause or be perceived to cause a loss of independence or objectivity in recommending investments or taking investment action.

External sources may try to influence the investment process by offering analysts and portfolio managers a variety of benefits. Corporations may seek expanded research coverage, issuers and underwriters may wish to promote new securities offerings, brokers may want to increase commission business, and independent rating agencies may be influenced by the company requesting the rating. Benefits may include gifts, invitations to lavish functions, tickets, favors, or job referrals. One type of benefit is the allocation of shares in oversubscribed IPOs to investment managers for their personal accounts. This practice affords managers the opportunity to make quick profits that may not be available to their clients. Such a practice is prohibited under Standard I(B). Modest gifts and entertainment are acceptable, but special care must be taken by members and candidates to resist subtle and not-so-subtle pressures to act in conflict with the interests of their clients. Best practice dictates that members and candidates reject any offer of gift or entertainment that could be expected to threaten their independence and objectivity.

Receiving a gift, benefit, or consideration from a *client* can be distinguished from gifts given by entities seeking to influence a member or candidate to the detriment of other clients. In a client relationship, the client has already entered some type of compensation arrangement with the member, candidate, or his or her firm. A gift from a client could be considered supplementary compensation. The potential for obtaining influence to the detriment of other clients, although present, is not as great as in situations where no compensation arrangement exists. When possible, prior to accepting "bonuses" or gifts from clients, members and candidates should disclose to their employers such benefits offered by clients. If notification is not possible prior to acceptance, members and candidates must disclose to their employer benefits previously accepted from clients. Disclosure allows the employer of a member or candidate to make an independent determination about the extent to which the gift may affect the member's or candidate's independence and objectivity.

Members and candidates may also come under pressure from their own firms to, for example, issue favorable research reports or recommendations for certain companies with potential or continuing business relationships with the firm. The situation may be aggravated if an executive of the company sits on the bank or investment firm's board and attempts to interfere in investment decision making. Members and candidates acting in a sales or marketing capacity must be especially mindful of their objectivity in promoting appropriate investments for their clients.

Left unmanaged, pressures that threaten independence place research analysts in a difficult position and may jeopardize their ability to act independently and objectively. One of the ways that research analysts have coped with these pressures in the past is to use subtle and ambiguous language in their recommendations or to temper the tone of their research reports. Such subtleties are lost on some investors, however, who reasonably expect research reports and recommendations to be straightforward and transparent and to communicate clearly an analyst's views based on unbiased analysis and independent judgment.

Members and candidates are personally responsible for maintaining independence and objectivity when preparing research reports, making investment recommendations, and taking investment action on behalf of clients. Recommendations must convey the member's or candidate's true opinions, free of bias from internal or external pressures, and be stated in clear and unambiguous language.

Members and candidates also should be aware that some of their professional or social activities within CFA Institute or its member societies may subtly threaten their independence or objectivity. When seeking corporate financial support for conventions, seminars, or even weekly society luncheons, the members or candidates responsible for the activities should evaluate both the actual effect of such solicitations on their independence and whether their objectivity might be perceived to be compromised in the eyes of their clients.

Buy-Side Clients

One source of pressure on sell-side analysts is buy-side clients. Institutional clients are traditionally the primary users of sell-side research, either directly or with soft dollar brokerage. Portfolio managers may have significant positions in the security of a company under review. A rating downgrade may adversely affect the portfolio's performance, particularly in the short term, because the sensitivity of stock prices to ratings changes has increased in recent years. A downgrade may also affect the manager's compensation, which is usually tied to portfolio performance. Moreover, portfolio performance is subject to media and public scrutiny, which may affect the manager's professional reputation. Consequently, some portfolio managers implicitly or explicitly support sell-side ratings inflation.

Portfolio managers have a responsibility to respect and foster the intellectual honesty of sell-side research. Therefore, it is improper for portfolio managers to threaten or engage in retaliatory practices, such as reporting sell-side analysts to the covered company in order to instigate negative corporate reactions. Although most portfolio managers do not engage in such practices, the perception by the research analyst that a reprisal is possible may cause concern and make it difficult for the analyst to maintain independence and objectivity.

Fund Manager and Custodial Relationships

Research analysts are not the only people who must be concerned with maintaining their independence. Members and candidates who are responsible for hiring and retaining outside managers and third-party custodians should not accepts gifts, entertainment, or travel funding that may be perceived as impairing their decisions. The use of secondary fund managers has evolved into a common practice to manage specific asset allocations. The use of third-party custodians is common practice for independent investment advisory firms and helps them with trading capabilities and reporting requirements. Primary and secondary fund managers, as well as third-party custodians, often arrange educational and marketing events to inform others about their business strategies, investment process, or custodial services. Members and candidates must review the merits of each offer individually in determining whether they may attend yet maintain their independence.

Investment Banking Relationships

Some sell-side firms may exert pressure on their analysts to issue favorable research reports on current or prospective investment banking clients. For many of these firms, income from investment banking has become increasingly important to overall firm profitability because brokerage income has declined as a result of price competition. Consequently, firms offering investment banking services work hard to develop and maintain relationships with investment banking clients and prospects. These companies are often covered by the firm's research analysts because companies often select their investment banks on the basis of the reputation of their research analysts, the quality of their work, and their standing in the industry.

In some countries, research analysts frequently work closely with their investment banking colleagues to help evaluate prospective investment banking clients. In other countries, because of past abuses in managing the obvious conflicts of interest, regulators have established clear rules prohibiting the interaction of these groups. Although collaboration between research analysts and investment banking colleagues may benefit the firm and enhance market efficiency (e.g., by allowing firms to assess risks more accurately and make better pricing assumptions), it requires firms to carefully balance the conflicts of interest inherent in the collaboration. Having analysts work with investment bankers is appropriate only when the conflicts are adequately and effectively managed and disclosed. Firm managers have a responsibility to provide an environment in which analysts are neither coerced nor enticed into issuing research that does not reflect their true opinions. Firms should require public disclosure of actual conflicts of interest to investors.

Members, candidates, and their firms must adopt and follow perceived best practices in maintaining independence and objectivity in the corporate culture and protecting analysts from undue pressure by their investment banking colleagues. The "firewalls" traditionally built between these two functions must be managed to minimize conflicts of interest; indeed, enhanced firewall policies may go as far as prohibiting all communications between these groups. A key element of an enhanced firewall is separate reporting structures for personnel on the research side and personnel on the investment banking side. For example, investment banking personnel should not have any authority to approve, disapprove, or make changes to research reports or recommendations. Another element should be a compensation arrangement that minimizes the pressures on research analysts and rewards objectivity and accuracy. Compensation arrangements should not link analyst remuneration directly to investment banking assignments in which the analyst may participate as a team member. Firms should also regularly review their policies and procedures to determine whether analysts are adequately safeguarded and to improve the transparency of disclosures relating to conflicts of interest. The highest level of transparency is achieved when disclosures are prominent and specific rather than marginalized and generic.

Performance Measurement and Attribution

Members and candidates working within a firm's investment performance measurement department may also be presented with situations that challenge their independence and objectivity. As performance analysts, their analyses may reveal instances where managers may appear to have strayed from their mandate. Additionally, the performance analyst may receive requests to alter the construction of composite indexes owing to negative results for a selected account or fund. The member or candidate must not allow internal or external influences to affect their independence and objectivity as they faithfully complete their performance calculation and analysis-related responsibilities.

Public Companies

Analysts may be pressured to issue favorable reports and recommendations by the companies they follow. Not every stock is a "buy," and not every research report is favorable—for many reasons, including the cyclical nature of many business activities and market fluctuations. For instance, a "good company" does not always translate into a "good stock" rating if the current stock price is fully valued. In making an investment recommendation, the analyst is responsible for anticipating, interpreting, and assessing a company's prospects and stock price performance in a factual manner. Many company managers, however, believe that their company's stock is undervalued, and these managers may find it difficult to accept critical research reports or ratings downgrades. Company managers' compensation may also be dependent on stock performance.

Due diligence in financial research and analysis involves gathering information from a wide variety of sources, including public disclosure documents (such as proxy statements, annual reports, and other regulatory filings) and also company management and investor-relations personnel, suppliers, customers, competitors, and other relevant sources. Research analysts may justifiably fear that companies will limit their ability to conduct thorough research by denying analysts who have "negative" views direct access to company managers and/or barring them from conference calls and other communication venues. Retaliatory practices include companies bringing legal action against analysts personally and/or their firms to seek monetary damages for the economic effects of negative reports and recommendations. Although few companies engage in such behavior, the perception that a reprisal is possible is a reasonable concern for analysts. This concern may make it difficult for them to conduct the comprehensive research needed to make objective recommendations. For further information and guidance, members and candidates should refer to the CFA Institute publication *Best Practice Guidelines Governing Analyst/Corporate Issuer Relations* (www.cfainstitute.org).

Credit Rating Agency Opinions

Credit rating agencies provide a service by grading the fixed-income products offered by companies. Analysts face challenges related to incentives and compensation schemes that may be tied to the final rating and successful placement of the product. Members and candidates employed at rating agencies should ensure that procedures and processes at the agencies prevent undue influences from a sponsoring company during the analysis. Members and candidates should abide by their agencies' and the industry's standards of conduct regarding the analytical process and the distribution of their reports.

The work of credit rating agencies also raises concerns similar to those inherent in investment banking relationships. Analysts may face pressure to issue ratings at a specific level because of other services the agency offers companies—namely, advising on the development of structured products. The rating agencies need to develop the necessary firewalls and protections to allow the independent operations of their different business lines.

When using information provided by credit rating agencies, members and candidates should be mindful of the potential conflicts of interest. And because of the potential conflicts, members and candidates may need to independently validate the rating granted.

Influence during the Manager Selection/Procurement Process

Members and candidates may find themselves on either side of the manager selection process. An individual may be on the hiring side as a representative of a pension organization or an investment committee member of an endowment or a charitable organization. Additionally, other members may be representing their organizations in attempts to earn new investment allocation mandates. The responsibility of members and candidates to maintain their independence and objectivity extends to the hiring or firing of those who provide business services beyond investment management.

When serving in a hiring capacity, members and candidates should not solicit gifts, contributions, or other compensation that may affect their independence and objectivity. Solicitations do not have to benefit members and candidates personally to conflict with Standard I(B). Requesting contributions to a favorite charity or political organization may also be perceived as an attempt to influence the decision-making process. Additionally, members and candidates serving in a hiring capacity should refuse gifts, donations, and other offered compensation that may be perceived to influence their decision-making process.

When working to earn a new investment allocation, members and candidates should not offer gifts, contributions, or other compensation to influence the decision of the hiring representative. The offering of these items with the intent to impair the independence and objectivity of another person would not comply with Standard I(B). Such prohibited actions may include offering donations to a charitable organization or political candidate referred by the hiring representative.

A clear example of improperly influencing hiring representatives was displayed in the "pay-to-play" scandal involving government-sponsored pension funds in the United States. Managers looking to gain lucrative allocations from the large funds made requested donations to the political campaigns of individuals directly responsible for the hiring decisions. This scandal and other similar events have led to new laws requiring additional reporting concerning political contributions and bans on hiring—or hiring delays for—managers that made campaign contributions to representatives associated with the decision-making process.

Issuer-Paid Research

In light of the recent reduction of sell-side research coverage, many companies, seeking to increase visibility both in the financial markets and with potential investors, have hired analysts to produce research reports analyzing their companies. These reports bridge the gap created by the lack of coverage and can be an effective method of communicating with investors.

Issuer-paid research conducted by independent analysts, however, is fraught with potential conflicts. Depending on how the research is written and distributed, investors may be misled into believing that the research is from an independent source when, in reality, it has been paid for by the subject company.

Members and candidates must adhere to strict standards of conduct that govern how the research is to be conducted and what disclosures must be made in the report. Analysts must engage in thorough, independent, and unbiased analysis and must fully disclose potential conflicts of interest, including the nature of their compensation. Otherwise, analysts risk misleading investors.

Investors need clear, credible, and thorough information about companies, and they need research based on independent thought. At a minimum, issuer-paid research should include a thorough analysis of the company's financial statements based on publicly disclosed information, benchmarking within a peer group, and industry analysis. Analysts must exercise diligence, independence, and thoroughness in conducting their research in an objective manner. Analysts must distinguish between fact and opinion in their reports. Conclusions must have a reasonable and adequate basis and must be supported by appropriate research.

Independent analysts must also strictly limit the type of compensation that they accept for conducting issuer-paid research. Otherwise, the content and conclusions of the reports could reasonably be expected to be determined or affected by compensation from the sponsoring companies. Compensation that might influence the research report could be direct, such as payment based on the conclusions of the report, or indirect, such as stock warrants or other equity instruments that could increase in value on the basis of positive coverage in the report. In such instances, the independent analyst has an incentive to avoid including negative information or making negative conclusions. Best practice is for independent analysts, prior to writing their reports, to negotiate only a flat fee for their work that is not linked to their conclusions or recommendations.

Travel Funding

The benefits related to accepting paid travel extend beyond the cost savings to the member or candidate and his firm, such as the chance to talk exclusively with the executives of a company or learning more about the investment options provided by an investment organization. Acceptance also comes with potential concerns; for example, members and candidates may be influenced by these discussions when flying on a corporate or chartered jet or attending sponsored conferences where many expenses, including airfare and lodging, are covered. To avoid the appearance of compromising their independence and objectivity, best practice dictates that members and candidates always use commercial transportation at their expense or at the expense of their firm rather than accept paid travel arrangements from an outside company. Should commercial transportation be unavailable, members and candidates may accept modestly arranged travel to participate in appropriate information-gathering events, such as a property tour.

STANDARD I(B): RECOMMENDED PROCEDURES

Members and candidates should adhere to the following practices and should encourage their firms to establish procedures to avoid violations of Standard I(B):

- *Protect the integrity of opinions*: Members, candidates, and their firms should establish policies stating that every research report concerning the securities of a corporate client should reflect the unbiased opinion of the analyst. Firms should also design compensation systems that protect the integrity of the investment decision process by maintaining the independence and objectivity of analysts.
- *Create a restricted list*: If the firm is unwilling to permit dissemination of adverse opinions about a corporate client, members and candidates should encourage the firm to remove the controversial company from the research universe and put it on a restricted list so that the firm disseminates only factual information about the company.
- *Restrict special cost arrangements*: When attending meetings at an issuer's headquarters, members and candidates should pay for commercial transportation and hotel charges. No corporate issuer should reimburse members or candidates for air transportation. Members and candidates should encourage issuers to limit the use of corporate aircraft to situations in which commercial transportation is not available or in which efficient movement could not otherwise be arranged. Members and candidates should take particular care that when frequent meetings are held between an individual issuer and an individual member or candidate, the issuer should not always host the member or candidate.
- *Limit gifts*: Members and candidates must limit the acceptance of gratuities and/or gifts to token items. Standard I(B) does not preclude customary, ordinary business-related entertainment as long as its purpose is not to influence or reward members or candidates. Firms should consider a strict value limit for acceptable gifts that is based on the local or regional customs and should address whether the limit is per gift or an aggregate annual value.
- *Restrict investments*: Members and candidates should encourage their investment firms to develop formal policies related to employee purchases of equity or equity-related IPOs. Firms should require prior approval for employee

participation in IPOs, with prompt disclosure of investment actions taken following the offering. Strict limits should be imposed on investment personnel acquiring securities in private placements.

- *Review procedures*: Members and candidates should encourage their firms to implement effective supervisory and review procedures to ensure that analysts and portfolio managers comply with policies relating to their personal investment activities.
- *Independence policy*: Members, candidates, and their firms should establish a formal written policy on the independence and objectivity of research and implement reporting structures and review procedures to ensure that research analysts do not report to and are not supervised or controlled by any department of the firm that could compromise the independence of the analyst. More detailed recommendations related to a firm's policies regarding research objectivity are set forth in the CFA Institute statement *Research Objectivity Standards* (www.cfainstitute.org).
- *Appointed officer*: Firms should appoint a senior officer with oversight responsibilities for compliance with the firm's code of ethics and all regulations concerning its business. Firms should provide every employee with the procedures and policies for reporting potentially unethical behavior, violations of regulations, or other activities that may harm the firm's reputation.

STANDARD I(B): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Travel Expenses):

Steven Taylor, a mining analyst with Bronson Brokers, is invited by Precision Metals to join a group of his peers in a tour of mining facilities in several western US states. The company arranges for chartered group flights from site to site and for accommodations in Spartan Motels, the only chain with accommodations near the mines, for three nights. Taylor allows Precision Metals to pick up his tab, as do the other analysts, with one exception—John Adams, an employee of a large trust company who insists on following his company's policy and paying for his hotel room himself.

> *Comment*: The policy of the company where Adams works complies closely with Standard I(B) by avoiding even the appearance of a conflict of interest, but Taylor and the other analysts were not necessarily violating Standard I(B). In general, when allowing companies to pay for travel and/or accommodations in these circumstances, members and candidates must use their judgment. They must be on guard that such arrangements not impinge on a member's or candidate's independence and objectivity. In this example, the trip was strictly for business and Taylor was not accepting irrelevant or lavish hospitality. The itinerary required chartered flights, for which analysts were not expected to pay. The accommodations were modest. These arrangements are not unusual and did not violate Standard I(B) as

> long as Taylor's independence and objectivity were not compromised. In the final analysis, members and candidates should consider both whether they can remain objective and whether their integrity might be perceived by their clients to have been compromised.

Example 2 (Research Independence):

Susan Dillon, an analyst in the corporate finance department of an investment services firm, is making a presentation to a potential new business client that includes the promise that her firm will provide full research coverage of the potential client.

> *Comment*: Dillon may agree to provide research coverage, but she must not commit her firm's research department to providing a favorable recommendation. The firm's recommendation (favorable, neutral, or unfavorable) must be based on an independent and objective investigation and analysis of the company and its securities.

Example 3 (Research Independence and Intrafirm Pressure):

Walter Fritz is an equity analyst with Hilton Brokerage who covers the mining industry. He has concluded that the stock of Metals & Mining is overpriced at its current level, but he is concerned that a negative research report will hurt the good relationship between Metals & Mining and the investment banking division of his firm. In fact, a senior manager of Hilton Brokerage has just sent him a copy of a proposal his firm has made to Metals & Mining to underwrite a debt offering. Fritz needs to produce a report right away and is concerned about issuing a less-than-favorable rating.

> *Comment*: Fritz's analysis of Metals & Mining must be objective and based solely on consideration of company fundamentals. Any pressure from other divisions of his firm is inappropriate. This conflict could have been eliminated if, in anticipation of the offering, Hilton Brokerage had placed Metals & Mining on a restricted list for its sales force.

Example 4 (Research Independence and Issuer Relationship Pressure):

As in Example 3, Walter Fritz has concluded that Metals & Mining stock is overvalued at its current level, but he is concerned that a negative research report might jeopardize a close rapport that he has nurtured over the years with Metals & Mining's CEO, chief finance officer, and investment relations officer. Fritz is concerned that a negative report might result also in management retaliation—for instance, cutting him off from participating in conference calls when a quarterly earnings release is made, denying him the ability to ask questions on such calls, and/or denying him access to top management for arranging group meetings between Hilton Brokerage clients and top Metals & Mining managers.

> *Comment*: As in Example 3, Fritz's analysis must be objective and based solely on consideration of company fundamentals. Any pressure from Metals & Mining is inappropriate. Fritz should reinforce the integrity of his conclusions by stressing that his investment recommendation is based on relative valuation, which may include qualitative issues with respect to Metals & Mining's management.

Example 5 (Research Independence and Sales Pressure):

As support for the sales effort of her corporate bond department, Lindsey Warner offers credit guidance to purchasers of fixed-income securities. Her compensation is closely linked to the performance of the corporate bond department. Near the quarter's end, Warner's firm has a large inventory position in the bonds of Milton, Ltd., and has been unable to sell the bonds because of Milton's recent announcement of an operating problem. Salespeople have asked her to contact large clients to push the bonds.

> *Comment*: Unethical sales practices create significant potential violations of the Code and Standards. Warner's opinion of the Milton bonds must not be affected by internal pressure or compensation. In this case, Warner must refuse to push the Milton bonds unless she is able to justify that the market price has already adjusted for the operating problem.

Example 6 (Research Independence and Prior Coverage):

Jill Jorund is a securities analyst following airline stocks and a rising star at her firm. Her boss has been carrying a "buy" recommendation on International Airlines and asks Jorund to take over coverage of that airline. He tells Jorund that under no circumstances should the prevailing buy recommendation be changed.

> *Comment*: Jorund must be independent and objective in her analysis of International Airlines. If she believes that her boss's instructions have compromised her, she has two options: She can tell her boss that she cannot cover the company under these constraints, or she can take over coverage of the company, reach her own independent conclusions, and if they conflict with her boss's opinion, share the conclusions with her boss or other supervisors in the firm so that they can make appropriate recommendations. Jorund must issue only recommendations that reflect her independent and objective opinion.

Example 7 (Gifts and Entertainment from Related Party):

Edward Grant directs a large amount of his commission business to a New York–based brokerage house. In appreciation for all the business, the brokerage house gives Grant two tickets to the World Cup in South Africa, two nights at a nearby resort, several meals, and transportation via limousine to the game. Grant fails to disclose receiving this package to his supervisor.

> *Comment*: Grant has violated Standard I(B) because accepting these substantial gifts may impede his independence and objectivity. Every member and candidate should endeavor to avoid situations that might cause or be perceived to cause a loss of independence or objectivity in recommending investments or taking investment action. By accepting the trip, Grant has opened himself up to the accusation that he may give the broker favored treatment in return.

Example 8 (Gifts and Entertainment from Client):

Theresa Green manages the portfolio of Ian Knowlden, a client of Tisbury Investments. Green achieves an annual return for Knowlden that is consistently better than that of the benchmark she and the client previously agreed to. As a reward, Knowlden offers Green two tickets to Wimbledon and the use of Knowlden's flat in London for a week. Green discloses this gift to her supervisor at Tisbury.

> *Comment*: Green is in compliance with Standard I(B) because she disclosed the gift from one of her clients in accordance with the firm's policies. Members and candidates may accept bonuses or gifts from clients as long as they disclose them to their employer because gifts in a client relationship are deemed less likely to affect a member's or candidate's objectivity and independence than gifts in other situations. Disclosure is required, however, so that supervisors can monitor such situations to guard against employees favoring a gift-giving client to the detriment of other fee-paying clients (such as by allocating a greater proportion of IPO stock to the gift-giving client's portfolio).
>
> Best practices for monitoring include comparing the transaction costs of the Knowlden account with the costs of other accounts managed by Green and other similar accounts within Tisbury. The supervisor could also compare the performance returns with the returns of other clients with the same mandate. This comparison will assist in determining whether a pattern of favoritism by Green is disadvantaging other Tisbury clients or the possibility that this favoritism could affect her future behavior.

Example 9 (Travel Expenses from External Manager):

Tom Wayne is the investment manager of the Franklin City Employees Pension Plan. He recently completed a successful search for a firm to manage the foreign equity allocation of the plan's diversified portfolio. He followed the plan's standard procedure of seeking presentations from a number of qualified firms and recommended that his board select Penguin Advisors because of its experience, well-defined investment strategy, and performance record. The firm claims compliance with the Global Investment Performance Standards (GIPS) and has been verified. Following the selection of Penguin, a reporter from the *Franklin City Record* calls to ask if there was any connection between this action and the fact that Penguin was one of the sponsors of an "investment fact-finding trip to Asia" that Wayne made earlier in the year. The trip was one of several conducted by the Pension Investment Academy, which had arranged the itinerary of meetings with economic, government, and corporate officials in major cities in several Asian countries. The Pension Investment Academy obtains support for the cost of these trips from a number of investment managers, including Penguin Advisors; the Academy then pays the travel expenses of the various pension plan managers on the trip and provides all meals and accommodations. The president of Penguin Advisors was also one of the travelers on the trip.

> *Comment*: Although Wayne can probably put to good use the knowledge he gained from the trip in selecting portfolio managers and in other areas of managing the pension plan, his recommendation of Penguin Advisors may be tainted by the possible conflict incurred when he participated in a trip partly paid for by Penguin Advisors and when he was in the daily company of the president of Penguin Advisors. To avoid violating Standard I(B), Wayne's basic expenses for travel and accommodations should have been paid by his employer or the pension plan; contact with the president of Penguin Advisors should have been limited to informational or educational

events only; and the trip, the organizer, and the sponsor should have been made a matter of public record. Even if his actions were not in violation of Standard I(B), Wayne should have been sensitive to the public perception of the trip when reported in the newspaper and the extent to which the subjective elements of his decision might have been affected by the familiarity that the daily contact of such a trip would encourage. This advantage would probably not be shared by firms competing with Penguin Advisors.

Example 10 (Research Independence and Compensation Arrangements):

Javier Herrero recently left his job as a research analyst for a large investment adviser. While looking for a new position, he was hired by an investor-relations firm to write a research report on one of its clients, a small educational software company. The investor-relations firm hopes to generate investor interest in the technology company. The firm will pay Herrero a flat fee plus a bonus if any new investors buy stock in the company as a result of Herrero's report.

Comment: If Herrero accepts this payment arrangement, he will be in violation of Standard I(B) because the compensation arrangement can reasonably be expected to compromise his independence and objectivity. Herrero will receive a bonus for attracting investors, which provides an incentive to draft a positive report regardless of the facts and to ignore or play down any negative information about the company. Herrero should accept only a flat fee that is not tied to the conclusions or recommendations of the report. Issuer-paid research that is objective and unbiased can be done under the right circumstances as long as the analyst takes steps to maintain his or her objectivity and includes in the report proper disclosures regarding potential conflicts of interest.

Example 11 (Recommendation Objectivity and Service Fees):

Two years ago, Bob Wade, trust manager for Central Midas Bank, was approached by Western Funds about promoting its family of funds, with special interest in the service-fee class of funds. To entice Central to promote this class, Western Funds offered to pay the bank a service fee of 0.25%. Without disclosing the fee being offered to the bank, Wade asked one of the investment managers to review Western's funds to determine whether they were suitable for clients of Central Midas Bank. The manager completed the normal due diligence review and determined that the new funds were fairly valued in the market with fee structures on a par with competitors. Wade decided to accept Western's offer and instructed the team of portfolio managers to exclusively promote these funds and the service-fee class to clients seeking to invest new funds or transfer from their current investments.

Now, two years later, the funds managed by Western begin to underperform their peers. Wade is counting on the fees to reach his profitability targets and continues to push these funds as acceptable investments for Central's clients.

Comment: Wade is violating Standard I(B) because the fee arrangement has affected the objectivity of his recommendations. Wade is relying on the fee as a component of the department's profitability and is unwilling to offer other products that may affect the fees received.

See also Standard VI(A)–Disclosure of Conflicts.

Example 12 (Recommendation Objectivity):

Bob Thompson has been doing research for the portfolio manager of the fixed-income department. His assignment is to do sensitivity analysis on securitized subprime mortgages. He has discussed with the manager possible scenarios to use to calculate expected returns. A key assumption in such calculations is housing price appreciation (HPA) because it drives "prepays" (prepayments of mortgages) and losses. Thompson is concerned with the significant appreciation experienced over the previous five years as a result of the increased availability of funds from subprime mortgages. Thompson insists that the analysis should include a scenario run with −10% for Year 1, −5% for Year 2, and then (to project a worst-case scenario) 0% for Years 3 through 5. The manager replies that these assumptions are too dire because there has never been a time in their available database when HPA was negative.

Thompson conducts his research to better understand the risks inherent in these securities and evaluates these securities in the worst-case scenario, an unlikely but possible environment. Based on the results of the enhanced scenarios, Thompson does not recommend the purchase of the securitization. Against the general market trends, the manager follows Thompson's recommendation and does not invest. The following year, the housing market collapses. In avoiding the subprime investments, the manager's portfolio outperforms its peer group that year.

> *Comment*: Thompson's actions in running the worst-case scenario against the protests of the portfolio manager are in alignment with the principles of Standard I(B). Thompson did not allow his research to be pressured by the general trends of the market or the manager's desire to limit the research to historical norms.
>
> See also Standard V(A)–Diligence and Reasonable Basis.

Example 13 (Influencing Manager Selection Decisions):

Adrian Mandel, CFA, is a senior portfolio manager for ZZYY Capital Management who oversees a team of investment professionals who manage labor union pension funds. A few years ago, ZZYY sought to win a competitive asset manager search to manage a significant allocation of the pension fund of the United Doughnut and Pretzel Bakers Union (UDPBU). UDPBU's investment board is chaired by a recognized key decision maker and long-time leader of the union, Ernesto Gomez. To improve ZZYY's chances of winning the competition, Mandel made significant monetary contributions to Gomez's union reelection campaign fund. Even after ZZYY was hired as a primary manager of the pension, Mandel believed that his firm's position was not secure. Mandel continued to contribute to Gomez's reelection campaign chest as well as to entertain lavishly the union leader and his family at top restaurants on a regular basis. All of Mandel's outlays were routinely handled as marketing expenses reimbursed by ZZYY's expense accounts and were disclosed to his senior management as being instrumental in maintaining a strong close relationship with an important client.

> *Comment*: Mandel not only offered but actually gave monetary gifts, benefits, and other considerations that reasonably could be expected to compromise Gomez's objectivity. Therefore, Mandel was in violation of Standard I(B).

Example 14 (Influencing Manager Selection Decisions):

Adrian Mandel, CFA, had heard about the manager search competition for the UDPBU Pension Fund through a broker/dealer contact. The contact told him that a well-known retired professional golfer, Bobby "The Bear" Finlay, who had become a licensed broker/dealer serving as a pension consultant, was orchestrating the UDPBU manager search. Finlay had gained celebrity status with several labor union pension fund boards by entertaining their respective board members and regaling them with colorful stories of fellow pro golfers' antics in clubhouses around the world. Mandel decided to improve ZZYY's chances of being invited to participate in the search competition by befriending Finlay to curry his favor. Knowing Finlay's love of entertainment, Mandel wined and dined Finlay at high-profile bistros where Finlay could glow in the fan recognition lavished on him by all the other patrons. Mandel's endeavors paid off handsomely when Finlay recommended to the UDPBU board that ZZYY be entered as one of three finalist asset management firms in its search.

> *Comment*: Similar to Example 13, Mandel lavished gifts, benefits, and other considerations in the form of expensive entertainment that could reasonably be expected to influence the consultant to recommend the hiring of his firm. Therefore, Mandel was in violation of Standard I(B).

Example 15 (Fund Manager Relationships):

Amie Scott is a performance analyst within her firm with responsibilities for analyzing the performance of external managers. While completing her quarterly analysis, Scott notices a change in one manager's reported composite construction. The change concealed the bad performance of a particularly large account by placing that account into a new residual composite. This change allowed the manager to remain at the top of the list of manager performance. Scott knows her firm has a large allocation to this manager, and the fund's manager is a close personal friend of the CEO. She needs to deliver her final report but is concerned with pointing out the composite change.

> *Comment*: Scott would be in violation of Standard I(B) if she did not disclose the change in her final report. The analysis of managers' performance should not be influenced by personal relationships or the size of the allocation to the outside managers. By not including the change, Scott would not be providing an independent analysis of the performance metrics for her firm.

Example 16 (Intrafirm Pressure):

Jill Stein is head of performance measurement for her firm. During the last quarter, many members of the organization's research department were removed because of the poor quality of their recommendations. The subpar research caused one larger account holder to experience significant underperformance, which resulted in the client withdrawing his money after the end of the quarter. The head of sales requests that Stein remove this account from the firm's performance composite because the performance decline can be attributed to the departed research team and not the client's adviser.

> *Comment*: Pressure from other internal departments can create situations that cause a member or candidate to violate the Code and Standards. Stein must maintain her independence and objectivity and refuse to exclude specific accounts from the firm's performance composites to which they belong. As long as the client invested under a strategy similar to that of the defined composite, it cannot be excluded because of the poor stock selections that led to the underperformance and asset withdrawal.

STANDARD I(C): PROFESSIONALISM – MISREPRESENTATION

Members and Candidates must not knowingly make any misrepresentations relating to investment analysis, recommendations, actions, or other professional activities.

Guidance

Highlights:

- *Impact on Investment Practice*
- *Performance Reporting*
- *Social Media*
- *Omissions*
- *Plagiarism*
- *Work Completed for Employer*

Trust is the foundation of the investment profession. Investors must be able to rely on the statements and information provided to them by those with whom the investors have trusted their financial well-being. Investment professionals who make false or misleading statements not only harm investors but also reduce the level of investor confidence in the investment profession and threaten the integrity of capital markets as a whole.

A misrepresentation is any untrue statement or omission of a fact or any statement that is otherwise false or misleading. A member or candidate must not knowingly omit or misrepresent information or give a false impression of a firm, organization, or security in the member's or candidate's oral representations, advertising (whether in the press or through brochures), electronic communications, or written materials (whether publicly disseminated or not). In this context, "knowingly" means that the member or candidate either knows or should have known that the misrepresentation was being made or that omitted information could alter the investment decision-making process.

Written materials include, but are not limited to, research reports, underwriting documents, company financial reports, market letters, newspaper columns, and books. Electronic communications include, but are not limited to, internet communications, webpages, mobile applications, and e-mails. Members and candidates who use webpages should regularly monitor materials posted on these sites to ensure that they contain current information. Members and candidates should also ensure that all reasonable precautions have been taken to protect the site's integrity and security and that the site does not misrepresent any information and does provide full disclosure.

Standard I(C) prohibits members and candidates from guaranteeing clients any specific return on volatile investments. Most investments contain some element of risk that makes their return inherently unpredictable. For such investments, guaranteeing either a particular rate of return or a guaranteed preservation of investment capital (e.g., "I can guarantee that you will earn 8% on equities this year" or "I can guarantee that you will not lose money on this investment") is misleading to investors.

Standard I(C) does not prohibit members and candidates from providing clients with information on investment products that have guarantees built into the structure of the products themselves or for which an institution has agreed to cover any losses.

Impact on Investment Practice

Members and candidates must not misrepresent any aspect of their practice, including (but not limited to) their qualifications or credentials, the qualifications or services provided by their firm, their performance record and the record of their firm, and the characteristics of an investment. Any misrepresentation made by a member or candidate relating to the member's or candidate's professional activities is a breach of this standard.

Members and candidates should exercise care and diligence when incorporating third-party information. Misrepresentations resulting from the use of the credit ratings, research, testimonials, or marketing materials of outside parties become the responsibility of the investment professional when it affects that professional's business practices.

Investing through outside managers continues to expand as an acceptable method of investing in areas outside a firm's core competencies. Members and candidates must disclose their intended use of external managers and must not represent those managers' investment practices as their own. Although the level of involvement of outside managers may change over time, appropriate disclosures by members and candidates are important in avoiding misrepresentations, especially if the primary activity is to invest directly with a single external manager. Standard V(B)–Communication with Clients and Prospective Clients discusses in further detail communicating the firm's investment practices.

Performance Reporting

The performance benchmark selection process is another area where misrepresentations may occur. Members and candidates may misrepresent the success of their performance record through presenting benchmarks that are not comparable to their strategies. Further, clients can be misled if the benchmark's results are not reported on a basis comparable to that of the fund's or client's results. Best practice is selecting the most appropriate available benchmark from a universe of available options. The transparent presentation of appropriate performance benchmarks is an important aspect in providing clients with information that is useful in making investment decisions.

However, Standard I(C) does not require that a benchmark always be provided in order to comply. Some investment strategies may not lend themselves to displaying an appropriate benchmark because of the complexity or diversity of the investments included. Furthermore, some investment strategies may use reference indexes that do not reflect the opportunity set of the invested assets—for example, a hedge fund comparing its performance with a "cash plus" basis. When such a benchmark is used, members and candidates should make reasonable efforts to ensure that they disclose the reasons behind the use of this reference index to avoid misrepresentations of their performance. Members and candidates should discuss with clients on a continuous basis the appropriate benchmark to be used for performance evaluations and related fee calculations.

Reporting misrepresentations may also occur when valuations for illiquid or nontraded securities are available from more than one source. When different options are available, members and candidates may be tempted to switch providers to obtain higher security valuations. The process of shopping for values may misrepresent a security's worth, lead to misinformed decisions to sell or hold an investment, and result in overcharging clients advisory fees.

Members and candidates should take reasonable steps to provide accurate and reliable security pricing information to clients on a consistent basis. Changing pricing providers should not be based solely on the justification that the new provider reports a higher current value of a security. Consistency in the reported information will improve the perception of the valuation process for illiquid securities. Clients will likely have additional confidence that they were able to make an informed decision about continuing to hold these securities in their portfolios.

Social Media

The advancement of online discussion forums and communication platforms, commonly referred to as "social media," is placing additional responsibilities on members and candidates. When communicating through social media channels, members and candidates should provide only the same information they are allowed to distribute to clients and potential clients through other traditional forms of communication. The online or interactive aspects of social media do not remove the need to be open and honest about the information being distributed.

Along with understanding and following existing and newly developing rules and regulations regarding the allowed use of social media, members and candidates should also ensure that all communications in this format adhere to the requirements of the Code and Standards. The perceived anonymity granted through these platforms may entice individuals to misrepresent their qualifications or abilities or those of their employer. Actions undertaken through social media that knowingly misrepresent investment recommendations or professional activities are considered a violation of Standard I(C).

Omissions

The omission of a fact or outcome can be misleading, especially given the growing use of models and technical analysis processes. Many members and candidates rely on such models and processes to scan for new investment opportunities, to develop investment vehicles, and to produce investment recommendations and ratings. When inputs are knowingly omitted, the resulting outcomes may provide misleading information to those who rely on it for making investment decisions. Additionally, the outcomes from models shall not be presented as fact because they represent the expected results based on the inputs and analysis process incorporated.

Omissions in the performance measurement and attribution process can also misrepresent a manager's performance and skill. Members and candidates should encourage their firms to develop strict policies for composite development to prevent cherry picking—situations in which selected accounts are presented as representative of the firm's abilities. The omission of any accounts appropriate for the defined composite may misrepresent to clients the success of the manager's implementation of its strategy.

Plagiarism

Standard I(C) also prohibits plagiarism in the preparation of material for distribution to employers, associates, clients, prospects, or the general public. Plagiarism is defined as copying or using in substantially the same form materials prepared by others without acknowledging the source of the material or identifying the author and publisher of such material. Members and candidates must not copy (or represent as their own) original ideas or material without permission and must acknowledge and identify the source of ideas or material that is not their own.

The investment profession uses a myriad of financial, economic, and statistical data in the investment decision-making process. Through various publications and presentations, the investment professional is constantly exposed to the work of others and to the temptation to use that work without proper acknowledgment.

Misrepresentation through plagiarism in investment management can take various forms. The simplest and most flagrant example is to take a research report or study done by another firm or person, change the names, and release the material as one's own original analysis. This action is a clear violation of Standard I(C). Other practices include (1) using excerpts from articles or reports prepared by others either verbatim or with only slight changes in wording without acknowledgment, (2) citing specific quotations as attributable to "leading analysts" and "investment experts" without naming the specific references, (3) presenting statistical estimates of forecasts prepared by others and identifying the sources but without including the qualifying statements or caveats that may have been used, (4) using charts and graphs without stating their sources, and (5) copying proprietary computerized spreadsheets or algorithms without seeking the cooperation or authorization of their creators.

In the case of distributing third-party, outsourced research, members and candidates may use and distribute such reports as long as they do not represent themselves as the report's authors. Indeed, the member or candidate may add value for the client by sifting through research and repackaging it for clients. In such cases, clients should be fully informed that they are paying for the ability of the member or candidate to find the best research from a wide variety of sources. Members and candidates must not misrepresent their abilities, the extent of their expertise, or the extent of their work in a way that would mislead their clients or prospective clients. Members and candidates should disclose whether the research being presented to clients comes from another source—from either within or outside the member's or candidate's firm. This allows clients to understand who has the expertise behind the report or whether the work is being done by the analyst, other members of the firm, or an outside party.

Standard I(C) also applies to plagiarism in oral communications, such as through group meetings; visits with associates, clients, and customers; use of audio/video media (which is rapidly increasing); and telecommunications, including electronic data transfer and the outright copying of electronic media.

One of the most egregious practices in violation of this standard is the preparation of research reports based on multiple sources of information without acknowledging the sources. Examples of information from such sources include ideas, statistical compilations, and forecasts combined to give the appearance of original work. Although there is no monopoly on ideas, members and candidates must give credit where it is clearly due. Analysts should not use undocumented forecasts, earnings projections, asset values, and so on. Sources must be revealed to bring the responsibility directly back to the author of the report or the firm involved.

Work Completed for Employer

The preceding paragraphs address actions that would constitute a violation of Standard I(C). In some situations, however, members or candidates may use research conducted or models developed by others within the same firm without committing a violation. The most common example relates to the situation in which one (or more) of the original analysts is no longer with the firm. Research and models developed while employed by a firm are the property of the firm. The firm retains the right to continue using the work completed after a member or candidate has left the organization. The firm may issue future reports without providing attribution to the prior analysts. A member or candidate cannot, however, reissue a previously released report solely under his or her name.

STANDARD I(C): RECOMMENDED PROCEDURES

Factual Presentations

Members and candidates can prevent unintentional misrepresentations of their qualifications or the services they or their firms provide if each member and candidate understands the limit of the firm's or individual's capabilities and the need to be accurate and complete in presentations. Firms can provide guidance for employees who make written or oral presentations to clients or potential clients by providing a written list of the firm's available services and a description of the firm's qualifications. This list should suggest ways of describing the firm's services, qualifications, and compensation that are both accurate and suitable for client or customer presentations. Firms can also help prevent misrepresentation by specifically designating which employees are authorized to speak on behalf of the firm. Regardless of whether the firm provides guidance, members and candidates should make certain that they understand the services the firm can perform and its qualifications.

Qualification Summary

In addition, to ensure accurate presentations to clients, each member and candidate should prepare a summary of his or her own qualifications and experience and a list of the services the member or candidate is capable of performing. Firms can assist member and candidate compliance by periodically reviewing employee correspondence and documents that contain representations of individual or firm qualifications.

Verify Outside Information

When providing information to clients from a third party, members and candidates share a responsibility for the accuracy of the marketing and distribution materials that pertain to the third party's capabilities, services, and products. Misrepresentation by third parties can damage the member's or candidate's reputation, the reputation of the firm, and the integrity of the capital markets. Members and candidates should encourage their employers to develop procedures for verifying information of third-party firms.

Maintain Webpages

Members and candidates who publish a webpage should regularly monitor materials posted on the site to ensure that the site contains current information. Members and candidates should also ensure that all reasonable precautions have been taken to protect the site's integrity, confidentiality, and security and that the site does not misrepresent any information and provides full disclosure.

Plagiarism Policy

To avoid plagiarism in preparing research reports or conclusions of analysis, members and candidates should take the following steps:

- *Maintain copies*: Keep copies of all research reports, articles containing research ideas, material with new statistical methodologies, and other materials that were relied on in preparing the research report.

- *Attribute quotations*: Attribute to their sources any direct quotations, including projections, tables, statistics, model/product ideas, and new methodologies prepared by persons other than recognized financial and statistical reporting services or similar sources.
- *Attribute summaries*: Attribute to their sources any paraphrases or summaries of material prepared by others. For example, to support his analysis of Brown Company's competitive position, the author of a research report on Brown might summarize another analyst's report on Brown's chief competitor, but the author of the Brown report must acknowledge in his own report the reliance on the other analyst's report.

STANDARD I(C): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Disclosure of Issuer-Paid Research):

Anthony McGuire is an issuer-paid analyst hired by publicly traded companies to electronically promote their stocks. McGuire creates a website that promotes his research efforts as a seemingly independent analyst. McGuire posts a profile and a strong buy recommendation for each company on the website indicating that the stock is expected to increase in value. He does not disclose the contractual relationships with the companies he covers on his website, in the research reports he issues, or in the statements he makes about the companies in internet chat rooms.

> *Comment*: McGuire has violated Standard I(C) because the website is misleading to potential investors. Even if the recommendations are valid and supported with thorough research, his omissions regarding the true relationship between himself and the companies he covers constitute a misrepresentation. McGuire has also violated Standard VI(A)–Disclosure of Conflicts by not disclosing the existence of an arrangement with the companies through which he receives compensation in exchange for his services.

Example 2 (Correction of Unintentional Errors):

Hijan Yao is responsible for the creation and distribution of the marketing materials for his firm, which claims compliance with the GIPS standards. Yao creates and distributes a presentation of performance by the firm's Asian equity composite that states the composite has ¥350 billion in assets. In fact, the composite has only ¥35 billion in assets, and the higher figure on the presentation is a result of a typographical error. Nevertheless, the erroneous material is distributed to a number of clients before Yao catches the mistake.

> *Comment*: Once the error is discovered, Yao must take steps to cease distribution of the incorrect material and correct the error by informing those who have received the erroneous information. Because Yao did not knowingly make the misrepresentation, however, he did not violate Standard I(C). Because his firm claims compliance with the GIPS standards, it must also comply with the GIPS Guidance Statement on Error Correction in relation to the error.

Example 3 (Noncorrection of Known Errors):

Syed Muhammad is the president of an investment management firm. The promotional material for the firm, created by the firm's marketing department, incorrectly claims that Muhammad has an advanced degree in finance from a prestigious business school in addition to the CFA designation. Although Muhammad attended the school for a short period of time, he did not receive a degree. Over the years, Muhammad and others in the firm have distributed this material to numerous prospective clients and consultants.

> *Comment*: Even though Muhammad may not have been directly responsible for the misrepresentation of his credentials in the firm's promotional material, he used this material numerous times over an extended period and should have known of the misrepresentation. Thus, Muhammad has violated Standard I(C).

Example 4 (Plagiarism):

Cindy Grant, a research analyst for a Canadian brokerage firm, has specialized in the Canadian mining industry for the past 10 years. She recently read an extensive research report on Jefferson Mining, Ltd., by Jeremy Barton, another analyst. Barton provided extensive statistics on the mineral reserves, production capacity, selling rates, and marketing factors affecting Jefferson's operations. He also noted that initial drilling results on a new ore body, which had not been made public, might show the existence of mineral zones that could increase the life of Jefferson's main mines, but Barton cited no specific data as to the initial drilling results. Grant called an officer of Jefferson, who gave her the initial drilling results over the telephone. The data indicated that the expected life of the main mines would be tripled. Grant added these statistics to Barton's report and circulated it within her firm as her own report.

> *Comment*: Grant plagiarized Barton's report by reproducing large parts of it in her own report without acknowledgment.

Example 5 (Misrepresentation of Information):

When Ricki Marks sells mortgage-backed derivatives called "interest-only strips" (IOs) to public pension plan clients, she describes them as "guaranteed by the US government." Purchasers of the IOs are entitled only to the interest stream generated by the mortgages, however, not the notional principal itself. One particular municipality's investment policies and local law require that securities purchased by its public pension plans be guaranteed by the US government. Although the underlying mortgages are guaranteed, neither the investor's investment nor the interest stream on the IOs is guaranteed. When interest rates decline, causing an increase in prepayment of mortgages, interest payments to the IOs' investors decline, and these investors lose a portion of their investment.

Comment: Marks violated Standard I(C) by misrepresenting the terms and character of the investment.

Example 6 (Potential Information Misrepresentation):

Khalouck Abdrabbo manages the investments of several high-net-worth individuals in the United States who are approaching retirement. Abdrabbo advises these individuals that a portion of their investments be moved from equity to bank-sponsored certificates of deposit and money market accounts so that the principal will be "guaranteed" up to a certain amount. The interest is not guaranteed.

Comment: Although there is risk that the institution offering the certificates of deposits and money market accounts could go bankrupt, in the United States, these accounts are insured by the US government through the Federal Deposit Insurance Corporation. Therefore, using the term "guaranteed" in this context is not inappropriate as long as the amount is within the government-insured limit. Abdrabbo should explain these facts to the clients.

Example 7 (Plagiarism):

Steve Swanson is a senior analyst in the investment research department of Ballard and Company. Apex Corporation has asked Ballard to assist in acquiring the majority ownership of stock in the Campbell Company, a financial consulting firm, and to prepare a report recommending that stockholders of Campbell agree to the acquisition. Another investment firm, Davis and Company, had already prepared a report for Apex analyzing both Apex and Campbell and recommending an exchange ratio. Apex has given the Davis report to Ballard officers, who have passed it on to Swanson. Swanson reviews the Davis report and other available material on Apex and Campbell. From his analysis, he concludes that the common stocks of Campbell and Apex represent good value at their current prices; he believes, however, that the Davis report does not consider all the factors a Campbell stockholder would need to know to make a decision. Swanson reports his conclusions to the partner in charge, who tells him to "use the Davis report, change a few words, sign your name, and get it out."

Comment: If Swanson does as requested, he will violate Standard I(C). He could refer to those portions of the Davis report that he agrees with if he identifies Davis as the source; he could then add his own analysis and conclusions to the report before signing and distributing it.

Example 8 (Plagiarism):

Claude Browning, a quantitative analyst for Double Alpha, Inc., returns from a seminar in great excitement. At that seminar, Jack Jorrely, a well-known quantitative analyst at a national brokerage firm, discussed one of his new models in great detail, and Browning is intrigued by the new concepts. He proceeds to test the model, making some minor mechanical changes but retaining the concepts, until he produces some very positive results. Browning quickly announces to his supervisors at Double Alpha that he has discovered a new model and that clients and prospective clients should be informed of this positive finding as ongoing proof of Double Alpha's continuing innovation and ability to add value.

Comment: Although Browning tested Jorrely's model on his own and even slightly modified it, he must still acknowledge the original source of the idea. Browning can certainly take credit for the final, practical results; he can also support his conclusions with his own test. The credit for the innovative thinking, however, must be awarded to Jorrely.

Example 9 (Plagiarism):

Fernando Zubia would like to include in his firm's marketing materials some "plain-language" descriptions of various concepts, such as the price-to-earnings (P/E) multiple and why standard deviation is used as a measure of risk. The descriptions come from other sources, but Zubia wishes to use them without reference to the original authors. Would this use of material be a violation of Standard I(C)?

Comment: Copying verbatim any material without acknowledgement, including plain-language descriptions of the P/E multiple and standard deviation, violates Standard I(C). Even though these concepts are general, best practice would be for Zubia to describe them in his own words or cite the sources from which the descriptions are quoted. Members and candidates would be violating Standard I(C) if they either were responsible for creating marketing materials without attribution or knowingly use plagiarized materials.

Example 10 (Plagiarism):

Through a mainstream media outlet, Erika Schneider learns about a study that she would like to cite in her research. Should she cite both the mainstream intermediary source as well as the author of the study itself when using that information?

Comment: In all instances, a member or candidate must cite the actual source of the information. Best practice for Schneider would be to obtain the information directly from the author and review it before citing it in a report. In that case, Schneider would not need to report how she found out about the information. For example, suppose Schneider read in the *Financial Times* about a study issued by CFA Institute; best practice for Schneider would be to obtain a copy of the study from CFA Institute, review it, and then cite it in her report. If she does not use any interpretation of the report from the *Financial Times* and the newspaper does not add value to the report itself, the newspaper is merely a conduit of the original information and does not need to be cited. If she does not obtain the report and review the information, Schneider runs the risk of relying on second-hand information that may misstate facts. If, for example, the *Financial Times* erroneously reported some information from the original CFA Institute study and Schneider copied that erroneous information without acknowledging CFA Institute, she could be the object of complaints. Best practice would be either to obtain the complete study from its original author and cite only that author or to use the information provided by the intermediary and cite both sources.

Example 11 (Misrepresentation of Information):

Paul Ostrowski runs a two-person investment management firm. Ostrowski's firm subscribes to a service from a large investment research firm that provides research reports that can be repackaged by smaller firms for those firms' clients. Ostrowski's firm distributes these reports to clients as its own work.

> *Comment*: Ostrowski can rely on third-party research that has a reasonable and adequate basis, but he cannot imply that he is the author of such research. If he does, Ostrowski is misrepresenting the extent of his work in a way that misleads the firm's clients or prospective clients.

Example 12 (Misrepresentation of Information):

Tom Stafford is part of a team within Appleton Investment Management responsible for managing a pool of assets for Open Air Bank, which distributes structured securities to offshore clients. He becomes aware that Open Air is promoting the structured securities as a much less risky investment than the investment management policy followed by him and the team to manage the original pool of assets. Also, Open Air has procured an independent rating for the pool that significantly overstates the quality of the investments. Stafford communicates his concerns to his supervisor, who responds that Open Air owns the product and is responsible for all marketing and distribution. Stafford's supervisor goes on to say that the product is outside of the US regulatory regime that Appleton follows and that all risks of the product are disclosed at the bottom of page 184 of the prospectus.

> *Comment*: As a member of the investment team, Stafford is qualified to recognize the degree of accuracy of the materials that characterize the portfolio, and he is correct to be worried about Appleton's responsibility for a misrepresentation of the risks. Thus, he should continue to pursue the issue of Open Air's inaccurate promotion of the portfolio according to the firm's policies and procedures.
>
> The Code and Standards stress protecting the reputation of the firm and the sustainability and integrity of the capital markets. Misrepresenting the quality and risks associated with the investment pool may lead to negative consequences for others well beyond the direct investors.

Example 13 (Avoiding a Misrepresentation):

Trina Smith is a fixed-income portfolio manager at a pension fund. She has observed that the market for highly structured mortgages is the focus of salespeople she meets and that these products represent a significant number of trading opportunities. In discussions about this topic with her team, Smith learns that calculating yields on changing cash flows within the deal structure requires very specialized vendor software. After more research, they find out that each deal is unique and that deals can have more than a dozen layers and changing cash flow priorities. Smith comes to the conclusion that, because of the complexity of these securities, the team cannot effectively distinguish between potentially good and bad investment options. To avoid misrepresenting their understanding, the team decides that the highly structured mortgage segment of the securitized market should not become part of the core of the fund's portfolio; they will allow some of the less complex securities to be part of the core.

> *Comment*: Smith is in compliance with Standard I(C) by not investing in securities that she and her team cannot effectively understand. Because she is not able to describe the risk and return profile of the securities to the pension fund beneficiaries and trustees, she appropriately limits the fund's exposure to this sector.

Example 14 (Misrepresenting Composite Construction):

Robert Palmer is head of performance for a fund manager. When asked to provide performance numbers to fund rating agencies, he avoids mentioning that the fund manager is quite liberal in composite construction. The reason accounts are included/excluded is not fully explained. The performance values reported to the rating agencies for the composites, although accurate for the accounts shown each period, may not present a true representation of the fund manager's ability.

> *Comment*: "Cherry picking" accounts to include in either published reports or information provided to rating agencies conflicts with Standard I(C). Moving accounts into or out of a composite to influence the overall performance results materially misrepresents the reported values over time. Palmer should work with his firm to strengthen its reporting practices concerning composite construction to avoid misrepresenting the firm's track record or the quality of the information being provided.

Example 15 (Presenting Out-of-Date Information):

David Finch is a sales director at a commercial bank, where he directs the bank's client advisers in the sale of third-party mutual funds. Each quarter, he holds a division-wide training session where he provides fact sheets on investment funds the bank is allowed to offer to clients. These fact sheets, which can be redistributed to potential clients, are created by the fund firms and contain information about the funds, including investment strategy and target distribution rates.

Finch knows that some of the fact sheets are out of date; for example, one long-only fund approved the use of significant leverage last quarter as a method to enhance returns. He continues to provide the sheets to the sales team without updates because the bank has no control over the marketing material released by the mutual fund firms.

> *Comment*: Finch is violating Standard I(C) by providing information that misrepresents aspects of the funds. By not providing the sales team and, ultimately, the clients with the updated information, he is misrepresenting the potential risks associated with the funds with outdated fact sheets. Finch can instruct the sales team to clarify the deficiencies in the fact sheets with clients and ensure they have the most recent fund prospectus document before accepting orders for investing in any fund.

Example 16 (Overemphasis of Firm Results):

Bob Anderson is chief compliance officer for Optima Asset Management Company, a firm currently offering eight funds to clients. Seven of the eight had 10-year returns below the median for their respective sectors. Anderson approves a recent advertisement, which includes this statement: "Optima Asset Management is achieving excellent returns for its investors. The Optima Emerging Markets Equity fund, for example, has 10-year returns that exceed the sector median by more than 10%."

Comment: From the information provided it is difficult to determine whether a violation has occurred as long as the sector outperformance is correct. Anderson may be attempting to mislead potential clients by citing the performance of the sole fund that achieved such results. Past performance is often used to demonstrate a firm's skill and abilities in comparison to funds in the same sectors.

However, if all the funds outperformed their respective benchmarks, then Anderson's assertion that the company "is achieving excellent returns" may be factual. Funds may exhibit positive returns for investors, exceed benchmarks, and yet have returns below the median in their sectors.

Members and candidates need to ensure that their marketing efforts do not include statements that misrepresent their skills and abilities to remain compliant with Standard I(C). Unless the returns of a single fund reflect the performance of a firm as a whole, the use of a singular fund for performance comparisons should be avoided.

STANDARD I(D): PROFESSIONALISM – MISCONDUCT

Members and Candidates must not engage in any professional conduct involving dishonesty, fraud, or deceit or commit any act that reflects adversely on their professional reputation, integrity, or competence.

Guidance

Whereas Standard I(A) addresses the obligation of members and candidates to comply with applicable law that governs their professional activities, Standard I(D) addresses *all* conduct that reflects poorly on the professional integrity, good reputation, or competence of members and candidates. Any act that involves lying, cheating, stealing, or other dishonest conduct is a violation of this standard if the offense reflects adversely on a member's or candidate's professional activities. Although CFA Institute discourages any sort of unethical behavior by members and candidates, the Code and Standards are primarily aimed at conduct and actions related to a member's or candidate's professional life.

Conduct that damages trustworthiness or competence may include behavior that, although not illegal, nevertheless negatively affects a member's or candidate's ability to perform his or her responsibilities. For example, abusing alcohol during business hours might constitute a violation of this standard because it could have a detrimental effect on the member's or candidate's ability to fulfill his or her professional responsibilities. Personal bankruptcy may not reflect on the integrity or trustworthiness of the person declaring bankruptcy, but if the circumstances of the bankruptcy involve fraudulent or deceitful business conduct, the bankruptcy may be a violation of this standard.

In some cases, the absence of appropriate conduct or the lack of sufficient effort may be a violation of Standard I(D). The integrity of the investment profession is built on trust. A member or candidate—whether an investment banker, rating or research analyst, or portfolio manager—is expected to conduct the necessary due diligence to properly understand the nature and risks of an investment before making an investment recommendation. By not taking these steps and, instead, relying on someone else in the process to perform them, members or candidates may violate the trust their clients have placed in them. This loss of trust may have a significant impact on the reputation of the member or candidate and the operations of the financial market as a whole.

Individuals may attempt to abuse the CFA Institute Professional Conduct Program by actively seeking CFA Institute enforcement of the Code and Standards, and Standard I(D) in particular, as a method of settling personal, political, or other disputes unrelated to professional ethics. CFA Institute is aware of this issue, and appropriate disciplinary policies, procedures, and enforcement mechanisms are in place to address misuse of the Code and Standards and the Professional Conduct Program in this way.

STANDARD I(D): RECOMMENDED PROCEDURES

In addition to ensuring that their own behavior is consistent with Standard I(D), to prevent general misconduct, members and candidates should encourage their firms to adopt the following policies and procedures to support the principles of Standard I(D):

- *Code of ethics*: Develop and/or adopt a code of ethics to which every employee must subscribe, and make clear that any personal behavior that reflects poorly on the individual involved, the institution as a whole, or the investment industry will not be tolerated.
- *List of violations*: Disseminate to all employees a list of potential violations and associated disciplinary sanctions, up to and including dismissal from the firm.
- *Employee references*: Check references of potential employees to ensure that they are of good character and not ineligible to work in the investment industry because of past infractions of the law.

STANDARD I(D): APPLICATION OF THE STANDARD

a demonstrate the application of the Code of Ethics and Standards of Professional Conduct to situations involving issues of professional integrity;

b identify conduct that conforms to the Code and Standards and conduct that violates the Code and Standards;

Example 1 (Professionalism and Competence):

Simon Sasserman is a trust investment officer at a bank in a small affluent town. He enjoys lunching every day with friends at the country club, where his clients have observed him having numerous drinks. Back at work after lunch, he clearly is intoxicated while making investment decisions. His colleagues make a point of handling any business with Sasserman in the morning because they distrust his judgment after lunch.

> *Comment*: Sasserman's excessive drinking at lunch and subsequent intoxication at work constitute a violation of Standard I(D) because this conduct has raised questions about his professionalism and competence. His behavior reflects poorly on him, his employer, and the investment industry.

Example 2 (Fraud and Deceit):

Howard Hoffman, a security analyst at ATZ Brothers, Inc., a large brokerage house, submits reimbursement forms over a two-year period to ATZ's self-funded health insurance program for more than two dozen bills, most of which have been altered to increase the amount due. An investigation by the firm's director of employee benefits uncovers the inappropriate conduct. ATZ subsequently terminates Hoffman's employment and notifies CFA Institute.

> *Comment*: Hoffman violated Standard I(D) because he engaged in intentional conduct involving fraud and deceit in the workplace that adversely reflected on his integrity.

Example 3 (Fraud and Deceit):

Jody Brink, an analyst covering the automotive industry, volunteers much of her spare time to local charities. The board of one of the charitable institutions decides to buy five new vans to deliver hot lunches to low-income elderly people. Brink offers to donate her time to handle purchasing agreements. To pay a long-standing debt to a friend who operates an automobile dealership—and to compensate herself for her trouble—she agrees to a price 20% higher than normal and splits the surcharge with her friend. The director of the charity ultimately discovers the scheme and tells Brink that her services, donated or otherwise, are no longer required.

> *Comment*: Brink engaged in conduct involving dishonesty, fraud, and misrepresentation and has violated Standard I(D).

Example 4 (Personal Actions and Integrity):

Carmen Garcia manages a mutual fund dedicated to socially responsible investing. She is also an environmental activist. As the result of her participation in nonviolent protests, Garcia has been arrested on numerous occasions for trespassing on the property of a large petrochemical plant that is accused of damaging the environment.

> *Comment*: Generally, Standard I(D) is not meant to cover legal transgressions resulting from acts of civil disobedience in support of personal beliefs because such conduct does not reflect poorly on the member's or candidate's professional reputation, integrity, or competence.

Example 5 (Professional Misconduct):

Meredith Rasmussen works on a buy-side trading desk of an investment management firm and concentrates on in-house trades for a hedge fund subsidiary managed by a team at the investment management firm. The hedge fund has been very successful and is marketed globally by the firm. From her experience as the trader for much of the activity of the fund, Rasmussen has become quite knowledgeable about the hedge fund's strategy, tactics, and performance. When a distinct break in the market occurs and many of the securities involved in the hedge fund's strategy decline markedly in value, Rasmussen observes that the reported performance of the hedge fund does not reflect this decline. In her experience, the lack of effect is a very unlikely occurrence. She approaches the head of trading about her concern and is told that she should not ask any questions and that the fund is big and successful and is not her concern. She is fairly sure something is not right, so she contacts the compliance officer, who also tells her to stay away from the issue of the hedge fund's reporting.

Comment: Rasmussen has clearly come across an error in policies, procedures, and compliance practices within the firm's operations. According to the firm's procedures for reporting potentially unethical activity, she should pursue the issue by gathering some proof of her reason for doubt. Should all internal communications within the firm not satisfy her concerns, Rasmussen should consider reporting the potential unethical activity to the appropriate regulator.

See also Standard IV(A) for guidance on whistleblowing and Standard IV(C) for the duties of a supervisor.

STANDARD II(A): INTEGRITY OF CAPITAL MARKETS - MATERIAL NONPUBLIC INFORMATION

Standard II(A) Material Nonpublic Information

Members and Candidates who possess material nonpublic information that could affect the value of an investment must not act or cause others to act on the information.

Guidance

Highlights:

- *What Is "Material" Information?*
- *What Constitutes "Nonpublic" Information?*
- *Mosaic Theory*
- *Social Media*
- *Using Industry Experts*
- *Investment Research Reports*

Trading or inducing others to trade on material nonpublic information erodes confidence in capital markets, institutions, and investment professionals by supporting the idea that those with inside information and special access can take unfair advantage of the general investing public. Although trading on inside information may lead to short-term profits, in the long run, individuals and the profession as a whole suffer from such trading. These actions have caused and will continue to cause investors to avoid capital markets because the markets are perceived to be "rigged" in favor of the knowledgeable insider. When the investing public avoids capital markets, the markets and capital allocation become less efficient and less supportive of strong and vibrant economies. Standard II(A) promotes and maintains a high level of confidence in market integrity, which is one of the foundations of the investment profession.

The prohibition on using this information goes beyond the direct buying and selling of individual securities or bonds. Members and candidates must not use material nonpublic information to influence their investment actions related to derivatives (e.g., swaps or option contracts), mutual funds, or other alternative investments. *Any* trading based on material nonpublic information constitutes a violation of Standard II(A).

The expansion of financial products and the increasing interconnectivity of financial markets globally have resulted in new potential opportunities for trading on material nonpublic information.

What Is "Material" Information?

Information is "material" if its disclosure would probably have an impact on the price of a security or if reasonable investors would want to know the information before making an investment decision. In other words, information is material if it would significantly alter the total mix of information currently available about a security in such a way that the price of the security would be affected.

The specificity of the information, the extent of its difference from public information, its nature, and its reliability are key factors in determining whether a particular piece of information fits the definition of material. For example, material information may include, but is not limited to, information on the following:

- earnings;
- mergers, acquisitions, tender offers, or joint ventures;
- changes in assets or asset quality;
- innovative products, processes, or discoveries (e.g., new product trials or research efforts);
- new licenses, patents, registered trademarks, or regulatory approval/rejection of a product;
- developments regarding customers or suppliers (e.g., the acquisition or loss of a contract);
- changes in management;
- change in auditor notification or the fact that the issuer may no longer rely on an auditor's report or qualified opinion;
- events regarding the issuer's securities (e.g., defaults on senior securities, calls of securities for redemption, repurchase plans, stock splits, changes in dividends, changes to the rights of security holders, and public or private sales of additional securities);
- bankruptcies;
- significant legal disputes;
- government reports of economic trends (employment, housing starts, currency information, etc.);
- orders for large trades before they are executed; and
- new or changing equity or debt ratings issued by a third party (e.g., sell-side recommendations and credit ratings).

In addition to the substance and specificity of the information, the source or relative reliability of the information also determines materiality. The less reliable a source, the less likely the information provided would be considered material. For example, factual information from a corporate insider regarding a significant new contract for a company is likely to be material, whereas an assumption based on speculation by a competitor about the same contract is likely to be less reliable and, therefore, not material. Additionally, information about trials of a new drug, product, or service under development from qualified personnel involved in the trials is likely to be material, whereas educated conjecture by subject experts not connected to the trials is unlikely to be material.

Also, the more ambiguous the effect of the information on price, the less material that information is considered. If it is unclear whether and to what extent the information will affect the price of a security, the information may not be considered material. The passage of time may also render information that was once important immaterial.

What Constitutes "Nonpublic" Information?

Information is "nonpublic" until it has been disseminated or is available to the marketplace in general (as opposed to a select group of investors). "Disseminated" can be defined as "made known." For example, a company report of profits that is posted on the internet and distributed widely through a press release or accompanied by a filing has been effectively disseminated to the marketplace. Members and candidates must have a reasonable expectation that people have received the information before it can be considered public. It is not necessary, however, to wait for the slowest method of delivery. Once the information is disseminated to the market, it is public information that is no longer covered by this standard.

Members and candidates must be particularly aware of information that is selectively disclosed by corporations to a small group of investors, analysts, or other market participants. Information that is made available to analysts remains nonpublic until it is made available to investors in general. Corporations that disclose information on a limited basis create the potential for insider-trading violations.

Issues of selective disclosure often arise when a corporate insider provides material information to analysts in a briefing or conference call before that information is released to the public. Analysts must be aware that a disclosure made to a room full of analysts does not necessarily make the disclosed information "public." Analysts should also be alert to the possibility that they are selectively receiving material nonpublic information when a company provides them with guidance or interpretation of such publicly available information as financial statements or regulatory filings.

A member or candidate may use insider information provided legitimately by the source company for the specific purpose of conducting due diligence according to the business agreement between the parties for such activities as mergers, loan underwriting, credit ratings, and offering engagements. In such instances, the investment professional would not be considered in violation of Standard II(A) by using the material information. However, the use of insider information provided by the source company for other purposes, especially to trade or entice others to trade the securities of the firm, conflicts with this standard.

Mosaic Theory

A financial analyst gathers and interprets large quantities of information from many sources. The analyst may use significant conclusions derived from the analysis of public and nonmaterial nonpublic information as the basis for investment recommendations and decisions even if those conclusions would have been material inside information had they been communicated directly to the analyst by a company. Under the "mosaic theory," financial analysts are free to act on this collection, or mosaic, of information without risking violation.

The practice of financial analysis depends on the free flow of information. For the fair and efficient operation of the capital markets, analysts and investors must have the greatest amount of information possible to facilitate making well-informed investment decisions about how and where to invest capital. Accurate, timely, and intelligible communication is essential if analysts and investors are to obtain the data needed to make informed decisions about how and where to invest capital. These disclosures must go beyond the information mandated by the reporting requirements of the securities laws and should include specific business information about items used to guide a company's future growth, such as new products, capital projects, and the competitive environment. Analysts seek and use such information to compare and contrast investment alternatives.

Much of the information used by analysts comes directly from companies. Analysts often receive such information through contacts with corporate insiders, especially investor-relations staff and financial officers. Information may be disseminated in the form of press releases, through oral presentations by company executives in analysts' meetings or conference calls, or during analysts' visits to company premises. In seeking to develop the most accurate and complete picture of a company, analysts should also reach beyond contacts with companies themselves and collect information from other sources, such as customers, contractors, suppliers, and the companies' competitors.

Analysts are in the business of formulating opinions and insights that are not obvious to the general investing public about the attractiveness of particular securities. In the course of their work, analysts actively seek out corporate information not generally known to the market for the express purpose of analyzing that information, forming an opinion on its significance, and informing their clients, who can be expected to trade on the basis of the recommendation. Analysts' initiatives to discover and analyze information and communicate their findings to their clients significantly enhance market efficiency, thus benefiting all investors (see *Dirks v. Securities and Exchange Commission*). Accordingly, violations of Standard II(A) will *not* result when a perceptive analyst reaches a conclusion about a corporate action or event through an analysis of public information and items of nonmaterial nonpublic information.

Investment professionals should note, however, that although analysts are free to use mosaic information in their research reports, they should save and document all their research [see Standard V(C)–Record Retention]. Evidence of the analyst's knowledge of public and nonmaterial nonpublic information about a corporation strengthens the assertion that the analyst reached his or her conclusions solely through appropriate methods rather than through the use of material nonpublic information.

Social Media

The continuing advancement in technology allows members, candidates, and the industry at large to exchange information at rates not previously available. It is important for investment professionals to understand the implications of using information from the internet and social media platforms because all such information may not actually be considered public.

Some social media platforms require membership in specific groups in order to access the published content. Members and candidates participating in groups with membership limitations should verify that material information obtained from these sources can also be accessed from a source that would be considered available to the public (e.g., company filings, webpages, and press releases).

Members and candidates may use social media platforms to communicate with clients or investors without conflicting with this standard. As long as the information reaches all clients or is open to the investing public, the use of these platforms would be comparable with other traditional forms of communications, such as e-mails and press releases. Members and candidates, as required by Standard I(A), should also complete all appropriate regulatory filings related to information distributed through social media platforms.

Using Industry Experts

The increased demand for insights for understanding the complexities of some industries has led to an expansion of engagement with outside experts. As the level of engagement increased, new businesses formed to connect analysts and investors with individuals who have specialized knowledge of their industry (e.g., technology or pharmaceuticals). These networks offer investors the opportunity to reach beyond their usual business circles to speak with experts regarding economic conditions, industry trends, and technical issues relating to specific products and services.

Members and candidates may provide compensation to individuals for their insights without violating this standard. However, members and candidates are ultimately responsible for ensuring that they are not requesting or acting on confidential information received from external experts, which is in violation of security regulations and laws or duties to others. As the recent string of insider-trading cases displayed, some experts are willing to provide confidential and protected information for the right incentive.

Firms connecting experts with members or candidates often require both parties to sign agreements concerning the disclosure of material nonpublic information. Even with the protections from such compliance practices, if an expert provides material nonpublic information, members and candidates would be prohibited from taking investment actions on the associated firm until the information became publicly known to the market.

Investment Research Reports

When a particularly well-known or respected analyst issues a report or makes changes to his or her recommendation, that information alone may have an effect on the market and thus may be considered material. Theoretically, under Standard II(A), such a report would have to be made public at the time it was distributed to clients. The analyst is not a company insider, however, and does not have access to inside information. Presumably, the analyst created the report from information available to the public (mosaic theory) and by using his or her expertise to interpret the information. The analyst's hard work, paid for by the client, generated the conclusions.

Simply because the public in general would find the conclusions material does not require that the analyst make his or her work public. Investors who are not clients of the analyst can either do the work themselves or become clients of the analyst to gain access to the analyst's expertise.

STANDARD II(A): RECOMMENDED PROCEDURES

Achieve Public Dissemination

If a member or candidate determines that information is material, the member or candidate should make reasonable efforts to achieve public dissemination of the information. These efforts usually entail encouraging the issuing company to make the information public. If public dissemination is not possible, the member or candidate must communicate the information only to the designated supervisory and compliance personnel within the member's or candidate's firm and must not take investment action or alter current investment recommendations on the basis of the information. Moreover, members and candidates must not knowingly engage in any conduct that may induce company insiders to privately disclose material nonpublic information.

Adopt Compliance Procedures

Members and candidates should encourage their firms to adopt compliance procedures to prevent the misuse of material nonpublic information. Particularly important is improving compliance in such areas as the review of employee and proprietary trading, the review of investment recommendations, documentation of firm procedures, and the supervision of interdepartmental communications in multiservice firms. Compliance procedures should suit the particular characteristics of a firm, including its size and the nature of its business.

Members and candidates are encouraged to inform their supervisor and compliance personnel of suspected inappropriate use of material nonpublic information as the basis for security trading activities or recommendations being made within their firm.

Adopt Disclosure Procedures

Members and candidates should encourage their firms to develop and follow disclosure policies designed to ensure that information is disseminated to the marketplace in an equitable manner. For example, analysts from small firms should receive the same information and attention from a company as analysts from large firms receive. Similarly, companies should not provide certain information to buy-side analysts but not to sell-side analysts, or vice versa. Furthermore, a company should not discriminate among analysts in the provision of information or "blackball" particular analysts who have given negative reports on the company in the past.

Within investment and research firms, members and candidates should encourage the development of and compliance with procedures for distributing new and updated investment opinions to clients. Recommendations of this nature may represent material market-moving information that needs to be communicated to all clients fairly.

Issue Press Releases

Companies should consider issuing press releases prior to analyst meetings and conference calls and scripting those meetings and calls to decrease the chance that further information will be disclosed. If material nonpublic information is disclosed for the first time in an analyst meeting or call, the company should promptly issue a press release or otherwise make the information publicly available.

Firewall Elements

An information barrier commonly referred to as a "firewall" is the most widely used approach for preventing the communication of material nonpublic information within firms. It restricts the flow of confidential information to those who need to know the information to perform their jobs effectively. The minimum elements of such a system include, but are not limited to, the following:

- substantial control of relevant interdepartmental communications, preferably through a clearance area within the firm in either the compliance or legal department;
- review of employee trading through the maintenance of "watch," "restricted," and "rumor" lists;
- documentation of the procedures designed to limit the flow of information between departments and of the actions taken to enforce those procedures; and
- heightened review or restriction of proprietary trading while a firm is in possession of material nonpublic information.

Appropriate Interdepartmental Communications

Although documentation requirements must, for practical reasons, take into account the differences between the activities of small firms and those of large, multiservice firms, firms of all sizes and types benefit by improving the documentation of their internal enforcement of firewall procedures. Therefore, even at small firms, procedures concerning interdepartmental communication, the review of trading activity, and the investigation of possible violations should be compiled and formalized.

Physical Separation of Departments

As a practical matter, to the greatest extent possible, firms should consider the physical separation of departments and files to prevent the communication of sensitive information that should not be shared. For example, the investment banking and corporate finance areas of a brokerage firm should be separated from the sales and research departments, and a bank's commercial lending department should be segregated from its trust and research departments.

Prevention of Personnel Overlap

There should be no overlap of personnel between the investment banking and corporate finance areas of a brokerage firm and the sales and research departments or between a bank's commercial lending department and its trust and research departments. For a firewall to be effective in a multiservice firm, an employee should be on only one side of the firewall at any time. Inside knowledge may not be limited to information about a specific offering or the current financial condition of a company. Analysts may be exposed to much information about the company, including new product developments or future budget projections that clearly constitute inside knowledge and thus preclude the analyst from returning to his or her research function. For example, an analyst who follows a particular company may provide limited assistance to the investment bankers under carefully controlled circumstances when the firm's investment banking department is involved in a deal with the company. That analyst must then be treated as though he or she were an investment banker; the analyst must remain on the investment banking side of the wall until any information he or she learns is publicly disclosed. In short, the analyst cannot use any information learned in the course of the project for research purposes and cannot share that information with colleagues in the research department.

A Reporting System

A primary objective of an effective firewall procedure is to establish a reporting system in which authorized people review and approve communications between departments. If an employee behind a firewall believes that he or she needs to share confidential information with someone on the other side of the wall, the employee should consult a designated compliance officer to determine whether sharing the information is necessary and how much information should be shared. If the sharing is necessary, the compliance officer should coordinate the process of "looking over the wall" so that the necessary information will be shared and the integrity of the procedure will be maintained.

A single supervisor or compliance officer should have the specific authority and responsibility of deciding whether information is material and whether it is sufficiently public to be used as the basis for investment decisions. Ideally, the supervisor or compliance officer responsible for communicating information to a firm's research or brokerage area would not be a member of that area.

Personal Trading Limitations

Firms should consider restrictions or prohibitions on personal trading by employees and should carefully monitor both proprietary trading and personal trading by employees. Firms should require employees to make periodic reports (to the extent that such reporting is not already required by securities laws) of their own transactions and transactions made for the benefit of family members. Securities should be placed on a restricted list when a firm has or may have material nonpublic information. The broad distribution of a restricted list often triggers the sort of trading the list was developed to avoid. Therefore, a watch list shown to only the few people responsible for compliance should be used to monitor transactions in specified securities. The use of a watch list in combination with a restricted list is an increasingly common means of ensuring effective control of personal trading.

Record Maintenance

Multiservice firms should maintain written records of the communications between various departments. Firms should place a high priority on training and should consider instituting comprehensive training programs, particularly for employees in sensitive areas.

Proprietary Trading Procedures

Procedures concerning the restriction or review of a firm's proprietary trading while the firm possesses material nonpublic information will necessarily depend on the types of proprietary trading in which the firm may engage. A prohibition on all types of proprietary activity when a firm comes into possession of material nonpublic information is *not* appropriate. For example, when a firm acts as a market maker, a prohibition on proprietary trading may be counterproductive to the goals of maintaining the confidentiality of information and market liquidity. This concern is particularly important in the relationships between small, regional broker/dealers and small issuers. In many situations, a firm will take a small issuer public with the understanding that the firm will continue to be a market maker in the stock. In such instances, a withdrawal by the firm from market-making activities would be a clear tip to outsiders. Firms that continue market-making activity while in the possession of material nonpublic information should, however, instruct their market makers to remain passive with respect to the market—that is, to take only the contra side of unsolicited customer trades.

In risk-arbitrage trading, the case for a trading prohibition is more compelling than it is in the case of market making. The impetus for arbitrage trading is neither passive nor reactive, and the potential for illegal profits is greater than in market making. The most prudent course for firms is to suspend arbitrage activity when a security is placed on the watch list. Those firms that continue arbitrage activity face a high hurdle in proving the adequacy of their internal procedures for preventing trading on material nonpublic information and must demonstrate a stringent review and documentation of firm trades.

Communication to All Employees

Members and candidates should encourage their employers to circulate written compliance policies and guidelines to all employees. Policies and guidelines should be used in conjunction with training programs aimed at enabling employees to recognize material nonpublic information. Such information is not always clearly identifiable.

Employees must be given sufficient training to either make an informed decision or to realize they need to consult a supervisor or compliance officer before engaging in questionable transactions. Appropriate policies reinforce that using material nonpublic information is illegal in many countries. Such trading activities based on material nonpublic information undermine the integrity of the individual, the firm, and the capital markets.

STANDARD II(A): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Acting on Nonpublic Information):

Frank Barnes, the president and controlling shareholder of the SmartTown clothing chain, decides to accept a tender offer and sell the family business at a price almost double the market price of its shares. He describes this decision to his sister (SmartTown's treasurer), who conveys it to her daughter (who owns no stock in the family company at present), who tells her husband, Staple. Staple, however, tells his stockbroker, Alex Halsey, who immediately buys SmartTown stock for himself.

> *Comment*: The information regarding the pending sale is both material and nonpublic. Staple has violated Standard II(A) by communicating the inside information to his broker. Halsey also has violated the standard by buying the shares on the basis of material nonpublic information.

Example 2 (Controlling Nonpublic Information):

Samuel Peter, an analyst with Scotland and Pierce Incorporated, is assisting his firm with a secondary offering for Bright Ideas Lamp Company. Peter participates, via telephone conference call, in a meeting with Scotland and Pierce investment banking employees and Bright Ideas' CEO. Peter is advised that the company's earnings projections for the next year have significantly dropped. Throughout the telephone conference call, several Scotland and Pierce salespeople and portfolio managers walk in and out of Peter's office, where the telephone call is taking place. As a result, they are aware of the drop in projected earnings for Bright Ideas. Before the conference call is concluded, the salespeople trade the stock of the company on behalf of the firm's clients and other firm personnel trade the stock in a firm proprietary account and in employees' personal accounts.

> *Comment*: Peter has violated Standard II(A) because he failed to prevent the transfer and misuse of material nonpublic information to others in his firm. Peter's firm should have adopted information barriers to prevent the communication of nonpublic information between departments of the firm. The salespeople and portfolio managers who traded on the information have also violated Standard II(A) by trading on inside information.

Example 3 (Selective Disclosure of Material Information):

Elizabeth Levenson is based in Hanoi and covers the Vietnamese market for her firm, which is based in Singapore. She is invited, together with the other 10 largest shareholders of a manufacturing company, to meet the finance director of that company. During the meeting, the finance director states that the company expects its workforce to strike next Friday, which will cripple productivity and distribution. Can Levenson use this information as a basis to change her rating on the company from "buy" to "sell"?

> *Comment*: Levenson must first determine whether the material information is public. According to Standard II(A), if the company has not made this information public (a small group forum does not qualify as a method of public dissemination), she cannot use the information.

Example 4 (Determining Materiality):

Leah Fechtman is trying to decide whether to hold or sell shares of an oil-and-gas exploration company that she owns in several of the funds she manages. Although the company has underperformed the index for some time already, the trends in the industry sector signal that companies of this type might become takeover targets. While she is considering her decision, her doctor, who casually follows the markets, mentions that she thinks that the company in question will soon be bought out by a large multinational conglomerate and that it would be a good idea to buy the stock right now. After talking to various investment professionals and checking their opinions on the company as well as checking industry trends, Fechtman decides the next day to accumulate more stock in the oil-and-gas exploration company.

> *Comment*: Although information on an expected takeover bid may be of the type that is generally material and nonpublic, in this case, the source of information is unreliable, so the information cannot be considered material. Therefore, Fechtman is not prohibited from trading the stock on the basis of this information.

Example 5 (Applying the Mosaic Theory):

Jagdish Teja is a buy-side analyst covering the furniture industry. Looking for an attractive company to recommend as a buy, he analyzes several furniture makers by studying their financial reports and visiting their operations. He also talks to some designers and retailers to find out which furniture styles are trendy and popular. Although none of the companies that he analyzes are a clear buy, he discovers that one of them, Swan Furniture Company (SFC), may be in financial trouble. SFC's extravagant new designs have been introduced at substantial cost. Even though these designs initially attracted attention, the public is now buying more conservative furniture from other makers. Based on this information and on a profit-and-loss analysis, Teja believes that SFC's next quarter earnings will drop substantially. He issues a sell recommendation for SFC. Immediately after receiving that recommendation, investment managers start reducing the SFC stock in their portfolios.

> *Comment*: Information on quarterly earnings data is material and nonpublic. Teja arrived at his conclusion about the earnings drop on the basis of public information and on pieces of nonmaterial nonpublic information (such as opinions of designers and retailers). Therefore, trading based on Teja's correct conclusion is not prohibited by Standard II(A).

Example 6 (Applying the Mosaic Theory):

Roger Clement is a senior financial analyst who specializes in the European automobile sector at Rivoli Capital. Because he has been repeatedly nominated by many leading industry magazines and newsletters as a "best analyst" for the automobile industry, he is widely regarded as an authority on the sector. After speaking with representatives of Turgot Chariots—a European auto manufacturer with sales primarily in South Korea—and after conducting interviews with salespeople, labor leaders, his firm's Korean currency analysts, and banking officials, Clement analyzed Turgot Chariots and concluded that (1) its newly introduced model will probably not meet sales expectations, (2) its corporate restructuring strategy may well face serious opposition from unions, (3) the depreciation of the Korean won should lead to pressure on margins for the industry in general and Turgot's market segment in particular, and (4) banks could take a tougher-than-expected stance in the upcoming round of credit renegotiations with the company. For these reasons, he changes his conclusion about the company from "market outperform" to "market underperform." Clement retains the support material used to reach his conclusion in case questions later arise.

> *Comment*: To reach a conclusion about the value of the company, Clement has pieced together a number of nonmaterial or public bits of information that affect Turgot Chariots. Therefore, under the mosaic theory, Clement has not violated Standard II(A) in drafting the report.

Example 7 (Analyst Recommendations as Material Nonpublic Information):

The next day, Clement is preparing to be interviewed on a global financial news television program where he will discuss his changed recommendation on Turgot Chariots for the first time in public. While preparing for the program, he mentions to the show's producers and Mary Zito, the journalist who will be interviewing him, the information he will be discussing. Just prior to going on the air, Zito sells her holdings in Turgot Chariots. She also phones her father with the information because she knows that he and other family members have investments in Turgot Chariots.

> *Comment*: When Zito receives advance notice of Clement's change of opinion, she knows it will have a material impact on the stock price, even if she is not totally aware of Clement's underlying reasoning. She is not a client of Clement but obtains early access to the material nonpublic information prior to publication. Her trades are thus based on material nonpublic information and violate Standard II(A).
>
> Zito further violates the Standard by relaying the information to her father. It would not matter if he or any other family member traded; the act of providing the information violates Standard II(A). The fact that the information is provided to a family member does not absolve someone of the prohibition of using or communicating material nonpublic information.

Example 8 (Acting on Nonpublic Information):

Ashton Kellogg is a retired investment professional who manages his own portfolio. He owns shares in National Savings, a large local bank. A close friend and golfing buddy, John Mayfield, is a senior executive at National. National has seen its stock price drop considerably, and the news and outlook are not good. In a conversation about the economy and the banking industry on the golf course, Mayfield relays the information that National will surprise the investment community in a few days when it announces excellent earnings for the quarter. Kellogg is pleasantly surprised by this information, and thinking that Mayfield, as a senior executive, knows the law and would not disclose inside information, he doubles his position in the bank. Subsequently, National announces that it had good operating earnings but had to set aside reserves for anticipated significant losses on its loan portfolio. The combined news causes the stock to go down 60%.

> *Comment*: Even though Kellogg believes that Mayfield would not break the law by disclosing inside information and money was lost on the purchase, Kellogg should not have purchased additional shares of National. It is the member's or candidate's responsibility to make sure, before executing investment actions, that comments about earnings are not material non-public information. Kellogg has violated Standard II(A).

Example 9 (Mosaic Theory):

John Doll is a research analyst for a hedge fund that also sells its research to a select group of paying client investment firms. Doll's focus is medical technology companies and products, and he has been in the business long enough and has been successful enough to build up a very credible network of friends and experts in the business. Doll has been working on a major research report recommending Boyce Health, a medical device manufacturer. He recently ran into an old acquaintance at a wedding who is a senior executive at Boyce, and Doll asked about the business. Doll was drawn to a statement that the executive, who has responsibilities in the new products area, made about a product: "I would not get too excited about the medium-term prospects; we have a lot of work to do first." Doll incorporated this and other information about the new Boyce product in his long-term recommendation of Boyce.

> *Comment*: Doll's conversation with the senior executive is part of the mosaic of information used in recommending Boyce. When holding discussions with a firm executive, Doll would need to guard against soliciting or obtaining material nonpublic information. Before issuing the report, the executive's statement about the continuing development of the product would need to be weighed against the other known public facts to determine whether it would be considered material.

Example 10 (Materiality Determination):

Larry Nadler, a trader for a mutual fund, gets a text message from another firm's trader, whom he has known for years. The message indicates a software company is going to report strong earnings when the firm publicly announces in two days. Nadler has a buy order from a portfolio manager within his firm to purchase several hundred thousand shares of the stock. Nadler is aggressive in placing the portfolio manager's order and completes the purchases by the following morning, a day ahead of the firm's planned earnings announcement.

> *Comment*: There are often rumors and whisper numbers before a release of any kind. The text message from the other trader would most likely be considered market noise. Unless Nadler knew that the trader had an ongoing business relationship with the public firm, he had no reason to suspect he was receiving material nonpublic information that would prevent him from completing the trading request of the portfolio manager.

Example 11 (Using an Expert Network):

Mary McCoy is the senior drug analyst at a mutual fund. Her firm hires a service that connects her to experts in the treatment of cancer. Through various phone conversations, McCoy enhances her understanding of the latest therapies for successful treatment. This information is critical to Mary making informed recommendations of the companies producing these drugs.

> *Comment*: McCoy is appropriately using the expert networks to enhance her evaluation process. She has neither asked for nor received information that may be considered material and nonpublic, such as preliminary trial results. McCoy is allowed to seek advice from professionals within the industry that she follows.

Example 12 (Using an Expert Network):

Tom Watson is a research analyst working for a hedge fund. To stay informed, Watson relies on outside experts for information on such industries as technology and pharmaceuticals, where new advancements occur frequently. The meetings with the industry experts often are arranged through networks or placement agents that have specific policies and procedures in place to deter the exchange of material nonpublic information.

Watson arranges a call to discuss future prospects for one of the fund's existing technology company holdings, a company that was testing a new semiconductor product. The scientist leading the tests indicates his disappointment with the performance of the new semiconductor. Following the call, Watson relays the insights he received to others at the fund. The fund sells its current position in the company and buys many put options because the market is anticipating the success of the new semiconductor and the share price reflects the market's optimism.

> *Comment*: Watson has violated Standard II(A) by passing along material nonpublic information concerning the ongoing product tests, which the fund used to trade in the securities and options of the related company. Watson cannot simply rely on the agreements signed by individuals who participate in expert networks that state that he has not received information that would prohibit his trading activity. He must make his own determination whether information he received through these arrangements reaches a materiality threshold that would affect his trading abilities.

STANDARD II(B): INTEGRITY OF CAPITAL MARKETS - MARKET MANIPULATION

> Members and Candidates must not engage in practices that distort prices or artificially inflate trading volume with the intent to mislead market participants.

Guidance

Highlights:

- *Information-Based Manipulation*
- *Transaction-Based Manipulation*

Standard II(B) requires that members and candidates uphold market integrity by prohibiting market manipulation. Market manipulation includes practices that distort security prices or trading volume with the intent to deceive people or entities that rely on information in the market. Market manipulation damages the interests of all investors by disrupting the smooth functioning of financial markets and lowering investor confidence.

Market manipulation may lead to a lack of trust in the fairness of the capital markets, resulting in higher risk premiums and reduced investor participation. A reduction in the efficiency of a local capital market may negatively affect the growth and economic health of the country and may also influence the operations of the globally interconnected capital markets. Although market manipulation may be less likely to occur in mature financial markets than in emerging markets, cross-border investing increasingly exposes all global investors to the potential for such practices.

Market manipulation includes (1) the dissemination of false or misleading information and (2) transactions that deceive or would be likely to mislead market participants by distorting the price-setting mechanism of financial instruments. The development of new products and technologies increases the incentives, means, and opportunities for market manipulation. Additionally, the increasing complexity and sophistication of the technologies used for communicating with market participants have created new avenues for manipulation.

Information-Based Manipulation

Information-based manipulation includes, but is not limited to, spreading false rumors to induce trading by others. For example, members and candidates must refrain from "pumping up" the price of an investment by issuing misleading positive information or overly optimistic projections of a security's worth only to later "dump" the investment (i.e., sell it) once the price, fueled by the misleading information's effect on other market participants, reaches an artificially high level.

Transaction-Based Manipulation

Transaction-based manipulation involves instances where a member or candidate knew or should have known that his or her actions could affect the pricing of a security. This type of manipulation includes, but is not limited to, the following:

- transactions that artificially affect prices or volume to give the impression of activity or price movement in a financial instrument, which represent a diversion from the expectations of a fair and efficient market, and
- securing a controlling, dominant position in a financial instrument to exploit and manipulate the price of a related derivative and/or the underlying asset.

Standard II(B) is not intended to preclude transactions undertaken on legitimate trading strategies based on perceived market inefficiencies. The intent of the action is critical to determining whether it is a violation of this standard.

STANDARD II(B): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Independent Analysis and Company Promotion):

The principal owner of Financial Information Services (FIS) entered into an agreement with two microcap companies to promote the companies' stock in exchange for stock and cash compensation. The principal owner caused FIS to disseminate e-mails, design and maintain several websites, and distribute an online investment newsletter—all of which recommended investment in the two companies. The systematic publication of purportedly independent analyses and recommendations containing inaccurate and highly promotional and speculative statements increased public investment in the companies and led to dramatically higher stock prices.

> *Comment*: The principal owner of FIS violated Standard II(B) by using inaccurate reporting and misleading information under the guise of independent analysis to artificially increase the stock price of the companies. Furthermore, the principal owner violated Standard V(A)–Diligence and Reasonable Basis by not having a reasonable and adequate basis for recommending the two companies and violated Standard VI(A)–Disclosure of Conflicts by not disclosing to investors the compensation agreements (which constituted a conflict of interest).

Example 2 (Personal Trading Practices and Price):

John Gray is a private investor in Belgium who bought a large position several years ago in Fame Pharmaceuticals, a German small-cap security with limited average trading volume. He has now decided to significantly reduce his holdings owing to the poor price performance. Gray is worried that the low trading volume for the stock may cause the price to decline further as he attempts to sell his large position.

Gray devises a plan to divide his holdings into multiple accounts in different brokerage firms and private banks in the names of family members, friends, and even a private religious institution. He then creates a rumor campaign on various blogs and social media outlets promoting the company.

Gray begins to buy and sell the stock using the accounts in hopes of raising the trading volume and the price. He conducts the trades through multiple brokers, selling slightly larger positions than he bought on a tactical schedule, and over time, he is able to reduce his holding as desired without negatively affecting the sale price.

> *Comment*: John violated Standard II(B) by fraudulently creating the appearance that there was a greater investor interest in the stock through the online rumors. Additionally, through his trading strategy, he created the appearance that there was greater liquidity in the stock than actually existed. He was able to manipulate the price through both misinformation and trading practices.

Example 3 (Creating Artificial Price Volatility):

Matthew Murphy is an analyst at Divisadero Securities & Co., which has a significant number of hedge funds among its most important brokerage clients. Some of the hedge funds hold short positions on Wirewolf Semiconductor. Two trading days before the publication of a quarter-end report, Murphy alerts his sales force that he is about to issue a research report on Wirewolf that will include the following opinions:

- quarterly revenues are likely to fall short of management's guidance,
- earnings will be as much as 5 cents per share (or more than 10%) below consensus, and
- Wirewolf's highly respected chief financial officer may be about to join another company.

Knowing that Wirewolf has already entered its declared quarter-end "quiet period" before reporting earnings (and thus would be reluctant to respond to rumors), Murphy times the release of his research report specifically to sensationalize the negative aspects of the message in order to create significant downward pressure on Wirewolf's stock—to the distinct advantage of Divisadero's hedge fund clients. The report's conclusions are based on speculation, not on fact. The next day, the research report is broadcast to all of Divisadero's clients and to the usual newswire services.

Before Wirewolf's investor-relations department can assess the damage on the final trading day of the quarter and refute Murphy's report, its stock opens trading sharply lower, allowing Divisadero's clients to cover their short positions at substantial gains.

> *Comment*: Murphy violated Standard II(B) by aiming to create artificial price volatility designed to have a material impact on the price of an issuer's stock. Moreover, by lacking an adequate basis for the recommendation, Murphy also violated Standard V(A)–Diligence and Reasonable Basis.

Example 4 (Personal Trading and Volume):

Rajesh Sekar manages two funds—an equity fund and a balanced fund—whose equity components are supposed to be managed in accordance with the same model. According to that model, the funds' holdings in stock of Digital Design Inc. (DD) are excessive. Reduction of the DD holdings would not be easy, however, because the stock has low liquidity in the stock market. Sekar decides to start trading larger portions of DD stock back and forth between his two funds to slowly increase the price; he believes market participants will see growing volume and increasing price and become interested in the stock. If other investors are willing to buy the DD stock because of such interest, then Sekar will be able to get rid of at least some of his overweight position without inducing price decreases. In this way, the whole transaction will be for the benefit of fund participants, even if additional brokers' commissions are incurred.

> *Comment*: Sekar's plan would be beneficial for his funds' participants but is based on artificial distortion of both trading volume and the price of the DD stock and thus constitutes a violation of Standard II(B).

Example 5 ("Pump-Priming" Strategy):

ACME Futures Exchange is launching a new bond futures contract. To convince investors, traders, arbitrageurs, hedgers, and so on, to use its contract, the exchange attempts to demonstrate that it has the best liquidity. To do so, it enters into agreements with members in which they commit to a substantial minimum trading volume on the new contract over a specific period in exchange for substantial reductions of their regular commissions.

> *Comment*: The formal liquidity of a market is determined by the obligations set on market makers, but the actual liquidity of a market is better estimated by the actual trading volume and bid–ask spreads. Attempts to mislead participants about the actual liquidity of the market constitute a violation of Standard II(B). In this example, investors have been intentionally misled to believe they chose the most liquid instrument for some specific purpose, but they could eventually see the actual liquidity of the contract significantly reduced after the term of the agreement expires. If the ACME Futures Exchange fully discloses its agreement with members to boost transactions over some initial launch period, it will not violate Standard II(B). ACME's intent is not to harm investors but, on the contrary, to give them a better service. For that purpose, it may engage in a liquidity-pumping strategy, but the strategy must be disclosed.

Example 6 (Creating Artificial Price Volatility):

Emily Gordon, an analyst of household products companies, is employed by a research boutique, Picador & Co. Based on information that she has gathered during a trip through Latin America, she believes that Hygene, Inc., a major marketer of personal care products, has generated better-than-expected sales from its new product initiatives in South America. After modestly boosting her projections for revenue and for gross profit margin in her worksheet models for Hygene, Gordon estimates that her earnings projection of US$2.00 per diluted share for the current year may be as much as 5% too low. She contacts the chief financial officer (CFO) of Hygene to try to gain confirmation of her findings from her trip and to get some feedback regarding her revised models. The CFO declines to comment and reiterates management's most recent guidance of US$1.95–US$2.05 for the year.

Gordon decides to try to force a comment from the company by telling Picador & Co. clients who follow a momentum investment style that consensus earnings projections for Hygene are much too low; she explains that she is considering raising her published estimate by an ambitious US$0.15 to US$2.15 per share. She believes that when word of an unrealistically high earnings projection filters back to Hygene's investor-relations department, the company will feel compelled to update its earnings guidance. Meanwhile, Gordon hopes that she is at least correct with respect to the earnings direction and that she will help clients who act on her insights to profit from a quick gain by trading on her advice.

> *Comment*: By exaggerating her earnings projections in order to try to fuel a quick gain in Hygene's stock price, Gordon is in violation of Standard II(B). Furthermore, by virtue of previewing her intentions of revising upward her earnings projections to only a select group of clients, she is in violation of Standard III(B)–Fair Dealing. However, it would have been acceptable for Gordon to write a report that
>
> - framed her earnings projection in a range of possible outcomes,
> - outlined clearly the assumptions used in her Hygene models that took into consideration the findings from her trip through Latin America, and
> - was distributed to all Picador & Co. clients in an equitable manner.

Example 7 (Pump and Dump Strategy):

In an effort to pump up the price of his holdings in Moosehead & Belfast Railroad Company, Steve Weinberg logs on to several investor chat rooms on the internet to start rumors that the company is about to expand its rail network in anticipation of receiving a large contract for shipping lumber.

> *Comment*: Weinberg has violated Standard II(B) by disseminating false information about Moosehead & Belfast with the intent to mislead market participants.

Example 8 (Manipulating Model Inputs):

Bill Mandeville supervises a structured financing team for Superior Investment Bank. His responsibilities include packaging new structured investment products and managing Superior's relationship with relevant rating agencies. To achieve the best rating possible, Mandeville uses mostly positive scenarios as model inputs—scenarios that reflect minimal downside risk in the assets underlying the structured products. The resulting output statistics in the rating request and underwriting prospectus support the idea that the new structured products have minimal potential downside risk. Additionally, Mandeville's compensation from Superior is partially based on both the level of the rating assigned and the successful sale of new structured investment products but does not have a link to the long-term performance of the instruments.

Mandeville is extremely successful and leads Superior as the top originator of structured investment products for the next two years. In the third year, the economy experiences difficulties and the values of the assets underlying structured products significantly decline. The subsequent defaults lead to major turmoil in the capital markets, the demise of Superior Investment Bank, and the loss of Mandeville's employment.

> *Comment*: Mandeville manipulates the inputs of a model to minimize associated risk to achieve higher ratings. His understanding of structured products allows him to skillfully decide which inputs to include in support of the desired rating and price. This information manipulation for short-term gain, which is in violation of Standard II(B), ultimately causes significant damage to many parties and the capital markets as a whole. Mandeville should have realized that promoting a rating and price with inaccurate information could cause not only a loss of price confidence in the particular structured product but also a loss of investor trust in the system. Such loss of confidence affects the ability of the capital markets to operate efficiently.

Example 9 (Information Manipulation):

Allen King is a performance analyst for Torrey Investment Funds. King believes that the portfolio manager for the firm's small- and microcap equity fund dislikes him because the manager never offers him tickets to the local baseball team's games but does offer tickets to other employees. To incite a potential regulatory review of the manager, King creates user profiles on several online forums under the portfolio manager's name and starts rumors about potential mergers for several of the smaller companies in the portfolio. As the prices of these companies' stocks increase, the portfolio manager sells the position, which leads to an investigation by the regulator as King desired.

Comment: King has violated Standard II(B) even though he did not personally profit from the market's reaction to the rumor. In posting the false information, King misleads others into believing the companies were likely to be acquired. Although his intent was to create trouble for the portfolio manager, his actions clearly manipulated the factual information that was available to the market.

STANDARD III(A): DUTIES TO CLIENTS - LOYALTY, PRUDENCE, AND CARE

Standard III(A) Loyalty, Prudence, and Care

Members and Candidates have a duty of loyalty to their clients and must act with reasonable care and exercise prudent judgment. Members and Candidates must act for the benefit of their clients and place their clients' interests before their employer's or their own interests.

Guidance

Highlights:

- *Understanding the Application of Loyalty, Prudence, and Care*
- *Identifying the Actual Investment Client*
- *Developing the Client's Portfolio*
- *Soft Commission Policies*
- *Proxy Voting Policies*

Standard III(A) clarifies that client interests are paramount. A member's or candidate's responsibility to a client includes a duty of loyalty and a duty to exercise reasonable care. Investment actions must be carried out for the sole benefit of the client and in a manner the member or candidate believes, given the known facts and circumstances, to be in the best interest of the client. Members and candidates must exercise the same level of prudence, judgment, and care that they would apply in the management and disposition of their own interests in similar circumstances.

Prudence requires caution and discretion. The exercise of prudence by investment professionals requires that they act with the care, skill, and diligence that a reasonable person acting in a like capacity and familiar with such matters would use. In the context of managing a client's portfolio, prudence requires following the investment parameters set forth by the client and balancing risk and return. Acting with care requires members and candidates to act in a prudent and judicious manner in avoiding harm to clients.

Standard III(A) sets minimum expectations for members and candidates when fulfilling their responsibilities to their clients. Regulatory and legal requirements for such duties can vary across the investment industry depending on a variety of factors, including job function of the investment professional, the existence of an adviser/client relationship, and the nature of the recommendations being offered. From the perspective of the end user of financial services, these different standards can be arcane and confusing, leaving investors unsure of what level of service to expect from investment professionals they employ. The single standard of conduct described in Standard III(A) benefits investors by establishing a benchmark for the duties of loyalty, prudence, and care and clarifies that all CFA Institute members and candidates, regardless of job title, local laws, or cultural differences, are required to comply with these fundamental responsibilities. Investors hiring members or candidates who must adhere to the duty of loyalty, prudence, and care set forth in this standard can be confident that these responsibilities are a requirement regardless of any legally imposed fiduciary duties.

Standard III(A), however, is not a substitute for a member's or candidate's legal or regulatory obligations. As stated in Standard I(A), members and candidates must abide by the most strict requirements imposed on them by regulators or the Code and Standards, including any legally imposed fiduciary duty. Members and candidates must also be aware of whether they have "custody" or effective control of client assets. If so, a heightened level of responsibility arises. Members and candidates are considered to have custody if they have any direct or indirect access to client funds. Members and candidates must manage any pool of assets in their control in accordance with the terms of the governing documents (such as trust documents and investment management agreements), which are the primary determinant of the manager's powers and duties. Whenever their actions are contrary to provisions of those instruments or applicable law, members and candidates are at risk of violating Standard III(A).

Understanding the Application of Loyalty, Prudence, and Care

Standard III(A) establishes a minimum benchmark for the duties of loyalty, prudence, and care that are required of all members and candidates regardless of whether a legal fiduciary duty applies. Although fiduciary duty often encompasses the principles of loyalty, prudence, and care, Standard III(A) does not render all members and candidates fiduciaries. The responsibilities of members and candidates for fulfilling their obligations under this standard depend greatly on the nature of their professional responsibilities and the relationships they have with clients. The conduct of members and candidates may or may not rise to the level of being a fiduciary, depending on the type of client, whether the member or candidate is giving investment advice, and the many facts and circumstances surrounding a particular transaction or client relationship.

Fiduciary duties are often imposed by law or regulation when an individual or institution is charged with the duty of acting for the benefit of another party, such as managing investment assets. The duty required in fiduciary relationships exceeds what is acceptable in many other business relationships because a fiduciary is in an enhanced position of trust. Although members and candidates must comply with any legally imposed fiduciary duty, the Code and Standards neither impose such a legal responsibility nor require all members or candidates to act as fiduciaries. However, Standard III(A) requires members and candidates to work in the client's best interest no matter what the job function.

A member or candidate who does not provide advisory services to a client but who acts only as a trade execution professional must prudently work in the client's interest when completing requested trades. Acting in the client's best interest requires these professionals to use their skills and diligence to execute trades in the most favorable terms that can be achieved. Members and candidates operating in such positions must use care to operate within the parameters set by the client's trading instructions.

Members and candidates may also operate in a blended environment where they execute client trades and offer advice on a limited set of investment options. The extent of the advisory arrangement and limitations should be outlined in the agreement with the client at the outset of the relationship. For instance, members and candidates should inform clients that the advice provided will be limited to the propriety products of the firm and not include other products available on the market. Clients who want access to a wider range of investment products would have the information necessary to decide not to engage with members or candidates working under these restrictions.

Members and candidates operating in this blended context would comply with their obligations by recommending the allowable products that are consistent with the client's objectives and risk tolerance. They would exercise care through diligently aligning the client's needs with the attributes of the products being recommended. Members and candidates should place the client's interests first by disregarding any firm or personal interest in motivating a recommended transaction.

There is a large variety of professional relationships that members and candidates have with their clients. Standard III(A) requires them to fulfill the obligations outlined explicitly or implicitly in the client agreements to the best of their abilities and with loyalty, prudence, and care. Whether a member or candidate is structuring a new securitization transaction, completing a credit rating analysis, or leading a public company, he or she must work with prudence and care in delivering the agreed-on services.

Identifying the Actual Investment Client

The first step for members and candidates in fulfilling their duty of loyalty to clients is to determine the identity of the "client" to whom the duty of loyalty is owed. In the context of an investment manager managing the personal assets of an individual, the client is easily identified. When the manager is responsible for the portfolios of pension plans or trusts, however, the client is not the person or entity who hires the manager but, rather, the beneficiaries of the plan or trust. The duty of loyalty is owed to the ultimate beneficiaries.

In some situations, an actual client or group of beneficiaries may not exist. Members and candidates managing a fund to an index or an expected mandate owe the duty of loyalty, prudence, and care to invest in a manner consistent with the stated mandate. The decisions of a fund's manager, although benefiting all fund investors, do not have to be based on an individual investor's requirements and risk profile. Client loyalty and care for those investing in the fund are the responsibility of members and candidates who have an advisory relationship with those individuals.

Situations involving potential conflicts of interest with respect to responsibilities to clients may be extremely complex because they may involve a number of competing interests. The duty of loyalty, prudence, and care applies to a large number of persons in varying capacities, but the exact duties may differ in many respects in accord with the relationship with each client or each type of account in which the assets are managed. Members and candidates must not only put their obligations to clients first in all dealings but also endeavor to avoid all real or potential conflicts of interest.

Members and candidates with positions whose responsibilities do not include direct investment management also have "clients" that must be considered. Just as there are various types of advisory relationships, members and candidates must look at their roles and responsibilities when making a determination of who their clients are. Sometimes the client is easily identifiable; such is the case in the relationship between a company executive and the firm's public shareholders. At other times, the client may be the investing public as a whole, in which case the goals of independence and objectivity of research surpass the goal of loyalty to a single organization.

Developing the Client's Portfolio

The duty of loyalty, prudence, and care owed to the individual client is especially important because the professional investment manager typically possesses greater knowledge in the investment arena than the client does. This disparity places the individual client in a vulnerable position; the client must trust the manager. The manager in these situations should ensure that the client's objectives and expectations for the performance of the account are realistic and suitable to the client's circumstances and that the risks involved are appropriate. In most circumstances, recommended investment strategies should relate to the long-term objectives and circumstances of the client.

Particular care must be taken to detect whether the goals of the investment manager or the firm in conducting business, selling products, and executing security transactions potentially conflict with the best interests and objectives of the client. When members and candidates cannot avoid potential conflicts between their firm and clients' interests, they must provide clear and factual disclosures of the circumstances to the clients.

Members and candidates must follow any guidelines set by their clients for the management of their assets. Some clients, such as charitable organizations and pension plans, have strict investment policies that limit investment options to certain types or classes of investment or prohibit investment in certain securities. Other organizations have aggressive policies that do not prohibit investments by type but, instead, set criteria on the basis of the portfolio's total risk and return.

Investment decisions must be judged in the context of the total portfolio rather than by individual investment within the portfolio. The member's or candidate's duty is satisfied with respect to a particular investment if the individual has thoroughly considered the investment's place in the overall portfolio, the risk of loss and opportunity for gains, tax implications, and the diversification, liquidity, cash flow, and overall return requirements of the assets or the portion of the assets for which the manager is responsible.

Soft Commission Policies

An investment manager often has discretion over the selection of brokers executing transactions. Conflicts may arise when an investment manager uses client brokerage to purchase research services, a practice commonly called "soft dollars" or "soft commissions." A member or candidate who pays a higher brokerage commission than he or she would normally pay to allow for the purchase of goods or services, without corresponding benefit to the client, violates the duty of loyalty to the client.

From time to time, a client will direct a manager to use the client's brokerage to purchase goods or services for the client, a practice that is commonly called "directed brokerage." Because brokerage commission is an asset of the client and is used to benefit that client, not the manager, such a practice does not violate any duty of loyalty. However, a member or candidate is obligated to seek "best price" and "best execution" and be assured by the client that the goods or services purchased from the brokerage will benefit the account beneficiaries. "Best execution" refers to a trading process that seeks to maximize the value of the client's portfolio within the client's stated investment objectives and constraints. In addition, the member or candidate should disclose to the client that the client may not be getting best execution from the directed brokerage.

Proxy Voting Policies

The duty of loyalty, prudence, and care may apply in a number of situations facing the investment professional besides those related directly to investing assets.

Part of a member's or candidate's duty of loyalty includes voting proxies in an informed and responsible manner. Proxies have economic value to a client, and members and candidates must ensure that they properly safeguard and maximize this value. An investment manager who fails to vote, casts a vote without considering the impact of the question, or votes blindly with management on nonroutine governance issues (e.g., a change in company capitalization) may violate this standard. Voting of proxies is an integral part of the management of investments.

A cost–benefit analysis may show that voting all proxies may not benefit the client, so voting proxies may not be necessary in all instances. Members and candidates should disclose to clients their proxy voting policies.

STANDARD III(A): RECOMMENDED PROCEDURES

Regular Account Information

Members and candidates with control of client assets (1) should submit to each client, at least quarterly, an itemized statement showing the funds and securities in the custody or possession of the member or candidate plus all debits, credits, and transactions that occurred during the period, (2) should disclose to the client where the assets are to be maintained, as well as where or when they are moved, and (3) should separate the client's assets from any other party's assets, including the member's or candidate's own assets.

Client Approval

If a member or candidate is uncertain about the appropriate course of action with respect to a client, the member or candidate should consider what he or she would expect or demand if the member or candidate were the client. If in doubt, a member or candidate should disclose the questionable matter in writing to the client and obtain client approval.

Firm Policies

Members and candidates should address and encourage their firms to address the following topics when drafting the statements or manuals containing their policies and procedures regarding responsibilities to clients:

- *Follow all applicable rules and laws*: Members and candidates must follow all legal requirements and applicable provisions of the Code and Standards.
- *Establish the investment objectives of the client*: Make a reasonable inquiry into a client's investment experience, risk and return objectives, and financial constraints prior to making investment recommendations or taking investment actions.
- *Consider all the information when taking actions*: When taking investment actions, members and candidates must consider the appropriateness and suitability of the investment relative to (1) the client's needs and circumstances, (2) the investment's basic characteristics, and (3) the basic characteristics of the total portfolio.
- *Diversify*: Members and candidates should diversify investments to reduce the risk of loss, unless diversification is not consistent with plan guidelines or is contrary to the account objectives.
- *Carry out regular reviews*: Members and candidates should establish regular review schedules to ensure that the investments held in the account adhere to the terms of the governing documents.
- *Deal fairly with all clients with respect to investment actions*: Members and candidates must not favor some clients over others and should establish policies for allocating trades and disseminating investment recommendations.

- *Disclose conflicts of interest*: Members and candidates must disclose all actual and potential conflicts of interest so that clients can evaluate those conflicts.
- *Disclose compensation arrangements*: Members and candidates should make their clients aware of all forms of manager compensation.
- *Vote proxies*: In most cases, members and candidates should determine who is authorized to vote shares and vote proxies in the best interests of the clients and ultimate beneficiaries.
- *Maintain confidentiality*: Members and candidates must preserve the confidentiality of client information.
- *Seek best execution*: Unless directed by the client as ultimate beneficiary, members and candidates must seek best execution for their clients. (Best execution is defined in the preceding text.)
- *Place client interests first*: Members and candidates must serve the best interests of clients.

STANDARD III(A): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Identifying the Client—Plan Participants):

First Country Bank serves as trustee for the Miller Company's pension plan. Miller is the target of a hostile takeover attempt by Newton, Inc. In attempting to ward off Newton, Miller's managers persuade Julian Wiley, an investment manager at First Country Bank, to purchase Miller common stock in the open market for the employee pension plan. Miller's officials indicate that such action would be favorably received and would probably result in other accounts being placed with the bank. Although Wiley believes the stock is overvalued and would not ordinarily buy it, he purchases the stock to support Miller's managers, to maintain Miller's good favor toward the bank, and to realize additional new business. The heavy stock purchases cause Miller's market price to rise to such a level that Newton retracts its takeover bid.

> *Comment*: Standard III(A) requires that a member or candidate, in evaluating a takeover bid, act prudently and solely in the interests of plan participants and beneficiaries. To meet this requirement, a member or candidate must carefully evaluate the long-term prospects of the company against the short-term prospects presented by the takeover offer and by the ability to invest elsewhere. In this instance, Wiley, acting on behalf of his employer, which was the trustee for a pension plan, clearly violated Standard III(A). He used the pension plan to perpetuate existing management, perhaps to the detriment of plan participants and the company's shareholders, and to benefit himself. Wiley's responsibilities to the plan participants and beneficiaries should have taken precedence over any ties of his bank to corporate managers and over his self-interest. Wiley had a duty to examine the takeover offer on its own merits and to make an independent decision.

The guiding principle is the appropriateness of the investment decision to the pension plan, not whether the decision benefited Wiley or the company that hired him.

Example 2 (Client Commission Practices):

JNI, a successful investment counseling firm, serves as investment manager for the pension plans of several large regionally based companies. Its trading activities generate a significant amount of commission-related business. JNI uses the brokerage and research services of many firms, but most of its trading activity is handled through a large brokerage company, Thompson, Inc., because the executives of the two firms have a close friendship. Thompson's commission structure is high in comparison with charges for similar brokerage services from other firms. JNI considers Thompson's research services and execution capabilities average. In exchange for JNI directing its brokerage to Thompson, Thompson absorbs a number of JNI overhead expenses, including those for rent.

Comment: JNI executives are breaching their responsibilities by using client brokerage for services that do not benefit JNI clients and by not obtaining best price and best execution for their clients. Because JNI executives are not upholding their duty of loyalty, they are violating Standard III(A).

Example 3 (Brokerage Arrangements):

Charlotte Everett, a struggling independent investment adviser, serves as investment manager for the pension plans of several companies. One of her brokers, Scott Company, is close to consummating management agreements with prospective new clients whereby Everett would manage the new client accounts and trade the accounts exclusively through Scott. One of Everett's existing clients, Crayton Corporation, has directed Everett to place securities transactions for Crayton's account exclusively through Scott. But to induce Scott to exert efforts to send more new accounts to her, Everett also directs transactions to Scott from other clients without their knowledge.

Comment: Everett has an obligation at all times to seek best price and best execution on all trades. Everett may direct new client trades exclusively through Scott Company as long as Everett receives best price and execution on the trades or receives a written statement from new clients that she is *not* to seek best price and execution and that they are aware of the consequence for their accounts. Everett may trade other accounts through Scott as a reward for directing clients to Everett only if the accounts receive best price and execution and the practice is disclosed to the accounts. Because Everett does not disclose the directed trading, Everett has violated Standard III(A).

Example 4 (Brokerage Arrangements):

Emilie Rome is a trust officer for Paget Trust Company. Rome's supervisor is responsible for reviewing Rome's trust account transactions and her monthly reports of personal stock transactions. Rome has been using Nathan Gray, a broker, almost exclusively for trust account brokerage transactions. When Gray makes a market in stocks, he has been giving Rome a lower price for personal purchases and a higher price for sales than he gives to Rome's trust accounts and other investors.

Comment: Rome is violating her duty of loyalty to the bank's trust accounts by using Gray for brokerage transactions simply because Gray trades Rome's personal account on favorable terms. Rome is placing her own interests before those of her clients.

Example 5 (Client Commission Practices):

Lauren Parker, an analyst with Provo Advisors, covers South American equities for her firm. She likes to travel to the markets for which she is responsible and decides to go on a trip to Chile, Argentina, and Brazil. The trip is sponsored by SouthAM, Inc., a research firm with a small broker/dealer affiliate that uses the clearing facilities of a larger New York brokerage house. SouthAM specializes in arranging South American trips for analysts during which they can meet with central bank officials, government ministers, local economists, and senior executives of corporations. SouthAM accepts commission dollars at a ratio of 2 to 1 against the hard-dollar costs of the research fee for the trip. Parker is not sure that SouthAM's execution is competitive, but without informing her supervisor, she directs the trading desk at Provo to start giving commission business to SouthAM so she can take the trip. SouthAM has conveniently timed the briefing trip to coincide with the beginning of Carnival season, so Parker also decides to spend five days of vacation in Rio de Janeiro at the end of the trip. Parker uses commission dollars to pay for the five days of hotel expenses.

Comment: Parker is violating Standard III(A) by not exercising her duty of loyalty to her clients. She should have determined whether the commissions charged by SouthAM are reasonable in relation to the benefit of the research provided by the trip. She also should have determined whether best execution and prices could be received from SouthAM. In addition, the five extra days are not part of the research effort because they do not assist in the investment decision making. Thus, the hotel expenses for the five days should not be paid for with client assets.

Example 6 (Excessive Trading):

Vida Knauss manages the portfolios of a number of high-net-worth individuals. A major part of her investment management fee is based on trading commissions. Knauss engages in extensive trading for each of her clients to ensure that she attains the minimum commission level set by her firm. Although the securities purchased and sold for the clients are appropriate and fall within the acceptable asset classes for the clients, the amount of trading for each account exceeds what is necessary to accomplish the client's investment objectives.

Comment: Knauss has violated Standard III(A) because she is using the assets of her clients to benefit her firm and herself.

Example 7 (Managing Family Accounts):

Adam Dill recently joined New Investments Asset Managers. To assist Dill in building a book of clients, both his father and brother opened new fee-paying accounts. Dill followed all the firm's procedures in noting his relationships with these clients and in developing their investment policy statements.

After several years, the number of Dill's clients has grown, but he still manages the original accounts of his family members. An IPO is coming to market that is a suitable investment for many of his clients, including his brother. Dill does not receive the amount of stock he requested, so to avoid any appearance of a conflict of interest, he does not allocate any shares to his brother's account.

> *Comment*: Dill has violated Standard III(A) because he is not acting for the benefit of his brother's account as well as his other accounts. The brother's account is a regular fee-paying account comparable to the accounts of his other clients. By not allocating the shares proportionately across *all* accounts for which he thought the IPO was suitable, Dill is disadvantaging specific clients.
>
> Dill would have been correct in not allocating shares to his brother's account if that account was being managed outside the normal fee structure of the firm.

Example 8 (Identifying the Client):

Donna Hensley has been hired by a law firm to testify as an expert witness. Although the testimony is intended to represent impartial advice, she is concerned that her work may have negative consequences for the law firm. If the law firm is Hensley's client, how does she ensure that her testimony will not violate the required duty of loyalty, prudence, and care to one's client?

> *Comment*: In this situation, the law firm represents Hensley's employer and the aspect of "who is the client" is not well defined. When acting as an expert witness, Hensley is bound by the standard of independence and objectivity in the same manner as an independent research analyst would be bound. Hensley must not let the law firm influence the testimony she provides in the legal proceedings.

Example 9 (Identifying the Client):

Jon Miller is a mutual fund portfolio manager. The fund is focused on the global financial services sector. Wanda Spears is a private wealth manager in the same city as Miller and is a friend of Miller. At a local CFA Institute society meeting, Spears mentions to Miller that her new client is an investor in Miller's fund. She states that the two of them now share a responsibility to this client.

> *Comment*: Spears' statement is not totally correct. Because she provides the advisory services to her new client, she alone is bound by the duty of loyalty to this client. Miller's responsibility is to manage the fund according to the investment policy statement of the fund. His actions should not be influenced by the needs of any particular fund investor.

Example 10 (Client Loyalty):

After providing client account investment performance to the external-facing departments but prior to it being finalized for release to clients, Teresa Nguyen, an investment performance analyst, notices the reporting system missed a trade. Correcting the omission resulted in a large loss for a client that had previously placed the firm on "watch" for potential termination owing to underperformance in prior periods. Nguyen knows this news is unpleasant but informs the appropriate individuals that the report needs to be updated before releasing it to the client.

> *Comment*: Nguyen's actions align with the requirements of Standard III(A). Even though the correction may lead to the firm's termination by the client, withholding information on errors would not be in the best interest of the client.

Example 11 (Execution-Only Responsibilities):

Baftija Sulejman recently became a candidate in the CFA Program. He is a broker who executes client-directed trades for several high-net-worth individuals. Sulejman does not provide any investment advice and only executes the trading decisions made by clients. He is concerned that the Code and Standards impose a fiduciary duty on him in his dealing with clients and sends an e-mail to the CFA Ethics Helpdesk (ethics@cfainstitute.org) to seek guidance on this issue.

> *Comment*: In this instance, Sulejman serves in an execution-only capacity and his duty of loyalty, prudence, and care is centered on the skill and diligence used when executing trades—namely, by seeking best execution and making trades within the parameters set by the clients (instructions on quantity, price, timing, etc.). Acting in the best interests of the client dictates that trades are executed on the most favorable terms that can be achieved for the client. Given this job function, the requirements of the Code and Standards for loyalty, prudence, and care clearly do not impose a fiduciary duty.

STANDARD III(B): DUTIES TO CLIENTS - FAIR DEALING

Members and Candidates must deal fairly and objectively with all clients when providing investment analysis, making investment recommendations, taking investment action, or engaging in other professional activities.

Guidance

Highlights:

- *Investment Recommendations*
- *Investment Action*

Standard III(B) requires members and candidates to treat all clients fairly when disseminating investment recommendations or making material changes to prior investment recommendations or when taking investment action with regard to general purchases, new issues, or secondary offerings. Only through the fair treatment of all parties can the investment management profession maintain the confidence of the investing public.

When an investment adviser has multiple clients, the potential exists for the adviser to favor one client over another. This favoritism may take various forms—from the quality and timing of services provided to the allocation of investment opportunities.

The term “fairly” implies that the member or candidate must take care not to discriminate against any clients when disseminating investment recommendations or taking investment action. Standard III(B) does not state “equally” because members and candidates could not possibly reach all clients at exactly the same time—whether by printed mail, telephone (including text messaging), computer (including internet updates and e-mail distribution), facsimile (fax), or wire. Each client has unique needs, investment criteria, and investment objectives, so not all investment opportunities are suitable for all clients. In addition, members and candidates may provide more personal, specialized, or in-depth service to clients who are willing to pay for premium services through higher management fees or higher levels of brokerage. Members and candidates may differentiate their services to clients, but different levels of service must not disadvantage or negatively affect clients. In addition, the different service levels should be disclosed to clients and prospective clients and should be available to everyone (i.e., different service levels should not be offered selectively).

Standard III(B) covers conduct in two broadly defined categories—investment recommendations and investment action.

Investment Recommendations

The first category of conduct involves members and candidates whose primary function is the preparation of investment recommendations to be disseminated either to the public or within a firm for the use of others in making investment decisions. This group includes members and candidates employed by investment counseling, advisory, or consulting firms as well as banks, brokerage firms, and insurance companies. The criterion is that the member’s or candidate’s primary responsibility is the preparation of recommendations to be acted on by others, including those in the member’s or candidate’s organization.

An investment recommendation is any opinion expressed by a member or candidate in regard to purchasing, selling, or holding a given security or other investment. The opinion may be disseminated to customers or clients through an initial detailed research report, through a brief update report, by addition to or deletion from a list of recommended securities, or simply by oral communication. A recommendation that is distributed to anyone outside the organization is considered a communication for general distribution under Standard III(B).

Standard III(B) addresses the manner in which investment recommendations or changes in prior recommendations are disseminated to clients. Each member or candidate is obligated to ensure that information is disseminated in such a manner that all clients have a fair opportunity to act on every recommendation. Communicating with all clients on a uniform basis presents practical problems for members and candidates because of differences in timing and methods of communication with various types of customers and clients. Members and candidates should encourage their firms to design an equitable system to prevent selective or discriminatory disclosure and should inform clients about what kind of communications they will receive.

The duty to clients imposed by Standard III(B) may be more critical when members or candidates change their recommendations than when they make initial recommendations. Material changes in a member’s or candidate’s prior investment recommendations because of subsequent research should be communicated to all current clients; particular care should be taken that the information reaches those clients who the member or candidate knows have acted on or been affected by the earlier advice. Clients who do not know that the member or candidate has changed a recommendation and who, therefore, place orders contrary to a current recommendation should be advised of the changed recommendation before the order is accepted.

Investment Action

The second category of conduct includes those members and candidates whose primary function is taking investment action (portfolio management) on the basis of recommendations prepared internally or received from external sources. Investment action, like investment recommendations, can affect market value. Consequently, Standard III(B) requires that members or candidates treat all clients fairly in light of their investment objectives and circumstances. For example, when making investments in new offerings or in secondary financings, members and candidates should distribute the issues to all customers for whom the investments are appropriate in a manner consistent with the policies of the firm for allocating blocks of stock. If the issue is oversubscribed, then the issue should be prorated to all subscribers. This action should be taken on a round-lot basis to avoid odd-lot distributions. In addition, if the issue is oversubscribed, members and candidates should forgo any sales to themselves or their immediate families in order to free up additional shares for clients. If the investment professional's family-member accounts are managed similarly to the accounts of other clients of the firm, however, the family-member accounts should not be excluded from buying such shares.

Members and candidates must make every effort to treat all individual and institutional clients in a fair and impartial manner. A member or candidate may have multiple relationships with an institution; for example, the member or candidate may be a corporate trustee, pension fund manager, manager of funds for individuals employed by the customer, loan originator, or creditor. A member or candidate must exercise care to treat all clients fairly.

Members and candidates should disclose to clients and prospective clients the documented allocation procedures they or their firms have in place and how the procedures would affect the client or prospect. The disclosure should be clear and complete so that the client can make an informed investment decision. Even when complete disclosure is made, however, members and candidates must put client interests ahead of their own. A member's or candidate's duty of fairness and loyalty to clients can never be overridden by client consent to patently unfair allocation procedures.

Treating clients fairly also means that members and candidates should not take advantage of their position in the industry to the detriment of clients. For instance, in the context of IPOs, members and candidates must make bona fide public distributions of "hot issue" securities (defined as securities of a public offering that are trading at a premium in the secondary market whenever such trading commences because of the great demand for the securities). Members and candidates are prohibited from withholding such securities for their own benefit and must not use such securities as a reward or incentive to gain benefit.

STANDARD III(B): RECOMMENDED PROCEDURES

Develop Firm Policies

Although Standard III(B) refers to a member's or candidate's responsibility to deal fairly and objectively with clients, members and candidates should also encourage their firms to establish compliance procedures requiring all employees who disseminate investment recommendations or take investment actions to treat customers and clients fairly. At the very least, a member or candidate should recommend appropriate procedures to management if none are in place. And the member or candidate should make management aware of possible violations of fair-dealing practices within the firm when they come to the attention of the member or candidate.

The extent of the formality and complexity of such compliance procedures depends on the nature and size of the organization and the type of securities involved. An investment adviser who is a sole proprietor and handles only discretionary accounts might not disseminate recommendations to the public, but that adviser should have formal written procedures to ensure that all clients receive fair investment action.

Good business practice dictates that initial recommendations be made available to all customers who indicate an interest. Although a member or candidate need not communicate a recommendation to all customers, the selection process by which customers receive information should be based on suitability and known interest, not on any preferred or favored status. A common practice to assure fair dealing is to communicate recommendations simultaneously within the firm and to customers.

Members and candidates should consider the following points when establishing fair-dealing compliance procedures:

- *Limit the number of people involved*: Members and candidates should make reasonable efforts to limit the number of people who are privy to the fact that a recommendation is going to be disseminated.
- *Shorten the time frame between decision and dissemination*: Members and candidates should make reasonable efforts to limit the amount of time that elapses between the decision to make an investment recommendation and the time the actual recommendation is disseminated. If a detailed institutional recommendation that might take two or three weeks to publish is in preparation, a short summary report including the conclusion might be published in advance. In an organization where both a research committee and an investment policy committee must approve a recommendation, the meetings should be held on the same day if possible. The process of reviewing reports and printing and mailing them, faxing them, or distributing them by e-mail necessarily involves the passage of time, sometimes long periods of time. In large firms with extensive review processes, the time factor is usually not within the control of the analyst who prepares the report. Thus, many firms and their analysts communicate to customers and firm personnel the new or changed recommendations by an update or "flash" report. The communication technique might be fax, e-mail, wire, or short written report.
- *Publish guidelines for pre-dissemination behavior*: Members and candidates should encourage firms to develop guidelines that prohibit personnel who have prior knowledge of an investment recommendation from discussing or taking any action on the pending recommendation.
- *Simultaneous dissemination*: Members and candidates should establish procedures for the timing of dissemination of investment recommendations so that all clients are treated fairly—that is, are informed at approximately the same time. For example, if a firm is going to announce a new recommendation, supervisory personnel should time the announcement to avoid placing any client or group of clients at an unfair advantage relative to other clients. A communication to all branch offices should be sent at the time of the general announcement. (When appropriate, the firm should accompany the announcement of a new recommendation with a statement that trading restrictions for the firm's employees are now in effect. The trading restrictions should stay in effect until the recommendation is widely distributed to all relevant clients.) Once this distribution has occurred, the member or candidate may follow up separately with individual clients, but members and candidates should not give favored clients advance information when such advance notification may disadvantage other clients.

- *Maintain a list of clients and their holdings*: Members and candidates should maintain a list of all clients and the securities or other investments each client holds in order to facilitate notification of customers or clients of a change in an investment recommendation. If a particular security or other investment is to be sold, such a list can be used to ensure that all holders are treated fairly in the liquidation of that particular investment.
- *Develop and document trade allocation procedures*: When formulating procedures for allocating trades, members and candidates should develop a set of guiding principles that ensure
  - fairness to advisory clients, both in priority of execution of orders and in the allocation of the price obtained in execution of block orders or trades,
  - timeliness and efficiency in the execution of orders, and
  - accuracy of the member's or candidate's records as to trade orders and client account positions.

With these principles in mind, members and candidates should develop or encourage their firm to develop written allocation procedures, with particular attention to procedures for block trades and new issues. Procedures to consider are as follows:

- requiring orders and modifications or cancellations of orders to be documented and time stamped;
- processing and executing orders on a first-in, first-out basis with consideration of bundling orders for efficiency as appropriate for the asset class or the security;
- developing a policy to address such issues as calculating execution prices and "partial fills" when trades are grouped, or in a block, for efficiency;
- giving all client accounts participating in a block trade the same execution price and charging the same commission;
- when the full amount of the block order is not executed, allocating partially executed orders among the participating client accounts pro rata on the basis of order size while not going below an established minimum lot size for some securities (e.g., bonds); and
- when allocating trades for new issues, obtaining advance indications of interest, allocating securities by client (rather than portfolio manager), and providing a method for calculating allocations.

Disclose Trade Allocation Procedures

Members and candidates should disclose to clients and prospective clients how they select accounts to participate in an order and how they determine the amount of securities each account will buy or sell. Trade allocation procedures must be fair and equitable, and disclosure of inequitable allocation methods does not relieve the member or candidate of this obligation.

Establish Systematic Account Review

Member and candidate supervisors should review each account on a regular basis to ensure that no client or customer is being given preferential treatment and that the investment actions taken for each account are suitable for each account's objectives. Because investments should be based on individual needs and circumstances, an investment manager may have good reasons for placing a given security or other investment in one account while selling it from another account and should fully document the reasons behind both sides of the transaction. Members and candidates should encourage firms to establish review procedures, however, to detect whether trading in one account is being used to benefit a favored client.

Disclose Levels of Service

Members and candidates should disclose to all clients whether the organization offers different levels of service to clients for the same fee or different fees. Different levels of service should not be offered to clients selectively.

STANDARD III(B): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Selective Disclosure):

Bradley Ames, a well-known and respected analyst, follows the computer industry. In the course of his research, he finds that a small, relatively unknown company whose shares are traded over the counter has just signed significant contracts with some of the companies he follows. After a considerable amount of investigation, Ames decides to write a research report on the small company and recommend purchase of its shares. While the report is being reviewed by the company for factual accuracy, Ames schedules a luncheon with several of his best clients to discuss the company. At the luncheon, he mentions the purchase recommendation scheduled to be sent early the following week to all the firm's clients.

> *Comment*: Ames has violated Standard III(B) by disseminating the purchase recommendation to the clients with whom he has lunch a week before the recommendation is sent to all clients.

Example 2 (Fair Dealing between Funds):

Spencer Rivers, president of XYZ Corporation, moves his company's growth-oriented pension fund to a particular bank primarily because of the excellent investment performance achieved by the bank's commingled fund for the prior five-year period. Later, Rivers compares the results of his pension fund with those of the bank's commingled fund. He is startled to learn that, even though the two accounts have the same investment objectives and similar portfolios, his company's pension fund has significantly underperformed the bank's commingled fund. Questioning this result at his next meeting with the pension fund's manager, Rivers is told that, as a matter of policy, when a new security is placed on the recommended list, Morgan Jackson, the pension fund manager, first purchases the security for the commingled account and then purchases it on a pro rata basis for all other pension fund accounts. Similarly, when a sale is recommended, the security is sold first from the commingled account and then sold on a pro rata basis from all other accounts. Rivers also learns that if the bank cannot get enough shares (especially of hot issues) to be meaningful to all the accounts, its policy is to place the new issues only in the commingled account.

Seeing that Rivers is neither satisfied nor pleased by the explanation, Jackson quickly adds that nondiscretionary pension accounts and personal trust accounts have a lower priority on purchase and sale recommendations than discretionary pension fund accounts. Furthermore, Jackson states, the company's pension fund had the opportunity to invest up to 5% in the commingled fund.

> *Comment*: The bank's policy does not treat all customers fairly, and Jackson has violated her duty to her clients by giving priority to the growth-oriented commingled fund over all other funds and to discretionary accounts over nondiscretionary accounts. Jackson must execute orders on a systematic basis that is fair to all clients. In addition, trade allocation procedures should be disclosed to all clients when they become clients. Of course, in this case, disclosure of the bank's policy would not change the fact that the policy is unfair.

Example 3 (Fair Dealing and IPO Distribution):

Dominic Morris works for a small regional securities firm. His work consists of corporate finance activities and investing for institutional clients. Arena, Ltd., is planning to go public. The partners have secured rights to buy an arena football league franchise and are planning to use the funds from the issue to complete the purchase. Because arena football is the current rage, Morris believes he has a hot issue on his hands. He has quietly negotiated some options for himself for helping convince Arena to do the financing through his securities firm. When he seeks expressions of interest, the institutional buyers oversubscribe the issue. Morris, assuming that the institutions have the financial clout to drive the stock up, then fills all orders (including his own) and decreases the institutional blocks.

> *Comment*: Morris has violated Standard III(B) by not treating all customers fairly. He should not have taken any shares himself and should have prorated the shares offered among all clients. In addition, he should have disclosed to his firm and to his clients that he received options as part of the deal [see Standard VI(A)–Disclosure of Conflicts].

Example 4 (Fair Dealing and Transaction Allocation):

Eleanor Preston, the chief investment officer of Porter Williams Investments (PWI), a medium-size money management firm, has been trying to retain a client, Colby Company. Management at Colby, which accounts for almost half of PWI's revenues, recently told Preston that if the performance of its account did not improve, it would find a new money manager. Shortly after this threat, Preston purchases mortgage-backed securities (MBSs) for several accounts, including Colby's. Preston is busy with a number of transactions that day, so she fails to allocate the trades immediately or write up the trade tickets. A few days later, when Preston is allocating trades, she notes that some of the MBSs have significantly increased in price and some have dropped. Preston decides to allocate the profitable trades to Colby and spread the losing trades among several other PWI accounts.

> *Comment*: Preston has violated Standard III(B) by failing to deal fairly with her clients in taking these investment actions. Preston should have allocated the trades prior to executing the orders, or she should have had a systematic

> approach to allocating the trades, such as pro rata, as soon as practical after they were executed. Among other things, Preston must disclose to the client that the adviser may act as broker for, receive commissions from, and have a potential conflict of interest regarding both parties in agency cross-transactions. After the disclosure, she should obtain from the client consent authorizing such transactions in advance.

Example 5 (Selective Disclosure):

Saunders Industrial Waste Management (SIWM) publicly indicates to analysts that it is comfortable with the somewhat disappointing earnings-per-share projection of US$1.16 for the quarter. Bernard Roberts, an analyst at Coffey Investments, is confident that SIWM management has understated the forecasted earnings so that the real announcement will cause an "upside surprise" and boost the price of SIWM stock. The "whisper number" (rumored) estimate based on extensive research and discussed among knowledgeable analysts is higher than US$1.16. Roberts repeats the US$1.16 figure in his research report to all Coffey clients but informally tells his large clients that he expects the earnings per share to be higher, making SIWM a good buy.

> *Comment*: By not sharing his opinion regarding the potential for a significant upside earnings surprise with all clients, Roberts is not treating all clients fairly and has violated Standard III(B).

Example 6 (Additional Services for Select Clients):

Jenpin Weng uses e-mail to issue a new recommendation to all his clients. He then calls his three largest institutional clients to discuss the recommendation in detail.

> *Comment*: Weng has not violated Standard III(B) because he widely disseminated the recommendation and provided the information to all his clients prior to discussing it with a select few. Weng's largest clients received additional personal service because they presumably pay higher fees or because they have a large amount of assets under Weng's management. If Weng had discussed the report with a select group of clients prior to distributing it to all his clients, he would have violated Standard III(B).

Example 7 (Minimum Lot Allocations):

Lynn Hampton is a well-respected private wealth manager in her community with a diversified client base. She determines that a new 10-year bond being offered by Healthy Pharmaceuticals is appropriate for five of her clients. Three clients request to purchase US$10,000 each, and the other two request US$50,000 each. The minimum lot size is established at US$5,000, and the issue is oversubscribed at the time of placement. Her firm's policy is that odd-lot allocations, especially those below the minimum, should be avoided because they may affect the liquidity of the security at the time of sale.

Hampton is informed she will receive only US$55,000 of the offering for all accounts. Hampton distributes the bond investments as follows: The three accounts that requested US$10,000 are allocated US$5,000 each, and the two accounts that requested US$50,000 are allocated US$20,000 each.

Comment: Hampton has not violated Standard III(B), even though the distribution is not on a completely pro rata basis because of the required minimum lot size. With the total allocation being significantly below the amount requested, Hampton ensured that each client received at least the minimum lot size of the issue. This approach allowed the clients to efficiently sell the bond later if necessary.

Example 8 (Excessive Trading):

Ling Chan manages the accounts for many pension plans, including the plan of his father's employer. Chan developed similar but not identical investment policies for each client, so the investment portfolios are rarely the same. To minimize the cost to his father's pension plan, he intentionally trades more frequently in the accounts of other clients to ensure the required brokerage is incurred to continue receiving free research for use by all the pensions.

Comment: Chan is violating Standard III(B) because his trading actions are disadvantaging his clients to enhance a relationship with a preferred client. All clients are benefiting from the research being provided and should incur their fair portion of the costs. This does not mean that additional trading should occur if a client has not paid an equal portion of the commission; trading should occur only as required by the strategy.

Example 9 (Limited Social Media Disclosures):

Mary Burdette was recently hired by Fundamental Investment Management (FIM) as a junior auto industry analyst. Burdette is expected to expand the social media presence of the firm because she is active with various networks, including Facebook, LinkedIn, and Twitter. Although Burdette's supervisor, Joe Graf, has never used social media, he encourages Burdette to explore opportunities to increase FIM's online presence and ability to share content, communicate, and broadcast information to clients. In response to Graf's encouragement, Burdette is working on a proposal detailing the advantages of getting FIM onto Twitter in addition to launching a company Facebook page.

As part of her auto industry research for FIM, Burdette is completing a report on the financial impact of Sun Drive Auto Ltd.'s new solar technology for compact automobiles. This research report will be her first for FIM, and she believes Sun Drive's technology could revolutionize the auto industry. In her excitement, Burdette sends a quick tweet to FIM Twitter followers summarizing her "buy" recommendation for Sun Drive Auto stock.

Comment: Burdette has violated Standard III(B) by sending an investment recommendation to a select group of contacts prior to distributing it to all clients. Burdette must make sure she has received the appropriate training about FIM's policies and procedures, including the appropriate business use of personal social media networks before engaging in such activities.

See Standard IV(C) for guidance related to the duties of the supervisor.

Example 10 (Fair Dealing between Clients):

Paul Rove, performance analyst for Alpha-Beta Investment Management, is describing to the firm's chief investment officer (CIO) two new reports he would like to develop to assist the firm in meeting its obligations to treat clients fairly. Because many of the firm's clients have similar investment objectives and portfolios, Rove suggests a report detailing securities owned across several clients and the percentage of the portfolio the security represents. The second report would compare the monthly performance of portfolios with similar strategies. The outliers within each report would be submitted to the CIO for review.

> *Comment*: As a performance analyst, Rove likely has little direct contact with clients and thus has limited opportunity to treat clients differently. The recommended reports comply with Standard III(B) while helping the firm conduct after-the-fact reviews of how effectively the firm's advisers are dealing with their clients' portfolios. Reports that monitor the fair treatment of clients are an important oversight tool to ensure that clients are treated fairly.

STANDARD III(C): DUTIES TO CLIENTS – SUITABILITY

1. When Members and Candidates are in an advisory relationship with a client, they must:
   a. Make a reasonable inquiry into a client's or prospective client's investment experience, risk and return objectives, and financial constraints prior to making any investment recommendation or taking investment action and must reassess and update this information regularly.
   b. Determine that an investment is suitable to the client's financial situation and consistent with the client's written objectives, mandates, and constraints before making an investment recommendation or taking investment action.
   c. Judge the suitability of investments in the context of the client's total portfolio.
2. When Members and Candidates are responsible for managing a portfolio to a specific mandate, strategy, or style, they must make only investment recommendations or take only investment actions that are consistent with the stated objectives and constraints of the portfolio.

Guidance

Highlights:

- *Developing an Investment Policy*
- *Understanding the Client's Risk Profile*
- *Updating an Investment Policy*
- *The Need for Diversification*
- *Addressing Unsolicited Trading Requests*
- *Managing to an Index or Mandate*

Standard III(C) requires that members and candidates who are in an investment advisory relationship with clients consider carefully the needs, circumstances, and objectives of the clients when determining the appropriateness and suitability of a given investment or course of investment action. An appropriate suitability determination will not, however, prevent some investments or investment actions from losing value.

In judging the suitability of a potential investment, the member or candidate should review many aspects of the client's knowledge, experience related to investing, and financial situation. These aspects include, but are not limited to, the risk profile of the investment as compared with the constraints of the client, the impact of the investment on the diversity of the portfolio, and whether the client has the means or net worth to assume the associated risk. The investment professional's determination of suitability should reflect only the investment recommendations or actions that a prudent person would be willing to undertake. Not every investment opportunity will be suitable for every portfolio, regardless of the potential return being offered.

The responsibilities of members and candidates to gather information and make a suitability analysis prior to making a recommendation or taking investment action fall on those members and candidates who provide investment advice in the course of an advisory relationship with a client. Other members and candidates may be simply executing specific instructions for retail clients when buying or selling securities, such as shares in mutual funds. These members and candidates and some others, such as sell-side analysts, may not have the opportunity to judge the suitability of a particular investment for the ultimate client.

Developing an Investment Policy

When an advisory relationship exists, members and candidates must gather client information at the inception of the relationship. Such information includes the client's financial circumstances, personal data (such as age and occupation) that are relevant to investment decisions, attitudes toward risk, and objectives in investing. This information should be incorporated into a written investment policy statement (IPS) that addresses the client's risk tolerance, return requirements, and all investment constraints (including time horizon, liquidity needs, tax concerns, legal and regulatory factors, and unique circumstances). Without identifying such client factors, members and candidates cannot judge whether a particular investment or strategy is suitable for a particular client. The IPS also should identify and describe the roles and responsibilities of the parties to the advisory relationship and investment process, as well as schedules for review and evaluation of the IPS. After formulating long-term capital market expectations, members and candidates can assist in developing an appropriate strategic asset allocation and investment program for the client, whether these are presented in separate documents or incorporated in the IPS or in appendices to the IPS.

Understanding the Client's Risk Profile

One of the most important factors to be considered in matching appropriateness and suitability of an investment with a client's needs and circumstances is measuring that client's tolerance for risk. The investment professional must consider the possibilities of rapidly changing investment environments and their likely impact on a client's holdings, both individual securities and the collective portfolio. The risk of many investment strategies can and should be analyzed and quantified in advance.

The use of synthetic investment vehicles and derivative investment products has introduced particular issues of risk. Members and candidates should pay careful attention to the leverage inherent in many of these vehicles or products when considering them for use in a client's investment program. Such leverage and limited liquidity, depending on the degree to which they are hedged, bear directly on the issue of suitability for the client.

Updating an Investment Policy

Updating the IPS should be repeated at least annually and also prior to material changes to any specific investment recommendations or decisions on behalf of the client. The effort to determine the needs and circumstances of each client is not a one-time occurrence. Investment recommendations or decisions are usually part of an ongoing process that takes into account the diversity and changing nature of portfolio and client characteristics. The passage of time is bound to produce changes that are important with respect to investment objectives.

For an individual client, important changes might include the number of dependents, personal tax status, health, liquidity needs, risk tolerance, amount of wealth beyond that represented in the portfolio, and extent to which compensation and other income provide for current income needs. With respect to an institutional client, such changes might relate to the magnitude of unfunded liabilities in a pension fund, the withdrawal privileges in an employee savings plan, or the distribution requirements of a charitable foundation. Without efforts to update information concerning client factors, one or more factors could change without the investment manager's knowledge.

Suitability review can be done most effectively when the client fully discloses his or her complete financial portfolio, including those portions not managed by the member or candidate. If clients withhold information about their financial portfolios, the suitability analysis conducted by members and candidates cannot be expected to be complete; it must be based on the information provided.

The Need for Diversification

The investment profession has long recognized that combining several different investments is likely to provide a more acceptable level of risk exposure than having all assets in a single investment. The unique characteristics (or risks) of an individual investment may become partially or entirely neutralized when it is combined with other individual investments within a portfolio. Some reasonable amount of diversification is thus the norm for many portfolios, especially those managed by individuals or institutions that have some degree of legal fiduciary responsibility.

An investment with high relative risk on its own may be a suitable investment in the context of the entire portfolio or when the client's stated objectives contemplate speculative or risky investments. The manager may be responsible for only a portion of the client's total portfolio, or the client may not have provided a full financial picture. Members and candidates can be responsible for assessing the suitability of an investment only on the basis of the information and criteria actually provided by the client.

Addressing Unsolicited Trading Requests

Members and candidates may receive requests from a client for trades that do not properly align with the risk and return objectives outlined in the client's investment policy statement. These transaction requests may be based on the client's individual biases or professional experience. Members and candidates will need to make reasonable efforts to balance their clients' trading requests with their responsibilities to follow the agreed-on investment policy statement.

In cases of unsolicited trade requests that a member or candidate knows are unsuitable for a client, the member or candidate should refrain from making the trade until he or she discusses the concerns with the client. The discussions and resulting actions may encompass a variety of scenarios depending on how the requested unsuitable investment relates to the client's full portfolio.

Many times, an unsolicited request may be expected to have only a minimum impact on the entire portfolio because the size of the requested trade is small or the trade would result in a limited change to the portfolio's risk profile. In discussing the trade, the member or candidate should focus on educating the investor on how the request deviates from the current policy statement. Following the discussion, the member or candidate may follow his or her firm's policies regarding the necessary client approval for executing unsuitable trades. At a minimum, the client should acknowledge the discussion and accept the conditions that make the recommendation unsuitable.

Should the unsolicited request be expected to have a material impact on the portfolio, the member or candidate should use this opportunity to update the investment policy statement. Doing so would allow the client to fully understand the potential effect of the requested trade on his or her current goals or risk levels.

Members and candidates may have some clients who decline to modify their policy statements while insisting an unsolicited trade be made. In such instances, members or candidates will need to evaluate the effectiveness of their services to the client. The options available to the members or candidates will depend on the services provided by their employer. Some firms may allow for the trade to be executed in a new unmanaged account. If alternative options are not available, members and candidates ultimately will need to determine whether they should continue the advisory arrangement with the client.

Managing to an Index or Mandate

Some members and candidates do not manage money for individuals but are responsible for managing a fund to an index or an expected mandate. The responsibility of these members and candidates is to invest in a manner consistent with the stated mandate. For example, a member or candidate who serves as the fund manager for a large-cap income fund would not be following the fund mandate by investing heavily in small-cap or start-up companies whose stock is speculative in nature. Members and candidates who manage pooled assets to a specific mandate are not responsible for determining the suitability of the *fund* as an investment for investors who may be purchasing shares in the fund. The responsibility for determining the suitability of an investment for clients can be conferred only on members and candidates who have an advisory relationship with clients.

STANDARD III(C): RECOMMENDED PROCEDURES

Investment Policy Statement

To fulfill the basic provisions of Standard III(C), a member or candidate should put the needs and circumstances of each client and the client's investment objectives into a written investment policy statement. In formulating an investment policy for the client, the member or candidate should take the following into consideration:

- client identification—(1) type and nature of client, (2) the existence of separate beneficiaries, and (3) approximate portion of total client assets that the member or candidate is managing;
- investor objectives—(1) return objectives (income, growth in principal, maintenance of purchasing power) and (2) risk tolerance (suitability, stability of values);
- investor constraints—(1) liquidity needs, (2) expected cash flows (patterns of additions and/or withdrawals), (3) investable funds (assets and liabilities or other commitments), (4) time horizon, (5) tax considerations, (6) regulatory and legal circumstances, (7) investor preferences, prohibitions, circumstances, and unique needs, and (8) proxy voting responsibilities and guidance; and
- performance measurement benchmarks.

Regular Updates

The investor's objectives and constraints should be maintained and reviewed periodically to reflect any changes in the client's circumstances. Members and candidates should regularly compare client constraints with capital market expectations to arrive at an appropriate asset allocation. Changes in either factor may result in a fundamental change in asset allocation. Annual review is reasonable unless business or other reasons, such as a major change in market conditions, dictate more frequent review. Members and candidates should document attempts to carry out such a review if circumstances prevent it.

Suitability Test Policies

With the increase in regulatory required suitability tests, members and candidates should encourage their firms to develop related policies and procedures. The procedures will differ according to the size of the firm and the scope of the services offered to its clients.

The test procedures should require the investment professional to look beyond the potential return of the investment and include the following:

- an analysis of the impact on the portfolio's diversification,
- a comparison of the investment risks with the client's assessed risk tolerance, and
- the fit of the investment with the required investment strategy.

STANDARD III(C): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Investment Suitability—Risk Profile):

Caleb Smith, an investment adviser, has two clients: Larry Robertson, 60 years old, and Gabriel Lanai, 40 years old. Both clients earn roughly the same salary, but Robertson has a much higher risk tolerance because he has a large asset base. Robertson is willing to invest part of his assets very aggressively; Lanai wants only to achieve a steady rate of return with low volatility to pay for his children's education. Smith recommends investing 20% of both portfolios in zero-yield, small-cap, high-technology equity issues.

> *Comment*: In Robertson's case, the investment may be appropriate because of his financial circumstances and aggressive investment position, but this investment is not suitable for Lanai. Smith is violating Standard III(C) by applying Robertson's investment strategy to Lanai because the two clients' financial circumstances and objectives differ.

Example 2 (Investment Suitability—Entire Portfolio):

Jessica McDowell, an investment adviser, suggests to Brian Crosby, a risk-averse client, that covered call options be used in his equity portfolio. The purpose would be to enhance Crosby's income and partially offset any untimely depreciation in the portfolio's value should the stock market or other circumstances affect his holdings unfavorably. McDowell educates Crosby about all possible outcomes, including the risk of incurring an added tax liability if a stock rises in price and is called away and, conversely, the risk of his holdings losing protection on the downside if prices drop sharply.

> *Comment*: When determining suitability of an investment, the primary focus should be the characteristics of the client's entire portfolio, not the characteristics of single securities on an issue-by-issue basis. The basic characteristics of the entire portfolio will largely determine whether investment recommendations are taking client factors into account. Therefore, the most important aspects of a particular investment are those that will affect the characteristics of the total portfolio. In this case, McDowell properly considers the investment in the context of the entire portfolio and thoroughly explains the investment to the client.

Example 3 (IPS Updating):

In a regular meeting with client Seth Jones, the portfolio managers at Blue Chip Investment Advisors are careful to allow some time to review his current needs and circumstances. In doing so, they learn that some significant changes have recently taken place in his life. A wealthy uncle left Jones an inheritance that increased his net worth fourfold, to US$1 million.

> *Comment*: The inheritance has significantly increased Jones's ability (and possibly his willingness) to assume risk and has diminished the average yield required to meet his current income needs. Jones's financial circumstances have definitely changed, so Blue Chip managers must update Jones's investment policy statement to reflect how his investment objectives have changed. Accordingly, the Blue Chip portfolio managers should consider a somewhat higher equity ratio for his portfolio than was called for by the previous circumstances, and the managers' specific common stock recommendations might be heavily tilted toward low-yield, growth-oriented issues.

Example 4 (Following an Investment Mandate):

Louis Perkowski manages a high-income mutual fund. He purchases zero-dividend stock in a financial services company because he believes the stock is undervalued and is in a potential growth industry, which makes it an attractive investment.

> *Comment*: A zero-dividend stock does not seem to fit the mandate of the fund that Perkowski is managing. Unless Perkowski's investment fits within the mandate or is within the realm of allowable investments the fund has made clear in its disclosures, Perkowski has violated Standard III(C).

Example 5 (IPS Requirements and Limitations):

Max Gubler, chief investment officer of a property/casualty insurance subsidiary of a large financial conglomerate, wants to improve the diversification of the subsidiary's investment portfolio and increase its returns. The subsidiary's investment policy statement provides for highly liquid investments, such as large-cap equities and government, supranational, and corporate bonds with a minimum credit rating of AA and maturity of no more than five years. In a recent presentation, a venture capital group offered very attractive prospective returns on some of its private equity funds that provide seed capital to ventures. An exit strategy was already contemplated, but investors would have to observe a minimum three-year lockup period and a subsequent laddered exit option for a maximum of one-third of their shares per year. Gubler does not want to miss this opportunity. After extensive analysis, with the intent to optimize the return on the equity assets within the subsidiary's current portfolio, he invests 4% in this seed fund, leaving the portfolio's total equity exposure still well below its upper limit.

> *Comment*: Gubler is violating Standard III(A)–Loyalty, Prudence, and Care as well as Standard III(C). His new investment locks up part of the subsidiary's assets for at least three years and up to as many as five years and possibly beyond. The IPS requires investments in highly liquid investments and describes accepted asset classes; private equity investments with a lockup period certainly do not qualify. Even without a lockup period, an asset class with only an occasional, and thus implicitly illiquid, market may not be suitable for the portfolio. Although an IPS typically describes objectives and constraints in great detail, the manager must also make every effort to understand the client's business and circumstances. Doing so should enable the manager to recognize, understand, and discuss with the client other factors that may be or may become material in the investment management process.

Example 6 (Submanager and IPS Reviews):

Paul Ostrowski's investment management business has grown significantly over the past couple of years, and some clients want to diversify internationally. Ostrowski decides to find a submanager to handle the expected international investments. Because this will be his first subadviser, Ostrowski uses the CFA Institute model "request for proposal" to design a questionnaire for his search. By his deadline, he receives seven completed questionnaires from a variety of domestic and international firms trying to gain his business. Ostrowski reviews all the applications in detail and decides to select the firm that charges the lowest fees because doing so will have the least impact on his firm's bottom line.

> *Comment*: When selecting an external manager or subadviser, Ostrowski needs to ensure that the new manager's services are appropriate for his clients. This due diligence includes comparing the risk profile of the clients with the investment strategy of the manager. In basing the decision on the fee structure alone, Ostrowski may be violating Standard III(C).
>
> When clients ask to diversify into international products, it is an appropriate time to review and update the clients' IPSs. Ostrowski's review may determine that the risk of international investments modifies the risk profiles of the clients or does not represent an appropriate investment.
>
> See also Standard V(A)–Diligence and Reasonable Basis for further discussion of the review process needed in selecting appropriate submanagers.

Example 7 (Investment Suitability—Risk Profile):

Samantha Snead, a portfolio manager for Thomas Investment Counsel, Inc., specializes in managing public retirement funds and defined benefit pension plan accounts, all of which have long-term investment objectives. A year ago, Snead's employer, in an attempt to motivate and retain key investment professionals, introduced a bonus compensation system that rewards portfolio managers on the basis of quarterly performance relative to their peers and to certain benchmark indexes. In an attempt to improve the short-term performance of her accounts, Snead changes her investment strategy and purchases several high-beta stocks for client portfolios. These purchases are seemingly contrary to the clients' investment policy statements. Following their purchase, an officer of Griffin Corporation, one of Snead's pension fund clients, asks why Griffin Corporation's portfolio seems to be dominated by high-beta stocks of companies that often appear among the most actively traded issues. No change in objective or strategy has been recommended by Snead during the year.

> *Comment*: Snead violated Standard III(C) by investing the clients' assets in high-beta stocks. These high-risk investments are contrary to the long-term risk profile established in the clients' IPSs. Snead has changed the investment strategy of the clients in an attempt to reap short-term rewards offered by her firm's new compensation arrangement, not in response to changes in clients' investment policy statements.
>
> See also Standard VI(A)–Disclosure of Conflicts.

Example 8 (Investment Suitability):

Andre Shrub owns and operates Conduit, an investment advisory firm. Prior to opening Conduit, Shrub was an account manager with Elite Investment, a hedge fund managed by his good friend Adam Reed. To attract clients to a new Conduit fund, Shrub offers lower-than-normal management fees. He can do so because the fund consists of two top-performing funds managed by Reed. Given his personal friendship with Reed and the prior performance record of these two funds, Shrub believes this new fund is a winning combination for all parties. Clients quickly invest with Conduit to gain access to the Elite funds. No one is turned away because Conduit is seeking to expand its assets under management.

> *Comment*: Shrub has violated Standard III(C) because the risk profile of the new fund may not be suitable for every client. As an investment adviser, Shrub needs to establish an investment policy statement for each client and recommend only investments that match each client's risk and return profile in the IPS. Shrub is required to act as more than a simple sales agent for Elite.
>
> Although Shrub cannot disobey the direct request of a client to purchase a specific security, he should fully discuss the risks of a planned purchase and provide reasons why it might not be suitable for a client. This requirement may lead members and candidates to decline new customers if those customers' requested investment decisions are significantly out of line with their stated requirements.
>
> See also Standard V(A)–Diligence and Reasonable Basis.

STANDARD III(D): DUTIES TO CLIENTS - PERFORMANCE PRESENTATION

When communicating investment performance information, Members and Candidates must make reasonable efforts to ensure that it is fair, accurate, and complete.

Guidance

Standard III(D) requires members and candidates to provide credible performance information to clients and prospective clients and to avoid misstating performance or misleading clients and prospective clients about the investment performance of members or candidates or their firms. This standard encourages full disclosure of investment performance data to clients and prospective clients.

Standard III(D) covers any practice that would lead to misrepresentation of a member's or candidate's performance record, whether the practice involves performance presentation or performance measurement. This standard prohibits misrepresentations of past performance or reasonably expected performance. A member or candidate must give a fair and complete presentation of performance information whenever communicating data with respect to the performance history of individual accounts, composites or groups of accounts, or composites of an analyst's or firm's performance results. Furthermore, members and candidates should not state or imply that clients will obtain or benefit from a rate of return that was generated in the past.

The requirements of this standard are not limited to members and candidates managing separate accounts. Whenever a member or candidate provides performance information for which the manager is claiming responsibility, such as for pooled funds, the history must be accurate. Research analysts promoting the success or accuracy of their recommendations must ensure that their claims are fair, accurate, and complete.

If the presentation is brief, the member or candidate must make available to clients and prospects, on request, the detailed information supporting that communication. Best practice dictates that brief presentations include a reference to the limited nature of the information provided.

STANDARD III(D): RECOMMENDED PROCEDURES

Apply the GIPS Standards

For members and candidates who are showing the performance history of the assets they manage, compliance with the GIPS standards is the best method to meet their obligations under Standard III(D). Members and candidates should encourage their firms to comply with the GIPS standards.

Compliance without Applying GIPS Standards

Members and candidates can also meet their obligations under Standard III(D) by

- considering the knowledge and sophistication of the audience to whom a performance presentation is addressed,

- presenting the performance of the weighted composite of similar portfolios rather than using a single representative account,
- including terminated accounts as part of performance history with a clear indication of when the accounts were terminated,
- including disclosures that fully explain the performance results being reported (for example, stating, when appropriate, that results are simulated when model results are used, clearly indicating when the performance record is that of a prior entity, or disclosing whether the performance is gross of fees, net of fees, or after tax), and
- maintaining the data and records used to calculate the performance being presented.

STANDARD III(D): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Performance Calculation and Length of Time):

Kyle Taylor of Taylor Trust Company, noting the performance of Taylor's common trust fund for the past two years, states in a brochure sent to his potential clients, "You can expect steady 25% annual compound growth of the value of your investments over the year." Taylor Trust's common trust fund did increase at the rate of 25% per year for the past year, which mirrored the increase of the entire market. The fund has never averaged that growth for more than one year, however, and the average rate of growth of all of its trust accounts for five years is 5% per year.

> *Comment*: Taylor's brochure is in violation of Standard III(D). Taylor should have disclosed that the 25% growth occurred only in one year. Additionally, Taylor did not include client accounts other than those in the firm's common trust fund. A general claim of firm performance should take into account the performance of all categories of accounts. Finally, by stating that clients can expect a steady 25% annual compound growth rate, Taylor is also violating Standard I(C)–Misrepresentation, which prohibits assurances or guarantees regarding an investment.

Example 2 (Performance Calculation and Asset Weighting):

Anna Judd, a senior partner of Alexander Capital Management, circulates a performance report for the capital appreciation accounts for the years 1988 through 2004. The firm claims compliance with the GIPS standards. Returns are not calculated in accordance with the requirements of the GIPS standards, however, because the composites are not asset weighted.

> *Comment*: Judd is in violation of Standard III(D). When claiming compliance with the GIPS standards, firms must meet *all* of the requirements, make mandatory disclosures, and meet any other requirements that apply to that firm's specific situation. Judd's violation is not from any misuse of the data but from a false claim of GIPS compliance.

Example 3 (Performance Presentation and Prior Fund/Employer):

Aaron McCoy is vice president and managing partner of the equity investment group of Mastermind Financial Advisors, a new business. Mastermind recruited McCoy because he had a proven six-year track record with G&P Financial. In developing Mastermind's advertising and marketing campaign, McCoy prepares an advertisement that includes the equity investment performance he achieved at G&P Financial. The advertisement for Mastermind does not identify the equity performance as being earned while at G&P. The advertisement is distributed to existing clients and prospective clients of Mastermind.

> *Comment*: McCoy has violated Standard III(D) by distributing an advertisement that contains material misrepresentations about the historical performance of Mastermind. Standard III(D) requires that members and candidates make every reasonable effort to ensure that performance information is a fair, accurate, and complete representation of an individual's or firm's performance. As a general matter, this standard does not prohibit showing past performance of funds managed at a prior firm as part of a performance track record as long as showing that record is accompanied by appropriate disclosures about where the performance took place and the person's specific role in achieving that performance. If McCoy chooses to use his past performance from G&P in Mastermind's advertising, he should make full disclosure of the source of the historical performance.

Example 4 (Performance Presentation and Simulated Results):

Jed Davis has developed a mutual fund selection product based on historical information from the 1990–95 period. Davis tested his methodology by applying it retroactively to data from the 1996–2003 period, thus producing simulated performance results for those years. In January 2004, Davis's employer decided to offer the product and Davis began promoting it through trade journal advertisements and direct dissemination to clients. The advertisements included the performance results for the 1996–2003 period but did not indicate that the results were simulated.

> *Comment*: Davis violated Standard III(D) by failing to clearly identify simulated performance results. Standard III(D) prohibits members and candidates from making any statements that misrepresent the performance achieved by them or their firms and requires members and candidates to make every reasonable effort to ensure that performance information presented to clients is fair, accurate, and complete. Use of simulated results should be accompanied by full disclosure as to the source of the performance data, including the fact that the results from 1995 through 2003 were the result of applying the model retroactively to that time period.

Example 5 (Performance Calculation and Selected Accounts Only):

In a presentation prepared for prospective clients, William Kilmer shows the rates of return realized over a five-year period by a “composite” of his firm’s discretionary accounts that have a “balanced” objective. This composite, however, consisted of only a few of the accounts that met the balanced criterion set by the firm, excluded accounts under a certain asset level without disclosing the fact of their exclusion, and included accounts that did not have the balanced mandate because those accounts would boost the investment results. In addition, to achieve better results, Kilmer manipulated the narrow range of accounts included in the composite by changing the accounts that made up the composite over time.

> *Comment*: Kilmer violated Standard III(D) by misrepresenting the facts in the promotional material sent to prospective clients, distorting his firm’s performance record, and failing to include disclosures that would have clarified the presentation.

Example 6 (Performance Attribution Changes):

Art Purell is reviewing the quarterly performance attribution reports for distribution to clients. Purell works for an investment management firm with a bottom-up, fundamentals-driven investment process that seeks to add value through stock selection. The attribution methodology currently compares each stock with its sector. The attribution report indicates that the value added this quarter came from asset allocation and that stock selection contributed negatively to the calculated return.

Through running several different scenarios, Purell discovers that calculating attribution by comparing each stock with its industry and then rolling the effect to the sector level improves the appearance of the manager’s stock selection activities. Because the firm defines the attribution terms and the results better reflect the stated strategy, Purell recommends that the client reports should use the revised methodology.

> *Comment*: Modifying the attribution methodology without proper notifications to clients would fail to meet the requirements of Standard III(D). Purrell’s recommendation is being done solely for the interest of the firm to improve its perceived ability to meet the stated investment strategy. Such changes are unfair to clients and obscure the facts regarding the firm’s abilities.
>
> Had Purell believed the new methodology offered improvements to the original model, then he would have needed to report the results of both calculations to the client. The report should also include the reasons why the new methodology is preferred, which would allow the client to make a meaningful comparison to prior results and provide a basis for comparing future attributions.

Example 7 (Performance Calculation Methodology Disclosure):

While developing a new reporting package for existing clients, Alisha Singh, a performance analyst, discovers that her company’s new system automatically calculates both time-weighted and money-weighted returns. She asks the head of client services and retention which value would be preferred given that the firm has various investment strategies that include bonds, equities, securities without leverage, and alternatives. Singh is told not to label the return value so that the firm may show whichever value is greatest for the period.

Comment: Following these instructions would lead to Singh violating Standard III(D). In reporting inconsistent return values, Singh would not be providing complete information to the firm's clients. Full information is provided when clients have sufficient information to judge the performance generated by the firm.

Example 8 (Performance Calculation Methodology Disclosure):

Richmond Equity Investors manages a long–short equity fund in which clients can trade once a week (on Fridays). For transparency reasons, a daily net asset value of the fund is calculated by Richmond. The monthly fact sheets of the fund report month-to-date and year-to-date performance. Richmond publishes the performance based on the higher of the last trading day of the month (typically, not the last business day) or the last business day of the month as determined by Richmond. The fact sheet mentions only that the data are as of the end of the month, without giving the exact date. Maggie Clark, the investment performance analyst in charge of the calculations, is concerned about the frequent changes and asks her supervisor whether they are appropriate.

Comment: Clark's actions in questioning the changing performance metric comply with Standard III(D). She has shown concern that these changes are not presenting an accurate and complete picture of the performance generated.

STANDARD III(E): DUTIES TO CLIENTS - PRESERVATION OF CONFIDENTIALITY

Members and Candidates must keep information about current, former, and prospective clients confidential unless:

1 The information concerns illegal activities on the part of the client;

2 Disclosure is required by law; or

3 The client or prospective client permits disclosure of the information.

Guidance

Highlights:

- *Status of Client*
- *Compliance with Laws*
- *Electronic Information and Security*
- *Professional Conduct Investigations by CFA Institute*

Standard III(E) requires that members and candidates preserve the confidentiality of information communicated to them by their clients, prospective clients, and former clients. This standard is applicable when (1) the member or candidate receives information because of his or her special ability to conduct a portion of the client's business or personal affairs and (2) the member or candidate receives information that arises from or is relevant to that portion of the client's business that is the subject of the special or confidential relationship. If disclosure of the information is required by law or the information concerns illegal activities by the client, however, the member or candidate may have an obligation to report the activities to the appropriate authorities.

Status of Client

This standard protects the confidentiality of client information even if the person or entity is no longer a client of the member or candidate. Therefore, members and candidates must continue to maintain the confidentiality of client records even after the client relationship has ended. If a client or former client expressly authorizes the member or candidate to disclose information, however, the member or candidate may follow the terms of the authorization and provide the information.

Compliance with Laws

As a general matter, members and candidates must comply with applicable law. If applicable law requires disclosure of client information in certain circumstances, members and candidates must comply with the law. Similarly, if applicable law requires members and candidates to maintain confidentiality, even if the information concerns illegal activities on the part of the client, members and candidates should not disclose such information. Additionally, applicable laws, such as inter-departmental communication restrictions within financial institutions, can impose limitations on information flow about a client within an entity that may lead to a violation of confidentiality. When in doubt, members and candidates should consult with their employer's compliance personnel or legal counsel before disclosing confidential information about clients.

Electronic Information and Security

Because of the ever-increasing volume of electronically stored information, members and candidates need to be particularly aware of possible accidental disclosures. Many employers have strict policies about how to electronically communicate sensitive client information and store client information on personal laptops, mobile devices, or portable disk/flash drives. In recent years, regulatory authorities have imposed stricter data security laws applying to the use of mobile remote digital communication, including the use of social media, that must be considered. Standard III(E) does not require members or candidates to become experts in information security technology, but they should have a thorough understanding of the policies of their employer. The size and operations of the firm will lead to differing policies for ensuring the security of confidential information maintained within the firm. Members and candidates should encourage their firm to conduct regular periodic training on confidentiality procedures for all firm personnel, including portfolio associates, receptionists, and other non-investment staff who have routine direct contact with clients and their records.

Professional Conduct Investigations by CFA Institute

The requirements of Standard III(E) are not intended to prevent members and candidates from cooperating with an investigation by the CFA Institute Professional Conduct Program (PCP). When permissible under applicable law, members and candidates shall consider the PCP an extension of themselves when requested to provide information about a client in support of a PCP investigation into their own conduct. Members and candidates are encouraged to cooperate with investigations into the conduct of others. Any information turned over to the PCP is kept in the strictest confidence. Members and candidates will not be considered in violation of this standard by forwarding confidential information to the PCP.

STANDARD III(E): RECOMMENDED PROCEDURES

The simplest, most conservative, and most effective way to comply with Standard III(E) is to avoid disclosing any information received from a client except to authorized fellow employees who are also working for the client. In some instances, however, a member or candidate may want to disclose information received from clients that is outside the scope of the confidential relationship and does not involve illegal activities. Before making such a disclosure, a member or candidate should ask the following:

- In what context was the information disclosed? If disclosed in a discussion of work being performed for the client, is the information relevant to the work?
- Is the information background material that, if disclosed, will enable the member or candidate to improve service to the client?

Members and candidates need to understand and follow their firm's electronic information communication and storage procedures. If the firm does not have procedures in place, members and candidates should encourage the development of procedures that appropriately reflect the firm's size and business operations.

Communicating with Clients

Technological changes are constantly enhancing the methods that are used to communicate with clients and prospective clients. Members and candidates should make reasonable efforts to ensure that firm-supported communication methods and compliance procedures follow practices designed for preventing accidental distribution of confidential information. Given the rate at which technology changes, a regular review of privacy protection measures is encouraged.

Members and candidates should be diligent in discussing with clients the appropriate methods for providing confidential information. It is important to convey to clients that not all firm-sponsored resources may be appropriate for such communications.

STANDARD III(E): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Possessing Confidential Information):

Sarah Connor, a financial analyst employed by Johnson Investment Counselors, Inc., provides investment advice to the trustees of City Medical Center. The trustees have given her a number of internal reports concerning City Medical's needs for physical plant renovation and expansion. They have asked Connor to recommend investments that would generate capital appreciation in endowment funds to meet projected capital expenditures. Connor is approached by a local businessman, Thomas Kasey, who is considering a substantial contribution either to City Medical Center or to another local hospital. Kasey wants to find out the building plans of both institutions before making a decision, but he does not want to speak to the trustees.

Comment: The trustees gave Connor the internal reports so she could advise them on how to manage their endowment funds. Because the information in the reports is clearly both confidential and within the scope of the confidential relationship, Standard III(E) requires that Connor refuse to divulge information to Kasey.

Example 2 (Disclosing Confidential Information):

Lynn Moody is an investment officer at the Lester Trust Company. She has an advisory customer who has talked to her about giving approximately US$50,000 to charity to reduce her income taxes. Moody is also treasurer of the Home for Indigent Widows (HIW), which is planning its annual giving campaign. HIW hopes to expand its list of prospects, particularly those capable of substantial gifts. Moody recommends that HIW's vice president for corporate gifts call on her customer and ask for a donation in the US$50,000 range.

Comment: Even though the attempt to help the Home for Indigent Widows was well intended, Moody violated Standard III(E) by revealing confidential information about her client.

Example 3 (Disclosing Possible Illegal Activity):

Government officials approach Casey Samuel, the portfolio manager for Garcia Company's pension plan, to examine pension fund records. They tell her that Garcia's corporate tax returns are being audited and the pension fund is being reviewed. Two days earlier, Samuel had learned in a regular investment review with Garcia officers that potentially excessive and improper charges were being made to the pension plan by Garcia. Samuel consults her employer's general counsel and is advised that Garcia has probably violated tax and fiduciary regulations and laws.

Comment: Samuel should inform her supervisor of these activities, and her employer should take steps, with Garcia, to remedy the violations. If that approach is not successful, Samuel and her employer should seek advice of legal counsel to determine the appropriate steps to be taken. Samuel may well have a duty to disclose the evidence she has of the continuing legal violations and to resign as asset manager for Garcia.

Example 4 (Disclosing Possible Illegal Activity):

David Bradford manages money for a family-owned real estate development corporation. He also manages the individual portfolios of several of the family members and officers of the corporation, including the chief financial officer (CFO). Based on the financial records of the corporation and some questionable practices of the CFO that Bradford has observed, Bradford believes that the CFO is embezzling money from the corporation and putting it into his personal investment account.

Comment: Bradford should check with his firm's compliance department or appropriate legal counsel to determine whether applicable securities regulations require reporting the CFO's financial records.

Example 5 (Accidental Disclosure of Confidential Information):

Lynn Moody is an investment officer at the Lester Trust Company (LTC). She has stewardship of a significant number of individually managed taxable accounts. In addition to receiving quarterly written reports, about a dozen high-net-worth individuals have indicated to Moody a willingness to receive communications about overall economic and financial market outlooks directly from her by way of a social media platform. Under the direction of her firm's technology and compliance departments, she established a new group page on an existing social media platform specifically for her clients. In the instructions provided to clients, Moody asked them to "join" the group so they may be granted access to the posted content. The instructions also advised clients that all comments posted would be available to the public and thus the platform was not an appropriate method for communicating personal or confidential information.

Six months later, in early January, Moody posted LTC's year-end "Market Outlook." The report outlined a new asset allocation strategy that the firm is adding to its recommendations in the new year. Moody introduced the publication with a note informing her clients that she would be discussing the changes with them individually in their upcoming meetings.

One of Moody's clients responded directly on the group page that his family recently experienced a major change in their financial profile. The client described highly personal and confidential details of the event. Unfortunately, all clients that were part of the group were also able to read the detailed posting until Moody was able to have the comment removed.

> *Comment*: Moody has taken reasonable steps for protecting the confidentiality of client information while using the social media platform. She provided instructions clarifying that all information posted to the site would be publically viewable to all group members and warned against using this method for communicating confidential information. The accidental disclosure of confidential information by a client is not under Moody's control. Her actions to remove the information promptly once she became aware further align with Standard III(E).
>
> In understanding the potential sensitivity clients express surrounding the confidentiality of personal information, this event highlights a need for further training. Moody might advocate for additional warnings or controls for clients when they consider using social media platforms for two-way communications.

STANDARD IV(A): DUTIES TO EMPLOYERS – LOYALTY

Standard IV(A) Loyalty

In matters related to their employment, Members and Candidates must act for the benefit of their employer and not deprive their employer of the advantage of their skills and abilities, divulge confidential information, or otherwise cause harm to their employer.

Guidance

Highlights:

- *Employer Responsibilities*
- *Independent Practice*
- *Leaving an Employer*
- *Use of Social Media*
- *Whistleblowing*
- *Nature of Employment*

Standard IV(A) requires members and candidates to protect the interests of their firm by refraining from any conduct that would injure the firm, deprive it of profit, or deprive it of the member's or candidate's skills and ability. Members and candidates must always place the interests of clients above the interests of their employer but should also consider the effects of their conduct on the sustainability and integrity of the employer firm. In matters related to their employment, members and candidates must not engage in conduct that harms the interests of their employer. Implicit in this standard is the obligation of members and candidates to comply with the policies and procedures established by their employers that govern the employer–employee relationship—to the extent that such policies and procedures do not conflict with applicable laws, rules, or regulations or the Code and Standards.

This standard is not meant to be a blanket requirement to place employer interests ahead of personal interests in all matters. The standard does not require members and candidates to subordinate important personal and family obligations to their work. Members and candidates should enter into a dialogue with their employer about balancing personal and employment obligations when personal matters may interfere with their work on a regular or significant basis.

Employer Responsibilities

The employer–employee relationship imposes duties and responsibilities on both parties. Employers must recognize the duties and responsibilities that they owe to their employees if they expect to have content and productive employees.

Members and candidates are encouraged to provide their employer with a copy of the Code and Standards. These materials will inform the employer of the responsibilities of a CFA Institute member or a candidate in the CFA Program. The Code and Standards also serve as a basis for questioning employer policies and practices that conflict with these responsibilities.

Employers are not obligated to adhere to the Code and Standards. In expecting to retain competent employees who are members and candidates, however, they should not develop conflicting policies and procedures. The employer is responsible for a positive working environment, which includes an ethical workplace. Senior management has the additional responsibility to devise compensation structures and incentive arrangements that do not encourage unethical behavior.

Independent Practice

Included in Standard IV(A) is the requirement that members and candidates abstain from independent competitive activity that could conflict with the interests of their employer. Although Standard IV(A) does not preclude members or candidates from entering into an independent business while still employed, members and candidates who plan to engage in independent practice for compensation must notify their employer and describe the types of services they will render to prospective independent clients, the expected duration of the services, and the compensation for the services. Members and candidates should not render services until they receive consent from their employer to all of the terms of the arrangement. “Practice” means any service that the employer currently makes available for remuneration. “Undertaking independent practice” means engaging in competitive business, as opposed to making preparations to begin such practice.

Leaving an Employer

When members and candidates are planning to leave their current employer, they must continue to act in the employer’s best interest. They must not engage in any activities that would conflict with this duty until their resignation becomes effective. It is difficult to define specific guidelines for those members and candidates who are planning to compete with their employer as part of a new venture. The circumstances of each situation must be reviewed to distinguish permissible preparations from violations of duty. Activities that might constitute a violation, especially in combination, include the following:

- misappropriation of trade secrets,
- misuse of confidential information,
- solicitation of the employer’s clients prior to cessation of employment,
- self-dealing (appropriating for one’s own property a business opportunity or information belonging to one’s employer), and
- misappropriation of clients or client lists.

A departing employee is generally free to make arrangements or preparations to go into a competitive business before terminating the relationship with his or her employer as long as such preparations do not breach the employee’s duty of loyalty. A member or candidate who is contemplating seeking other employment must not contact existing clients or potential clients prior to leaving his or her employer for purposes of soliciting their business for the new employer. Once notice is provided to the employer of the intent to resign, the member or candidate must follow the employer’s policies and procedures related to notifying clients of his or her planned departure. In addition, the member or candidate must not take records or files to a new employer without the written permission of the previous employer.

Once an employee has left the firm, the skills and experience that an employee obtained while employed are not “confidential” or “privileged” information. Similarly, simple knowledge of the names and existence of former clients is generally not confidential information unless deemed such by an agreement or by law. Standard IV(A) does not prohibit experience or knowledge gained at one employer from being used at another employer. Firm records or work performed on behalf of the firm that is stored in paper copy or electronically for the member’s or candidate’s convenience while employed, however, should be erased or returned to the employer unless the firm gives permission to keep those records after employment ends.

The standard does not prohibit former employees from contacting clients of their previous firm as long as the contact information does not come from the records of the former employer or violate an applicable “noncompete agreement.” Members and candidates are free to use public information after departing to contact former clients without violating Standard IV(A) as long as there is no specific agreement not to do so.

Employers often require employees to sign noncompete agreements that preclude a departing employee from engaging in certain conduct. Members and candidates should take care to review the terms of any such agreement when leaving their employer to determine what, if any, conduct those agreements may prohibit.

In some markets, there are agreements between employers within an industry that outline information that departing employees are permitted to take upon resignation, such as the "Protocol for Broker Recruiting" in the United States. These agreements ease individuals' transition between firms that have agreed to follow the outlined procedures. Members and candidates who move between firms that sign such agreements may rely on the protections provided as long as they faithfully adhere to all the procedures outlined.

For example, under the agreement between many US brokers, individuals are allowed to take some general client contact information when departing. To be protected, a copy of the information the individual is taking must be provided to the local management team for review. Additionally, the specific client information may only be used by the departing employee and not others employed by the new firm.

Use of Social Media

The growth in various online networking platforms, such as LinkedIn, Twitter, and Facebook (commonly referred to as social media platforms), is providing new opportunities and challenges for businesses. Members and candidates should understand and abide by all applicable firm policies and regulations as to the acceptable use of social media platforms to interact with clients and prospective clients. This is especially important when a member or candidate is planning to leave an employer.

Social media use makes determining how and when departure notification is delivered to clients more complex. Members and candidates may have developed profiles on these platforms that include connections with individuals who are clients of the firm, as well as individuals unrelated to their employer. Communications through social media platforms that potentially reach current clients should adhere to the employer's policies and procedures regarding notification of departing employees.

Social media connections with clients are also raising questions concerning the differences between public information and firm property. Specific accounts and user profiles of members and candidates may be created for solely professional reasons, including firm-approved accounts for client engagements. Such firm-approved business-related accounts would be considered part of the firm's assets, thus requiring members and candidates to transfer or delete the accounts as directed by their firm's policies and procedures. Best practice for members and candidates is to maintain separate accounts for their personal and professional social media activities. Members and candidates should discuss with their employers how profiles should be treated when a single account includes personal connections and also is used to conduct aspects of their professional activities.

Whistleblowing

A member's or candidate's personal interests, as well as the interests of his or her employer, are secondary to protecting the integrity of capital markets and the interests of clients. Therefore, circumstances may arise (e.g., when an employer is engaged in illegal or unethical activity) in which members and candidates must act contrary to their employer's interests in order to comply with their duties to the market and clients. In such instances, activities that would normally violate a member's or candidate's duty to his or her employer (such as contradicting employer instructions, violating certain policies and procedures, or preserving a record by copying employer records) may be justified. Such action would be permitted only if the intent is clearly aimed at protecting clients or the integrity of the market, not for personal gain.

Nature of Employment

A wide variety of business relationships exists within the investment industry. For instance, a member or candidate may be an employee or an independent contractor. Members and candidates must determine whether they are employees or independent contractors in order to determine the applicability of Standard IV(A). This issue will be decided largely by the degree of control exercised by the employing entity over the member or candidate. Factors determining control include whether the member's or candidate's hours, work location, and other parameters of the job are set; whether facilities are provided to the member or candidate; whether the member's or candidate's expenses are reimbursed; whether the member or candidate seeks work from other employers; and the number of clients or employers the member or candidate works for.

A member's or candidate's duties within an independent contractor relationship are governed by the oral or written agreement between the member and the client. Members and candidates should take care to define clearly the scope of their responsibilities and the expectations of each client within the context of each relationship. Once a member or candidate establishes a relationship with a client, the member or candidate has a duty to abide by the terms of the agreement.

STANDARD IV(A): RECOMMENDED PROCEDURES

Employers may establish codes of conduct and operating procedures for their employees to follow. Members and candidates should fully understand the policies to ensure that they are not in conflict with the Code and Standards. The following topics identify policies that members and candidates should encourage their firms to adopt if the policies are not currently in place.

Competition Policy

A member or candidate must understand any restrictions placed by the employer on offering similar services outside the firm while employed by the firm. The policy may outline the procedures for requesting approval to undertake the outside service or may be a strict prohibition of such service. If a member's or candidate's employer elects to have its employees sign a noncompete agreement as part of the employment agreement, the member or candidate should ensure that the details are clear and fully explained prior to signing the agreement.

Termination Policy

Members and candidates should clearly understand the termination policies of their employer. Termination policies should establish clear procedures regarding the resignation process, including addressing how the termination will be disclosed to clients and staff and whether updates posted through social media platforms will be allowed. The firm's policy may also outline the procedures for transferring ongoing research and account management responsibilities. Finally, the procedures should address agreements that allow departing employees to remove specific client-related information upon resignation.

Incident-Reporting Procedures

Members and candidates should be aware of their firm's policies related to whistleblowing and encourage their firm to adopt industry best practices in this area. Many firms are required by regulatory mandates to establish confidential and anonymous reporting procedures that allow employees to report potentially unethical and illegal activities in the firm.

Employee Classification

Members and candidates should understand their status within their employer firm. Firms are encouraged to adopt a standardized classification structure (e.g., part time, full time, outside contractor) for their employees and indicate how each of the firm's policies applies to each employee class.

STANDARD IV(A): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Soliciting Former Clients):

Samuel Magee manages pension accounts for Trust Assets, Inc., but has become frustrated with the working environment and has been offered a position with Fiduciary Management. Before resigning from Trust Assets, Magee asks four big accounts to leave that firm and open accounts with Fiduciary. Magee also persuades several prospective clients to sign agreements with Fiduciary Management. Magee had previously made presentations to these prospects on behalf of Trust Assets.

> *Comment*: Magee violated the employee–employer principle requiring him to act solely for his employer's benefit. Magee's duty is to Trust Assets as long as he is employed there. The solicitation of Trust Assets' current clients and prospective clients is unethical and violates Standard IV(A).

Example 2 (Former Employer's Documents and Files):

James Hightower has been employed by Jason Investment Management Corporation for 15 years. He began as an analyst but assumed increasing responsibilities and is now a senior portfolio manager and a member of the firm's investment policy committee. Hightower has decided to leave Jason Investment and start his own investment management business. He has been careful not to tell any of Jason's clients that he is leaving; he does not want to be accused of breaching his duty to Jason by soliciting Jason's clients before his departure. Hightower is planning to copy and take with him the following documents and information he developed or worked on while at Jason: (1) the client list, with addresses, telephone numbers, and other pertinent client information; (2) client account statements; (3) sample marketing presentations to prospective clients containing Jason's performance record; (4) Jason's recommended list of securities; (5) computer models to determine asset allocations for accounts with various objectives; (6) computer models for stock selection; and (7) personal computer spreadsheets for Hightower's major corporate recommendations, which he developed when he was an analyst.

Comment: Except with the consent of their employer, departing members and candidates may not take employer property, which includes books, records, reports, and other materials, because taking such materials may interfere with their employer's business opportunities. Taking any employer records, even those the member or candidate prepared, violates Standard IV(A). Employer records include items stored in hard copy or any other medium (e.g., home computers, portable storage devices, cell phones).

Example 3 (Addressing Rumors):

Reuben Winston manages all-equity portfolios at Target Asset Management (TAM), a large, established investment counselor. Ten years previously, Philpott & Company, which manages a family of global bond mutual funds, acquired TAM in a diversification move. After the merger, the combined operations prospered in the fixed-income business but the equity management business at TAM languished. Lately, a few of the equity pension accounts that had been with TAM before the merger have terminated their relationships with TAM. One day, Winston finds on his voice mail the following message from a concerned client: "Hey! I just heard that Philpott is close to announcing the sale of your firm's equity management business to Rugged Life. What is going on?" Not being aware of any such deal, Winston and his associates are stunned. Their internal inquiries are met with denials from Philpott management, but the rumors persist. Feeling left in the dark, Winston contemplates leading an employee buyout of TAM's equity management business.

Comment: An employee-led buyout of TAM's equity asset management business would be consistent with Standard IV(A) because it would rest on the permission of the employer and, ultimately, the clients. In this case, however, in which employees suspect the senior managers or principals are not truthful or forthcoming, Winston should consult legal counsel to determine appropriate action.

Example 4 (Ownership of Completed Prior Work):

Laura Clay, who is unemployed, wants part-time consulting work while seeking a full-time analyst position. During an interview at Bradley Associates, a large institutional asset manager, Clay is told that the firm has no immediate research openings but would be willing to pay her a flat fee to complete a study of the wireless communications industry within a given period of time. Clay would be allowed unlimited access to Bradley's research files and would be welcome to come to the offices and use whatever support facilities are available during normal working hours. Bradley's research director does not seek any exclusivity for Clay's output, and the two agree to the arrangement on a handshake. As Clay nears completion of the study, she is offered an analyst job in the research department of Winston & Company, a brokerage firm, and she is pondering submitting the draft of her wireless study for publication by Winston.

Comment: Although she is under no written contractual obligation to Bradley, Clay has an obligation to let Bradley act on the output of her study before Winston & Company or Clay uses the information to their advantage. That is, unless Bradley gives permission to Clay and waives its rights to her wireless report, Clay would be in violation of Standard IV(A) if she were to immediately recommend to Winston the same transactions recommended in the report to Bradley. Furthermore, Clay must not take from Bradley any research file material or other property that she may have used.

Example 5 (Ownership of Completed Prior Work):

Emma Madeline, a recent college graduate and a candidate in the CFA Program, spends her summer as an unpaid intern at Murdoch and Lowell. The senior managers at Murdoch are attempting to bring the firm into compliance with the GIPS standards, and Madeline is assigned to assist in its efforts. Two months into her internship, Madeline applies for a job at McMillan & Company, which has plans to become GIPS compliant. Madeline accepts the job with McMillan. Before leaving Murdoch, she copies the firm's software that she helped develop because she believes this software will assist her in her new position.

Comment: Even though Madeline does not receive monetary compensation for her services at Murdoch, she has used firm resources in creating the software and is considered an employee because she receives compensation and benefits in the form of work experience and knowledge. By copying the software, Madeline violated Standard IV(A) because she misappropriated Murdoch's property without permission.

Example 6 (Soliciting Former Clients):

Dennis Elliot has hired Sam Chisolm, who previously worked for a competing firm. Chisolm left his former firm after 18 years of employment. When Chisolm begins working for Elliot, he wants to contact his former clients because he knows them well and is certain that many will follow him to his new employer. Is Chisolm in violation of Standard IV(A) if he contacts his former clients?

Comment: Because client records are the property of the firm, contacting former clients for any reason through the use of client lists or other information taken from a former employer without permission would be a violation of Standard IV(A). In addition, the nature and extent of the contact with former clients may be governed by the terms of any noncompete agreement signed by the employee and the former employer that covers contact with former clients after employment.

Simple knowledge of the names and existence of former clients is not confidential information, just as skills or experience that an employee obtains while employed are not "confidential" or "privileged" information. The Code and Standards do not impose a prohibition on the use of experience or knowledge gained at one employer from being used at another employer. The Code and Standards also do not prohibit former employees from contacting clients of their previous firm, in the absence of a noncompete agreement. Members and candidates are free to use public information about their former firm after departing to contact former clients without violating Standard IV(A).

In the absence of a noncompete agreement, as long as Chisolm maintains his duty of loyalty to his employer before joining Elliot's firm, does not take steps to solicit clients until he has left his former firm, and does not use material from his former employer without its permission after he has left, he is not in violation of the Code and Standards.

Example 7 (Starting a New Firm):

Geraldine Allen currently works at a registered investment company as an equity analyst. Without notice to her employer, she registers with government authorities to start an investment company that will compete with her employer, but she does not actively seek clients. Does registration of this competing company with the appropriate regulatory authorities constitute a violation of Standard IV(A)?

> *Comment*: Allen's preparation for the new business by registering with the regulatory authorities does not conflict with the work for her employer if the preparations have been done on Allen's own time outside the office and if Allen will not be soliciting clients for the business or otherwise operating the new company until she has left her current employer.

Example 8 (Competing with Current Employer):

Several employees are planning to depart their current employer within a few weeks and have been careful to not engage in any activities that would conflict with their duty to their current employer. They have just learned that one of their employer's clients has undertaken a request for proposal (RFP) to review and possibly hire a new investment consultant. The RFP has been sent to the employer and all of its competitors. The group believes that the new entity to be formed would be qualified to respond to the RFP and be eligible for the business. The RFP submission period is likely to conclude before the employees' resignations are effective. Is it permissible for the group of departing employees to respond to the RFP for their anticipated new firm?

> *Comment*: A group of employees responding to an RFP that their employer is also responding to would lead to direct competition between the employees and the employer. Such conduct violates Standard IV(A) unless the group of employees receives permission from their employer as well as the entity sending out the RFP.

Example 9 (Externally Compensated Assignments):

Alfonso Mota is a research analyst with Tyson Investments. He works part time as a mayor for his hometown, a position for which he receives compensation. Must Mota seek permission from Tyson to serve as mayor?

> *Comment*: If Mota's mayoral duties are so extensive and time-consuming that they might detract from his ability to fulfill his responsibilities at Tyson, he should discuss his outside activities with his employer and come to a mutual agreement regarding how to manage his personal commitments with his responsibilities to his employer.

Example 10 (Soliciting Former Clients):

After leaving her employer, Shawna McQuillen establishes her own money management business. While with her former employer, she did not sign a noncompete agreement that would have prevented her from soliciting former clients. Upon her departure, she does not take any of her client lists or contact information and she clears her personal computer of any employer records, including client contact information. She obtains the phone numbers of her former clients through public records and contacts them to solicit their business.

Comment: McQuillen is not in violation of Standard IV(A) because she has not used information or records from her former employer and is not prevented by an agreement with her former employer from soliciting her former clients.

Example 11 (Whistleblowing Actions):

Meredith Rasmussen works on a buy-side trading desk and concentrates on in-house trades for a hedge fund subsidiary managed by a team at the investment management firm. The hedge fund has been very successful and is marketed globally by the firm. From her experience as the trader for much of the activity of the fund, Rasmussen has become quite knowledgeable about the hedge fund's strategy, tactics, and performance. When a distinct break in the market occurs, however, and many of the securities involved in the hedge fund's strategy decline markedly in value, Rasmussen observes that the reported performance of the hedge fund does not reflect this decline. In her experience, the lack of any effect is a very unlikely occurrence. She approaches the head of trading about her concern and is told that she should not ask any questions and that the fund is big and successful and is not her concern. She is fairly sure something is not right, so she contacts the compliance officer, who also tells her to stay away from the issue of this hedge fund's reporting.

Comment: Rasmussen has clearly come upon an error in policies, procedures, and compliance practices in the firm's operations. Having been unsuccessful in finding a resolution with her supervisor and the compliance officer, Rasmussen should consult the firm's whistleblowing policy to determine the appropriate next step toward informing management of her concerns. The potentially unethical actions of the investment management division are appropriate grounds for further disclosure, so Rasmussen's whistleblowing would not represent a violation of Standard IV(A).

See also Standard I(D)–Misconduct and Standard IV(C)–Responsibilities of Supervisors.

Example 12 (Soliciting Former Clients):

Angel Crome has been a private banker for YBSafe Bank for the past eight years. She has been very successful and built a considerable client portfolio during that time but is extremely frustrated by the recent loss of reputation by her current employer and subsequent client insecurity. A locally renowned headhunter contacted Crome a few days ago and offered her an interesting job with a competing private bank. This bank offers a substantial signing bonus for advisers with their own client portfolios. Crome figures that she can solicit at least 70% of her clients to follow her and gladly enters into the new employment contract.

Comment: Crome may contact former clients upon termination of her employment with YBSafe Bank, but she is prohibited from using client records built by and kept with her in her capacity as an employee of YBSafe Bank. Client lists are proprietary information of her former employer and must not be used for her or her new employer's benefit. The use of written, electronic, or any other form of records other than publicly available information to contact her former clients at YBSafe Bank will be a violation of Standard IV(A).

Example 13 (Notification of Code and Standards):

Krista Smith is a relatively new assistant trader for the fixed-income desk of a major investment bank. She is on a team responsible for structuring collateralized debt obligations (CDOs) made up of securities in the inventory of the trading desk. At a meeting of the team, senior executives explain the opportunity to eventually separate the CDO into various risk-rated tranches to be sold to the clients of the firm. After the senior executives leave the meeting, the head trader announces various responsibilities of each member of the team and then says, "This is a good time to unload some of the junk we have been stuck with for a while and disguise it with ratings and a thick, unreadable prospectus, so don't be shy in putting this CDO together. Just kidding." Smith is worried by this remark and asks some of her colleagues what the head trader meant. They all respond that he was just kidding but that there is some truth in the remark because the CDO is seen by management as an opportunity to improve the quality of the securities in the firm's inventory.

Concerned about the ethical environment of the workplace, Smith decides to talk to her supervisor about her concerns and provides the head trader with a copy of the Code and Standards. Smith discusses the principle of placing the client above the interest of the firm and the possibility that the development of the new CDO will not adhere to this responsibility. The head trader assures Smith that the appropriate analysis will be conducted when determining the appropriate securities for collateral. Furthermore, the ratings are assigned by an independent firm and the prospectus will include full and factual disclosures. Smith is reassured by the meeting, but she also reviews the company's procedures and requirements for reporting potential violations of company policy and securities laws.

> *Comment*: Smith's review of the company policies and procedures for reporting violations allows her to be prepared to report through the appropriate whistleblower process if she decides that the CDO development process involves unethical actions by others. Smith's actions comply with the Code and Standards principles of placing the client's interests first and being loyal to her employer. In providing her supervisor with a copy of the Code and Standards, Smith is highlighting the high level of ethical conduct she is required to adhere to in her professional activities.

Example 14 (Leaving an Employer):

Laura Webb just left her position as portfolio analyst at Research Systems, Inc. (RSI). Her employment contract included a non-solicitation agreement that requires her to wait two years before soliciting RSI clients for any investment-related services. Upon leaving, Webb was informed that RSI would contact clients immediately about her departure and introduce her replacement.

While working at RSI, Webb connected with clients, other industry associates, and friends through her LinkedIn network. Her business and personal relationships were intermingled because she considered many of her clients to be personal friends. Realizing that her LinkedIn network would be a valuable resource for new employment opportunities, she updated her profile several days following her departure from RSI. LinkedIn automatically sent a notification to Webb's entire network that her employment status had been changed in her profile.

> *Comment*: Prior to her departure, Webb should have discussed any client information contained in her social media networks. By updating her LinkedIn profile after RSI notified clients and after her employment ended, she has appropriately placed her employer's interests ahead of her own

personal interests. In addition, she has not violated the non-solicitation agreement with RSI, unless it prohibited any contact with clients during the two-year period.

Example 15 (Confidential Firm Information):

Sanjay Gupta is a research analyst at Naram Investment Management (NIM). NIM uses a team-based research process to develop recommendations on investment opportunities covered by the team members. Gupta, like others, provides commentary for NIM's clients through the company blog, which is posted weekly on the NIM password-protected website. According to NIM's policy, every contribution to the website must be approved by the company's compliance department before posting. Any opinions expressed on the website are disclosed as representing the perspective of NIM.

Gupta also writes a personal blog to share his experiences with friends and family. As with most blogs, Gupta's personal blog is widely available to interested readers through various internet search engines. Occasionally, when he disagrees with the team-based research opinions of NIM, Gupta uses his personal blog to express his own opinions as a counterpoint to the commentary posted on the NIM website. Gupta believes this provides his readers with a more complete perspective on these investment opportunities.

> *Comment*: Gupta is in violation of Standard IV(A) for disclosing confidential firm information through his personal blog. The recommendations on the firm's blog to clients are not freely available across the internet, but his personal blog post indirectly provides the firm's recommendations.
>
> Additionally, by posting research commentary on his personal blog, Gupta is using firm resources for his personal advantage. To comply with Standard IV(A), members and candidates must receive consent from their employer prior to using company resources.

STANDARD IV(B): DUTIES TO EMPLOYERS - ADDITIONAL COMPENSATION ARRANGEMENTS

> Members and Candidates must not accept gifts, benefits, compensation, or consideration that competes with or might reasonably be expected to create a conflict of interest with their employer's interest unless they obtain written consent from all parties involved.

Guidance

Standard IV(B) requires members and candidates to obtain permission from their employer before accepting compensation or other benefits from third parties for the services rendered to the employer or for any services that might create a conflict with their employer's interest. Compensation and benefits include direct compensation by the client and any indirect compensation or other benefits received from third parties. "Written consent" includes any form of communication that can be documented (for example, communication via e-mail that can be retrieved and documented).

Members and candidates must obtain permission for additional compensation/benefits because such arrangements may affect loyalties and objectivity and create potential conflicts of interest. Disclosure allows an employer to consider the outside arrangements when evaluating the actions and motivations of members and candidates. Moreover, the employer is entitled to have full knowledge of all compensation/benefit arrangements so as to be able to assess the true cost of the services members or candidates are providing.

There may be instances in which a member or candidate is hired by an employer on a "part-time" basis. "Part-time" status applies to employees who do not commit the full number of hours required for a normal work week. Members and candidates should discuss possible limitations to their abilities to provide services that may be competitive with their employer during the negotiation and hiring process. The requirements of Standard IV(B) would be applicable to limitations identified at that time.

STANDARD IV(B): RECOMMENDED PROCEDURES

Members and candidates should make an immediate written report to their supervisor and compliance officer specifying any compensation they propose to receive for services in addition to the compensation or benefits received from their employer. The details of the report should be confirmed by the party offering the additional compensation, including performance incentives offered by clients. This written report should state the terms of any agreement under which a member or candidate will receive additional compensation; "terms" include the nature of the compensation, the approximate amount of compensation, and the duration of the agreement.

STANDARD IV(B): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Notification of Client Bonus Compensation):

Geoff Whitman, a portfolio analyst for Adams Trust Company, manages the account of Carol Cochran, a client. Whitman is paid a salary by his employer, and Cochran pays the trust company a standard fee based on the market value of assets in her portfolio. Cochran proposes to Whitman that "any year that my portfolio achieves at least a 15% return before taxes, you and your wife can fly to Monaco at my expense and use my condominium during the third week of January." Whitman does not inform his employer of the arrangement and vacations in Monaco the following January as Cochran's guest.

> *Comment*: Whitman violated Standard IV(B) by failing to inform his employer in writing of this supplemental, contingent compensation arrangement. The nature of the arrangement could have resulted in partiality to

Cochran's account, which could have detracted from Whitman's performance with respect to other accounts he handles for Adams Trust. Whitman must obtain the consent of his employer to accept such a supplemental benefit.

Example 2 (Notification of Outside Compensation):

Terry Jones sits on the board of directors of Exercise Unlimited, Inc. In return for his services on the board, Jones receives unlimited membership privileges for his family at all Exercise Unlimited facilities. Jones purchases Exercise Unlimited stock for the client accounts for which it is appropriate. Jones does not disclose this arrangement to his employer because he does not receive monetary compensation for his services to the board.

Comment: Jones has violated Standard IV(B) by failing to disclose to his employer benefits received in exchange for his services on the board of directors. The nonmonetary compensation may create a conflict of interest in the same manner as being paid to serve as a director.

Example 3 (Prior Approval for Outside Compensation):

Jonathan Hollis is an analyst of oil-and-gas companies for Specialty Investment Management. He is currently recommending the purchase of ABC Oil Company shares and has published a long, well-thought-out research report to substantiate his recommendation. Several weeks after publishing the report, Hollis receives a call from the investor-relations office of ABC Oil saying that Thomas Andrews, CEO of the company, saw the report and really liked the analyst's grasp of the business and his company. The investor-relations officer invites Hollis to visit ABC Oil to discuss the industry further. ABC Oil offers to send a company plane to pick Hollis up and arrange for his accommodations while visiting. Hollis, after gaining the appropriate approvals, accepts the meeting with the CEO but declines the offered travel arrangements.

Several weeks later, Andrews and Hollis meet to discuss the oil business and Hollis's report. Following the meeting, Hollis joins Andrews and the investment relations officer for dinner at an upscale restaurant near ABC Oil's headquarters.

Upon returning to Specialty Investment Management, Hollis provides a full review of the meeting to the director of research, including a disclosure of the dinner attended.

Comment: Hollis's actions did not violate Standard IV(B). Through gaining approval before accepting the meeting and declining the offered travel arrangements, Hollis sought to avoid any potential conflicts of interest between his company and ABC Oil. Because the location of the dinner was not available prior to arrival and Hollis notified his company of the dinner upon his return, accepting the dinner should not impair his objectivity. By disclosing the dinner, Hollis has enabled Specialty Investment Management to assess whether it has any impact on future reports and recommendations by Hollis related to ABC Oil.

STANDARD IV(C): DUTIES TO EMPLOYERS - RESPONSIBILITIES OF SUPERVISORS

Members and Candidates must make reasonable efforts to ensure that anyone subject to their supervision or authority complies with applicable laws, rules, regulations, and the Code and Standards.

Guidance

Highlights:

- *System for Supervision*
- *Supervision Includes Detection*

Standard IV(C) states that members and candidates must promote actions by all employees under their supervision and authority to comply with applicable laws, rules, regulations, and firm policies and the Code and Standards.

Any investment professional who has employees subject to her or his control or influence—whether or not the employees are CFA Institute members, CFA charterholders, or candidates in the CFA Program—exercises supervisory responsibility. Members and candidates acting as supervisors must also have in-depth knowledge of the Code and Standards so that they can apply this knowledge in discharging their supervisory responsibilities.

The conduct that constitutes reasonable supervision in a particular case depends on the number of employees supervised and the work performed by those employees. Members and candidates with oversight responsibilities for large numbers of employees may not be able to personally evaluate the conduct of these employees on a continuing basis. These members and candidates may delegate supervisory duties to subordinates who directly oversee the other employees. A member's or candidate's responsibilities under Standard IV(C) include instructing those subordinates to whom supervision is delegated about methods to promote compliance, including preventing and detecting violations of laws, rules, regulations, firm policies, and the Code and Standards.

At a minimum, Standard IV(C) requires that members and candidates with supervisory responsibility make reasonable efforts to prevent and detect violations by ensuring the establishment of effective compliance systems. However, an effective compliance system goes beyond enacting a code of ethics, establishing policies and procedures to achieve compliance with the code and applicable law, and reviewing employee actions to determine whether they are following the rules.

To be effective supervisors, members and candidates should implement education and training programs on a recurring or regular basis for employees under their supervision. Such programs will assist the employees with meeting their professional obligations to practice in an ethical manner within the applicable legal system. Further, establishing incentives—monetary or otherwise—for employees not only to meet business goals but also to reward ethical behavior offers supervisors another way to assist employees in complying with their legal and ethical obligations.

Often, especially in large organizations, members and candidates may have supervisory responsibility but not the authority to establish or modify firm-wide compliance policies and procedures or incentive structures. Such limitations should not prevent a member or candidate from working with his or her own superiors and within the firm structure to develop and implement effective compliance tools, including but not limited to:

- a code of ethics,
- compliance policies and procedures,
- education and training programs,
- an incentive structure that rewards ethical conduct, and
- adoption of firm-wide best practice standards (e.g., the GIPS standards, the CFA Institute Asset Manager Code of Professional Conduct).

A member or candidate with supervisory responsibility should bring an inadequate compliance system to the attention of the firm's senior managers and recommend corrective action. If the member or candidate clearly cannot discharge supervisory responsibilities because of the absence of a compliance system or because of an inadequate compliance system, the member or candidate should decline in writing to accept supervisory responsibility until the firm adopts reasonable procedures to allow adequate exercise of supervisory responsibility.

System for Supervision

Members and candidates with supervisory responsibility must understand what constitutes an adequate compliance system for their firms and make reasonable efforts to see that appropriate compliance procedures are established, documented, communicated to covered personnel, and followed. "Adequate" procedures are those designed to meet industry standards, regulatory requirements, the requirements of the Code and Standards, and the circumstances of the firm. Once compliance procedures are established, the supervisor must also make reasonable efforts to ensure that the procedures are monitored and enforced.

To be effective, compliance procedures must be in place prior to the occurrence of a violation of the law or the Code and Standards. Although compliance procedures cannot be designed to anticipate every potential violation, they should be designed to anticipate the activities most likely to result in misconduct. Compliance programs must be appropriate for the size and nature of the organization. The member or candidate should review model compliance procedures or other industry programs to ensure that the firm's procedures meet the minimum industry standards.

Once a supervisor learns that an employee has violated or may have violated the law or the Code and Standards, the supervisor must promptly initiate an assessment to determine the extent of the wrongdoing. Relying on an employee's statements about the extent of the violation or assurances that the wrongdoing will not reoccur is not enough. Reporting the misconduct up the chain of command and warning the employee to cease the activity are also not enough. Pending the outcome of the investigation, a supervisor should take steps to ensure that the violation will not be repeated, such as placing limits on the employee's activities or increasing the monitoring of the employee's activities.

Supervision Includes Detection

Members and candidates with supervisory responsibility must also make reasonable efforts to detect violations of laws, rules, regulations, firm policies, and the Code and Standards. The supervisors exercise reasonable supervision by establishing and implementing written compliance procedures and ensuring that those procedures are followed through periodic review. If a member or candidate has adopted reasonable procedures and taken steps to institute an effective compliance program, then the member or candidate may not be in violation of Standard IV(C) if he or she does not detect violations that occur despite these efforts. The fact that violations do occur may indicate, however, that the compliance procedures are inadequate. In addition, in some cases, merely enacting such procedures may not be sufficient to fulfill the duty required by Standard IV(C). A member or candidate may be in violation of Standard IV(C) if he or she knows or should know that the procedures designed to promote compliance, including detecting and preventing violations, are not being followed.

STANDARD IV(C): RECOMMENDED PROCEDURES

Codes of Ethics or Compliance Procedures

Members and candidates are encouraged to recommend that their employers adopt a code of ethics. Adoption of a code of ethics is critical to establishing a strong ethical foundation for investment advisory firms and their employees. Codes of ethics formally emphasize and reinforce the client loyalty responsibilities of investment firm personnel, protect investing clients by deterring misconduct, and protect the firm's reputation for integrity.

There is a distinction, however, between codes of ethics and the specific policies and procedures needed to ensure compliance with the codes and with securities laws and regulations. Although both are important, codes of ethics should consist of fundamental, principle-based ethical and fiduciary concepts that are applicable to all of the firm's employees. In this way, firms can best convey to employees and clients the ethical ideals that investment advisers strive to achieve. These concepts need to be implemented, however, by detailed, firm-wide compliance policies and procedures. Compliance procedures assist the firm's personnel in fulfilling the responsibilities enumerated in the code of ethics and make probable that the ideals expressed in the code of ethics will be adhered to in the day-to-day operation of the firm.

Stand-alone codes of ethics should be written in plain language and should address general fiduciary concepts. They should be unencumbered by numerous detailed procedures. Codes presented in this way are the most effective in stressing to employees that they are in positions of trust and must act with integrity at all times. Mingling compliance procedures in the firm's code of ethics goes against the goal of reinforcing the ethical obligations of employees.

Separating the code of ethics from compliance procedures will also reduce, if not eliminate, the legal terminology and "boilerplate" language that can make the underlying ethical principles incomprehensible to the average person. Above all, to ensure the creation of a culture of ethics and integrity rather than one that merely focuses on following the rules, the principles in the code of ethics must be stated in a way that is accessible and understandable to everyone in the firm.

Members and candidates should encourage their employers to provide their codes of ethics to clients. In this case also, a simple, straightforward code of ethics will be best understood by clients. Unencumbered by the compliance procedures, the code of ethics will be effective in conveying that the firm is committed to conducting business in an ethical manner and in the best interests of the clients.

Adequate Compliance Procedures

A supervisor complies with Standard IV(C) by identifying situations in which legal violations or violations of the Code and Standards are likely to occur and by establishing and enforcing compliance procedures to prevent such violations. Adequate compliance procedures should

- be contained in a clearly written and accessible manual that is tailored to the firm's operations,
- be drafted so that the procedures are easy to understand,
- designate a compliance officer whose authority and responsibility are clearly defined and who has the necessary resources and authority to implement the firm's compliance procedures,
- describe the hierarchy of supervision and assign duties among supervisors,
- implement a system of checks and balances,
- outline the scope of the procedures,
- outline procedures to document the monitoring and testing of compliance procedures,
- outline permissible conduct, and
- delineate procedures for reporting violations and sanctions.

Once a compliance program is in place, a supervisor should

- disseminate the contents of the program to appropriate personnel,
- periodically update procedures to ensure that the measures are adequate under the law,
- continually educate personnel regarding the compliance procedures,
- issue periodic reminders of the procedures to appropriate personnel,
- incorporate a professional conduct evaluation as part of an employee's performance review,
- review the actions of employees to ensure compliance and identify violators, and
- take the necessary steps to enforce the procedures once a violation has occurred.

Once a violation is discovered, a supervisor should

- respond promptly,
- conduct a thorough investigation of the activities to determine the scope of the wrongdoing,
- increase supervision or place appropriate limitations on the wrongdoer pending the outcome of the investigation, and
- review procedures for potential changes necessary to prevent future violations from occurring.

Implementation of Compliance Education and Training

No amount of ethics education and awareness will deter someone determined to commit fraud for personal enrichment. But the vast majority of investment professionals strive to achieve personal success with dedicated service to their clients and employers.

Regular ethics and compliance training, in conjunction with adoption of a code of ethics, is critical to investment firms seeking to establish a strong culture of integrity and to provide an environment in which employees routinely engage in ethical conduct in compliance with the law. Training and education assist individuals in both recognizing areas that are prone to ethical and legal pitfalls and identifying those circumstances and influences that can impair ethical judgment.

By implementing educational programs, supervisors can train their subordinates to put into practice what the firm's code of ethics requires. Education helps employees make the link between legal and ethical conduct and the long-term success of the business; a strong culture of compliance signals to clients and potential clients that the firm has truly embraced ethical conduct as fundamental to the firm's mission to serve its clients.

Establish an Appropriate Incentive Structure

Even if individuals want to make the right choices and follow an ethical course of conduct and are aware of the obstacles that may trip them up, they can still be influenced to act improperly by a corporate culture that embraces a "succeed at all costs" mentality, stresses results regardless of the methods used to achieve those results, and does not reward ethical behavior. Supervisors can reinforce an individual's natural desire to "do the right thing" by building a culture of integrity in the workplace.

Supervisors and firms must look closely at their incentive structure to determine whether the structure encourages profits and returns at the expense of ethically appropriate conduct. Reward structures may turn a blind eye to how desired outcomes are achieved and encourage dysfunctional or counterproductive behavior. Only when compensation and incentives are firmly tied to client interests and *how* outcomes are achieved, rather than *how much* is generated for the firm, will employees work to achieve a culture of integrity.

STANDARD IV(C): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Supervising Research Activities):

Jane Mattock, senior vice president and head of the research department of H&V, Inc., a regional brokerage firm, has decided to change her recommendation for Timber Products from buy to sell. In line with H&V's procedures, she orally advises certain other H&V executives of her proposed actions before the report is prepared for publication. As a result of Mattock's conversation with Dieter Frampton, one of the H&V executives accountable to Mattock, Frampton immediately sells Timber's stock from his own account and from certain discretionary client accounts. In addition, other personnel inform certain institutional customers of the changed recommendation before it is printed and disseminated to all H&V customers who have received previous Timber reports.

Comment: Mattock has violated Standard IV(C) by failing to reasonably and adequately supervise the actions of those accountable to her. She did not prevent or establish reasonable procedures designed to prevent dissemination of or trading on the information by those who knew of her changed recommendation. She must ensure that her firm has procedures for reviewing or recording any trading in the stock of a corporation that has been the subject of an unpublished change in recommendation. Adequate procedures would have informed the subordinates of their duties and detected sales by Frampton and selected customers.

Example 2 (Supervising Research Activities):

Deion Miller is the research director for Jamestown Investment Programs. The portfolio managers have become critical of Miller and his staff because the Jamestown portfolios do not include any stock that has been the subject of a merger or tender offer. Georgia Ginn, a member of Miller's staff, tells Miller that she has been studying a local company, Excelsior, Inc., and recommends its purchase. Ginn adds that the company has been widely rumored to be the subject of a merger study by a well-known conglomerate and discussions between them are under way. At Miller's request, Ginn prepares a memo recommending the stock. Miller passes along Ginn's memo to the portfolio managers prior to leaving for vacation, and he notes that he has not reviewed the memo. As a result of the memo, the portfolio managers buy Excelsior stock immediately. The day Miller returns to the office, he learns that Ginn's only sources for the report were her brother, who is an acquisitions analyst with Acme Industries, the "well-known conglomerate," and that the merger discussions were planned but not held.

Comment: Miller violated Standard IV(C) by not exercising reasonable supervision when he disseminated the memo without checking to ensure that Ginn had a reasonable and adequate basis for her recommendations and that Ginn was not relying on material nonpublic information.

Example 3 (Supervising Trading Activities):

David Edwards, a trainee trader at Wheeler & Company, a major national brokerage firm, assists a customer in paying for the securities of Highland, Inc., by using anticipated profits from the immediate sale of the same securities. Despite the fact that Highland is not on Wheeler's recommended list, a large volume of its stock is traded through Wheeler in this manner. Roberta Ann Mason is a Wheeler vice president responsible for supervising compliance with the securities laws in the trading department. Part of her compensation from Wheeler is based on commission revenues from the trading department. Although she notices the increased trading activity, she does nothing to investigate or halt it.

Comment: Mason's failure to adequately review and investigate purchase orders in Highland stock executed by Edwards and her failure to supervise the trainee's activities violate Standard IV(C). Supervisors should be especially sensitive to actual or potential conflicts between their own self-interests and their supervisory responsibilities.

Example 4 (Supervising Trading Activities and Record Keeping):

Samantha Tabbing is senior vice president and portfolio manager for Crozet, Inc., a registered investment advisory and registered broker/dealer firm. She reports to Charles Henry, the president of Crozet. Crozet serves as the investment adviser and principal underwriter for ABC and XYZ public mutual funds. The two funds' prospectuses allow Crozet to trade financial futures for the funds for the limited purpose of hedging against market risks. Henry, extremely impressed by Tabbing's performance in the past two years, directs Tabbing to act as portfolio manager for the funds. For the benefit of its employees, Crozet has also organized the Crozet Employee Profit-Sharing Plan (CEPSP), a defined contribution retirement plan. Henry assigns Tabbing to manage 20% of the assets of CEPSP. Tabbing's investment objective for her portion of CEPSP's assets is aggressive growth. Unbeknownst to Henry, Tabbing frequently places S&P 500 Index purchase and sale orders for the funds and the CEPSP without providing the futures commission merchants (FCMs) who take the orders with any prior or simultaneous designation of the account for which the trade has been placed. Frequently, neither Tabbing nor anyone else at Crozet completes an internal trade ticket to record the time an order was placed or the specific account for which the order was intended. FCMs often designate a specific account only after the trade, when Tabbing provides such designation. Crozet has no written operating procedures or compliance manual concerning its futures trading, and its compliance department does not review such trading. After observing the market's movement, Tabbing assigns to CEPSP the S&P 500 positions with more favorable execution prices and assigns positions with less favorable execution prices to the funds.

> *Comment*: Henry violated Standard IV(C) by failing to adequately supervise Tabbing with respect to her S&P 500 trading. Henry further violated Standard IV(C) by failing to establish record-keeping and reporting procedures to prevent or detect Tabbing's violations. Henry must make a reasonable effort to determine that adequate compliance procedures covering all employee trading activity are established, documented, communicated, and followed.

Example 5 (Accepting Responsibility):

Meredith Rasmussen works on a buy-side trading desk and concentrates on in-house trades for a hedge fund subsidiary managed by a team at the investment management firm. The hedge fund has been very successful and is marketed globally by the firm. From her experience as the trader for much of the activity of the fund, Rasmussen has become quite knowledgeable about the hedge fund's strategy, tactics, and performance. When a distinct break in the market occurs and many of the securities involved in the hedge fund's strategy decline markedly in value, however, Rasmussen observes that the reported performance of the hedge fund does not at all reflect this decline. From her experience, this lack of an effect is a very unlikely occurrence. She approaches the head of trading about her concern and is told that she should not ask any questions and that the fund is too big and successful and is not her concern. She is fairly sure something is not right, so she contacts the compliance officer and is again told to stay away from the hedge fund reporting issue.

> *Comment*: Rasmussen has clearly come upon an error in policies, procedures, and compliance practices within the firm's operations. According to Standard IV(C), the supervisor and the compliance officer have the responsibility to review the concerns brought forth by Rasmussen. Supervisors

have the responsibility of establishing and encouraging an ethical culture in the firm. The dismissal of Rasmussen's question violates Standard IV(C) and undermines the firm's ethical operations.

See also Standard I(D)–Misconduct and, for guidance on whistleblowing, Standard IV(A)–Loyalty.

Example 6 (Inadequate Procedures):

Brendan Witt, a former junior sell-side technology analyst, decided to return to school to earn an MBA. To keep his research skills and industry knowledge sharp, Witt accepted a position with On-line and Informed, an independent internet-based research company. The position requires the publication of a recommendation and report on a different company every month. Initially, Witt is a regular contributor of new research and a participant in the associated discussion boards that generally have positive comments on the technology sector. Over time, his ability to manage his educational requirements and his work requirements begin to conflict with one another. Knowing a recommendation is due the next day for On-line, Witt creates a report based on a few news articles and what the conventional wisdom of the markets has deemed the "hot" security of the day.

Comment: Allowing the report submitted by Witt to be posted highlights a lack of compliance procedures by the research firm. Witt's supervisor needs to work with the management of On-line to develop an appropriate review process to ensure that all contracted analysts comply with the requirements.

See also Standard V(A)–Diligence and Reasonable Basis because it relates to Witt's responsibility for substantiating a recommendation.

Example 7 (Inadequate Supervision):

Michael Papis is the chief investment officer of his state's retirement fund. The fund has always used outside advisers for the real estate allocation, and this information is clearly presented in all fund communications. Thomas Nagle, a recognized sell-side research analyst and Papis's business school classmate, recently left the investment bank he worked for to start his own asset management firm, Accessible Real Estate. Nagle is trying to build his assets under management and contacts Papis about gaining some of the retirement fund's allocation. In the previous few years, the performance of the retirement fund's real estate investments was in line with the fund's benchmark but was not extraordinary. Papis decides to help out his old friend and also to seek better returns by moving the real estate allocation to Accessible. The only notice of the change in adviser appears in the next annual report in the listing of associated advisers.

Comment: Papis's actions highlight the need for supervision and review at all levels in an organization. His responsibilities may include the selection of external advisers, but the decision to change advisers appears arbitrary. Members and candidates should ensure that their firm has appropriate policies and procedures in place to detect inappropriate actions, such as the action taken by Papis.

See also Standard V(A)–Diligence and Reasonable Basis, Standard V(B)–Communication with Clients and Prospective Clients, and Standard VI(A)–Disclosure of Conflicts.

Example 8 (Supervising Research Activities):

Mary Burdette was recently hired by Fundamental Investment Management (FIM) as a junior auto industry analyst. Burdette is expected to expand the social media presence of the firm because she is active with various networks, including Facebook, LinkedIn, and Twitter. Although Burdette's supervisor, Joe Graf, has never used social media, he encourages Burdette to explore opportunities to increase FIM's online presence and ability to share content, communicate, and broadcast information to clients. In response to Graf's encouragement, Burdette is working on a proposal detailing the advantages of getting FIM onto Twitter in addition to launching a company Facebook page.

As part of her auto industry research for FIM, Burdette is completing a report on the financial impact of Sun Drive Auto Ltd.'s new solar technology for compact automobiles. This research report will be her first for FIM, and she believes Sun Drive's technology could revolutionize the auto industry. In her excitement, Burdette sends a quick tweet to FIM Twitter followers summarizing her "buy" recommendation for Sun Drive Auto stock.

> *Comment*: Graf has violated Standard IV(C) by failing to reasonably supervise Burdette with respect to the contents of her tweet. He did not establish reasonable procedures to prevent the unauthorized dissemination of company research through social media networks. Graf must make sure all employees receive regular training about FIM's policies and procedures, including the appropriate business use of personal social media networks.
>
> See Standard III(B) for additional guidance.

Example 9 (Supervising Research Activities):

Chen Wang leads the research department at YYRA Retirement Planning Specialists. Chen supervises a team of 10 analysts in a fast-paced and understaffed organization. He is responsible for coordinating the firm's approved process to review all reports before they are provided to the portfolio management team for use in rebalancing client portfolios.

One of Chen's direct reports, Huang Mei, covers the banking industry. Chen must submit the latest updates to the portfolio management team tomorrow morning. Huang has yet to submit her research report on ZYX Bank because she is uncomfortable providing a "buy" or "sell" opinion of ZYX on the basis of the completed analysis. Pressed for time and concerned that Chen will reject a "hold" recommendation, she researches various websites and blogs on the banking sector for whatever she can find on ZYX. One independent blogger provides a new interpretation of the recently reported data Huang has analyzed and concludes with a strong "sell" recommendation for ZYX. She is impressed by the originality and resourcefulness of this blogger's report.

Very late in the evening, Huang submits her report and "sell" recommendation to Chen without any reference to the independent blogger's report. Given the late time of the submission and the competence of Huang's prior work, Chen compiles this report with the recommendations from each of the other analysts and meets with the portfolio managers to discuss implementation.

> *Comment*: Chen has violated Standard IV(C) by neglecting to reasonably and adequately follow the firm's approved review process for Huang's research report. The delayed submission and the quality of prior work do not remove Chen's requirement to uphold the designated review process. A member or candidate with supervisory responsibility must make reasonable efforts to see that appropriate procedures are established, documented, communicated to covered personnel, and followed.

STANDARD V(A): INVESTMENT ANALYSIS, RECOMMENDATIONS, AND ACTIONS - DILIGENCE AND REASONABLE BASIS

Standard V(A) Diligence and Reasonable Basis

Members and Candidates must:

1 Exercise diligence, independence, and thoroughness in analyzing investments, making investment recommendations, and taking investment actions.

2 Have a reasonable and adequate basis, supported by appropriate research and investigation, for any investment analysis, recommendation, or action.

Guidance

Highlights:

- *Defining Diligence and Reasonable Basis*
- *Using Secondary or Third-Party Research*
- *Using Quantitatively Oriented Research*
- *Developing Quantitatively Oriented Techniques*
- *Selecting External Advisers and Subadvisers*
- *Group Research and Decision Making*

The application of Standard V(A) depends on the investment philosophy the member, candidate, or firm is following, the role of the member or candidate in the investment decision-making process, and the support and resources provided by the member's or candidate's employer. These factors will dictate the nature of the diligence and thoroughness of the research and the level of investigation required by Standard V(A).

The requirements for issuing conclusions based on research will vary in relation to the member's or candidate's role in the investment decision-making process, but the member or candidate must make reasonable efforts to cover all pertinent issues when arriving at a recommendation. Members and candidates enhance transparency by providing or offering to provide supporting information to clients when recommending a purchase or sale or when changing a recommendation.

Defining Diligence and Reasonable Basis

Every investment decision is based on a set of facts known and understood at the time. Clients turn to members and candidates for advice and expect these advisers to have more information and knowledge than they do. This information and knowledge is the basis from which members and candidates apply their professional judgment in taking investment actions and making recommendations.

At a basic level, clients want assurance that members and candidates are putting forth the necessary effort to support the recommendations they are making. Communicating the level and thoroughness of the information reviewed before the member or candidate makes a judgment allows clients to understand the reasonableness of the recommended investment actions.

As with determining the suitability of an investment for the client, the necessary level of research and analysis will differ with the product, security, or service being offered. In providing an investment service, members and candidates typically use a variety of resources, including company reports, third-party research, and results from quantitative models. A reasonable basis is formed through a balance of these resources appropriate for the security or decision being analyzed.

The following list provides some, but definitely not all, examples of attributes to consider while forming the basis for a recommendation:

- global, regional, and country macroeconomic conditions,
- a company's operating and financial history,
- the industry's and sector's current conditions and the stage of the business cycle,
- a mutual fund's fee structure and management history,
- the output and potential limitations of quantitative models,
- the quality of the assets included in a securitization, and
- the appropriateness of selected peer-group comparisons.

Even though an investment recommendation may be well informed, downside risk remains for any investment. Members and candidates can base their decisions only on the information available at the time decisions are made. The steps taken in developing a diligent and reasonable recommendation should minimize unexpected downside events.

Using Secondary or Third-Party Research

If members and candidates rely on secondary or third-party research, they must make reasonable and diligent efforts to determine whether such research is sound. Secondary research is defined as research conducted by someone else in the member's or candidate's firm. Third-party research is research conducted by entities outside the member's or candidate's firm, such as a brokerage firm, bank, or research firm. If a member or candidate has reason to suspect that either secondary or third-party research or information comes from a source that lacks a sound basis, the member or candidate must not rely on that information.

Members and candidates should make reasonable enquiries into the source and accuracy of all data used in completing their investment analysis and recommendations. The sources of the information and data will influence the level of the review a member or candidate must undertake. Information and data taken from internet sources, such as personal blogs, independent research aggregation websites, or social media websites, likely require a greater level of review than information from more established research organizations.

Criteria that a member or candidate can use in forming an opinion on whether research is sound include the following:

- assumptions used,
- rigor of the analysis performed,
- date/timeliness of the research, and
- evaluation of the objectivity and independence of the recommendations.

A member or candidate may rely on others in his or her firm to determine whether secondary or third-party research is sound and use the information in good faith unless the member or candidate has reason to question its validity or the processes and procedures used by those responsible for the research. For example, a portfolio manager may not have a choice of a data source because the firm's senior managers conducted due diligence to determine which vendor would provide services; the member or candidate can use the information in good faith assuming the due diligence process was deemed adequate.

A member or candidate should verify that the firm has a policy about the timely and consistent review of approved research providers to ensure that the quality of the research continues to meet the necessary standards. If such a policy is not in place at the firm, the member or candidate should encourage the development and adoption of a formal review practice.

Using Quantitatively Oriented Research

Standard V(A) applies to the rapidly expanding use of quantitatively oriented research models and processes, such as computer-generated modeling, screening, and ranking of investment securities; the creation or valuation of derivative instruments; and quantitative portfolio construction techniques. These models and processes are being used for much more than the back testing of investment strategies, especially with continually advancing technology and techniques. The continued broad development of quantitative methods and models is an important part of capital market developments.

Members and candidates need to have an understanding of the parameters used in models and quantitative research that are incorporated into their investment recommendations. Although they are not required to become experts in every technical aspect of the models, they must understand the assumptions and limitations inherent in any model and how the results were used in the decision-making process.

The reliance on and potential limitations of financial models became clear through the investment crisis that unfolded in 2007 and 2008. In some cases, the financial models used to value specific securities and related derivative products did not adequately demonstrate the level of associated risks. Members and candidates should make reasonable efforts to test the output of investment models and other pre-programed analytical tools they use. Such validation should occur before incorporating the process into their methods, models, or analyses.

Although not every model can test for every factor or outcome, members and candidates should ensure that their analyses incorporate a broad range of assumptions sufficient to capture the underlying characteristics of investments. The omission from the analysis of potentially negative outcomes or of levels of risk outside the norm may misrepresent the true economic value of an investment. The possible scenarios for analysis should include factors that are likely to have a substantial influence on the investment value and may include extremely positive and negative scenarios.

Developing Quantitatively Oriented Techniques

Individuals who create new quantitative models and services must exhibit a higher level of diligence in reviewing new products than the individuals who ultimately use the analytical output. Members and candidates involved in the development and oversight of quantitatively oriented models, methods, and algorithms must understand the technical aspects of the products they provide to clients. A thorough testing of the model and resulting analysis should be completed prior to product distribution.

Members and candidates need to consider the source and time horizon of the data used as inputs in financial models. The information from many commercially available databases may not effectively incorporate both positive and negative market cycles. In the development of a recommendation, the member or candidate may need to test the models by using volatility and performance expectations that represent scenarios outside the observable databases. In reviewing the computer models or the resulting output, members and candidates need to pay particular attention to the assumptions used in the analysis and the rigor of the analysis to ensure that the model incorporates a wide range of possible input expectations, including negative market events.

Selecting External Advisers and Subadvisers

Financial instruments and asset allocation techniques continue to develop and evolve. This progression has led to the use of specialized managers to invest in specific asset classes or diversification strategies that complement a firm's in-house expertise. Standard V(A) applies to the level of review necessary in selecting an external adviser or subadviser to manage a specifically mandated allocation. Members and candidates must review managers as diligently as they review individual funds and securities.

Members and candidates who are directly involved with the use of external advisers need to ensure that their firms have standardized criteria for reviewing these selected external advisers and managers. Such criteria would include, but would not be limited to, the following:

- reviewing the adviser's established code of ethics,
- understanding the adviser's compliance and internal control procedures,
- assessing the quality of the published return information, and
- reviewing the adviser's investment process and adherence to its stated strategy.

Codes, standards, and guides to best practice published by CFA Institute provide members and candidates with examples of acceptable practices for external advisers and advice in selecting a new adviser. The following guides are available at the CFA Institute website (www.cfainstitute.org): Asset Manager Code of Professional Conduct, Global Investment Performance Standards, and Model Request for Proposal (for equity, credit, or real estate managers).

Group Research and Decision Making

Commonly, members and candidates are part of a group or team that is collectively responsible for producing investment analysis or research. The conclusions or recommendations of the group report represent the consensus of the group and are not necessarily the views of the member or candidate, even though the name of the member or candidate is included on the report. In some instances, a member or candidate will not agree with the view of the group. If, however, the member or candidate believes that the consensus opinion has a reasonable and adequate basis and is independent and objective, the member or candidate need not decline to be identified with the report. If the member or candidate is confident in the process, the member or candidate does not need to dissociate from the report even if it does not reflect his or her opinion.

STANDARD V(A): RECOMMENDED PROCEDURES

Members and candidates should encourage their firms to consider the following policies and procedures to support the principles of Standard V(A):

- Establish a policy requiring that research reports, credit ratings, and investment recommendations have a basis that can be substantiated as reasonable and adequate. An individual employee (a supervisory analyst) or a group of employees (a review committee) should be appointed to review and approve such items prior to external circulation to determine whether the criteria established in the policy have been met.
- Develop detailed, written guidance for analysts (research, investment, or credit), supervisory analysts, and review committees that establishes the due diligence procedures for judging whether a particular recommendation has a reasonable and adequate basis.

- Develop measurable criteria for assessing the quality of research, the reasonableness and adequacy of the basis for any recommendation or rating, and the accuracy of recommendations over time. In some cases, firms may consider implementing compensation arrangements that depend on these measurable criteria and that are applied consistently to all related analysts.
- Develop detailed, written guidance that establishes minimum levels of scenario testing of all computer-based models used in developing, rating, and evaluating financial instruments. The policy should contain criteria related to the breadth of the scenarios tested, the accuracy of the output over time, and the analysis of cash flow sensitivity to inputs.
- Develop measurable criteria for assessing outside providers, including the quality of information being provided, the reasonableness and adequacy of the provider's collection practices, and the accuracy of the information over time. The established policy should outline how often the provider's products are reviewed.
- Adopt a standardized set of criteria for evaluating the adequacy of external advisers. The policy should include how often and on what basis the allocation of funds to the adviser will be reviewed.

STANDARD V(A): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Sufficient Due Diligence):

Helen Hawke manages the corporate finance department of Sarkozi Securities, Ltd. The firm is anticipating that the government will soon close a tax loophole that currently allows oil-and-gas exploration companies to pass on drilling expenses to holders of a certain class of shares. Because market demand for this tax-advantaged class of stock is currently high, Sarkozi convinces several companies to undertake new equity financings at once, before the loophole closes. Time is of the essence, but Sarkozi lacks sufficient resources to conduct adequate research on all the prospective issuing companies. Hawke decides to estimate the IPO prices on the basis of the relative size of each company and to justify the pricing later when her staff has time.

> *Comment*: Sarkozi should have taken on only the work that it could adequately handle. By categorizing the issuers by general size, Hawke has bypassed researching all the other relevant aspects that should be considered when pricing new issues and thus has not performed sufficient due diligence. Such an omission can result in investors purchasing shares at prices that have no actual basis. Hawke has violated Standard V(A).

Example 2 (Sufficient Scenario Testing):

Babu Dhaliwal works for Heinrich Brokerage in the corporate finance group. He has just persuaded Feggans Resources, Ltd., to allow his firm to do a secondary equity financing at Feggans Resources' current stock price. Because the stock has been trading at higher multiples than similar companies with equivalent production, Dhaliwal presses the Feggans Resources managers to project what would be the maximum production they could achieve in an optimal scenario. Based on these numbers, he is able to justify the price his firm will be asking for the secondary issue. During a sales pitch to the brokers, Dhaliwal then uses these numbers as the base-case production levels that Feggans Resources will achieve.

> *Comment*: When presenting information to the brokers, Dhaliwal should have given a range of production scenarios and the probability of Feggans Resources achieving each level. By giving the maximum production level as the likely level of production, he has misrepresented the chances of achieving that production level and seriously misled the brokers. Dhaliwal has violated Standard V(A).

Example 3 (Developing a Reasonable Basis):

Brendan Witt, a former junior sell-side technology analyst, decided to return to school to earn an MBA. To keep his research skills and industry knowledge sharp, Witt accepted a position with On-line and Informed, an independent internet-based research company. The position requires the publication of a recommendation and report on a different company every month. Initially, Witt is a regular contributor of new research and a participant in the associated discussion boards that generally have positive comments on the technology sector. Over time, his ability to manage his educational requirements and his work requirements begin to conflict with one another. Knowing a recommendation is due the next day for On-line, Witt creates a report based on a few news articles and what the conventional wisdom of the markets has deemed the "hot" security of the day.

> *Comment*: Witt's knowledge of and exuberance for technology stocks, a few news articles, and the conventional wisdom of the markets do not constitute, without more information, a reasonable and adequate basis for a stock recommendation that is supported by appropriate research and investigation. Therefore, Witt has violated Standard V(A).
>
> See also Standard IV(C)–Responsibilities of Supervisors because it relates to the firm's inadequate procedures.

Example 4 (Timely Client Updates):

Kristen Chandler is an investment consultant in the London office of Dalton Securities, a major global investment consultant firm. One of her UK pension funds has decided to appoint a specialist US equity manager. Dalton's global manager of research relies on local consultants to cover managers within their regions and, after conducting thorough due diligence, puts their views and ratings in Dalton's manager database. Chandler accesses Dalton's global manager research database and conducts a screen of all US equity managers on the basis of a match with the client's desired philosophy/style, performance, and tracking-error targets. She selects the five managers that meet these criteria and puts them in a briefing report that is delivered to the client 10 days later. Between the time of Chandler's database search and the delivery of the report to the client, Chandler is told that Dalton has updated the database with the information that one of the firms that Chandler has recommended for consideration lost its chief investment officer, the head of its US equity research, and the majority of its portfolio managers on the US equity product—all of whom have left to establish their own firm. Chandler does not revise her report with this updated information.

> *Comment*: Chandler has failed to satisfy the requirement of Standard V(A). Although Dalton updated the manager ratings to reflect the personnel turnover at one of the firms, Chandler did not update her report to reflect the new information.

Example 5 (Group Research Opinions):

Evelyn Mastakis is a junior analyst who has been asked by her firm to write a research report predicting the expected interest rate for residential mortgages over the next six months. Mastakis submits her report to the fixed-income investment committee of her firm for review, as required by firm procedures. Although some committee members support Mastakis's conclusion, the majority of the committee disagrees with her conclusion, and the report is significantly changed to indicate that interest rates are likely to increase more than originally predicted by Mastakis. Should Mastakis ask that her name be taken off the report when it is disseminated?

> *Comment*: The results of research are not always clear, and different people may have different opinions based on the same factual evidence. In this case, the committee may have valid reasons for issuing a report that differs from the analyst's original research. The firm can issue a report that is different from the original report of an analyst as long as there is a reasonable and adequate basis for its conclusions.
>
> Generally, analysts must write research reports that reflect their own opinion and can ask the firm not to put their name on reports that ultimately differ from that opinion. When the work is a group effort, however, not all members of the team may agree with all aspects of the report. Ultimately, members and candidates can ask to have their names removed from the report, but if they are satisfied that the process has produced results or conclusions that have a reasonable and adequate basis, members and candidates do not have to dissociate from the report even when they do not agree with its contents. If Mastakis is confident in the process, she does not need to dissociate from the report even if it does not reflect her opinion.

Example 6 (Reliance on Third-Party Research):

Gary McDermott runs a two-person investment management firm. McDermott's firm subscribes to a service from a large investment research firm that provides research reports. McDermott's firm makes investment recommendations on the basis of these reports.

> *Comment*: Members and candidates can rely on third-party research but must make reasonable and diligent efforts to determine that such research is sound. If McDermott undertakes due diligence efforts on a regular basis to ensure that the research produced by the large firm is objective and reasonably based, McDermott can rely on that research when making investment recommendations to clients.

Example 7 (Due Diligence in Submanager Selection):

Paul Ostrowski's business has grown significantly over the past couple of years, and some clients want to diversify internationally. Ostrowski decides to find a submanager to handle the expected international investments. Because this will be his first subadviser, Ostrowski uses the CFA Institute model "request for proposal" to design a questionnaire for his search. By his deadline, he receives seven completed questionnaires from a variety of domestic and international firms trying to gain his business. Ostrowski reviews all the applications in detail and decides to select the firm that charges the lowest fees because doing so will have the least impact on his firm's bottom line.

> *Comment*: The selection of an external adviser or subadviser should be based on a full and complete review of the adviser's services, performance history, and cost structure. In basing the decision on the fee structure alone, Ostrowski may be violating Standard V(A).
>
> See also Standard III(C)–Suitability because it relates to the ability of the selected adviser to meet the needs of the clients.

Example 8 (Sufficient Due Diligence):

Michael Papis is the chief investment officer of his state's retirement fund. The fund has always used outside advisers for the real estate allocation, and this information is clearly presented in all fund communications. Thomas Nagle, a recognized sell-side research analyst and Papis's business school classmate, recently left the investment bank he worked for to start his own asset management firm, Accessible Real Estate. Nagle is trying to build his assets under management and contacts Papis about gaining some of the retirement fund's allocation. In the previous few years, the performance of the retirement fund's real estate investments was in line with the fund's benchmark but was not extraordinary. Papis decides to help out his old friend and also to seek better returns by moving the real estate allocation to Accessible. The only notice of the change in adviser appears in the next annual report in the listing of associated advisers.

> *Comment*: Papis violated Standard V(A). His responsibilities may include the selection of the external advisers, but the decision to change advisers appears to have been arbitrary. If Papis was dissatisfied with the current real estate adviser, he should have conducted a proper solicitation to select the most appropriate adviser.
>
> See also Standard IV(C)–Responsibilities of Supervisors, Standard V(B)–Communication with Clients and Prospective Clients, and Standard VI(A)–Disclosure of Conflicts.

Example 9 (Sufficient Due Diligence):

Andre Shrub owns and operates Conduit, an investment advisory firm. Prior to opening Conduit, Shrub was an account manager with Elite Investment, a hedge fund managed by his good friend Adam Reed. To attract clients to a new Conduit fund, Shrub offers lower-than-normal management fees. He can do so because the fund consists of two top-performing funds managed by Reed. Given his personal friendship with Reed and the prior performance record of these two funds, Shrub believes this new fund is a winning combination for all parties. Clients quickly invest with Conduit to gain access to the Elite funds. No one is turned away because Conduit is seeking to expand its assets under management.

Comment: Shrub violated Standard V(A) by not conducting a thorough analysis of the funds managed by Reed before developing the new Conduit fund. Shrub's reliance on his personal relationship with Reed and his prior knowledge of Elite are insufficient justification for the investments. The funds may be appropriately considered, but a full review of their operating procedures, reporting practices, and transparency are some elements of the necessary due diligence.

See also Standard III(C)–Suitability.

Example 10 (Sufficient Due Diligence):

Bob Thompson has been doing research for the portfolio manager of the fixed-income department. His assignment is to do sensitivity analysis on securitized subprime mortgages. He has discussed with the manager possible scenarios to use to calculate expected returns. A key assumption in such calculations is housing price appreciation (HPA) because it drives "prepays" (prepayments of mortgages) and losses. Thompson is concerned with the significant appreciation experienced over the previous five years as a result of the increased availability of funds from subprime mortgages. Thompson insists that the analysis should include a scenario run with −10% for Year 1, −5% for Year 2, and then (to project a worst-case scenario) 0% for Years 3 through 5. The manager replies that these assumptions are too dire because there has never been a time in their available database when HPA was negative.

Thompson conducts his research to better understand the risks inherent in these securities and evaluates these securities in the worst-case scenario, a less likely but possible environment. Based on the results of the enhanced scenarios, Thompson does not recommend the purchase of the securitization. Against the general market trends, the manager follows Thompson's recommendation and does not invest. The following year, the housing market collapses. In avoiding the subprime investments, the manager's portfolio outperforms its peer group that year.

Comment: Thompson's actions in running the scenario test with inputs beyond the historical trends available in the firm's databases adhere to the principles of Standard V(A). His concerns over recent trends provide a sound basis for further analysis. Thompson understands the limitations of his model, when combined with the limited available historical information, to accurately predict the performance of the funds if market conditions change negatively.

See also Standard I(B)–Independence and Objectivity.

Example 11 (Use of Quantitatively Oriented Models):

Espacia Liakos works in sales for Hellenica Securities, a firm specializing in developing intricate derivative strategies to profit from particular views on market expectations. One of her clients is Eugenie Carapalis, who has become convinced that commodity prices will become more volatile over the coming months. Carapalis asks Liakos to quickly engineer a strategy that will benefit from this expectation. Liakos turns to Hellenica's modeling group to fulfill this request. Because of the tight deadline, the modeling group outsources parts of the work to several trusted third parties. Liakos implements the disparate components of the strategy as the firms complete them.

Within a month, Carapalis is proven correct: Volatility across a range of commodities increases sharply. But her derivatives position with Hellenica returns huge losses, and the losses increase daily. Liakos investigates and realizes that although each of the various components of the strategy had been validated, they had never been evaluated as an integrated whole. In extreme conditions, portions of the model worked at cross-purposes with other portions, causing the overall strategy to fail dramatically.

> *Comment*: Liakos violated Standard V(A). Members and candidates must understand the statistical significance of the results of the models they recommend and must be able to explain them to clients. Liakos did not take adequate care to ensure a thorough review of the whole model; its components were evaluated only individually. Because Carapalis clearly intended to implement the strategy as a whole rather than as separate parts, Liakos should have tested how the components of the strategy interacted as well as how they performed individually.

Example 12 (Successful Due Diligence/Failed Investment):

Alton Newbury is an investment adviser to high-net-worth clients. A client with an aggressive risk profile in his investment policy statement asks about investing in the Top Shelf hedge fund. This fund, based in Calgary, Alberta, Canada, has reported 20% returns for the first three years. The fund prospectus states that its strategy involves long and short positions in the energy sector and extensive leverage. Based on his analysis of the fund's track record, the principals involved in managing the fund, the fees charged, and the fund's risk profile, Newbury recommends the fund to the client and secures a position in it. The next week, the fund announces that it has suffered a loss of 60% of its value and is suspending operations and redemptions until after a regulatory review. Newbury's client calls him in a panic and asks for an explanation.

> *Comment*: Newbury's actions were consistent with Standard V(A). Analysis of an investment that results in a reasonable basis for recommendation does not guarantee that the investment has no downside risk. Newbury should discuss the analysis process with the client while reminding him or her that past performance does not lead to guaranteed future gains and that losses in an aggressive investment portfolio should be expected.

Example 13 (Quantitative Model Diligence):

Barry Cannon is the lead quantitative analyst at CityCenter Hedge Fund. He is responsible for the development, maintenance, and enhancement of the proprietary models the fund uses to manage its investors' assets. Cannon reads several high-level mathematical publications and blogs to stay informed of current developments. One blog, run by Expert CFA, presents some intriguing research that may benefit one of CityCenter's current models. Cannon is under pressure from firm executives to improve the model's predictive abilities, and he incorporates the factors discussed in the online research. The updated output recommends several new investments to the fund's portfolio managers.

> *Comment*: Cannon has violated Standard V(A) by failing to have a reasonable basis for the new recommendations made to the portfolio managers. He needed to diligently research the effect of incorporating the new factors before offering the output recommendations. Cannon may use the blog for ideas, but it is his responsibility to determine the effect on the firm's proprietary models.
>
> See Standard VII(B) regarding the violation by "Expert CFA" in the use of the CFA designation.

Example 14 (Selecting a Service Provider):

Ellen Smith is a performance analyst at Artic Global Advisors, a firm that manages global equity mandates for institutional clients. She was asked by her supervisor to review five new performance attribution systems and recommend one that would more appropriately explain the firm's investment strategy to clients. On the list was a system she recalled learning about when visiting an exhibitor booth at a recent conference. The system is highly quantitative and something of a "black box" in how it calculates the attribution values. Smith recommended this option without researching the others because the sheer complexity of the process was sure to impress the clients.

> *Comment*: Smith's actions do not demonstrate a sufficient level of diligence in reviewing this product to make a recommendation for selecting the service. Besides not reviewing or considering the other four potential systems, she did not determine whether the "black box" attribution process aligns with the investment practices of the firm, including its investments in different countries and currencies. Smith must review and understand the process of any software or system before recommending its use as the firm's attribution system.

Example 15 (Subadviser Selection):

Craig Jackson is working for Adams Partners, Inc., and has been assigned to select a hedge fund subadviser to improve the diversification of the firm's large fund-of-funds product. The allocation must be in place before the start of the next quarter. Jackson uses a consultant database to find a list of suitable firms that claim compliance with the GIPS standards. He calls more than 20 firms on the list to confirm their potential interest and to determine their most recent quarterly and annual total return values. Because of the short turnaround, Jackson recommends the firm with the greatest total return values for selection.

> *Comment*: By considering only performance and GIPS compliance, Jackson has not conducted sufficient review of potential firms to satisfy the requirements of Standard V(A). A thorough investigation of the firms and their operations should be conducted to ensure that their addition would increase the diversity of clients' portfolios and that they are suitable for the fund-of-funds product.

Example 16 (Manager Selection):

Timothy Green works for Peach Asset Management, where he creates proprietary models that analyze data from the firm request for proposal questionnaires to identify managers for possible inclusion in the firm's fund-of-funds investment platform. Various criteria must be met to be accepted to the platform. Because of the number of respondents to the questionnaires, Green uses only the data submitted to make a recommendation for adding a new manager.

> *Comment*: By failing to conduct any additional outside review of the information to verify what was submitted through the request for proposal, Green has likely not satisfied the requirements of Standard V(A). The amount of information requested from outside managers varies among firms. Although the requested information may be comprehensive, Green should ensure sufficient effort is undertaken to verify the submitted information before recommending a firm for inclusion. This requires that he goes beyond the

information provided by the manager on the request for proposal questionnaire and may include interviews with interested managers, reviews of regulatory filings, and discussions with the managers' custodian or auditor.

Example 17 (Technical Model Requirements):

Jérôme Dupont works for the credit research group of XYZ Asset Management, where he is in charge of developing and updating credit risk models. In order to perform accurately, his models need to be regularly updated with the latest market data.

Dupont does not interact with or manage money for any of the firm's clients. He is in contact with the firm's US corporate bond fund manager, John Smith, who has only very superficial knowledge of the model and who from time to time asks very basic questions regarding the output recommendations. Smith does not consult Dupont with respect to finalizing his clients' investment strategies.

Dupont's recently assigned objective is to develop a new emerging market corporate credit risk model. The firm is planning to expand into emerging credit, and the development of such a model is a critical step in this process. Because Smith seems to follow the model's recommendations without much concern for its quality as he develops his clients' investment strategies, Dupont decides to focus his time on the development of the new emerging market model and neglects to update the US model.

After several months without regular updates, Dupont's diagnostic statistics start to show alarming signs with respect to the quality of the US credit model. Instead of conducting the long and complicated data update, Dupont introduces new codes into his model with some limited new data as a quick "fix." He thinks this change will address the issue without needing to complete the full data update, so he continues working on the new emerging market model.

Several months following the quick "fix," another set of diagnostic statistics reveals nonsensical results and Dupont realizes that his earlier change contained an error. He quickly corrects the error and alerts Smith. Smith realizes that some of the prior trades he performed were due to erroneous model results. Smith rebalances the portfolio to remove the securities purchased on the basis of the questionable results without reporting the issue to anyone else.

Comment: Smith violated standard V(A) because exercising "diligence, independence, and thoroughness in analyzing investments, making investment recommendations, and taking investment actions" means that members and candidates must understand the technical aspects of the products they provide to clients. Smith does not understand the model he is relying on to manage money. Members and candidates should also make reasonable enquiries into the source and accuracy of all data used in completing their investment analysis and recommendations.

Dupont violated V(A) even if he does not trade securities or make investment decisions. Dupont's models give investment recommendations, and Dupont is accountable for the quality of those recommendations. Members and candidates should make reasonable efforts to test the output of preprogramed analytical tools they use. Such validation should occur before incorporating the tools into their decision-making process.

See also Standard V(B)–Communication with Clients and Prospective Clients.

STANDARD V(B): INVESTMENT ANALYSIS, RECOMMENDATIONS, AND ACTIONS - COMMUNICATION WITH CLIENTS AND PROSPECTIVE CLIENTS

Members and Candidates must:

1 Disclose to clients and prospective clients the basic format and general principles of the investment processes they use to analyze investments, select securities, and construct portfolios and must promptly disclose any changes that might materially affect those processes.

2 Disclose to clients and prospective clients significant limitations and risks associated with the investment process.

3 Use reasonable judgment in identifying which factors are important to their investment analyses, recommendations, or actions and include those factors in communications with clients and prospective clients.

4 Distinguish between fact and opinion in the presentation of investment analyses and recommendations.

Guidance

Highlights:

- *Informing Clients of the Investment Process*
- *Different Forms of Communication*
- *Identifying Risk and Limitations*
- *Report Presentation*
- *Distinction between Facts and Opinions in Reports*

Standard V(B) addresses member and candidate conduct with respect to communicating with clients. Developing and maintaining clear, frequent, and thorough communication practices is critical to providing high-quality financial services to clients. When clients understand the information communicated to them, they also can understand exactly how members and candidates are acting on their behalf, which gives clients the opportunity to make well-informed decisions about their investments. Such understanding can be accomplished only through clear communication.

Standard V(B) states that members and candidates should communicate in a recommendation the factors that were instrumental in making the investment recommendation. A critical part of this requirement is to distinguish clearly between opinions and facts. In preparing a research report, the member or candidate must present the basic characteristics of the security(ies) being analyzed, which will allow the reader to evaluate the report and incorporate information the reader deems relevant to his or her investment decision-making process.

Similarly, in preparing a recommendation about, for example, an asset allocation strategy, alternative investment vehicle, or structured investment product, the member or candidate should include factors that are relevant to the asset classes that are being discussed. Follow-up communication of significant changes in the risk characteristics of a security or asset strategy is required. Providing regular updates to any changes in the risk characteristics is recommended.

Informing Clients of the Investment Process

Members and candidates must adequately describe to clients and prospective clients the manner in which they conduct the investment decision-making process. Such disclosure should address factors that have positive and negative influences on the recommendations, including significant risks and limitations of the investment process used. The member or candidate must keep clients and other interested parties informed on an ongoing basis about changes to the investment process, especially newly identified significant risks and limitations. Only by thoroughly understanding the nature of the investment product or service can a client determine whether changes to that product or service could materially affect his or her investment objectives.

Understanding the basic characteristics of an investment is of great importance in judging the suitability of that investment on a standalone basis, but it is especially important in determining the impact each investment will have on the characteristics of a portfolio. Although the risk and return characteristics of a common stock might seem to be essentially the same for any investor when the stock is viewed in isolation, the effects of those characteristics greatly depend on the other investments held. For instance, if the particular stock will represent 90% of an individual's investments, the stock's importance in the portfolio is vastly different from what it would be to an investor with a highly diversified portfolio for whom the stock will represent only 2% of the holdings.

A firm's investment policy may include the use of outside advisers to manage various portions of clients' assets under management. Members and candidates should inform the clients about the specialization or diversification expertise provided by the external adviser(s). This information allows clients to understand the full mix of products and strategies being applied that may affect their investment objectives.

Different Forms of Communication

For purposes of Standard V(B), communication is not confined to a written report of the type traditionally generated by an analyst researching a security, company, or industry. A presentation of information can be made via any means of communication, including in-person recommendation or description, telephone conversation, media broadcast, or transmission by computer (e.g., on the internet).

Computer and mobile device communications have rapidly evolved over the past few years. Members and candidates using any social media service to communicate business information must be diligent in their efforts to avoid unintended problems because these services may not be available to all clients. When providing information to clients through new technologies, members and candidates should take reasonable steps to ensure that such delivery would treat all clients fairly and, if necessary, be considered publicly disseminated.

The nature of client communications is highly diverse—from one word ("buy" or "sell") to in-depth reports of more than 100 pages. A communication may contain a general recommendation about the market, asset allocations, or classes of investments (e.g., stocks, bonds, real estate) or may relate to a specific security. If recommendations are contained in capsule form (such as a recommended stock list), members and candidates should notify clients that additional information and analyses are available from the producer of the report.

Identifying Risks and Limitations

Members and candidates must outline to clients and prospective clients significant risks and limitations of the analysis contained in their investment products or recommendations. The type and nature of significant risks will depend on the investment process that members and candidates are following and on the personal circumstances of the client. In general, the use of leverage constitutes a significant risk and should be disclosed.

Members and candidates must adequately disclose the general market-related risks and the risks associated with the use of complex financial instruments that are deemed significant. Other types of risks that members and candidates may consider disclosing include, but are not limited to, counterparty risk, country risk, sector or industry risk, security-specific risk, and credit risk.

Investment securities and vehicles may have limiting factors that influence a client's or potential client's investment decision. Members and candidates must report to clients and prospective clients the existence of limitations significant to the decision-making process. Examples of such factors and attributes include, but are not limited to, investment liquidity and capacity. Liquidity is the ability to liquidate an investment on a timely basis at a reasonable cost. Capacity is the investment amount beyond which returns will be negatively affected by new investments.

The appropriateness of risk disclosure should be assessed on the basis of what was known at the time the investment action was taken (often called an *ex ante* basis). Members and candidates must disclose significant risks known to them at the time of the disclosure. Members and candidates cannot be expected to disclose risks they are unaware of at the time recommendations or investment actions are made. In assessing compliance with Standard V(B), it is important to establish knowledge of a purported significant risk or limitation. A one-time investment loss that occurs after the disclosure does not constitute a pertinent factor in assessing whether significant risks and limitations were properly disclosed. Having no knowledge of a risk or limitation that subsequently triggers a loss may reveal a deficiency in the diligence and reasonable basis of the research of the member or candidate but may not reveal a breach of Standard V(B).

Report Presentation

Once the analytical process has been completed, the member or candidate who prepares the report must include those elements that are important to the analysis and conclusions of the report so that the reader can follow and challenge the report's reasoning. A report writer who has done adequate investigation may emphasize certain areas, touch briefly on others, and omit certain aspects deemed unimportant. For instance, a report may dwell on a quarterly earnings release or new-product introduction and omit other matters as long as the analyst clearly stipulates the limits to the scope of the report.

Investment advice based on quantitative research and analysis must be supported by readily available reference material and should be applied in a manner consistent with previously applied methodology. If changes in methodology are made, they should be highlighted.

Distinction between Facts and Opinions in Reports

Standard V(B) requires that opinion be separated from fact. Violations often occur when reports fail to separate the past from the future by not indicating that earnings estimates, changes in the outlook for dividends, or future market price information are *opinions* subject to future circumstances.

In the case of complex quantitative analyses, members and candidates must clearly separate fact from statistical conjecture and should identify the known limitations of an analysis. Members and candidates may violate Standard V(B) by failing to identify the limits of statistically developed projections because such omission leaves readers unaware of the limits of the published projections.

Members and candidates should explicitly discuss with clients and prospective clients the assumptions used in the investment models and processes to generate the analysis. Caution should be used in promoting the perceived accuracy of any model or process to clients because the ultimate output is merely an estimate of future results and not a certainty.

STANDARD V(B): RECOMMENDED PROCEDURES

Because the selection of relevant factors is an analytical skill, determination of whether a member or candidate has used reasonable judgment in excluding and including information in research reports depends heavily on case-by-case review rather than a specific checklist.

Members and candidates should encourage their firms to have a rigorous methodology for reviewing research that is created for publication and dissemination to clients.

To assist in the after-the-fact review of a report, the member or candidate must maintain records indicating the nature of the research and should, if asked, be able to supply additional information to the client (or any user of the report) covering factors not included in the report.

STANDARD V(B): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Sufficient Disclosure of Investment System):

Sarah Williamson, director of marketing for Country Technicians, Inc., is convinced that she has found the perfect formula for increasing Country Technicians' income and diversifying its product base. Williamson plans to build on Country Technicians' reputation as a leading money manager by marketing an exclusive and expensive investment advice letter to high-net-worth individuals. One hitch in the plan is the complexity of Country Technicians' investment system—a combination of technical trading rules (based on historical price and volume fluctuations) and portfolio construction rules designed to minimize risk. To simplify the newsletter, she decides to include only each week's top five "buy" and "sell" recommendations and to leave out details of the valuation models and the portfolio structuring scheme.

> *Comment*: Williamson's plans for the newsletter violate Standard V(B). Williamson need not describe the investment system in detail in order to implement the advice effectively, but she must inform clients of Country

Technicians' basic process and logic. Without understanding the basis for a recommendation, clients cannot possibly understand its limitations or its inherent risks.

Example 2 (Providing Opinions as Facts):

Richard Dox is a mining analyst for East Bank Securities. He has just finished his report on Boisy Bay Minerals. Included in his report is his own assessment of the geological extent of mineral reserves likely to be found on the company's land. Dox completed this calculation on the basis of the core samples from the company's latest drilling. According to Dox's calculations, the company has more than 500,000 ounces of gold on the property. Dox concludes his research report as follows: "Based on the fact that the company has 500,000 ounces of gold to be mined, I recommend a strong BUY."

> *Comment*: If Dox issues the report as written, he will violate Standard V(B). His calculation of the total gold reserves for the property based on the company's recent sample drilling is a quantitative opinion, not a fact. Opinion must be distinguished from fact in research reports.

Example 3 (Proper Description of a Security):

Olivia Thomas, an analyst at Government Brokers, Inc., which is a brokerage firm specializing in government bond trading, has produced a report that describes an investment strategy designed to benefit from an expected decline in US interest rates. The firm's derivative products group has designed a structured product that will allow the firm's clients to benefit from this strategy. Thomas's report describing the strategy indicates that high returns are possible if various scenarios for declining interest rates are assumed. Citing the proprietary nature of the structured product underlying the strategy, the report does not describe in detail how the firm is able to offer such returns or the related risks in the scenarios, nor does the report address the likely returns of the strategy if, contrary to expectations, interest rates rise.

> *Comment*: Thomas has violated Standard V(B) because her report fails to describe properly the basic characteristics of the actual and implied risks of the investment strategy, including how the structure was created and the degree to which leverage was embedded in the structure. The report should include a balanced discussion of how the strategy would perform in the case of rising as well as falling interest rates, preferably illustrating how the strategies might be expected to perform in the event of a reasonable variety of interest rate and credit risk–spread scenarios. If liquidity issues are relevant with regard to the valuation of either the derivatives or the underlying securities, provisions the firm has made to address those risks should also be disclosed.

Example 4 (Notification of Fund Mandate Change):

May & Associates is an aggressive growth manager that has represented itself since its inception as a specialist at investing in small-cap US stocks. One of May's selection criteria is a maximum capitalization of US$250 million for any given company. After a string of successful years of superior performance relative to its peers, May has expanded its client base significantly, to the point at which assets under management now exceed US$3 billion. For liquidity purposes, May's chief investment officer (CIO) decides to lift the maximum permissible market-cap ceiling to US$500 million and change the firm's sales and marketing literature accordingly to inform prospective clients and third-party consultants.

> *Comment*: Although May's CIO is correct about informing potentially interested parties as to the change in investment process, he must also notify May's existing clients. Among the latter group might be a number of clients who not only retained May as a small-cap manager but also retained mid-cap and large-cap specialists in a multiple-manager approach. Such clients could regard May's change of criteria as a style change that distorts their overall asset allocations.

Example 5 (Notification of Fund Mandate Change):

Rather than lifting the ceiling for its universe from US$250 million to US$500 million, May & Associates extends its small-cap universe to include a number of non-US companies.

> *Comment*: Standard V(B) requires that May's CIO advise May's clients of this change because the firm may have been retained by some clients specifically for its prowess at investing in US small-cap stocks. Other changes that require client notification are introducing derivatives to emulate a certain market sector or relaxing various other constraints, such as portfolio beta. In all such cases, members and candidates must disclose changes to all interested parties.

Example 6 (Notification of Changes to the Investment Process):

RJZ Capital Management is an active value-style equity manager that selects stocks by using a combination of four multifactor models. The firm has found favorable results when back testing the most recent 10 years of available market data in a new dividend discount model (DDM) designed by the firm. This model is based on projected inflation rates, earnings growth rates, and interest rates. The president of RJZ decides to replace its simple model that uses price to trailing 12-month earnings with the new DDM.

> *Comment*: Because the introduction of a new and different valuation model represents a material change in the investment process, RJZ's president must communicate the change to the firm's clients. RJZ is moving away from a model based on hard data toward a new model that is at least partly dependent on the firm's forecasting skills. Clients would likely view such a model as a significant change rather than a mere refinement of RJZ's process.

Example 7 (Notification of Changes to the Investment Process):

RJZ Capital Management loses the chief architect of its multifactor valuation system. Without informing its clients, the president of RJZ decides to redirect the firm's talents and resources toward developing a product for passive equity management—a product that will emulate the performance of a major market index.

> *Comment*: By failing to disclose to clients a substantial change to its investment process, the president of RJZ has violated Standard V(B).

Example 8 (Notification of Changes to the Investment Process):

At Fundamental Asset Management, Inc., the responsibility for selecting stocks for addition to the firm's "approved" list has just shifted from individual security analysts to a committee consisting of the research director and three senior portfolio managers. Eleanor Morales, a portfolio manager with Fundamental Asset Management, thinks this change is not important enough to communicate to her clients.

> *Comment*: Morales must disclose the process change to all her clients. Some of Fundamental's clients might be concerned about the morale and motivation among the firm's best research analysts after such a change. Moreover, clients might challenge the stock-picking track record of the portfolio managers and might even want to monitor the situation closely.

Example 9 (Sufficient Disclosure of Investment System):

Amanda Chinn is the investment director for Diversified Asset Management, which manages the endowment of a charitable organization. Because of recent staff departures, Diversified has decided to limit its direct investment focus to large-cap securities and supplement the needs for small-cap and mid-cap management by hiring outside fund managers. In describing the planned strategy change to the charity, Chinn's update letter states, "As investment director, I will directly oversee the investment team managing the endowment's large-capitalization allocation. I will coordinate the selection and ongoing review of external managers responsible for allocations to other classes." The letter also describes the reasons for the change and the characteristics external managers must have to be considered.

> *Comment*: Standard V(B) requires the disclosure of the investment process used to construct the portfolio of the fund. Changing the investment process from managing all classes of investments within the firm to the use of external managers is one example of information that needs to be communicated to clients. Chinn and her firm have embraced the principles of Standard V(B) by providing their client with relevant information. The charity can now make a reasonable decision about whether Diversified Asset Management remains the appropriate manager for its fund.

Example 10 (Notification of Changes to the Investment Process):

Michael Papis is the chief investment officer of his state's retirement fund. The fund has always used outside advisers for the real estate allocation, and this information is clearly presented in all fund communications. Thomas Nagle, a recognized sell-side research analyst and Papis's business school classmate, recently left the investment bank he worked for to start his own asset management firm, Accessible Real Estate. Nagle is trying to build his assets under management and contacts Papis about gaining some of the retirement fund's allocation. In the previous few years, the performance of the retirement fund's real estate investments was in line with the fund's benchmark but was not extraordinary. Papis decides to help out his old friend and also to seek better returns by moving the real estate allocation to Accessible. The only notice of the change in adviser appears in the next annual report in the listing of associated advisers.

Comment: Papis has violated Standard V(B). He attempted to hide the nature of his decision to change external managers by making only a limited disclosure. The plan recipients and the fund's trustees need to be aware when changes are made to ensure that operational procedures are being followed.

See also Standard IV(C)–Responsibilities of Supervisors, Standard V(A)–Diligence and Reasonable Basis, and Standard VI(A)–Disclosure of Conflicts.

Example 11 (Notification of Errors):

Jérôme Dupont works for the credit research group of XYZ Asset Management, where he is in charge of developing and updating credit risk models. In order to perform accurately, his models need to be regularly updated with the latest market data.

Dupont does not interact with or manage money for any of the firm's clients. He is in contact with the firm's US corporate bond fund manager, John Smith, who has only very superficial knowledge of the model and who from time to time asks very basic questions regarding the output recommendations. Smith does not consult Dupont with respect to finalizing his clients' investment strategies.

Dupont's recently assigned objective is to develop a new emerging market corporate credit risk model. The firm is planning to expand into emerging credit, and the development of such a model is a critical step in this process. Because Smith seems to follow the model's recommendations without much concern for its quality as he develops his clients' investment strategies, Dupont decides to focus his time on the development of the new emerging market model and neglects to update the US model.

After several months without regular updates, Dupont's diagnostic statistics start to show alarming signs with respect to the quality of the US credit model. Instead of conducting the long and complicated data update, Dupont introduces new codes into his model with some limited new data as a quick "fix." He thinks this change will address the issue without needing to complete the full data update, so he continues working on the new emerging market model.

Several months following the quick "fix," another set of diagnostic statistics reveals nonsensical results and Dupont realizes that his earlier change contained an error. He quickly corrects the error and alerts Smith. Smith realizes that some of the prior trades he performed were due to erroneous model results. Smith rebalances the portfolio to remove the securities purchased on the basis of the questionable results without reporting the issue to anyone else.

Comment: Smith violated V(B) by not disclosing a material error in the investment process. Clients should have been informed about the error and the corrective actions the firm was undertaking on their behalf.

See also Standard V(A)–Diligence and Reasonable Basis.

Example 12 (Notification of Risks and Limitations):

Quantitative analyst Yuri Yakovlev has developed an investment strategy that selects small-cap stocks on the basis of quantitative signals. Yakovlev's strategy typically identifies only a small number of stocks (10–20) that tend to be illiquid, but according to his backtests, the strategy generates significant risk-adjusted returns. The partners at Yakovlev's firm, QSC Capital, are impressed by these results. After a thorough examination of the strategy's risks, stress testing, historical back testing, and scenario analysis, QSC decides to seed the strategy with US$10 million of internal capital in order for Yakovlev to create a track record for the strategy.

After two years, the strategy has generated performance returns greater than the appropriate benchmark and the Sharpe ratio of the fund is close to 1.0. On the basis of these results, QSC decides to actively market the fund to large institutional investors. While creating the offering materials, Yakovlev informs the marketing team that the capacity of the strategy is limited. The extent of the limitation is difficult to ascertain with precision; it depends on market liquidity and other factors in his model that can evolve over time. Yakovlev indicates that given the current market conditions, investments in the fund beyond US$100 million of capital could become more difficult and negatively affect expected fund returns.

Alan Wellard, the manager of the marketing team, is a partner with 30 years of marketing experience and explains to Yakovlev that these are complex technical issues that will muddy the marketing message. According to Wellard, the offering material should focus solely on the great track record of the fund. Yakovlev does not object because the fund has only US$12 million of capital, very far from the US$100 million threshold.

> *Comment*: Yakovlev and Wellard have not appropriately disclosed a significant limitation associated with the investment product. Yakovlev believes this limitation, once reached, will materially affect the returns of the fund. Although the fund is currently far from the US$100 million mark, current and prospective investors must be made aware of this capacity issue. If significant limitations are complicated to grasp and clients do not have the technical background required to understand them, Yakovlev and Wellard should either educate the clients or ascertain whether the fund is suitable for each client.

Example 13 (Notification of Risks and Limitations):

Brickell Advisers offers investment advisory services mainly to South American clients. Julietta Ramon, a risk analyst at Brickell, describes to clients how the firm uses value at risk (VaR) analysis to track the risk of its strategies. Ramon assures clients that calculating a VaR at a 99% confidence level, using a 20-day holding period, and applying a methodology based on an *ex ante* Monte Carlo simulation is extremely effective. The firm has never had losses greater than those predicted by this VaR analysis.

> *Comment*: Ramon has not sufficiently communicated the risks associated with the investment process to satisfy the requirements of Standard V(B). The losses predicted by a VaR analysis depend greatly on the inputs used in the model. The size and probability of losses can differ significantly from what an individual model predicts. Ramon must disclose how the inputs were selected and the potential limitations and risks associated with the investment strategy.

Example 14 (Notification of Risks and Limitations):

Lily Smith attended an industry conference and noticed that John Baker, an investment manager with Baker Associates, attracted a great deal of attention from the conference participants. On the basis of her knowledge of Baker's reputation and the interest he received at the conference, Smith recommends adding Baker Associates to the approved manager platform. Her recommendation to the approval committee included the statement "John Baker is well respected in the industry, and his insights are consistently sought after by investors. Our clients are sure to benefit from investing with Baker Associates."

Comment: Smith is not appropriately separating facts from opinions in her recommendation to include the manager within the platform. Her actions conflict with the requirements of Standard V(B). Smith is relying on her opinions about Baker's reputation and the fact that many attendees were talking with him at the conference. Smith should also review the requirements of Standard V(A) regarding reasonable basis to determine the level of review necessary to recommend Baker Associates.

STANDARD V(C): INVESTMENT ANALYSIS, RECOMMENDATIONS, AND ACTIONS - RECORD RETENTION

Members and Candidates must develop and maintain appropriate records to support their investment analyses, recommendations, actions, and other investment-related communications with clients and prospective clients.

Guidance

Highlights:

- *New Media Records*
- *Records Are Property of the Firm*
- *Local Requirements*

Members and candidates must retain records that substantiate the scope of their research and reasons for their actions or conclusions. The retention requirement applies to decisions to buy or sell a security as well as reviews undertaken that do not lead to a change in position. Which records are required to support recommendations or investment actions depends on the role of the member or candidate in the investment decision-making process. Records may be maintained either in hard copy or electronic form.

Some examples of supporting documentation that assists the member or candidate in meeting the requirements for retention are as follows:

- personal notes from meetings with the covered company,
- press releases or presentations issued by the covered company,
- computer-based model outputs and analyses,
- computer-based model input parameters,
- risk analyses of securities' impacts on a portfolio,
- selection criteria for external advisers,
- notes from clients from meetings to review investment policy statements, and
- outside research reports.

New Media Records

The increased use of new and evolving technological formats (e.g., social media) for gathering and sharing information creates new challenges in maintaining the appropriate records and files. The nature or format of the information does not remove a member's or candidate's responsibility to maintain a record of information used in his or her analysis or communicated to clients.

Members and candidates should understand that although employers and local regulators are developing digital media retention policies, these policies may lag behind the advent of new communication channels. Such lag places greater responsibility on the individual for ensuring that all relevant information is retained. Examples of non-print media formats that should be retained include, but are not limited to,

- e-mails,
- text messages,
- blog posts, and
- Twitter posts.

Records Are Property of the Firm

As a general matter, records created as part of a member's or candidate's professional activity on behalf of his or her employer are the property of the firm. When a member or candidate leaves a firm to seek other employment, the member or candidate cannot take the property of the firm, including original forms or copies of supporting records of the member's or candidate's work, to the new employer without the express consent of the previous employer. The member or candidate cannot use historical recommendations or research reports created at the previous firm because the supporting documentation is unavailable. For future use, the member or candidate must re-create the supporting records at the new firm with information gathered through public sources or directly from the covered company and not from memory or sources obtained at the previous employer.

Local Requirements

Local regulators often impose requirements on members, candidates, and their firms related to record retention that must be followed. Firms may also implement policies detailing the applicable time frame for retaining research and client communication records. Fulfilling such regulatory and firm requirements satisfies the requirements of Standard V(C). In the absence of regulatory guidance or firm policies, CFA Institute recommends maintaining records for at least seven years.

STANDARD V(C): RECOMMENDED PROCEDURES

The responsibility to maintain records that support investment action generally falls with the firm rather than individuals. Members and candidates must, however, archive research notes and other documents, either electronically or in hard copy, that support their current investment-related communications. Doing so will assist their firms in complying with requirements for preservation of internal or external records.

STANDARD V(C): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Record Retention and IPS Objectives and Recommendations):

One of Nikolas Lindstrom's clients is upset by the negative investment returns of his equity portfolio. The investment policy statement for the client requires that the portfolio manager follow a benchmark-oriented approach. The benchmark for the client includes a 35% investment allocation in the technology sector. The client acknowledges that this allocation was appropriate, but over the past three years, technology stocks have suffered severe losses. The client complains to the investment manager for allocating so much money to this sector.

> *Comment*: For Lindstrom, having appropriate records is important to show that over the past three years, the portion of technology stocks in the benchmark index was 35%, as called for in the IPS. Lindstrom should also have the client's IPS stating that the benchmark was appropriate for the client's investment objectives. He should also have records indicating that the investment has been explained appropriately to the client and that the IPS was updated on a regular basis. Taking these actions, Lindstrom would be in compliance with Standard V(C).

Example 2 (Record Retention and Research Process):

Malcolm Young is a research analyst who writes numerous reports rating companies in the luxury retail industry. His reports are based on a variety of sources, including interviews with company managers, manufacturers, and economists; on-site company visits; customer surveys; and secondary research from analysts covering related industries.

> *Comment*: Young must carefully document and keep copies of all the information that goes into his reports, including the secondary or third-party research of other analysts. Failure to maintain such files would violate Standard V(C).

Example 3 (Records as Firm, Not Employee, Property):

Martin Blank develops an analytical model while he is employed by Green Partners Investment Management, LLP (GPIM). While at the firm, he systematically documents the assumptions that make up the model as well as his reasoning behind the assumptions. As a result of the success of his model, Blank is hired to be the head of the research department of one of GPIM's competitors. Blank takes copies of the records supporting his model to his new firm.

Comment: The records created by Blank supporting the research model he developed at GPIM are the records of GPIM. Taking the documents with him to his new employer without GPIM's permission violates Standard V(C). To use the model in the future, Blank must re-create the records supporting his model at the new firm.

STANDARD VI(A): CONFLICTS OF INTEREST - DISCLOSURE OF CONFLICTS

Standard VI(A) Disclosure of Conflicts

Members and Candidates must make full and fair disclosure of all matters that could reasonably be expected to impair their independence and objectivity or interfere with respective duties to their clients, prospective clients, and employer. Members and Candidates must ensure that such disclosures are prominent, are delivered in plain language, and communicate the relevant information effectively.

Guidance

Highlights:

- *Disclosure of Conflicts to Employers*
- *Disclosure to Clients*
- *Cross-Departmental Conflicts*
- *Conflicts with Stock Ownership*
- *Conflicts as a Director*

Best practice is to avoid actual conflicts or the appearance of conflicts of interest when possible. Conflicts of interest often arise in the investment profession. Conflicts can occur between the interests of clients, the interests of employers, and the member's or candidate's own personal interests. Common sources for conflict are compensation structures, especially incentive and bonus structures that provide immediate returns for members and candidates with little or no consideration of long-term value creation.

Identifying and managing these conflicts is a critical part of working in the investment industry and can take many forms. When conflicts cannot be reasonably avoided, clear and complete disclosure of their existence is necessary.

Standard VI(A) protects investors and employers by requiring members and candidates to fully disclose to clients, potential clients, and employers all actual and potential conflicts of interest. Once a member or candidate has made full disclosure, the member's or candidate's employer, clients, and prospective clients will have the information needed to evaluate the objectivity of the investment advice or action taken on their behalf.

To be effective, disclosures must be prominent and must be made in plain language and in a manner designed to effectively communicate the information. Members and candidates have the responsibility of determining how often, in what manner, and in what particular circumstances the disclosure of conflicts must be made. Best practices dictate updating disclosures when the nature of a conflict of interest changes materially—for example, if the nature of a conflict of interest worsens through the introduction of bonuses based on each quarter's profits as to opposed annual profits. In making and updating disclosures of conflicts of interest, members and candidates should err on the side of caution to ensure that conflicts are effectively communicated.

Disclosure of Conflicts to Employers

Disclosure of conflicts to employers may be appropriate in many instances. When reporting conflicts of interest to employers, members and candidates must give their employers enough information to assess the impact of the conflict. By complying with employer guidelines, members and candidates allow their employers to avoid potentially embarrassing and costly ethical or regulatory violations.

Reportable situations include conflicts that would interfere with rendering unbiased investment advice and conflicts that would cause a member or candidate to act not in the employer's best interest. The same circumstances that generate conflicts to be reported to clients and prospective clients also would dictate reporting to employers. Ownership of stocks analyzed or recommended, participation on outside boards, and financial or other pressures that could influence a decision are to be promptly reported to the employer so that their impact can be assessed and a decision on how to resolve the conflict can be made.

The mere appearance of a conflict of interest may create problems for members, candidates, and their employers. Therefore, many of the conflicts previously mentioned could be explicitly prohibited by an employer. For example, many employers restrict personal trading, outside board membership, and related activities to prevent situations that might not normally be considered problematic from a conflict-of-interest point of view but that could give the appearance of a conflict of interest. Members and candidates must comply with these restrictions. Members and candidates must take reasonable steps to avoid conflicts and, if they occur inadvertently, must report them promptly so that the employer and the member or candidate can resolve them as quickly and effectively as possible.

Standard VI(A) also deals with a member's or candidate's conflicts of interest that might be detrimental to the employer's business. Any potential conflict situation that could prevent clear judgment about or full commitment to the execution of a member's or candidate's duties to the employer should be reported to the member's or candidate's employer and promptly resolved.

Disclosure to Clients

Members and candidates must maintain their objectivity when rendering investment advice or taking investment action. Investment advice or actions may be perceived to be tainted in numerous situations. Can a member or candidate remain objective if, on behalf of the firm, the member or candidate obtains or assists in obtaining fees for services? Can a member or candidate give objective advice if he or she owns stock in the company that is the subject of an investment recommendation or if the member or candidate has a close personal relationship with the company managers? Requiring members and candidates to disclose all matters that reasonably could be expected to impair the member's or candidate's objectivity allows clients and prospective clients to judge motives and possible biases for themselves.

Often in the investment industry, a conflict, or the perception of a conflict, cannot be avoided. The most obvious conflicts of interest, which should always be disclosed, are relationships between an issuer and the member, the candidate, or his or her firm (such as a directorship or consultancy by a member; investment banking, underwriting, and financial relationships; broker/dealer market-making activities; and material beneficial ownership of stock). For the purposes of Standard VI(A), members and candidates beneficially own securities or other investments if they have a direct or indirect pecuniary interest in the securities, have the power to vote or direct the voting of the shares of the securities or investments, or have the power to dispose or direct the disposition of the security or investment.

A member or candidate must take reasonable steps to determine whether a conflict of interest exists and disclose to clients any known conflicts of the member's or candidate's firm. Disclosure of broker/dealer market-making activities alerts clients that a purchase or sale might be made from or to the firm's principal account and that the firm has a special interest in the price of the stock.

Additionally, disclosures should be made to clients regarding fee arrangements, subadvisory agreements, or other situations involving nonstandard fee structures. Equally important is the disclosure of arrangements in which the firm benefits directly from investment recommendations. An obvious conflict of interest is the rebate of a portion of the service fee some classes of mutual funds charge to investors. Members and candidates should ensure that their firms disclose such relationships so clients can fully understand the costs of their investments and the benefits received by their investment manager's employer.

Cross-Departmental Conflicts

Other circumstances can give rise to actual or potential conflicts of interest. For instance, a sell-side analyst working for a broker/dealer may be encouraged, not only by members of her or his own firm but by corporate issuers themselves, to write research reports about particular companies. The buy-side analyst is likely to be faced with similar conflicts as banks exercise their underwriting and security-dealing powers. The marketing division may ask an analyst to recommend the stock of a certain company in order to obtain business from that company.

The potential for conflicts of interest also exists with broker-sponsored limited partnerships formed to invest venture capital. Increasingly, members and candidates are expected not only to follow issues from these partnerships once they are offered to the public but also to promote the issues in the secondary market after public offerings. Members, candidates, and their firms should attempt to resolve situations presenting potential conflicts of interest or disclose them in accordance with the principles set forth in Standard VI(A).

Conflicts with Stock Ownership

The most prevalent conflict requiring disclosure under Standard VI(A) is a member's or candidate's ownership of stock in companies that he or she recommends to clients or that clients hold. Clearly, the easiest method for preventing a conflict is to prohibit members and candidates from owning any such securities, but this approach is overly burdensome and discriminates against members and candidates.

Therefore, sell-side members and candidates should disclose any materially beneficial ownership interest in a security or other investment that the member or candidate is recommending. Buy-side members and candidates should disclose their procedures for reporting requirements for personal transactions. Conflicts arising from personal investing are discussed more fully in the guidance for Standard VI(B).

Conflicts as a Director

Service as a director poses three basic conflicts of interest. First, a conflict may exist between the duties owed to clients and the duties owed to shareholders of the company. Second, investment personnel who serve as directors may receive the securities or options to purchase securities of the company as compensation for serving on the board, which could raise questions about trading actions that might increase the value of those securities. Third, board service creates the opportunity to receive material nonpublic information involving the company. Even though the information is confidential, the perception could be that information not available to the public is being communicated to a director's firm—whether a broker, investment adviser, or other type of organization. When members or candidates providing investment services also serve as directors, they should be isolated from those making investment decisions by the use of firewalls or similar restrictions.

STANDARD VI(A): RECOMMENDED PROCEDURES

Members or candidates should disclose special compensation arrangements with the employer that might conflict with client interests, such as bonuses based on short-term performance criteria, commissions, incentive fees, performance fees, and referral fees. If the member's or candidate's firm does not permit such disclosure, the member or candidate should document the request and may consider dissociating from the activity.

Members' and candidates' firms are encouraged to include information on compensation packages in firms' promotional literature. If a member or candidate manages a portfolio for which the fee is based on capital gains or capital appreciation (a performance fee), this information should be disclosed to clients. If a member, a candidate, or a member's or candidate's firm has outstanding agent options to buy stock as part of the compensation package for corporate financing activities, the amount and expiration date of these options should be disclosed as a footnote to any research report published by the member's or candidate's firm.

STANDARD VI(A): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Conflict of Interest and Business Relationships):

Hunter Weiss is a research analyst with Farmington Company, a broker and investment banking firm. Farmington's merger and acquisition department has represented Vimco, a conglomerate, in all of Vimco's acquisitions for 20 years. From time to time, Farmington officers sit on the boards of directors of various Vimco subsidiaries. Weiss is writing a research report on Vimco.

> *Comment*: Weiss must disclose in his research report Farmington's special relationship with Vimco. Broker/dealer management of and participation in public offerings must be disclosed in research reports. Because the position of underwriter to a company entails a special past and potential future relationship with a company that is the subject of investment advice, it threatens the independence and objectivity of the report writer and must be disclosed.

Example 2 (Conflict of Interest and Business Stock Ownership):

The investment management firm of Dover & Roe sells a 25% interest in its partnership to a multinational bank holding company, First of New York. Immediately after the sale, Margaret Hobbs, president of Dover & Roe, changes her recommendation for First of New York's common stock from "sell" to "buy" and adds First of New York's commercial paper to Dover & Roe's approved list for purchase.

> *Comment*: Hobbs must disclose the new relationship with First of New York to all Dover & Roe clients. This relationship must also be disclosed to clients by the firm's portfolio managers when they make specific investment recommendations or take investment actions with respect to First of New York's securities.

Example 3 (Conflict of Interest and Personal Stock Ownership):

Carl Fargmon, a research analyst who follows firms producing office equipment, has been recommending purchase of Kincaid Printing because of its innovative new line of copiers. After his initial report on the company, Fargmon's wife inherits from a distant relative US$3 million of Kincaid stock. He has been asked to write a follow-up report on Kincaid.

> *Comment*: Fargmon must disclose his wife's ownership of the Kincaid stock to his employer and in his follow-up report. Best practice would be to avoid the conflict by asking his employer to assign another analyst to draft the follow-up report.

Example 4 (Conflict of Interest and Personal Stock Ownership):

Betty Roberts is speculating in penny stocks for her own account and purchases 100,000 shares of Drew Mining, Inc., for US$0.30 a share. She intends to sell these shares at the sign of any substantial upward price movement of the stock. A week later, her employer asks her to write a report on penny stocks in the mining industry to be published in two weeks. Even without owning the Drew stock, Roberts would recommend it in her report as a "buy." A surge in the price of the stock to the US$2 range is likely to result once the report is issued.

> *Comment*: Although this holding may not be material, Roberts must disclose it in the report and to her employer before writing the report because the gain for her will be substantial if the market responds strongly to her recommendation. The fact that she has only recently purchased the stock adds to the appearance that she is not entirely objective.

Example 5 (Conflict of Interest and Compensation Arrangements):

Samantha Snead, a portfolio manager for Thomas Investment Counsel, Inc., specializes in managing public retirement funds and defined benefit pension plan accounts, all of which have long-term investment objectives. A year ago, Snead's employer, in an attempt to motivate and retain key investment professionals, introduced a bonus compensation system that rewards portfolio managers on the basis of quarterly performance relative to their peers and to certain benchmark indexes. In an attempt to improve the short-term performance of her accounts, Snead changes her investment strategy and purchases several high-beta stocks for client portfolios. These purchases are seemingly contrary to the clients' investment policy statements. Following their purchase, an officer of Griffin Corporation, one of Snead's pension fund clients, asks why Griffin Corporation's portfolio seems to be dominated by high-beta stocks of companies that often appear among the most actively traded issues. No change in objective or strategy has been recommended by Snead during the year.

> *Comment*: Snead has violated Standard VI(A) by failing to inform her clients of the changes in her compensation arrangement with her employer, which created a conflict of interest between her compensation and her clients' IPSs. Firms may pay employees on the basis of performance, but pressure by Thomas Investment Counsel to achieve short-term performance goals is in basic conflict with the objectives of Snead's accounts.
>
> See also Standard III(C)–Suitability.

Example 6 (Conflict of Interest, Options, and Compensation Arrangements):

Wayland Securities works with small companies doing IPOs or secondary offerings. Typically, these deals are in the US$10 million to US$50 million range, and as a result, the corporate finance fees are quite small. To compensate for the small fees, Wayland Securities usually takes "agent options"—that is, rights (exercisable within a two-year time frame) to acquire up to an additional 10% of the current offering. Following an IPO performed by Wayland for Falk Resources, Ltd., Darcy Hunter, the head of corporate finance at Wayland, is concerned about receiving value for her Falk Resources options. The options are due to expire in one month, and the stock is not doing well. She contacts John Fitzpatrick in the research department of Wayland Securities, reminds him that he is eligible for 30% of these options, and indicates that now would be a good time to give some additional coverage to Falk Resources. Fitzpatrick agrees and immediately issues a favorable report.

> *Comment*: For Fitzpatrick to avoid being in violation of Standard VI(A), he must indicate in the report the volume and expiration date of agent options outstanding. Furthermore, because he is personally eligible for some of the options, Fitzpatrick must disclose the extent of this compensation. He also must be careful to not violate his duty of independence and objectivity under Standard I(B).

Example 7 (Conflict of Interest and Compensation Arrangements):

Gary Carter is a representative with Bengal International, a registered broker/dealer. Carter is approached by a stock promoter for Badger Company, who offers to pay Carter additional compensation for sales of Badger Company's stock to Carter's clients. Carter accepts the stock promoter's offer but does not disclose the arrangements to his clients or to his employer. Carter sells shares of the stock to his clients.

> *Comment*: Carter has violated Standard VI(A) by failing to disclose to clients that he is receiving additional compensation for recommending and selling Badger stock. Because he did not disclose the arrangement with Badger to his clients, the clients were unable to evaluate whether Carter's recommendations to buy Badger were affected by this arrangement. Carter's conduct also violated Standard VI(A) by failing to disclose to his employer monetary compensation received in addition to the compensation and benefits

conferred by his employer. Carter was required by Standard VI(A) to disclose the arrangement with Badger to his employer so that his employer could evaluate whether the arrangement affected Carter's objectivity and loyalty.

Example 8 (Conflict of Interest and Directorship):

Carol Corky, a senior portfolio manager for Universal Management, recently became involved as a trustee with the Chelsea Foundation, a large not-for-profit foundation in her hometown. Universal is a small money manager (with assets under management of approximately US$100 million) that caters to individual investors. Chelsea has assets in excess of US$2 billion. Corky does not believe informing Universal of her involvement with Chelsea is necessary.

Comment: By failing to inform Universal of her involvement with Chelsea, Corky violated Standard VI(A). Given the large size of the endowment at Chelsea, Corky's new role as a trustee can reasonably be expected to be time consuming, to the possible detriment of Corky's portfolio responsibilities with Universal. Also, as a trustee, Corky may become involved in the investment decisions at Chelsea. Therefore, Standard VI(A) obligates Corky to discuss becoming a trustee at Chelsea with her compliance officer or supervisor at Universal before accepting the position, and she should have disclosed the degree to which she would be involved in investment decisions at Chelsea.

Example 9 (Conflict of Interest and Personal Trading):

Bruce Smith covers eastern European equities for Marlborough Investments, an investment management firm with a strong presence in emerging markets. While on a business trip to Russia, Smith learns that investing in Russian equities directly is difficult but that equity-linked notes that replicate the performance of underlying Russian equities can be purchased from a New York–based investment bank. Believing that his firm would not be interested in such a security, Smith purchases a note linked to a Russian telecommunications company for his own account without informing Marlborough. A month later, Smith decides that the firm should consider investing in Russian equities by way of the equity-linked notes. He prepares a write-up on the market that concludes with a recommendation to purchase several of the notes. One note he recommends is linked to the same Russian telecom company that Smith holds in his personal account.

Comment: Smith has violated Standard VI(A) by failing to disclose his purchase and ownership of the note linked to the Russian telecom company. Smith is required by the standard to disclose the investment opportunity to his employer and look to his company's policies on personal trading to determine whether it was proper for him to purchase the note for his own account. By purchasing the note, Smith may or may not have impaired his ability to make an unbiased and objective assessment of the appropriateness of the derivative instrument for his firm, but Smith's failure to disclose the purchase to his employer impaired his employer's ability to decide whether his ownership of the security is a conflict of interest that might affect Smith's future recommendations. Then, when he recommended the particular telecom notes to his firm, Smith compounded his problems by not disclosing that he owned the notes in his personal account—a clear conflict of interest.

Example 10 (Conflict of Interest and Requested Favors):

Michael Papis is the chief investment officer of his state's retirement fund. The fund has always used outside advisers for the real estate allocation, and this information is clearly presented in all fund communications. Thomas Nagle, a recognized sell-side research analyst and Papis's business school classmate, recently left the investment bank he worked for to start his own asset management firm, Accessible Real Estate. Nagle is trying to build his assets under management and contacts Papis about gaining some of the retirement fund's allocation. In the previous few years, the performance of the retirement fund's real estate investments was in line with the fund's benchmark but was not extraordinary. Papis decides to help out his old friend and also to seek better returns by moving the real estate allocation to Accessible. The only notice of the change in adviser appears in the next annual report in the listing of associated advisers.

> *Comment*: Papis has violated Standard VI(A) by not disclosing to his employer his personal relationship with Nagle. Disclosure of his past history with Nagle would allow his firm to determine whether the conflict may have impaired Papis's independence in deciding to change managers.
>
> See also Standard IV(C)–Responsibilities of Supervisors, Standard V(A)–Diligence and Reasonable Basis, and Standard V(B)–Communication with Clients and Prospective Clients.

Example 11 (Conflict of Interest and Business Relationships):

Bob Wade, trust manager for Central Midas Bank, was approached by Western Funds about promoting its family of funds, with special interest in the service-fee class. To entice Central to promote this class, Western Funds offered to pay the bank a service fee of 0.25%. Without disclosing the fee being offered to the bank, Wade asked one of the investment managers to review the Western Funds family of funds to determine whether they were suitable for clients of Central. The manager completed the normal due diligence review and determined that the funds were fairly valued in the market with fee structures on a par with their competitors. Wade decided to accept Western's offer and instructed the team of portfolio managers to exclusively promote these funds and the service-fee class to clients seeking to invest new funds or transfer from their current investments. So as to not influence the investment managers, Wade did not disclose the fee offer and allowed that income to flow directly to the bank.

> *Comment*: Wade is violating Standard VI(A) by not disclosing the portion of the service fee being paid to Central. Although the investment managers may not be influenced by the fee, neither they nor the client have the proper information about Wade's decision to exclusively market this fund family and class of investments. Central may come to rely on the new fee as a component of the firm's profitability and may be unwilling to offer other products in the future that could affect the fees received.
>
> See also Standard I(B)–Independence and Objectivity.

Example 12 (Disclosure of Conflicts to Employers):

Yehudit Dagan is a portfolio manager for Risk Management Bank (RMB), whose clients include retirement plans and corporations. RMB provides a defined contribution retirement plan for its employees that offers 20 large diversified mutual fund investment options, including a mutual fund managed by Dagan's RMB colleagues. After being employed for six months, Dagan became eligible to participate in the retirement plan, and she intends to allocate her retirement plan assets in six of the investment options, including the fund managed by her RMB colleagues. Dagan is concerned that joining the plan will lead to a potentially significant amount of paperwork for her (e.g., disclosure of her retirement account holdings and needing preclearance for her transactions), especially with her investing in the in-house fund.

> *Comment*: Standard VI(A) would not require Dagan to disclose her personal or retirement investments in large diversified mutual funds, unless specifically required by her employer. For practical reasons, the standard does not require Dagan to gain preclearance for ongoing payroll deduction contributions to retirement plan account investment options.
>
> Dagan should ensure that her firm does not have a specific policy regarding investment—whether personal or in the retirement account—for funds managed by the company's employees. These mutual funds may be subject to the company's disclosure, preclearance, and trading restriction procedures to identify possible conflicts prior to the execution of trades.

STANDARD VI(B): CONFLICTS OF INTEREST - PRIORITY OF TRANSACTIONS

Investment transactions for clients and employers must have priority over investment transactions in which a Member or Candidate is the beneficial owner.

Guidance

Highlights:

- *Avoiding Potential Conflicts*
- *Personal Trading Secondary to Trading for Clients*
- *Standards for Nonpublic Information*
- *Impact on All Accounts with Beneficial Ownership*

Standard VI(B) reinforces the responsibility of members and candidates to give the interests of their clients and employers priority over their personal financial interests. This standard is designed to prevent any potential conflict of interest or the appearance of a conflict of interest with respect to personal transactions. Client interests have priority. Client transactions must take precedence over transactions made on behalf of the member's or candidate's firm or personal transactions.

Avoiding Potential Conflicts

Conflicts between the client's interest and an investment professional's personal interest may occur. Although conflicts of interest exist, nothing is inherently unethical about individual managers, advisers, or mutual fund employees making money from personal investments as long as (1) the client is not disadvantaged by the trade, (2) the investment professional does not benefit personally from trades undertaken for clients, and (3) the investment professional complies with applicable regulatory requirements.

Some situations occur where a member or candidate may need to enter a personal transaction that runs counter to current recommendations or what the portfolio manager is doing for client portfolios. For example, a member or candidate may be required at some point to sell an asset to make a college tuition payment or a down payment on a home, to meet a margin call, or so on. The sale may be contrary to the long-term advice the member or candidate is currently providing to clients. In these situations, the same three criteria given in the preceding paragraph should be applied in the transaction so as to not violate Standard VI(B).

Personal Trading Secondary to Trading for Clients

Standard VI(B) states that transactions for clients and employers must have priority over transactions in securities or other investments for which a member or candidate is the beneficial owner. The objective of the standard is to prevent personal transactions from adversely affecting the interests of clients or employers. A member or candidate having the same investment positions or being co-invested with clients does not always create a conflict. Some clients in certain investment situations require members or candidates to have aligned interests. Personal investment positions or transactions of members or candidates or their firm should never, however, adversely affect client investments.

Standards for Nonpublic Information

Standard VI(B) covers the activities of members and candidates who have knowledge of pending transactions that may be made on behalf of their clients or employers, who have access to nonpublic information during the normal preparation of research recommendations, or who take investment actions. Members and candidates are prohibited from conveying nonpublic information to any person whose relationship to the member or candidate makes the member or candidate a beneficial owner of the person's securities. Members and candidates must not convey this information to any other person if the nonpublic information can be deemed material.

Impact on All Accounts with Beneficial Ownership

Members or candidates may undertake transactions in accounts for which they are a beneficial owner only after their clients and employers have had adequate opportunity to act on a recommendation. Personal transactions include those made for the member's or candidate's own account, for family (including spouse, children, and other immediate family members) accounts, and for accounts in which the member or candidate has a direct or indirect pecuniary interest, such as a trust or retirement account. Family accounts that are client accounts should be treated like any other firm account and should neither be given special treatment nor be disadvantaged because of the family relationship. If a member or candidate has a beneficial ownership in the account, however, the member or candidate may be subject to preclearance or reporting requirements of the employer or applicable law.

STANDARD VI(B): RECOMMENDED PROCEDURES

Policies and procedures designed to prevent potential conflicts of interest, and even the appearance of a conflict of interest, with respect to personal transactions are critical to establishing investor confidence in the securities industry. Therefore, members and candidates should urge their firms to establish such policies and procedures. Because investment firms vary greatly in assets under management, types of clients, number of employees, and so on, each firm should have policies regarding personal investing that are best suited to the firm. Members and candidates should then prominently disclose these policies to clients and prospective clients.

The specific provisions of each firm's standards will vary, but all firms should adopt certain basic procedures to address the conflict areas created by personal investing. These procedures include the following:

- *Limited participation in equity IPOs*: Some eagerly awaited IPOs rise significantly in value shortly after the issue is brought to market. Because the new issue may be highly attractive and sought after, the opportunity to participate in the IPO may be limited. Therefore, purchases of IPOs by investment personnel create conflicts of interest in two principal ways. First, participation in an IPO may have the appearance of taking away an attractive investment opportunity from clients for personal gain—a clear breach of the duty of loyalty to clients. Second, personal purchases in IPOs may have the appearance that the investment opportunity is being bestowed as an incentive to make future investment decisions for the benefit of the party providing the opportunity. Members and candidates can avoid these conflicts or appearances of conflicts of interest by not participating in IPOs.

  Reliable and systematic review procedures should be established to ensure that conflicts relating to IPOs are identified and appropriately dealt with by supervisors. Members and candidates should preclear their participation in IPOs, even in situations without any conflict of interest between a member's or candidate's participation in an IPO and the client's interests. Members and candidates should not benefit from the position that their clients occupy in the marketplace—through preferred trading, the allocation of limited offerings, or oversubscription.

- *Restrictions on private placements*: Strict limits should be placed on investment personnel acquiring securities in private placements, and appropriate supervisory and review procedures should be established to prevent noncompliance.

  Firms do not routinely use private placements for clients (e.g., venture capital deals) because of the high risk associated with them. Conflicts related to private placements are more significant to members and candidates who manage large pools of assets or act as plan sponsors because these managers may be offered special opportunities, such as private placements, as a reward or an enticement for continuing to do business with a particular broker.

  Participation in private placements raises conflict-of-interest issues that are similar to issues surrounding IPOs. Investment personnel should not be involved in transactions, including (but not limited to) private placements, that could be perceived as favors or gifts that seem designed to influence future judgment or to reward past business deals.

  Whether the venture eventually proves to be good or bad, managers have an immediate conflict concerning private placement opportunities. If and when the investments go public, participants in private placements have an incentive to recommend the investments to clients regardless of the suitability of the investments for their clients. Doing so increases the value of the participants' personal portfolios.

- *Establish blackout/restricted periods*: Investment personnel involved in the investment decision-making process should establish blackout periods prior to trades for clients so that managers cannot take advantage of their knowledge of client activity by "front-running" client trades (trading for one's personal account before trading for client accounts).

Individual firms must decide who within the firm should be required to comply with the trading restrictions. At a minimum, all individuals who are involved in the investment decision-making process should be subject to the same restricted period. Each firm must determine specific requirements related to blackout and restricted periods that are most relevant to the firm while ensuring that the procedures are governed by the guiding principles set forth in the Code and Standards. Size of firm and type of securities purchased are relevant factors. For example, in a large firm, a blackout requirement is, in effect, a total trading ban because the firm is continually trading in most securities. In a small firm, the blackout period is more likely to prevent the investment manager from front-running.

- *Reporting requirements*: Supervisors should establish reporting procedures for investment personnel, including disclosure of personal holdings/beneficial ownerships, confirmations of trades to the firm and the employee, and preclearance procedures. Once trading restrictions are in place, they must be enforced. The best method for monitoring and enforcing procedures to eliminate conflicts of interest in personal trading is through reporting requirements, including the following:
    - **Disclosure of holdings in which the employee has a beneficial interest**. Disclosure by investment personnel to the firm should be made upon commencement of the employment relationship and at least annually thereafter. To address privacy considerations, disclosure of personal holdings should be handled in a confidential manner by the firm.
    - **Providing duplicate confirmations of transactions**. Investment personnel should be required to direct their brokers to supply to firms duplicate copies or confirmations of all their personal securities transactions and copies of periodic statements for all securities accounts. The duplicate confirmation requirement has two purposes: (1) The requirement sends a message that there is independent verification, which reduces the likelihood of unethical behavior, and (2) it enables verification of the accounting of the flow of personal investments that cannot be determined from merely looking at holdings.
    - **Preclearance procedures**. Investment personnel should examine all planned personal trades to identify possible conflicts prior to the execution of the trades. Preclearance procedures are designed to identify possible conflicts before a problem arises.
- *Disclosure of policies*: Members and candidates should fully disclose to investors their firm's policies regarding personal investing. The information about employees' personal investment activities and policies will foster an atmosphere of full and complete disclosure and calm the public's legitimate concerns about the conflicts of interest posed by investment personnel's personal trading. The disclosure must provide helpful information to investors; it should not be simply boilerplate language, such as "investment personnel are subject to policies and procedures regarding their personal trading."

STANDARD VI(B): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Personal Trading):

Research analyst Marlon Long does not recommend purchase of a common stock for his employer's account because he wants to purchase the stock personally and does not want to wait until the recommendation is approved and the stock is purchased by his employer.

> *Comment*: Long has violated Standard VI(B) by taking advantage of his knowledge of the stock's value before allowing his employer to benefit from that information.

Example 2 (Trading for Family Member Account):

Carol Baker, the portfolio manager of an aggressive growth mutual fund, maintains an account in her husband's name at several brokerage firms with which the fund and a number of Baker's other individual clients do a substantial amount of business. Whenever a hot issue becomes available, she instructs the brokers to buy it for her husband's account. Because such issues normally are scarce, Baker often acquires shares in hot issues but her clients are not able to participate in them.

> *Comment*: To avoid violating Standard VI(B), Baker must acquire shares for her mutual fund first and acquire them for her husband's account only after doing so, even though she might miss out on participating in new issues via her husband's account. She also must disclose the trading for her husband's account to her employer because this activity creates a conflict between her personal interests and her employer's interests.

Example 3 (Family Accounts as Equals):

Erin Toffler, a portfolio manager at Esposito Investments, manages the retirement account established with the firm by her parents. Whenever IPOs become available, she first allocates shares to all her other clients for whom the investment is appropriate; only then does she place any remaining portion in her parents' account, if the issue is appropriate for them. She has adopted this procedure so that no one can accuse her of favoring her parents.

> *Comment*: Toffler has violated Standard VI(B) by breaching her duty to her parents by treating them differently from her other accounts simply because of the family relationship. As fee-paying clients of Esposito Investments, Toffler's parents are entitled to the same treatment as any other client of the firm. If Toffler has beneficial ownership in the account, however, and Esposito Investments has preclearance and reporting requirements for personal transactions, she may have to preclear the trades and report the transactions to Esposito.

Example 4 (Personal Trading and Disclosure):

Gary Michaels is an entry-level employee who holds a low-paying job serving both the research department and the investment management department of an active investment management firm. He purchases a sports car and begins to wear expensive clothes after only a year of employment with the firm. The director of the investment management department, who has responsibility for monitoring the personal stock transactions of all employees, investigates and discovers that Michaels has made substantial investment gains by purchasing stocks just before they were put on the firm's recommended "buy" list. Michaels was regularly given the firm's quarterly personal transaction form but declined to complete it.

> *Comment*: Michaels violated Standard VI(B) by placing personal transactions ahead of client transactions. In addition, his supervisor violated Standard IV(C)–Responsibilities of Supervisors by permitting Michaels to continue to perform his assigned tasks without having signed the quarterly personal transaction form. Note also that if Michaels had communicated information about the firm's recommendations to a person who traded the security, that action would be a misappropriation of the information and a violation of Standard II(A)–Material Nonpublic Information.

Example 5 (Trading Prior to Report Dissemination):

A brokerage's insurance analyst, Denise Wilson, makes a closed-circuit TV report to her firm's branches around the country. During the broadcast, she includes negative comments about a major company in the insurance industry. The following day, Wilson's report is printed and distributed to the sales force and public customers. The report recommends that both short-term traders and intermediate investors take profits by selling that insurance company's stock. Seven minutes after the broadcast, however, Ellen Riley, head of the firm's trading department, had closed out a long "call" position in the stock. Shortly thereafter, Riley established a sizable "put" position in the stock. When asked about her activities, Riley claimed she took the actions to facilitate anticipated sales by institutional clients.

> *Comment*: Riley did not give customers an opportunity to buy or sell in the options market before the firm itself did. By taking action before the report was disseminated, Riley's firm may have depressed the price of the calls and increased the price of the puts. The firm could have avoided a conflict of interest if it had waited to trade for its own account until its clients had an opportunity to receive and assimilate Wilson's recommendations. As it is, Riley's actions violated Standard VI(B).

STANDARD VI(C): CONFLICTS OF INTEREST - REFERRAL FEES

Members and Candidates must disclose to their employer, clients, and prospective clients, as appropriate, any compensation, consideration, or benefit received from or paid to others for the recommendation of products or services.

Guidance

Standard VI(C) states the responsibility of members and candidates to inform their employer, clients, and prospective clients of any benefit received for referrals of customers and clients. Such disclosures allow clients or employers to evaluate (1) any partiality shown in any recommendation of services and (2) the full cost of the services. Members and candidates must disclose when they pay a fee or provide compensation to others who have referred prospective clients to the member or candidate.

Appropriate disclosure means that members and candidates must advise the client or prospective client, before entry into any formal agreement for services, of any benefit given or received for the recommendation of any services provided by the member or candidate. In addition, the member or candidate must disclose the nature of the consideration or benefit—for example, flat fee or percentage basis, one-time or continuing benefit, based on performance, benefit in the form of provision of research or other noncash benefit—together with the estimated dollar value. Consideration includes all fees, whether paid in cash, in soft dollars, or in kind.

STANDARD VI(C): RECOMMENDED PROCEDURES

Members and candidates should encourage their employers to develop procedures related to referral fees. The firm may completely restrict such fees. If the firm does not adopt a strict prohibition of such fees, the procedures should indicate the appropriate steps for requesting approval.

Employers should have investment professionals provide to the clients notification of approved referral fee programs and provide the employer regular (at least quarterly) updates on the amount and nature of compensation received.

STANDARD VI(C): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Disclosure of Referral Arrangements and Outside Parties):

Brady Securities, Inc., a broker/dealer, has established a referral arrangement with Lewis Brothers, Ltd., an investment counseling firm. In this arrangement, Brady Securities refers all prospective tax-exempt accounts, including pension, profit-sharing, and endowment accounts, to Lewis Brothers. In return, Lewis Brothers makes available to Brady Securities on a regular basis the security recommendations and reports of its research staff, which registered representatives of Brady Securities use in serving customers. In addition, Lewis Brothers conducts monthly economic and market reviews for Brady Securities personnel and directs all stock commission business generated by referral accounts to Brady Securities.

Willard White, a partner in Lewis Brothers, calculates that the incremental costs involved in functioning as the research department of Brady Securities are US$20,000 annually.

Referrals from Brady Securities last year resulted in fee income of US$200,000 for Lewis Brothers, and directing all stock trades through Brady Securities resulted in additional costs to Lewis Brothers' clients of US$10,000.

Diane Branch, the chief financial officer of Maxwell Inc., contacts White and says that she is seeking an investment manager for Maxwell's profit-sharing plan. She adds, "My friend Harold Hill at Brady Securities recommended your firm without qualification, and that's good enough for me. Do we have a deal?" White accepts the new account but does not disclose his firm's referral arrangement with Brady Securities.

> *Comment*: White has violated Standard VI(C) by failing to inform the prospective customer of the referral fee payable in services and commissions for an indefinite period to Brady Securities. Such disclosure could have caused Branch to reassess Hill's recommendation and make a more critical evaluation of Lewis Brothers' services.

Example 2 (Disclosure of Interdepartmental Referral Arrangements):

James Handley works for the trust department of Central Trust Bank. He receives compensation for each referral he makes to Central Trust's brokerage department and personal financial management department that results in a sale. He refers several of his clients to the personal financial management department but does not disclose the arrangement within Central Trust to his clients.

> *Comment*: Handley has violated Standard VI(C) by not disclosing the referral arrangement at Central Trust Bank to his clients. Standard VI(C) does not distinguish between referral payments paid by a third party for referring clients to the third party and internal payments paid within the firm to attract new business to a subsidiary. Members and candidates must disclose all such referral fees. Therefore, Handley is required to disclose, at the time of referral, any referral fee agreement in place among Central Trust Bank's departments. The disclosure should include the nature and the value of the benefit and should be made in writing.

Example 3 (Disclosure of Referral Arrangements and Informing Firm):

Katherine Roberts is a portfolio manager at Katama Investments, an advisory firm specializing in managing assets for high-net-worth individuals. Katama's trading desk uses a variety of brokerage houses to execute trades on behalf of its clients. Roberts asks the trading desk to direct a large portion of its commissions to Naushon, Inc., a small broker/dealer run by one of Roberts' business school classmates. Katama's traders have found that Naushon is not very competitive on pricing, and although Naushon generates some research for its trading clients, Katama's other analysts have found most of Naushon's research to be not especially useful. Nevertheless, the traders do as Roberts asks, and in return for receiving a large portion of Katama's business, Naushon recommends the investment services of Roberts and Katama to its wealthiest clients. This arrangement is not disclosed to either Katama or the clients referred by Naushon.

> *Comment*: Roberts is violating Standard VI(C) by failing to inform her employer of the referral arrangement.

Example 4 (Disclosure of Referral Arrangements and Outside Organizations):

Alex Burl is a portfolio manager at Helpful Investments, a local investment advisory firm. Burl is on the advisory board of his child's school, which is looking for ways to raise money to purchase new playground equipment for the school. Burl discusses a plan with his supervisor in which he will donate to the school a portion of his service fee from new clients referred by the parents of students at the school. Upon getting the approval from Helpful, Burl presents the idea to the school's advisory board and directors. The school agrees to announce the program at the next parent event and asks Burl to provide the appropriate written materials to be distributed. A week following the distribution of the flyers, Burl receives the first school-related referral. In establishing the client's investment policy statement, Burl clearly discusses the school's referral and outlines the plans for distributing the donation back to the school.

> *Comment*: Burl has not violated Standard VI(C) because he secured the permission of his employer, Helpful Investments, and the school prior to beginning the program and because he discussed the arrangement with the client at the time the investment policy statement was designed.

Example 5 (Disclosure of Referral Arrangements and Outside Parties):

The sponsor of a state employee pension is seeking to hire a firm to manage the pension plan's emerging market allocation. To assist in the review process, the sponsor has hired Thomas Arrow as a consultant to solicit proposals from various advisers. Arrow is contracted by the sponsor to represent its best interest in selecting the most appropriate new manager. The process runs smoothly, and Overseas Investments is selected as the new manager.

The following year, news breaks that Arrow is under investigation by the local regulator for accepting kickbacks from investment managers after they are awarded new pension allocations. Overseas Investments is included in the list of firms allegedly making these payments. Although the sponsor is happy with the performance of Overseas since it has been managing the pension plan's emerging market funds, the sponsor still decides to have an independent review of the proposals and the selection process to ensure that Overseas was the appropriate firm for its needs. This review confirms that, even though Arrow was being paid by both parties, the recommendation of Overseas appeared to be objective and appropriate.

> *Comment*: Arrow has violated Standard VI(C) because he did not disclose the fee being paid by Overseas. Withholding this information raises the question of a potential lack of objectivity in the recommendation of Overseas by Arrow; this aspect is in addition to questions about the legality of having firms pay to be considered for an allocation.
>
> Regulators and governmental agencies may adopt requirements concerning allowable consultant activities. Local regulations sometimes include having a consultant register with the regulatory agency's ethics board. Regulator policies may include a prohibition on acceptance of payments from investment managers receiving allocations and require regular reporting of contributions made to political organizations and candidates. Arrow would have to adhere to these requirements as well as the Code and Standards.

STANDARD VII(A): RESPONSIBILITIES AS A CFA INSTITUTE MEMBER OR CFA CANDIDATE - CONDUCT AS PARTICIPANTS IN CFA INSTITUTE PROGRAMS

Standard VII(A) Conduct as Participants in CFA Institute Programs

> Members and Candidates must not engage in any conduct that compromises the reputation or integrity of CFA Institute or the CFA designation or the integrity, validity, or security of CFA Institute programs.

Guidance

Highlights:

- *Confidential Program Information*
- *Additional CFA Program Restrictions*
- *Expressing an Opinion*

Standard VII(A) covers the conduct of CFA Institute members and candidates involved with the CFA Program and prohibits any conduct that undermines the public's confidence that the CFA charter represents a level of achievement based on merit and ethical conduct. There is an array of CFA Institute programs beyond the CFA Program that provide additional educational and credentialing opportunities, including the Certificate in Investment Performance Measurement (CIPM) Program and the CFA

Institute Investment Foundations™ Program. The standard's function is to hold members and candidates to a high ethical criterion while they are participating in or involved with any CFA Institute program. Conduct covered includes but is not limited to

- giving or receiving assistance (cheating) on any CFA Institute examinations;
- violating the rules, regulations, and testing policies of CFA Institute programs;
- providing confidential program or exam information to candidates or the public;
- disregarding or attempting to circumvent security measures established for any CFA Institute examinations;
- improperly using an association with CFA Institute to further personal or professional goals; and
- misrepresenting information on the Professional Conduct Statement or in the CFA Institute Continuing Education Program.

Confidential Program Information

CFA Institute is vigilant about protecting the integrity of CFA Institute programs' content and examination processes. CFA Institute program rules, regulations, and policies prohibit candidates from disclosing confidential material gained during the exam process.

Examples of information that cannot be disclosed by candidates sitting for an exam include but are not limited to

- specific details of questions appearing on the exam and
- broad topical areas and formulas tested or not tested on the exam.

All aspects of the exam, including questions, broad topical areas, and formulas, tested or not tested, are considered confidential until such time as CFA Institute elects to release them publicly. This confidentiality requirement allows CFA Institute to maintain the integrity and rigor of exams for future candidates. Standard VII(A) does not prohibit candidates from discussing nonconfidential information or curriculum material with others or in study groups in preparation for the exam.

Candidates increasingly use online forums and new technology as part of their exam preparations. CFA Institute actively polices blogs, forums, and related social networking groups for information considered confidential. The organization works with both individual candidates and the sponsors of online or offline services to promptly remove any and all violations. As noted in the discussion of Standard I(A)–Knowledge of the Law, candidates, members, and the public are encouraged to report suspected violations to CFA Institute.

Additional CFA Program Restrictions

The CFA Program rules, regulations, and policies define additional allowed and disallowed actions concerning the exams. Violating any of the testing policies, such as the calculator policy, personal belongings policy, or the Candidate Pledge, constitutes a violation of Standard VII(A). Candidates will find all of these policies on the CFA Program portion of the CFA Institute website (www.cfainstitute.org). Exhibit 2 provides the Candidate Pledge, which highlights the respect candidates must have for the integrity, validity, and security of the CFA exam.

Members may participate as volunteers in various aspects of the CFA Program. Standard VII(A) prohibits members from disclosing and/or soliciting confidential material gained prior to or during the exam and grading processes with those outside the CFA exam development process.

Examples of information that cannot be shared by members involved in developing, administering, or grading the exams include but are not limited to

- questions appearing on the exam or under consideration,
- deliberation related to the exam process, and
- information related to the scoring of questions.

Members may also be asked to offer assistance with other CFA Institute programs, including but not limited to the CIPM and Investment Foundations programs. Members participating in any CFA Institute program should do so with the same level of integrity and confidentiality as is required of participation in the CFA Program.

Expressing an Opinion

Standard VII(A) does *not* cover expressing opinions regarding CFA Institute, the CFA Program, or other CFA Institute programs. Members and candidates are free to disagree and express their disagreement with CFA Institute on its policies, its procedures, or any advocacy positions taken by the organization. When expressing a personal opinion, a candidate is prohibited from disclosing content-specific information, including any actual exam question and the information as to subject matter covered or not covered in the exam.

Exhibit 2 Sample of CFA Program Testing Policies

Candidate Pledge	As a candidate in the CFA Program, I am obligated to follow Standard VII(A) of the CFA Institute Standards of Professional Conduct, which states that members and candidates must not engage in any conduct that compromises the reputation or integrity of CFA Institute or the CFA designation or the integrity, validity, or security of the CFA exam. ■ Prior to this exam, I have not given or received information regarding the content of this exam. During this exam, I will not give or receive any information regarding the content of this exam. ■ After this exam, I will not disclose **ANY** portion of this exam and I will not remove **ANY** exam materials from the testing room in original or copied form. I understand that all exam materials, including my answers, are the property of CFA Institute and will not be returned to me in any form. ■ I will follow **ALL** rules of the CFA Program as stated on the CFA Institute website and the back cover of the exam book. My violation of any rules of the CFA Program will result in CFA Institute voiding my exam results and may lead to suspension or termination of my candidacy in the CFA Program.

STANDARD VII(A): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Sharing Exam Questions):

Travis Nero serves as a proctor for the administration of the CFA examination in his city. In the course of his service, he reviews a copy of the Level II exam on the evening prior to the exam's administration and provides information concerning the exam questions to two candidates who use it to prepare for the exam.

> *Comment*: Nero and the two candidates have violated Standard VII(A). By giving information about the exam questions to two candidates, Nero provided an unfair advantage to the two candidates and undermined the integrity and validity of the Level II exam as an accurate measure of the knowledge, skills, and abilities necessary to earn the right to use the CFA designation. By accepting the information, the candidates also compromised the integrity and validity of the Level II exam and undermined the ethical framework that is a key part of the designation.

Example 2 (Bringing Written Material into Exam Room):

Loren Sullivan is enrolled to take the Level II CFA examination. He has been having difficulty remembering a particular formula, so prior to entering the exam room, he writes the formula on the palm of his hand. During the afternoon section of the exam, a proctor notices Sullivan looking at the palm of his hand. She asks to see his hand and finds the formula.

> *Comment*: Because Sullivan wrote down information from the Candidate Body of Knowledge (CBOK) and took that written information into the exam room, his conduct compromised the validity of his exam performance and violated Standard VII(A). Sullivan's conduct was also in direct contradiction with the rules and regulations of the CFA Program, the Candidate Pledge, and the CFA Institute Code and Standards.

Example 3 (Writing after Exam Period End):

At the conclusion of the morning section of the Level I CFA examination, the proctors announce, "Stop writing now." John Davis has not completed the exam, so he continues to randomly fill in ovals on his answer sheet. A proctor approaches Davis's desk and reminds him that he should stop writing immediately. Davis, however, continues to complete the answer sheet. After the proctor asks him to stop writing two additional times, Davis finally puts down his pencil.

Comment: By continuing to complete his exam after time was called, Davis has violated Standard VII(A). By continuing to write, Davis took an unfair advantage over other candidates, and his conduct compromised the validity of his exam performance. Additionally, by not heeding the proctor's repeated instructions, Davis violated the rules and regulations of the CFA Program.

Example 4 (Sharing Exam Content):

After completing Level II of the CFA exam, Annabelle Rossi posts on her blog about her experience. She posts the following: "Level II is complete! I think I did fairly well on the exam. It was really difficult, but fair. I think I did especially well on the derivatives questions. And there were tons of them! I think I counted 18! The ethics questions were really hard. I'm glad I spent so much time on the Code and Standards. I was surprised to see there were no questions at all about IPO allocations. I expected there to be a couple. Well, off to celebrate getting through it. See you tonight?"

Comment: Rossi did not violate Standard VII(A) when she wrote about how difficult she found the exam or how well she thinks she may have done. By revealing portions of the CBOK covered on the exam and areas not covered, however, she did violate Standard VII(A) and the Candidate Pledge. Depending on the time frame in which the comments were posted, Rossi not only may have assisted future candidates but also may have provided an unfair advantage to candidates yet to sit for the same exam, thereby undermining the integrity and validity of the Level II exam.

Example 5 (Sharing Exam Content):

Level I candidate Etienne Gagne has been a frequent visitor to an internet forum designed specifically for CFA Program candidates. The week after completing the Level I examination, Gagne and several others begin a discussion thread on the forum about the most challenging questions and attempt to determine the correct answers.

Comment: Gagne has violated Standard VII(A) by providing and soliciting confidential exam information, which compromises the integrity of the exam process and violates the Candidate Pledge. In trying to determine correct answers to specific questions, the group's discussion included question-specific details considered to be confidential to the CFA Program.

Example 6 (Sharing Exam Content):

CFA4Sure is a company that produces test-preparation materials for CFA Program candidates. Many candidates register for and use the company's products. The day after the CFA examination, CFA4Sure sends an e-mail to all its customers asking them to share with the company the hardest questions from the exam so that CFA4Sure can better prepare its customers for the next exam administration. Marisol Pena e-mails a summary of the questions she found most difficult on the exam.

Comment: Pena has violated Standard VII(A) by disclosing a portion of the exam questions. The information provided is considered confidential until publicly released by CFA Institute. CFA4Sure is likely to use such feedback to refine its review materials for future candidates. Pena's sharing of the specific questions undermines the integrity of the exam while potentially making the exam easier for future candidates.

If the CFA4Sure employees who participated in the solicitation of confidential CFA Program information are CFA Institute members or candidates, they also have violated Standard VII(A).

Example 7 (Discussion of Exam Grading Guidelines and Results):

Prior to participating in grading CFA examinations, Wesley Whitcomb is required to sign a CFA Institute Grader Agreement. As part of the Grader Agreement, Whitcomb agrees not to reveal or discuss the exam materials with anyone except CFA Institute staff or other graders. Several weeks after the conclusion of the CFA exam grading, Whitcomb tells several colleagues who are candidates in the CFA Program which question he graded. He also discusses the guideline answer and adds that few candidates scored well on the question.

Comment: Whitcomb violated Standard VII(A) by breaking the Grader Agreement and disclosing information related to a specific question on the exam, which compromised the integrity of the exam process.

Example 8 (Compromising CFA Institute Integrity as a Volunteer):

Jose Ramirez is an investor-relations consultant for several small companies that are seeking greater exposure to investors. He is also the program chair for the CFA Institute society in the city where he works. Ramirez schedules only companies that are his clients to make presentations to the society and excludes other companies.

Comment: Ramirez, by using his volunteer position at CFA Institute to benefit himself and his clients, compromises the reputation and integrity of CFA Institute and thus violates Standard VII(A).

Example 9 (Compromising CFA Institute Integrity as a Volunteer):

Marguerite Warrenski is a member of the CFA Institute GIPS Executive Committee, which oversees the creation, implementation, and revision of the GIPS standards. As a member of the Executive Committee, she has advance knowledge of confidential information regarding the GIPS standards, including any new or revised standards the committee is considering. She tells her clients that her Executive Committee membership will allow her to better assist her clients in keeping up with changes to the Standards and facilitating their compliance with the changes.

Comment: Warrenski is using her association with the GIPS Executive Committee to promote her firm's services to clients and potential clients. In defining her volunteer position at CFA Institute as a strategic business advantage over competing firms and implying to clients that she would use confidential information to further their interests, Warrenski is compromising the reputation and integrity of CFA Institute and thus violating Standard VII(A). She may factually state her involvement with the Executive Committee but cannot infer any special advantage to her clients from such participation.

STANDARD VII(B): RESPONSIBILITIES AS A CFA INSTITUTE MEMBER OR CFA CANDIDATE - REFERENCE TO CFA INSTITUTE, THE CFA DESIGNATION, AND THE CFA PROGRAM

When referring to CFA Institute, CFA Institute membership, the CFA designation, or candidacy in the CFA Program, Members and Candidates must not misrepresent or exaggerate the meaning or implications of membership in CFA Institute, holding the CFA designation, or candidacy in the CFA Program.

Guidance

Highlights:

- *CFA Institute Membership*
- *Using the CFA Designation*
- *Referring to Candidacy in the CFA Program*

Standard VII(B) is intended to prevent promotional efforts that make promises or guarantees that are tied to the CFA designation. Individuals must not exaggerate the meaning or implications of membership in CFA Institute, holding the CFA designation, or candidacy in the CFA Program.

Standard VII(B) is not intended to prohibit factual statements related to the positive benefit of earning the CFA designation. However, statements referring to CFA Institute, the CFA designation, or the CFA Program that overstate the competency of an individual or imply, either directly or indirectly, that superior performance can be expected from someone with the CFA designation are not allowed under the standard.

Statements that highlight or emphasize the commitment of CFA Institute members, CFA charterholders, and CFA candidates to ethical and professional conduct or mention the thoroughness and rigor of the CFA Program are appropriate. Members and candidates may make claims about the relative merits of CFA Institute, the CFA Program, or the Code and Standards as long as those statements are implicitly or explicitly stated as the opinion of the speaker. Statements that do not express opinions have to be supported by facts.

Standard VII(B) applies to any form of communication, including but not limited to communications made in electronic or written form (such as on firm letterhead, business cards, professional biographies, directory listings, printed advertising, firm brochures, or personal resumes) and oral statements made to the public, clients, or prospects.

CFA Institute Membership

The term "CFA Institute member" refers to "regular" and "affiliate" members of CFA Institute who have met the membership requirements as defined in the CFA Institute Bylaws. Once accepted as a CFA Institute member, the member must satisfy the following requirements to maintain his or her status:

- remit annually to CFA Institute a completed Professional Conduct Statement, which renews the commitment to abide by the requirements of the Code and Standards and the CFA Institute Professional Conduct Program, and
- pay applicable CFA Institute membership dues on an annual basis.

If a CFA Institute member fails to meet any of these requirements, the individual is no longer considered an active member. Until membership is reactivated, individuals must not present themselves to others as active members. They may state, however, that they were CFA Institute members in the past or refer to the years when their membership was active.

Using the CFA Designation

Those who have earned the right to use the Chartered Financial Analyst designation are encouraged to do so but only in a manner that does not misrepresent or exaggerate the meaning or implications of the designation. The use of the designation may be accompanied by an accurate explanation of the requirements that have been met to earn the right to use the designation.

"CFA charterholders" are those individuals who have earned the right to use the CFA designation granted by CFA Institute. These people have satisfied certain requirements, including completion of the CFA Program and required years of acceptable work experience. Once granted the right to use the designation, individuals must also satisfy the CFA Institute membership requirements (see above) to maintain their right to use the designation.

If a CFA charterholder fails to meet any of the membership requirements, he or she forfeits the right to use the CFA designation. Until membership is reactivated, individuals must not present themselves to others as CFA charterholders. They may state, however, that they were charterholders in the past.

Given the growing popularity of social media, where individuals may anonymously express their opinions, pseudonyms or online profile names created to hide a member's identity should not be tagged with the CFA designation.

Use of the CFA designation by a CFA charterholder is governed by the terms and conditions of the annual Professional Conduct Statement Agreement, entered into between CFA Institute and its membership prior to commencement of use of the CFA designation and reaffirmed annually.

Referring to Candidacy in the CFA Program

Candidates in the CFA Program may refer to their participation in the CFA Program, but such references must clearly state that an individual is a *candidate* in the CFA Program and must not imply that the candidate has achieved any type of partial designation. A person is a candidate in the CFA Program if

- the person's application for registration in the CFA Program has been accepted by CFA Institute, as evidenced by issuance of a notice of acceptance, and the person is enrolled to sit for a specified examination or
- the registered person has sat for a specified examination but exam results have not yet been received.

If an individual is registered for the CFA Program but declines to sit for an exam or otherwise does not meet the definition of a candidate as described in the CFA Institute Bylaws, then that individual is no longer considered an active candidate. Once the person is enrolled to sit for a future examination, his or her CFA Program candidacy resumes.

CFA Program candidates must never state or imply that they have a partial designation as a result of passing one or more levels or cite an expected completion date of any level of the CFA Program. Final award of the charter is subject to meeting the CFA Program requirements and approval by the CFA Institute Board of Governors.

If a candidate passes each level of the exam in consecutive years and wants to state that he or she did so, that is not a violation of Standard VII(B) because it is a statement of fact. If the candidate then goes on to claim or imply superior ability by obtaining the designation in only three years, however, he or she is in violation of Standard VII(B).

Exhibit 3 provides examples of proper and improper references to the CFA designation.

Exhibit 3 Proper and Improper References to the CFA Designation

Proper References	Improper References
"Completion of the CFA Program has enhanced my portfolio management skills."	"CFA charterholders achieve better performance results."
"John Smith passed all three CFA Program examinations in three consecutive years."	"John Smith is among the elite, having passed all three CFA examinations in three consecutive attempts."
"The CFA designation is globally recognized and attests to a charterholder's success in a rigorous and comprehensive study program in the field of investment management and research analysis."	"As a CFA charterholder, I am the most qualified to manage client investments."
"The credibility that the CFA designation affords and the skills the CFA Program cultivates are key assets for my future career development."	"As a CFA charterholder, Jane White provides the best value in trade execution."
"I enrolled in the CFA Program to obtain the highest set of credentials in the global investment management industry."	"Enrolling as a candidate in the CFA Program ensures one of becoming better at valuing debt securities."
"I passed Level I of the CFA Program."	"CFA, Level II"
"I am a 2010 Level III candidate in the CFA Program."	"CFA, Expected 2011"
"I passed all three levels of the CFA Program and may be eligible for the CFA charter upon completion of the required work experience."	"CFA, Expected 2011" "John Smith, Charter Pending"

STANDARD VII(B): RECOMMENDED PROCEDURES

Misuse of a member's CFA designation or CFA candidacy or improper reference to it is common by those in a member's or candidate's firm who do not possess knowledge of the requirements of Standard VII(B). As an appropriate step to reduce this risk, members and candidates should disseminate written information about Standard VII(B) and the accompanying guidance to their firm's legal, compliance, public relations, and marketing departments (see www.cfainstitute.org).

For materials that refer to employees' affiliation with CFA Institute, members and candidates should encourage their firms to create templates that are approved by a central authority (such as the compliance department) as being consistent with Standard VII(B). This practice promotes consistency and accuracy in the firm of references to CFA Institute membership, the CFA designation, and CFA candidacy.

STANDARD VII(B): APPLICATION OF THE STANDARD

a demonstrate a thorough knowledge of the CFA Institute Code of Ethics and Standards of Professional Conduct by applying the Code and Standards to specific situations;

b recommend practices and procedures designed to prevent violations of the Code of Ethics and Standards of Professional Conduct.

Example 1 (Passing Exams in Consecutive Years):

An advertisement for AZ Investment Advisors states that all the firm's principals are CFA charterholders and all passed the three examinations on their first attempt. The advertisement prominently links this fact to the notion that AZ's mutual funds have achieved superior performance.

> *Comment*: AZ may state that all principals passed the three examinations on the first try as long as this statement is true, but it must not be linked to performance or imply superior ability. Implying that (1) CFA charterholders achieve better investment results and (2) those who pass the exams on the first try may be more successful than those who do not violates Standard VII(B).

Example 2 (Right to Use CFA Designation):

Five years after receiving his CFA charter, Louis Vasseur resigns his position as an investment analyst and spends the next two years traveling abroad. Because he is not actively engaged in the investment profession, he does not file a completed Professional Conduct Statement with CFA Institute and does not pay his CFA Institute membership dues. At the conclusion of his travels, Vasseur becomes a self-employed analyst accepting assignments as an independent contractor. Without reinstating his CFA Institute membership by filing his Professional Conduct Statement and paying his dues, he prints business cards that display "CFA" after his name.

> *Comment*: Vasseur has violated Standard VII(B) because his right to use the CFA designation was suspended when he failed to file his Professional Conduct Statement and stopped paying dues. Therefore, he no longer is able

to state or imply that he is an active CFA charterholder. When Vasseur files his Professional Conduct Statement, resumes paying CFA Institute dues to activate his membership, and completes the CFA Institute reinstatement procedures, he will be eligible to use the CFA designation.

Example 3 ("Retired" CFA Institute Membership Status):

After a 25-year career, James Simpson retires from his firm. Because he is not actively engaged in the investment profession, he does not file a completed Professional Conduct Statement with CFA Institute and does not pay his CFA Institute membership dues. Simpson designs a plain business card (without a corporate logo) to hand out to friends with his new contact details, and he continues to put "CFA" after his name.

> *Comment*: Simpson has violated Standard VII(B). Because he failed to file his Professional Conduct Statement and ceased paying dues, his membership has been suspended and he has given up the right to use the CFA designation. CFA Institute has procedures, however, for reclassifying a member and charterholder as "retired" and reducing the annual dues. If he wants to obtain retired status, he needs to file the appropriate paperwork with CFA Institute. When Simpson receives his notification from CFA Institute that his membership has been reclassified as retired and he resumes paying reduced dues, his membership will be reactivated and his right to use the CFA designation will be reinstated.

Example 4 (Stating Facts about CFA Designation and Program):

Rhonda Reese has been a CFA charterholder since 2000. In a conversation with a friend who is considering enrolling in the CFA Program, she states that she has learned a great deal from the CFA Program and that many firms require their employees to be CFA charterholders. She would recommend the CFA Program to anyone pursuing a career in investment management.

> *Comment*: Reese's comments comply with Standard VII(B). Her statements refer to facts: The CFA Program enhanced her knowledge, and many firms require the CFA designation for their investment professionals.

Example 5 (Order of Professional and Academic Designations):

Tatiana Prittima has earned both her CFA designation and a PhD in finance. She would like to cite both her accomplishments on her business card but is unsure of the proper method for doing so.

> *Comment*: The order of designations cited on such items as resumes and business cards is a matter of personal preference. Prittima is free to cite the CFA designation either before or after citing her PhD. Multiple designations must be separated by a comma.

Example 6 (Use of Fictitious Name):

Barry Glass is the lead quantitative analyst at CityCenter Hedge Fund. Glass is responsible for the development, maintenance, and enhancement of the proprietary models the fund uses to manage its investors' assets. Glass reads several high-level mathematical publications and blogs to stay informed on current developments.

One blog, run by Expert CFA, presents some intriguing research that may benefit one of CityCenter's current models. Glass is under pressure from firm executives to improve the model's predictive abilities, and he incorporates the factors discussed in the online research. The updated output recommends several new investments to the fund's portfolio managers.

> *Comment*: "Expert CFA" has violated Standard VII(B) by using the CFA designation inappropriately. As with any research report, authorship of online comments must include the charterholder's full name along with any reference to the CFA designation.
>
> See also Standard V(A), which Glass has violated for guidance on diligence and reasonable basis.

PRACTICE PROBLEMS

Unless otherwise stated in the question, all individuals in the following questions are CFA Institute members or candidates in the CFA Program and, therefore, are subject to the CFA Institute Code of Ethics and Standards of Professional Conduct.

1 Smith, a research analyst with a brokerage firm, decides to change his recommendation for the common stock of Green Company, Inc., from a "buy" to a "sell." He mails this change in investment advice to all the firm's clients on Wednesday. The day after the mailing, a client calls with a buy order for 500 shares of Green Company. In this circumstance, Smith should:

   A Accept the order.

   B Advise the customer of the change in recommendation before accepting the order.

   C Not accept the order because it is contrary to the firm's recommendation.

2 Which statement about a manager's use of client brokerage commissions violates the Code and Standards?

   A A client may direct a manager to use that client's brokerage commissions to purchase goods and services for that client.

   B Client brokerage commissions should be used to benefit the client and should be commensurate with the value of the brokerage and research services received.

   C Client brokerage commissions may be directed to pay for the investment manager's operating expenses.

3 Jamison is a junior research analyst with Howard & Howard, a brokerage and investment banking firm. Howard & Howard's mergers and acquisitions department has represented the Britland Company in all of its acquisitions for the past 20 years. Two of Howard & Howard's senior officers are directors of various Britland subsidiaries. Jamison has been asked to write a research report on Britland. What is the best course of action for her to follow?

   A Jamison may write the report but must refrain from expressing any opinions because of the special relationships between the two companies.

   B Jamison should not write the report because the two Howard & Howard officers serve as directors for subsidiaries of Britland.

   C Jamison may write the report if she discloses the special relationships with the company in the report.

4 Which of the following statements clearly *conflicts* with the recommended procedures for compliance presented in the CFA Institute *Standards of Practice Handbook*?

   A Firms should disclose to clients the personal investing policies and procedures established for their employees.

   B Prior approval must be obtained for the personal investment transactions of all employees.

   C For confidentiality reasons, personal transactions and holdings should not be reported to employers unless mandated by regulatory organizations.

5 Bronson provides investment advice to the board of trustees of a private university endowment fund. The trustees have provided Bronson with the fund's financial information, including planned expenditures. Bronson receives a

© 2014 CFA Institute. All rights reserved.

phone call on Friday afternoon from Murdock, a prominent alumnus, requesting that Bronson fax him comprehensive financial information about the fund. According to Murdock, he has a potential contributor but needs the information that day to close the deal and cannot contact any of the trustees. Based on the CFA Institute Standards, Bronson should:

A Send Murdock the information because disclosure would benefit the client.

B Not send Murdock the information to preserve confidentiality.

C Send Murdock the information, provided Bronson promptly notifies the trustees.

6 Willier is the research analyst responsible for following Company X. All the information he has accumulated and documented suggests that the outlook for the company's new products is poor, so the stock should be rated a weak "hold." During lunch, however, Willier overhears a financial analyst from another firm whom he respects offer opinions that conflict with Willier's forecasts and expectations. Upon returning to his office, Willier releases a strong "buy" recommendation to the public. Willier:

A Violated the Standards by failing to distinguish between facts and opinions in his recommendation.

B Violated the Standards because he did not have a reasonable and adequate basis for his recommendation.

C Was in full compliance with the Standards.

7 An investment management firm has been hired by ETV Corporation to work on an additional public offering for the company. The firm's brokerage unit now has a "sell" recommendation on ETV, but the head of the investment banking department has asked the head of the brokerage unit to change the recommendation from "sell" to "buy." According to the Standards, the head of the brokerage unit would be permitted to:

A Increase the recommendation by no more than one increment (in this case, to a "hold" recommendation).

B Place the company on a restricted list and give only factual information about the company.

C Assign a new analyst to decide if the stock deserves a higher rating.

8 Albert and Tye, who recently started their own investment advisory business, have registered to take the Level III CFA examination. Albert's business card reads, "Judy Albert, CFA Level II." Tye has not put anything about the CFA designation on his business card, but promotional material that he designed for the business describes the CFA requirements and indicates that Tye participates in the CFA Program and has completed Levels I and II. According to the Standards:

A Albert has violated the Standards, but Tye has not.

B Tye has violated the Standards, but Albert has not.

C Both Albert and Tye have violated the Standards.

9 Scott works for a regional brokerage firm. He estimates that Walkton Industries will increase its dividend by US$1.50 a share during the next year. He realizes that this increase is contingent on pending legislation that would, if enacted, give Walkton a substantial tax break. The US representative for Walkton's home district has told Scott that, although she is lobbying hard for the bill and prospects for its passage are favorable, concern of the US Congress over the federal deficit could cause the tax bill to be voted down. Walkton Industries has not made any statements about a change in dividend policy. Scott writes in his research report, "We expect Walkton's stock price to rise by at least US$8.00 a share by the end of the year because the dividend will increase by US$1.50 a share. Investors buying the stock at the current time should expect to realize a total return of at least 15% on the stock." According to the Standards:

- **A** Scott violated the Standards because he used material inside information.
- **B** Scott violated the Standards because he failed to separate opinion from fact.
- **C** Scott violated the Standards by basing his research on uncertain predictions of future government action.

10 Which one of the following actions will help to ensure the fair treatment of brokerage firm clients when a new investment recommendation is made?

- **A** Informing all people in the firm in advance that a recommendation is to be disseminated.
- **B** Distributing recommendations to institutional clients prior to individual accounts.
- **C** Minimizing the time between the decision and the dissemination of a recommendation.

11 The mosaic theory holds that an analyst:

- **A** Violates the Code and Standards if the analyst fails to have knowledge of and comply with applicable laws.
- **B** Can use material public information and nonmaterial nonpublic information in the analyst's analysis.
- **C** Should use all available and relevant information in support of an investment recommendation.

12 Jurgen is a portfolio manager. One of her firm's clients has told Jurgen that he will compensate her beyond the compensation provided by her firm on the basis of the capital appreciation of his portfolio each year. Jurgen should:

- **A** Turn down the additional compensation because it will result in conflicts with the interests of other clients' accounts.
- **B** Turn down the additional compensation because it will create undue pressure on her to achieve strong short-term performance.
- **C** Obtain permission from her employer prior to accepting the compensation arrangement.

13 One of the discretionary accounts managed by Farnsworth is the Jones Corporation employee profit-sharing plan. Jones, the company president, recently asked Farnsworth to vote the shares in the profit-sharing plan in favor of the slate of directors nominated by Jones Corporation and against the directors sponsored by a dissident stockholder group. Farnsworth does not want to lose this account because he directs all the account's trades to a brokerage firm that provides Farnsworth with useful information about tax-free investments. Although this information is not of value in managing the Jones Corporation account, it does help in managing several other accounts. The brokerage firm providing this information also offers the lowest commissions for trades and provides best execution. Farnsworth investigates the director issue, concludes that the management-nominated slate is better for the long-run performance of the company than the dissident group's slate, and votes accordingly. Farnsworth:

- **A** Violated the Standards in voting the shares in the manner requested by Jones but not in directing trades to the brokerage firm.
- **B** Did not violate the Standards in voting the shares in the manner requested by Jones or in directing trades to the brokerage firm.

C Violated the Standards in directing trades to the brokerage firm but not in voting the shares as requested by Jones.

14 Brown works for an investment counseling firm. Green, a new client of the firm, is meeting with Brown for the first time. Green used another counseling firm for financial advice for years, but she has switched her account to Brown's firm. After spending a few minutes getting acquainted, Brown explains to Green that she has discovered a highly undervalued stock that offers large potential gains. She recommends that Green purchase the stock. Brown has committed a violation of the Standards. What should she have done differently?

A Brown should have determined Green's needs, objectives, and tolerance for risk before making a recommendation of any type of security.

B Brown should have thoroughly explained the characteristics of the company to Green, including the characteristics of the industry in which the company operates.

C Brown should have explained her qualifications, including her education, training, and experience and the meaning of the CFA designation.

15 Grey recommends the purchase of a mutual fund that invests solely in long-term US Treasury bonds. He makes the following statements to his clients:

I. "The payment of the bonds is guaranteed by the US government; therefore, the default risk of the bonds is virtually zero."

II. "If you invest in the mutual fund, you will earn a 10% rate of return each year for the next several years based on historical performance of the market."

Did Grey's statements violate the CFA Institute Code and Standards?

A Neither statement violated the Code and Standards.

B Only statement I violated the Code and Standards.

C Only statement II violated the Code and Standards.

16 Anderb, a portfolio manager for XYZ Investment Management Company—a registered investment organization that advises investment firms and private accounts—was promoted to that position three years ago. Bates, her supervisor, is responsible for reviewing Anderb's portfolio account transactions and her required monthly reports of personal stock transactions. Anderb has been using Jonelli, a broker, almost exclusively for brokerage transactions for the portfolio account. For securities in which Jonelli's firm makes a market, Jonelli has been giving Anderb lower prices for personal purchases and higher prices for personal sales than Jonelli gives to Anderb's portfolio accounts and other investors. Anderb has been filing monthly reports with Bates only for those months in which she has no personal transactions, which is about every fourth month. Which of the following is *most likely* to be a violation of the Code and Standards?

A Anderb failed to disclose to her employer her personal transactions.

B Anderb owned the same securities as those of her clients.

C Bates allowed Anderb to use Jonelli as her broker for personal trades.

17 Which of the following is a correct statement of a member's or candidate's duty under the Code and Standards?

A In the absence of specific applicable law or other regulatory requirements, the Code and Standards govern the member's or candidate's actions.

- **B** A member or candidate is required to comply only with applicable local laws, rules, regulations, or customs, even though the Code and Standards may impose a higher degree of responsibility or a higher duty on the member or candidate.
- **C** A member or candidate who trades securities in a securities market where no applicable local laws or stock exchange rules regulate the use of material nonpublic information may take investment action based on material nonpublic information.

18 Ward is scheduled to visit the corporate headquarters of Evans Industries. Ward expects to use the information he obtains there to complete his research report on Evans stock. Ward learns that Evans plans to pay all of Ward's expenses for the trip, including costs of meals, hotel room, and air transportation. Which of the following actions would be the *best* course for Ward to take under the Code and Standards?

- **A** Accept the expense-paid trip and write an objective report.
- **B** Pay for all travel expenses, including costs of meals and incidental items.
- **C** Accept the expense-paid trip but disclose the value of the services accepted in the report.

19 Which of the following statements is *correct* under the Code and Standards?

- **A** CFA Institute members and candidates are prohibited from undertaking independent practice in competition with their employer.
- **B** Written consent from the employer is necessary to permit independent practice that could result in compensation or other benefits in competition with a member's or candidate's employer.
- **C** Members and candidates are prohibited from making arrangements or preparations to go into a competitive business before terminating their relationship with their employer.

20 Smith is a financial analyst with XYZ Brokerage Firm. She is preparing a purchase recommendation on JNI Corporation. Which of the following situations is *most likely* to represent a conflict of interest for Smith that would have to be disclosed?

- **A** Smith frequently purchases items produced by JNI.
- **B** XYZ holds for its own account a substantial common stock position in JNI.
- **C** Smith's brother-in-law is a supplier to JNI.

21 Michelieu tells a prospective client, "I may not have a long-term track record yet, but I'm sure that you'll be very pleased with my recommendations and service. In the three years that I've been in the business, my equity-oriented clients have averaged a total return of more than 26% a year." The statement is true, but Michelieu only has a few clients, and one of his clients took a large position in a penny stock (against Michelieu's advice) and realized a huge gain. This large return caused the average of all of Michelieu's clients to exceed 26% a year. Without this one investment, the average gain would have been 8% a year. Has Michelieu violated the Standards?

- **A** No, because Michelieu is not promising that he can earn a 26% return in the future.
- **B** No, because the statement is a true and accurate description of Michelieu's track record.
- **C** Yes, because the statement misrepresents Michelieu's track record.

22 An investment banking department of a brokerage firm often receives material nonpublic information that could have considerable value if used in advising the firm's brokerage clients. In order to conform to the Code and Standards, which one of the following is the best policy for the brokerage firm?

- **A** Permanently prohibit both "buy" and "sell" recommendations of the stocks of clients of the investment banking department.
- **B** Establish physical and informational barriers within the firm to prevent the exchange of information between the investment banking and brokerage operations.
- **C** Monitor the exchange of information between the investment banking department and the brokerage operation.

23 Stewart has been hired by Goodner Industries, Inc., to manage its pension fund. Stewart's duty of loyalty, prudence, and care is owed to:

- **A** The management of Goodner.
- **B** The participants and beneficiaries of Goodner's pension plan.
- **C** The shareholders of Goodner.

24 Which of the following statements is a stated purpose of disclosure in Standard VI(C)–Referral Fees?

- **A** Disclosure will allow the client to request discounted service fees.
- **B** Disclosure will help the client evaluate any possible partiality shown in the recommendation of services.
- **C** Disclosure means advising a prospective client about the referral arrangement once a formal client relationship has been established.

25 Rose, a portfolio manager for a local investment advisory firm, is planning to sell a portion of his personal investment portfolio to cover the costs of his child's academic tuition. Rose wants to sell a portion of his holdings in Household Products, but his firm recently upgraded the stock to "strong buy." Which of the following describes Rose's options under the Code and Standards?

- **A** Based on his firm's "buy" recommendation, Rose cannot sell the shares because he would be improperly prospering from the inflated recommendation.
- **B** Rose is free to sell his personal holdings once his firm is properly informed of his intentions.
- **C** Rose can sell his personal holdings but only when a client of the firm places an order to buy shares of Household.

26 A former hedge fund manager, Jackman, has decided to launch a new private wealth management firm. From his prior experiences, he believes the new firm needs to achieve US$1 million in assets under management in the first year. Jackman offers a $10,000 incentive to any adviser who joins his firm with the minimum of $200,000 in committed investments. Jackman places notice of the opening on several industry web portals and career search sites. Which of the following is *correct* according to the Code and Standards?

- **A** A member or candidate is eligible for the new position and incentive if he or she can arrange for enough current clients to switch to the new firm and if the member or candidate discloses the incentive fee.
- **B** A member or candidate may not accept employment with the new firm because Jackman's incentive offer violates the Code and Standards.

C A member or candidate is not eligible for the new position unless he or she is currently unemployed because soliciting the clients of the member's or candidate's current employer is prohibited.

27 Carter works for Invest Today, a local asset management firm. A broker that provides Carter with proprietary research through client brokerage arrangements is offering a new trading service. The broker is offering low-fee, execution-only trades to complement its traditional full-service, execution-and-research trades. To entice Carter and other asset managers to send additional business its way, the broker will apply the commissions paid on the new service toward satisfying the brokerage commitment of the prior full-service arrangements. Carter has always been satisfied with the execution provided on the full-service trades, and the new low-fee trades are comparable to the fees of other brokers currently used for the accounts that prohibit soft dollar arrangements.

A Carter can trade for his accounts that prohibit soft dollar arrangements under the new low-fee trading scheme.

B Carter cannot use the new trading scheme because the commissions are prohibited by the soft dollar restrictions of the accounts.

C Carter should trade only through the new low-fee scheme and should increase his trading volume to meet his required commission commitment.

28 Stafford is a portfolio manager for a specialized real estate mutual fund. Her firm clearly describes in the fund's prospectus its soft dollar policies. Stafford decides that entering the CFA Program will enhance her investment decision-making skill and decides to use the fund's soft dollar account to pay the registration and exam fees for the CFA Program. Which of the following statements is *most likely* correct?

A Stafford did not violate the Code and Standards because the prospectus informed investors of the fund's soft dollar policies.

B Stafford violated the Code and Standards because improving her investment skills is not a reasonable use of the soft dollar account.

C Stafford violated the Code and Standards because the CFA Program does not meet the definition of research allowed to be purchased with brokerage commissions.

29 Long has been asked to be the keynote speaker at an upcoming investment conference. The event is being hosted by one of the third-party investment managers currently used by his pension fund. The manager offers to cover all conference and travel costs for Long and make the conference registrations free for three additional members of his investment management team. To ensure that the conference obtains the best speakers, the host firm has arranged for an exclusive golf outing for the day following the conference on a local championship-caliber course. Which of the following is *least likely* to violate Standard I(B)?

A Long may accept only the offer to have his conference-related expenses paid by the host firm.

B Long may accept the offer to have his conference-related expenses paid and may attend the exclusive golf outing at the expense of the hosting firm.

C Long may accept the entire package of incentives offered to speak at this conference.

30 Andrews, a private wealth manager, is conducting interviews for a new research analyst for his firm. One of the candidates is Wright, an analyst with a local investment bank. During the interview, while Wright is describing his analytical skills, he mentions a current merger in which his firm is acting as the adviser.

Andrews has heard rumors of a possible merger between the two companies, but no releases have been made by the companies concerned. Which of the following actions by Andrews is *least likely* a violation of the Code and Standards?

- **A** Waiting until the next day before trading on the information to allow time for it to become public.
- **B** Notifying all investment managers in his firm of the new information so none of their clients are disadvantaged.
- **C** Placing the securities mentioned as part of the merger on the firm's restricted trading list.

31 Pietro, president of Local Bank, has hired the bank's market maker, Vogt, to seek a merger partner. Local is currently listed on a stock exchange and has not reported that it is seeking strategic alternatives. Vogt has discussed the possibility of a merger with several firms, but they have all decided to wait until after the next period's financial data are available. The potential buyers believe the results will be worse than the results of prior periods and will allow them to pay less for Local Bank.

Pietro wants to increase the likelihood of structuring a merger deal quickly. Which of the following actions would *most likely* be a violation of the Code and Standards?

- **A** Pietro could instruct Local Bank to issue a press release announcing that it has retained Vogt to find a merger partner.
- **B** Pietro could place a buy order for 2,000 shares (or four times the average weekly volume) through Vogt for his personal account.
- **C** After confirming with Local's chief financial officer, Pietro could instruct Local to issue a press release reaffirming the firm's prior announced earnings guidance for the full fiscal year.

32 ABC Investment Management acquires a new, very large account with two concentrated positions. The firm's current policy is to add new accounts for the purpose of performance calculation after the first full month of management. Cupp is responsible for calculating the firm's performance returns. Before the end of the initial month, Cupp notices that one of the significant holdings of the new accounts is acquired by another company, causing the value of the investment to double. Because of this holding, Cupp decides to account for the new portfolio as of the date of transfer, thereby allowing ABC Investment to reap the positive impact of that month's portfolio return.

- **A** Cupp did not violate the Code and Standards because the GIPS standards allow composites to be updated on the date of large external cash flows.
- **B** Cupp did not violate the Code and Standards because companies are allowed to determine when to incorporate new accounts into their composite calculation.
- **C** Cupp violated the Code and Standards because the inclusion of the new account produces an inaccurate calculation of the monthly results according to the firm's stated policies.

33 Cannan has been working from home on weekends and occasionally saves correspondence with clients and completed work on her home computer. Because of worsening market conditions, Cannan is one of several employees released by her firm. While Cannan is looking for a new job, she uses the files she saved at home to request letters of recommendation from former clients. She also provides to prospective clients some of the reports as examples of her abilities.

A Cannan violated the Code and Standards because she did not receive permission from her former employer to keep or use the files after her employment ended.

B Cannan did not violate the Code and Standards because the files were created and saved on her own time and computer.

C Cannan violated the Code and Standards because she is prohibited from saving files on her home computer.

34 Quinn sat for the Level III CFA exam this past weekend. He updates his resume with the following statement: "In finishing the CFA Program, I improved my skills related to researching investments and managing portfolios. I will be eligible for the CFA charter upon completion of the required work experience."

A Quinn violated the Code and Standards by claiming he improved his skills through the CFA Program.

B Quinn violated the Code and Standards by incorrectly stating that he is eligible for the CFA charter.

C Quinn did not violate the Code and Standards with his resume update.

35 During a round of golf, Rodriguez, chief financial officer of Mega Retail, mentions to Hart, a local investment adviser and long-time personal friend, that Mega is having an exceptional sales quarter. Rodriguez expects the results to be almost 10% above the current estimates. The next day, Hart initiates the purchase of a large stake in the local exchange-traded retail fund for her personal account.

A Hart violated the Code and Standards by investing in the exchange-traded fund that included Mega Retail.

B Hart did not violate the Code and Standards because she did not invest directly in securities of Mega Retail.

C Rodriguez did not violate the Code and Standards because the comments made to Hart were not intended to solicit an investment in Mega Retail.

36 Park is very frustrated after taking her Level II exam. While she was studying for the exam, to supplement the curriculum provided, she ordered and used study material from a third-party provider. Park believes the additional material focused her attention on specific topic areas that were not tested while ignoring other areas. She posts the following statement on the provider's discussion board: "I am very dissatisfied with your firm's CFA Program Level II material. I found the exam extremely difficult and myself unprepared for specific questions after using your product. How could your service provide such limited instructional resources on the analysis of inventories and taxes when the exam had multiple questions about them? I will not recommend your products to other candidates."

A Park violated the Code and Standards by purchasing third-party review material.

B Park violated the Code and Standards by providing her opinion on the difficulty of the exam.

C Park violated the Code and Standards by providing specific information on topics tested on the exam.

37 Paper was recently terminated as one of a team of five managers of an equity fund. The fund had two value-focused managers and terminated one of them to reduce costs. In a letter sent to prospective employers, Paper presents, with written permission of the firm, the performance history of the fund to demonstrate his past success.

A Paper did not violate the Code and Standards.

B Paper violated the Code and Standards by claiming the performance of the entire fund as his own.

C Paper violated the Code and Standards by including the historical results of his prior employer.

38 Townsend was recently appointed to the board of directors of a youth golf program that is the local chapter of a national not-for-profit organization. The program is beginning a new fund-raising campaign to expand the number of annual scholarships it provides. Townsend believes many of her clients make annual donations to charity. The next week in her regular newsletter to all clients, she includes a small section discussing the fund-raising campaign and her position on the organization's board.

A Townsend did not violate the Code and Standards.

B Townsend violated the Code and Standards by soliciting donations from her clients through the newsletter.

C Townsend violated the Code and Standards by not getting approval of the organization before soliciting her clients.

The following information relates to Questions 39–44

Anne Boswin, CFA, is a senior fixed-income analyst at Greenfield Financial Corporation. Boswin develops financial models for predicting changes in bond prices. On the premise that bonds of firms targeted for leveraged buyouts (LBOs) often decline in value, Boswin develops a model to predict which firms are likely to be subject to LBOs.

Boswin works closely with another analyst, Robert Acertado, CFA. Acertado uses Boswin's model frequently to identify potential LBO targets for further research. Using the model and his extensive research skills, Acertado makes timely investment recommendations and develops a strong track record.

Based on this record, Acertado receives an employment offer from the asset management division of Smith & Garner Investments, Inc., a diversified financial services firm. With Boswin's consent, Acertado downloads the model before leaving Greenfield.

At Smith & Garner, Acertado presents the idea of predicting LBO targets as a way to identify bonds that might decline in value and thus be good sell recommendations. After Acertado walks his boss through the model, the supervisor comments, "I like your idea and your model, Robert. I can see that we made the right decision in hiring you."

Because Smith & Garner has both an Investment Banking (IB) and Asset Management (AM) division, Acertado's supervisor reminds him that he should not attempt to contact or engage in conversation with anyone from the Investment Banking division. The supervisor also directs him to eat in the East end of the company cafeteria. "The West end is reserved for the IB folks, and you may laugh at this, but we actually put up a wall between the two ends. If anyone were to accuse us of not having a firewall, we could actually point to it!" Robert's supervisor also tells him, "There should be absolutely no conversation about divisional business while in the hall and elevator that serves as a common access to the cafeteria for both divisions. We are very strict about this."

The following week, Acertado is riding alone in the elevator when it stops on an IB floor. As the doors begin to slide open, Acertado hears a voice whispering, "I am so pleased that we were able to put the financing together for Country Industries. I was concerned because the leverage will go to 80%—higher than our typical deal." As soon as the doors open enough to reveal that the elevator is occupied, all conversation stops.

Late that afternoon, Acertado uses the LBO model to measure the probability of Country Industries receiving an LBO offer. According to the model, the probability is 62%—slightly more than the 60% Acertado generally requires before conducting additional research. It is late in the afternoon and Acertado has little time to research the matter fully before the end of the trading day. He checks his inputs to the model. In the interest of time, Acertado immediately recommends selling Country Industries' senior bonds held in any long-only accounts. He also recommends establishing positions in derivatives contracts that will benefit from a decline in the value of Country Industries' bonds.

The next morning, after the firm has established the derivatives positions he recommended, Acertado calls Boswin. Knowing that his former associate will be preparing Greenfield's monthly newsletter, he tells her, "I ran Country Industries through your model and I think it is likely that they will receive an LBO offer." Acertado explains some of the inputs he used in the model. At the conclusion of the conversation Boswin responds, "You may be right. Country Industries sounds like a possible LBO candidate, and thus, a sell rating on their senior bonds would be in order. If I'm lucky, I can finish researching the issue in time to include the recommendation in the upcoming newsletter. Thanks. It was good talking with you, Robert."

After the conversation with Acertado, Boswin quickly runs Country Industries through the model. Based on her inputs, the model calculates that the probability of an LBO is 40%—not enough, in Boswin's opinion, to justify further research. She wonders if there is a discrepancy between her inputs and Acertado's. Pressed for time, Boswin resumes her work on the upcoming newsletter rather than investigating the matter.

Acertado soon begins searching the internet for information on companies that the model predicts have more than a 60% probability of an LBO offer. He scours blogs and company websites looking for signs of a potential offer. He uses evidence of rumored offers in developing sell recommendations on various corporations' bonds.

39 When downloading the model from Greenfield Financial Corporation, does Acertado violate any CFA Institute Standards of Practice and Professional Conduct?

A No.

B Yes, because he does not have written permission from Boswin.

C Yes, because he does not have permission from Greenfield Financial Corporation.

40 When using the model at Smith & Garner, Acertado is *least likely* to violate the Standard relating to:

A misrepresentation.

B loyalty to employer.

C material nonpublic information.

41 When making the recommendation regarding Country Industries, does Acertado violate any CFA Institute Standards?

A No.

B Yes, relating to diligence and reasonable basis.

C Yes, relating to material nonpublic information.

42 In his phone conversation with Boswin, Acertado *least likely* violates the CFA Institute Standard relating to:

A suitability.

B integrity of capital markets.

C preservation of confidentiality.

43 When analyzing the probability of an LBO of Country Industries, does Boswin violate any CFA Institute Standards?

A No.

B Yes, relating to independence and objectivity.

C Yes, relating to diligence and reasonable basis.

44 When searching blogs, does Acertado violate any CFA Institute Standards?

A No.

B Yes, because he misuses company resources.

C Yes, because he seeks inside information on the blogs.

The following information relates to Questions 45–49

Erik Brecksen, CFA, a portfolio manager at Apfelbaum Kapital, recently recruited Hans Grohl, a CFA candidate and recent MBA graduate from a top university with excellent quantitative analysis skills. Apfelbaum Kapital stresses "top-down" fundamental analysis and uses a team approach to investment management. The firm's investment professionals, all of whom are CFA charterholders or candidates, attend weekly investment committee meetings. At the meetings, analysts responsible for different industrial sectors present their research and recommendations. Following each presentation, the investment committee, consisting of senior portfolio managers, questions the analyst about the recommendation. If the majority of the committee agrees with the recommendation, the recommendation is approved and the stock is placed on a restricted list while the firm executes the necessary trades.

Apfelbaum considers its research proprietary. It is intended for the sole use of its investment professionals and is not distributed outside the firm. The names of all the investment personnel associated with the sector or investment class are listed on each research report regardless of their actual level of contribution to the report.

On Grohl's first day of work, Brecksen assigns him responsibility for a company that Brecksen covered previously. He provides Grohl with his past research including all of his files and reports. Brecksen instructs Grohl to report back when he has finished his research and is ready to submit his own research report on the company.

Grohl reads Brecksen's old reports before studying the financial statements of the company and its competitors. Taking advantage of his quantitative analysis skills, Grohl then conducts a detailed multi-factor analysis. Afterward, he produces a written buy recommendation using Brecksen's old research reports as a guide for format and submits a draft to Brecksen for review.

Brecksen reviews the work and indicates that he is not familiar with multi-factor analysis. He tells Grohl that he agrees with the buy recommendation, but instructs Grohl to omit the multi-factor analysis from the report. Grohl attempts to defend his research methodology, but is interrupted when Brecksen accepts a phone call. Grohl follows Brecksen's instructions and removes all mention of the multi-factor analysis from the final report. Brecksen presents the completed report at the weekly meeting with both his and Grohl's names listed on the document. After Brecksen's initial presentation, the committee turns to Grohl and asks about his research. Grohl takes the opportunity to mention the multi-factor analysis. Satisfied, the committee votes in favor of the recommendation and congratulates Grohl on his work.

Ottie Zardt, CFA, has worked as a real estate analyst for Apfelbaum for the past 18 months. A new independent rating service has determined that Zardt's recommendations have resulted in an excess return of 12% versus the industry's return of 2.7% for the past twelve months. After learning about the rating service, Zardt immediately updates the promotional material he is preparing for distribution at an upcoming industry conference. He includes a reference to the rating service and quotes its returns results and other information. Before distributing the material at the conference, he adds a footnote stating "Past performance is no guarantee of future success."

45 When preparing the initial draft for Brecksen's review, does Grohl violate any CFA Standards?

- **A** No.
- **B** Yes, because he used Brecksen's research reports without permission.
- **C** Yes, because he did not use reasonable judgment in identifying which factors were important to the analysis.

46 When instructing Grohl to eliminate the multi-factor analysis from the research report, does Brecksen violate any CFA Standards?

- **A** No.
- **B** Yes, relating to record retention.
- **C** Yes, relating to diligence and reasonable basis.

47 When removing the multi-factor analysis from his research report, does Grohl violate any CFA Standards?

- **A** No.
- **B** Yes, because he no longer has a reasonable basis for his recommendation.
- **C** Yes, because he is required to make full and fair disclosure of all relevant information.

48 When listing their names on the research report, do Brecksen and Grohl violate any CFA Standards?

- **A** No.
- **B** Yes, because Brecksen misrepresents his authorship.
- **C** Yes, because Grohl should dissociate from the report.

49 When distributing the material at the industry conference, does Zardt violate any CFA Standards?

- **A** No.
- **B** Yes, because Zardt does not verify the accuracy of the information.
- **C** Yes, because analysts cannot claim performance or promote the accuracy of their recommendations.

The following information relates to Questions 50–55

Samuel Telline, CFA, is a portfolio manager at Aiklin Investments with discretionary authority over all of his accounts. One of his clients, Alan Caper, Chief Executive Officer (CEO) of Ellipse Manufacturing, invites Telline to lunch.

At the restaurant, the CEO reveals the reason for the lunch. "As you know Reinhold Partners has made an unsolicited cash offer for all outstanding shares of Ellipse Manufacturing. Reinhold has made it clear that I will not be CEO if they are successful. I can assure you that our shareholders will be better off in the long term if I'm in charge." Caper then shows Telline his projections for a new plan designed to boost both sales and operating margins.

"I know that your firm is the trustee for our firm's Employee Stock Ownership Plan (ESOP). I hope that the trustee will vote in the best interest of our shareholders—and that would be a vote against the takeover offer."

After looking through Caper's business plans, Telline says, "This plan looks good. I will recommend that the trustee vote against the offer."

Caper responds, "I remember my friend Karen Leighton telling me that the Leighton Family's Trust is managed by your firm. Perhaps the trustee could vote those shares against the acquisition as well. Karen Leighton is a close friend. I am sure that she would agree."

Telline responds, "The Family Trust is no longer managed by Aiklin." He adds, "I understand that the Trust is very conservatively managed. I doubt it that it would have holdings in Ellipse Manufacturing." Telline does not mention that although the Family Trust has changed investment managers, Karen Leighton remains an important client at Aiklin with significant personal holdings in Ellipse.

After lunch, Telline meets with Sydney Brown, CFA, trustee of the Ellipse ESOP. He shows her Caper's plan for improvements. "I think the plan is a good one and Caper is one of the firm's most profitable accounts. We don't want to lose him." Brown agrees to analyze the plan. After thoroughly analyzing both the plan and the takeover offer, Brown concludes that the takeover offer is best for the shareholders in the ESOP and votes the plan's shares in favor of the takeover offer.

A few months later the acquisition of Ellipse by Reinhold Partners is completed. Caper again meets Telline for lunch. "I received a generous severance package and I'm counting on you to manage my money well for me. While we are on the subject, I would like to be more aggressive with my portfolio. With my severance package, I can take additional risk." Telline and Caper discuss his current financial situation, risk tolerance, and financial objectives throughout lunch. Telline agrees to adjust Caper's investment policy statement (IPS) to reflect his greater appetite for risk and his increased wealth.

Back at the office, Telline realizes that with the severance package, Caper is now his wealthiest client. He also realizes that Caper's increased appetite for risk gives him a risk profile similar to that of another client. He pulls a copy of the other client's investment policy statement (IPS) and reviews it quickly before realizing that the two clients have very different tax situations. Telline quickly revises Caper's IPS to reflect the changes in his financial situation. He uses the other client's IPS as a reference when revising the section relating to Caper's risk tolerance. He then files the revised IPS in Caper's file.

The following week, an Aiklin analyst issues a buy recommendation on a small technology company with a promising software product. Telline reads the report carefully and concludes it would be suitable under Caper's new IPS. Telline places an order for 10,000 shares in Caper's account and then calls Caper to discuss the stock in more detail. Telline does not purchase the stock for any other clients. Although the one client has the same risk profile as Caper, that client does not have cash available in his account and Telline determines that selling existing holdings does not make sense.

In a subsequent telephone conversation, Caper expresses his lingering anger over the takeover. "You didn't do enough to persuade Aiklin's clients to vote against the takeover. Maybe I should look for an investment manager who is more loyal." Telline tries to calm Caper but is unsuccessful. In an attempt to change the topic of conversation, Telline states, "The firm was just notified of our allocation of a long-awaited

IPO. Your account should receive a significant allocation. I would hate to see you lose out by moving your account." Caper seems mollified and concludes the phone call, "I look forward to a long-term relationship with you and your firm."

Aiklin distributes a copy of its firm policies regarding IPO allocations to all clients annually. According to the policy, Aiklin allocates IPO shares to each investment manager and each manager has responsibility for allocating shares to accounts for which the IPO is suitable. The statement also discloses that Aiklin offers different levels of service for different fees.

After carefully reviewing the proposed IPO and his client accounts, Telline determines that the IPO is suitable for 11 clients including Caper. Because the deal is oversubscribed, he receives only half of the shares he expected. Telline directs 50% of his allocation to Caper's account and divides the remaining 50% between the other ten accounts, each with a value equal to half of Caper's account.

50 When discussing the Leighton Family Trust, does Telline violate any CFA Institute Standards of Professional Conduct?

A No.

B Yes, relating to duties to clients.

C Yes, relating to misrepresentation.

51 When deciding how to vote the ESOP shares, does Brown violate any CFA Institute Standards?

A No.

B Yes, relating to loyalty, prudence, and care.

C Yes, relating to diligence and reasonable basis.

52 The Standard *least likely* to provide guidance for Telline when working with the clients' investment policy statements would be the Standard relating to:

A suitability.

B fair dealing.

C loyalty, prudence, and care.

53 Does Telline violate any CFA Institute Standards when he places the buy order for shares in the technology company for Caper's account?

A No.

B Yes, relating to fair dealing.

C Yes, relating to diligence and reasonable basis.

54 Is Aiklin's policy with respect to IPO allocations consistent with required and recommended CFA Institute Standards?

A Yes.

B No, because the IPO policy disadvantages certain clients.

C No, because the different levels of service disadvantage certain clients.

55 Does Telline violate any CFA Institute Standards in his allocation of IPO shares to Caper's account?

A No.

B Yes, because the IPO is not suitable for Caper.

C Yes, because he does not treat all his clients fairly.

The following information relates to Questions 56–61

Adam Craw, CFA, is chief executive officer (CEO) of Crawfood, a European private equity firm specializing in food retailers. The retail food industry has been consolidating during the past two years as private equity funds have closed numerous deals and taken many companies private.

Crawfood recently hired Lillian Voser, a CFA Level II candidate, as a controller. On Voser's first day of work, the head of personnel informs her that by signing the employment contract, Voser agrees to comply with the company's code of ethics and compliance manual. She hands Voser copies of the code and compliance manual without further comment. Voser spends the next hour reading both documents. An excerpt from the compliance manual appears in Exhibit 1.

Exhibit 1 Crawfood Company Compliance Manual Excerpts

1 Employees must not accept gifts, benefits, compensation, or consideration that competes with, or might reasonably be expected to create a conflict of interest with their employer's interest unless they obtain written consent from all parties involved.

2 Officers have responsibility for ensuring that their direct reports—that is, employees whom they directly supervise—adhere to applicable laws, rules, and regulations.

3 Employees in possession of material nonpublic information should make reasonable efforts to achieve public dissemination of the information if such actions would not breach a duty.

4 Employees shall not trade or cause others to trade in securities of food retailers that may be potential takeover targets of their employer.

When she enters her new office that afternoon, Voser finds a large gift basket sent by her sister. The card reads "Congratulations on your new position." The basket is filled with expensive high-quality food items from Greenhornfood—a local small, publicly-traded food retailer, which produces many delicatessen products under its own brand name.

During the next two weeks, Voser meets with all of Crawfood's upper management, including the CEO. In his office, Craw praises Voser's efforts to complete the CFA program. "The program is demanding, but it is worthwhile." Craw then explains his investment strategy for choosing Crawfood's acquisition targets. He points to a large map on the wall with multi-colored pins marking Crawfood's previous takeovers. The map shows acquisitions in all the major cities of Germany with one exception—the home of Crawfood headquarters. Craw remarks, "We are currently in talks for another purchase. Confidentiality prohibits me from discussing it any further, but you will hear more about it soon."

Introduced to Greenhornfood by her sister, Voser quickly becomes a loyal customer. She considers it the best food retailer in the vicinity and she frequently purchases its products.

The following week, the local newspaper features an article about Greenhornfood and its young founders. The article describes the company's loyal and growing customer base as well as its poor quarterly financial results. Voser notes that the stock has steadily declined during the past twelve months. She concludes that the company has an inexperienced management team, but its popular product line and loyal customer base make the company a potential acquisition target. Voser calls her sister and recommends that she purchase Greenhornfood shares because "it would be an attractive acquisition for a larger company." Based on Voser's recommendation, her sister buys €3,000 worth of shares.

During the following two weeks the stock price of Greenhornfood continues to decline. Voser's sister is uncertain of what she should do with her position. She seeks Voser's advice. Voser recommends that her sister wait another few days before making her decision and promises to analyze the situation in the meantime.

While walking by Craw's office the following day, Voser sees a document with Greenhornfood's distinctive logo and overhears the company's name through an open office door. That evening, Voser tells her sister, "with the price decline, the stock is even more attractive." She recommends that her sister increase her position. Based on her recommendation her sister buys an additional €3,000 worth of Greenhornfood shares.

One month later, Crawfood publicly announces the acquisition of Greenhornfood Company at a 20% premium to the previous day's closing price. Following the announcement, Voser's sister boasts about Voser's excellent recommendation and timing to her broker.

Regulatory authorities initiate an investigation into suspicious trading in Greenhornfood shares and options preceding the formal announcement of the acquisition. Craw receives a letter from regulatory authorities stating that he is the subject of a formal investigation into his professional conduct surrounding the acquisition. He learns from the compliance officer that Voser is also under investigation. The compliance officer provides no details and out of respect for Voser's privacy, Craw makes no inquiries.

The situation remains unchanged and the matter is still pending with regulatory authorities several months later when Craw receives his annual Professional Conduct Statement (PCS) from CFA Institute. He reviews the text asking "In the last two years, have you been . . . the subject of . . . any investigation . . . in which your professional conduct, in either a direct or supervisory capacity, was at issue?"

56 Are Excerpts 2 and 3 of Crawfood's compliance procedures consistent with the CFA Institute Standards of Professional Conduct?

- **A** Yes.
- **B** No, because Excerpt 2 applies only to officers and their direct reports.
- **C** No, because Excerpt 3 does not require employees to achieve public dissemination.

57 According to the CFA Institute Standards, must Voser obtain permission from her supervisor before accepting the Greenhornfood gift basket?

- **A** No.
- **B** Yes, because the value of the basket is higher than €50.
- **C** Yes, because consent is required by the company's compliance procedures.

58 When making her initial recommendation to purchase Greenhornfood company shares, Voser *most likely* violates the Standard relating to:

- **A** loyalty to employer.
- **B** integrity of capital markets.
- **C** diligence and reasonable basis.

59 When recommending the purchase of additional Greenhornfood company shares, Voser *least likely* violates the Standard relating to:

- **A** loyalty to employer.
- **B** integrity of capital markets.

C diligence and reasonable basis.

60 Does Craw violate any CFA Institute Standards?

A No.

B Yes, because he passes material nonpublic information to Voser.

C Yes, because he does not make reasonable efforts to prevent violations of applicable law.

61 According to the CFA Standards, Craw must disclose to CFA Institute the investigation into:

A his conduct.

B Voser's conduct.

C neither his conduct nor Voser's conduct.

SOLUTIONS

1 The correct answer is B. This question involves Standard III(B)–Fair Dealing. Smith disseminated a change in the stock recommendation to his clients but then received a request contrary to that recommendation from a client who probably had not yet received the recommendation. Prior to executing the order, Smith should take additional steps to ensure that the customer has received the change of recommendation. Answer A is incorrect because the client placed the order prior to receiving the recommendation and, therefore, does not have the benefit of Smith's most recent recommendation. Answer C is also incorrect; simply because the client request is contrary to the firm's recommendation does not mean a member can override a direct request by a client. After Smith contacts the client to ensure that the client has received the changed recommendation, if the client still wants to place a buy order for the shares, Smith is obligated to comply with the client's directive.

2 The correct answer is C. This question involves Standard III(A)–Loyalty, Prudence, and Care and the specific topic of soft dollars or soft commissions. Answer C is the correct choice because client brokerage commissions may not be directed to pay for the investment manager's operating expenses. Answer B describes how members and candidates should determine how to use brokerage commissions—that is, if the use is in the best interests of clients and is commensurate with the value of the services provided. Answer A describes a practice that is commonly referred to as "directed brokerage." Because brokerage is an asset of the client and is used to benefit the client, not the manager, such practice does not violate a duty of loyalty to the client. Members and candidates are obligated in all situations to disclose to clients their practices in the use of client brokerage commissions.

3 The correct answer is C. This question involves Standard VI(A)–Disclosure of Conflicts. The question establishes a conflict of interest in which an analyst, Jamison, is asked to write a research report on a company that is a client of the analyst's employer. In addition, two directors of the company are senior officers of Jamison's employer. Both facts establish that there are conflicts of interest that must be disclosed by Jamison in her research report. Answer B is incorrect because an analyst is not prevented from writing a report simply because of the special relationship the analyst's employer has with the company as long as that relationship is disclosed. Answer A is incorrect because whether or not Jamison expresses any opinions in the report is irrelevant to her duty to disclose a conflict of interest. Not expressing opinions does not relieve the analyst of the responsibility to disclose the special relationships between the two companies.

4 The correct answer is C. This question asks about compliance procedures relating to personal investments of members and candidates. The statement in answer C clearly conflicts with the recommended procedures in the *Standards of Practice Handbook*. Employers should compare personal transactions of employees with those of clients on a regular basis regardless of the existence of a requirement by any regulatory organization. Such comparisons ensure that employees' personal trades do not conflict with their duty to their clients, and the comparisons can be conducted in a confidential manner. The statement in answer A does not conflict with the procedures in the *Handbook*. Disclosure of such policies will give full information to clients regarding potential conflicts of interest on the part of those entrusted to manage their money. Answer B is incorrect because firms are encouraged to establish policies whereby employees clear their personal holdings and transactions with their employers.

5 The correct answer is B. This question relates to Standard III(A)–Loyalty, Prudence, and Care and Standard III(E)–Preservation of Confidentiality. In this case, the member manages funds of a private endowment. Clients, who are, in this case, the trustees of the fund, must place some trust in members and candidates. Bronson cannot disclose confidential financial information to anyone without the permission of the fund, regardless of whether the disclosure may benefit the fund. Therefore, answer A is incorrect. Answer C is incorrect because Bronson must notify the fund and obtain the fund's permission before publicizing the information.

6 The correct answer is B. This question relates to Standard V(A)–Diligence and Reasonable Basis. The opinion of another financial analyst is not an adequate basis for Willier's action in changing the recommendation. Answer C is thus incorrect. So is answer A because, although it is true that members and candidates must distinguish between facts and opinions in recommendations, the question does not illustrate a violation of that nature. If the opinion overheard by Willier had sparked him to conduct additional research and investigation that justified a change of opinion, then a changed recommendation would be appropriate.

7 The correct answer is B. This question relates to Standard I(B)–Independence and Objectivity. When asked to change a recommendation on a company stock to gain business for the firm, the head of the brokerage unit must refuse in order to maintain his independence and objectivity in making recommendations. He must not yield to pressure by the firm's investment banking department. To avoid the appearance of a conflict of interest, the firm should discontinue issuing recommendations about the company. Answer A is incorrect; changing the recommendation in any manner that is contrary to the analyst's opinion violates the duty to maintain independence and objectivity. Answer C is incorrect because merely assigning a new analyst to decide whether the stock deserves a higher rating will not address the conflict of interest.

8 The correct answer is A. Standard VII(B)–Reference to CFA Institute, the CFA Designation, and the CFA Program is the subject of this question. The reference on Albert's business card implies that there is a "CFA Level II" designation; Tye merely indicates in promotional material that he is participating in the CFA Program and has completed Levels I and II. Candidates may not imply that there is some sort of partial designation earned after passing a level of the CFA exam. Therefore, Albert has violated Standard VII(B). Candidates may communicate that they are participating in the CFA Program, however, and may state the levels that they have completed. Therefore, Tye has not violated Standard VII(B).

9 The correct answer is B. This question relates to Standard V(B)–Communication with Clients and Prospective Clients. Scott has issued a research report stating that he expects the price of Walkton Industries stock to rise by US$8 a share "because the dividend will increase" by US$1.50 per share. He has made this statement knowing that the dividend will increase only if Congress enacts certain legislation, an uncertain prospect. By stating that the dividend will increase, Scott failed to separate fact from opinion.

The information regarding passage of legislation is not material nonpublic information because it is conjecture, and the question does not state whether the US representative gave Scott her opinion on the passage of the legislation in confidence. She could have been offering this opinion to anyone who asked. Therefore, statement A is incorrect. It may be acceptable to base a recommendation, in part, on an expectation of future events, even though they may be uncertain. Therefore, answer C is incorrect.

10 The correct answer is C. This question, which relates to Standard III(B)–Fair Dealing, tests the knowledge of the procedures that will assist members and candidates in treating clients fairly when making investment recommendations. The step listed in C will help ensure the fair treatment of clients. Answer A may have negative effects on the fair treatment of clients. The more people who know about a pending change, the greater the chance that someone will inform some clients before the information's release. The firm should establish policies that limit the number of people who are aware in advance that a recommendation is to be disseminated. Answer B, distributing recommendations to institutional clients before distributing them to individual accounts, discriminates among clients on the basis of size and class of assets and is a violation of Standard III(B).

11 The correct answer is B. This question deals with Standard II(A)–Material Nonpublic Information. The mosaic theory states that an analyst may use material public information and nonmaterial nonpublic information in creating a larger picture than shown by any individual piece of information and the conclusions the analyst reaches become material only after the pieces are assembled. Answers A and C are accurate statements relating to the Code and Standards but do not describe the mosaic theory.

12 The correct answer is C. This question involves Standard IV(B)–Additional Compensation Arrangements. The arrangement described in the question—whereby Jurgen would be compensated beyond the compensation provided by her firm, on the basis of an account's performance—is not a violation of the Standards as long as Jurgen discloses the arrangement in writing to her employer and obtains permission from her employer prior to entering into the arrangement. Answers A and B are incorrect; although the private compensation arrangement could conflict with the interests of other clients and lead to short-term performance pressures, members and candidates may enter into such agreements as long as they have disclosed the arrangements to their employer and obtained permission for the arrangement from their employer.

13 The correct answer is B. This question relates to Standard III(A)–Loyalty, Prudence, and Care—specifically, a member's or candidate's responsibility for voting proxies and the use of client brokerage. According to the facts stated in the question, Farnsworth did not violate Standard III(A). Although the company president asked Farnsworth to vote the shares of the Jones Corporation profit-sharing plan a certain way, Farnsworth investigated the issue and concluded, independently, the best way to vote. Therefore, even though his decision coincided with the wishes of the company president, Farnsworth is not in violation of his responsibility to be loyal and to provide care to his clients. In this case, the participants and the beneficiaries of the profit-sharing plan are the clients, not the company's management. Had Farnsworth not investigated the issue or had he yielded to the president's wishes and voted for a slate of directors that he had determined was not in the best interest of the company, Farnsworth would have violated his responsibilities to the beneficiaries of the plan. In addition, because the brokerage firm provides the lowest commissions and best execution for securities transactions, Farnsworth has met his obligations to the client in using this brokerage firm. It does not matter that the brokerage firm also provides research information that is not useful for the account generating the commission because Farnsworth is not paying extra money of the client's for that information.

14 The correct answer is A. In this question, Brown is providing investment recommendations before making inquiries about the client's financial situation, investment experience, or investment objectives. Brown is thus violating

Standard III(C)–Suitability. Answers B and C provide examples of information members and candidates should discuss with their clients at the outset of the relationship, but these answers do not constitute a complete list of those factors. Answer A is the best answer.

15 The correct answer i*s* C. This question involves Standard I(C)–Misrepresentation. Statement I is a factual statement that discloses to clients and prospects accurate information about the terms of the investment instrument. Statement II, which guarantees a specific rate of return for a mutual fund, is an opinion stated as a fact and, therefore, violates Standard I(C). If statement II were rephrased to include a qualifying statement, such as "in my opinion, investors may earn . . ., " it would not be in violation of the Standards.

16 The correct answer is A. This question involves three of the Standards. Anderb, the portfolio manager, has been obtaining more favorable prices for her personal securities transactions than she gets for her clients, which is a breach of Standard III(A)–Loyalty, Prudence, and Care. In addition, she violated Standard I(D)–Misconduct by failing to adhere to company policy and by hiding her personal transactions from her firm. Anderb's supervisor, Bates, violated Standard IV(C)–Responsibilities of Supervisors; although the company had requirements for reporting personal trading, Bates failed to adequately enforce those requirements. Answer B does not represent a violation because Standard VI(B)–Priority of Transactions requires that personal trading in a security be conducted after the trading in that security of clients and the employer. The Code and Standards do not prohibit owning such investments, although firms may establish policies that limit the investment opportunities of members and candidates. Answer C does not represent a violation because the Code and Standards do not contain a prohibition against employees using the same broker for their personal accounts that they use for their client accounts. This arrangement should be disclosed to the employer so that the employer may determine whether a conflict of interest exists.

17 The correct answer is A because this question relates to Standard I(A)–Knowledge of the Law—specifically, global application of the Code and Standards. Members and candidates who practice in multiple jurisdictions may be subject to various securities laws and regulations. If applicable law is more strict than the requirements of the Code and Standards, members and candidates must adhere to applicable law; otherwise, members and candidates must adhere to the Code and Standards. Therefore, answer A is correct. Answer B is incorrect because members and candidates must adhere to the higher standard set by the Code and Standards if local applicable law is less strict. Answer C is incorrect because when no applicable law exists, members and candidates are required to adhere to the Code and Standards, and the Code and Standards prohibit the use of material nonpublic information.

18 The correct answer is B. The best course of action under Standard I(B)–Independence and Objectivity is to avoid a conflict of interest whenever possible. Therefore, for Ward to pay for all his expenses is the correct answer. Answer C details a course of action in which the conflict would be disclosed, but the solution is not as appropriate as avoiding the conflict of interest. Answer A would not be the best course because it would not remove the appearance of a conflict of interest; even though the report would not be affected by the reimbursement of expenses, it could appear to be.

19 The correct answer is B. Under Standard IV(A)–Loyalty, members and candidates may undertake independent practice that may result in compensation or other benefit in competition with their employer as long as they obtain consent from their employer. Answer C is not consistent with the Standards because the Standards allow members and candidates to make arrangements or preparations to go into competitive business as long as those arrangements do not interfere with their duty to their current employer. Answer A is not consistent with the Standards because the Standards do not include a complete prohibition against undertaking independent practice.

20 The correct answer is B. This question involves Standard VI(A)–Disclosure of Conflicts—specifically, the holdings of an analyst's employer in company stock. Answers A and C do not describe conflicts of interest that Smith would have to disclose. Answer A describes the use of a firm's products, which would not be a required disclosure. In answer C, the relationship between the analyst and the company through a relative is so tangential that it does not create a conflict of interest necessitating disclosure.

21 The correct answer is C. This question relates to Standard I(C)–Misrepresentation. Although Michelieu's statement about the total return of his clients' accounts on average may be technically true, it is misleading because the majority of the gain resulted from one client's large position taken against Michelieu's advice. Therefore, this statement misrepresents the investment performance the member is responsible for. He has not taken steps to present a fair, accurate, and complete presentation of performance. Answer B is thus incorrect. Answer A is incorrect because although Michelieu is not guaranteeing future results, his words are still a misrepresentation of his performance history.

22 The correct answer is B. The best policy to prevent violation of Standard II(A)–Material Nonpublic Information is the establishment of firewalls in a firm to prevent exchange of insider information. The physical and informational barrier of a firewall between the investment banking department and the brokerage operation prevents the investment banking department from providing information to analysts on the brokerage side who may be writing recommendations on a company stock. Prohibiting recommendations of the stock of companies that are clients of the investment banking department is an alternative, but answer A states that this prohibition would be permanent, which is not the best answer. Once an offering is complete and the material nonpublic information obtained by the investment banking department becomes public, resuming publishing recommendations on the stock is not a violation of the Code and Standards because the information of the investment banking department no longer gives the brokerage operation an advantage in writing the report. Answer C is incorrect because no exchange of information should be occurring between the investment banking department and the brokerage operation, so monitoring of such exchanges is not an effective compliance procedure for preventing the use of material nonpublic information.

23 The correct answer is B. Under Standard III(A)–Loyalty, Prudence, and Care, members and candidates who manage a company's pension fund owe these duties to the participants and beneficiaries of the pension plan, not the management of the company or the company's shareholders.

24 The correct answer is B. Answer B gives one of the two primary reasons listed in the *Handbook* for disclosing referral fees to clients under Standard VI(C)–Referral Fees. (The other is to allow clients and employers to evaluate the full cost of the services.) Answer A is incorrect because Standard VI(C) does not require members or candidates to discount their fees when they receive referral fees. Answer C is inconsistent with Standard VI(C) because disclosure of referral fees, to be effective, should be made to prospective clients before entering into a formal client relationship with them.

25 The correct answer is B. Standard VI(B)–Priority of Transactions does not limit transactions of company employees that differ from current recommendations as long as the sale does not disadvantage current clients. Thus, answer A is incorrect. Answer C is incorrect because the Standard does not require the matching of personal and client trades.

26 Answer C is correct. Standard IV(A)–Loyalty discusses activities permissible to members and candidates when they are leaving their current employer; soliciting clients is strictly prohibited. Thus, answer A is inconsistent with the Code and Standards even with the required disclosure. Answer B is incorrect because the offer does not directly violate the Code and Standards. There may be out-of-work members and candidates who can arrange the necessary commitments without violating the Code and Standards.

27 Answer A is correct. The question relates to Standard III(A)–Loyalty, Prudence, and Care. Carter believes the broker offers effective execution at a fee that is comparable with those of other brokers, so he is free to use the broker for all accounts. Answer B is incorrect because the accounts that prohibit soft dollar arrangements do not want to fund the purchase of research by Carter. The new trading scheme does not incur additional commissions from clients, so it would not go against the prohibitions. Answer C is incorrect because Carter should not incur unnecessary or excessive "churning" of the portfolios (excessive trading) for the purpose of meeting the brokerage commitments of soft dollar arrangements.

28 Answer C is correct. According to Standard III(A)–Loyalty, Prudence, and Care, the CFA Program would be considered a personal or firm expense and should not be paid for with the fund's brokerage commissions. Soft dollar accounts should be used only to purchase research services that directly assist the investment manager in the investment decision-making process, not to assist the management of the firm or to further education. Thus, answer A is incorrect. Answer B is incorrect because the reasonableness of how the money is used is not an issue; the issue is that educational expense is not research.

29 Answer A is correct. Standard I(B)–Independence and Objectivity emphasizes the need for members and candidates to maintain their independence and objectivity. Best practices dictate that firms adopt a strict policy not to accept compensation for travel arrangements. At times, however, accepting paid travel would not compromise one's independence and objectivity. Answers B and C are incorrect because the added benefits—free conference admission for additional staff members and an exclusive golf retreat for the speaker—could be viewed as inducements related to the firm's working arrangements and not solely related to the speaking engagement. Should Long wish to bring other team members or participate in the golf outing, he or his firm should be responsible for the associated fees.

30 Answer C is correct. The guidance to Standard II(A)–Material Nonpublic Information recommends adding securities to the firm's restricted list when the firm has or may have material nonpublic information. By adding these securities to this list, Andrews would uphold this standard. Because waiting until the next day will not ensure that news of the merger is made public, answer A is incorrect. Negotiations may take much longer between the two companies, and the merger may never happen. Andrews must wait until the information is disseminated to the market before he trades on that information. Answer B is incorrect because Andrews should not disclose the information to other managers; no trading is allowed on material nonpublic information.

31 Answer B is correct. Through placing a personal purchase order that is significantly greater than the average volume, Pietro is violating Standard IIB–Market Manipulation. He is attempting to manipulate an increase in the share price and thus bring a buyer to the negotiating table. The news of a possible merger and confirmation of the firm's earnings guidance may also have positive effects on the price of Local Bank, but Pietro's actions in instructing the release of the information does not represent a violation through market manipulation. Announcements of this nature are common and practical to keep investors informed. Thus, answers A and C are incorrect.

32 Answer C is correct. Cupp violated Standard III(D)–Performance Presentations when he deviated from the firm's stated policies solely to capture the gain from the holding being acquired. Answer A is incorrect because the firm does not claim GIPS compliance and the GIPS standards require external cash flows to be treated in a consistent manner with the firm's documented policies. Answer B is incorrect because the firm does not state that it is updating its composite policies. If such a change were to occur, all cash flows for the month would have to be reviewed to ensure their consistent treatment under the new policy.

33 Answer A is correct. According to Standard V(C)–Record Retention, Cannan needed the permission of her employer to maintain the files at home after her employment ended. Without that permission, she should have deleted the files. All files created as part of a member's or candidate's professional activity are the property of the firm, even those created outside normal work hours. Thus, answer B is incorrect. Answer C is incorrect because the Code and Standards do not prohibit using one's personal computer to complete work for one's employer.

34 Answer B is correct. According to Standard VII(B)–Reference to CFA Institute, the CFA Designation, and the CFA Program, Quinn cannot claim to have finished the CFA Program or be eligible for the CFA charter until he officially learns that he has passed the Level III exam. Until the results for the most recent exam are released, those who sat for the exam should continue to refer to themselves as "candidates." Thus, answer C is incorrect. Answer A is incorrect because members and candidates may discuss areas of practice in which they believe the CFA Program improved their personal skills.

35 Answer A is correct. Hart's decision to invest in the retail fund appears directly correlated with Rodriguez's statement about the successful quarter of Mega Retail and thus violates Standard II(A)–Material Nonpublic Information. Rodriguez's information would be considered material because it would influence the share price of Mega Retail and probably influence the price of the entire exchange-traded retail fund. Thus, answer B is incorrect. Answer C is also incorrect because Rodriguez shared information that was both material and nonpublic. Company officers regularly have such knowledge about their firms, which is not a violation. The sharing of such information, however, even in a conversation between friends, does violate Standard II(A).

36 Answer C is correct. Standard VII(A)–Conduct as Members and Candidates in the CFA Program prohibits providing information to candidates or the public that is considered confidential to the CFA Program. In revealing that questions related to the analysis of inventories and analysis of taxes were on the exam, Park has violated this standard. Answer B is incorrect because the guidance for the standard explicitly acknowledges that members and candidates are allowed to offer their opinions about the CFA Program. Answer A is incorrect because candidates are not prohibited from using outside resources.

37 Answer B is correct. Paper has violated Standard III(D)–Performance Presentation by not disclosing that he was part of a team of managers that achieved the results shown. If he had also included the return of the portion he directly managed, he would not have violated the standard. Thus, answer A is incorrect. Answer C is incorrect because Paper received written permission from his prior employer to include the results.

38 Answer A is correct. Townsend has not provided any information about her clients to the leaders or managers of the golf program; thus, she has not violated Standard III(E)–Preservation of Confidentiality. Providing contact information about her clients for a direct-mail solicitation would have been a violation. Answer B is incorrect because the notice in the newsletter does not violate Standard III(E). Answer C is incorrect because the golf program's fund-raising campaign had already begun, so discussing the opportunity to donate was appropriate.

39 C is correct. Boswin, as an employee, developed the model on behalf of Greenfield. Therefore, Greenfield, not Boswin, is the owner of the model. Acertado violates Standard IV(A) Duties to Employers: Loyalty when he downloads the model without proper written permission from Greenfield Financial. Acertado is misappropriating employer assets.

40 C is correct. Acertado is least likely to violate Standard II(A) regarding Material Nonpublic Information when using the model at Smith and Garner. Acertado likely violated Standard IV(A), Loyalty, when he used the model. The Standard prohibits members who leave an employer from taking records or files—such as the model—without the written permission of the employer. Acertado also likely violated Standard I(C)–Misrepresentation when he failed to correct his supervisor's impression that the investment idea and the model were Acertado's creation.

41 C is correct. Acertado violates Standard II(A)–Material Nonpublic Information. He has a reasonable belief that the conversation that he overhears is from a reliable source and would have a material impact on security prices. According to the CFA Standards, he must not act, nor cause others to act on the information. Acertado does not violate the Standard relating to Diligence and Reasonable Basis because he bases the recommendation on a reliable model and checks his inputs prior to making the recommendation.

42 A is correct. Acertado least likely violates Standard III(C), which relates to suitability during his phone conversation with Boswin. According to the Standard, members in an advisory relationship with a client must determine an investment's suitability within the context of the client's portfolio. The Standard also requires that members make reasonable inquiries into a client or prospective client's investment experience; risk and return objectives; and financial constraints prior to making investment recommendations. Boswin is neither a client nor a prospective client, thus Acertado is not bound by the Standard of Suitability during their conversation. Acertado is, however, in jeopardy of violating other Standards—specifically those relating to Integrity of Capital Markets and Preservation of Confidentiality by revealing material nonpublic information about a Smith & Garner client. According to Standard II(A), Acertado, who is in possession of material nonpublic information, must not act, nor cause others to act on the information. According to Standard III(E), members must keep information about current, former, and prospective clients confidential.

43 A is correct. Boswin uses her usual process in researching Country Industries. She is not in possession of material nonpublic information and she maintains her objectivity. Her use of the model provides a reasonable basis for the decision not to pursue additional research or make an investment recommendation regarding Country Industries.

44 A is correct. Blogs and company websites are in the public domain and thus do not constitute inside information. Acertado's use of blog sites to supplement his current research process is acceptable.

45 A is correct. Grohl exercised diligence, independence, and thoroughness in analyzing the company and its competitors. Brecksen provided his research reports for Grohl's use and using the reports as a guide was appropriate. Standard V(A) requires that members distinguish between fact and opinion in communicating investment recommendations to clients. The Standard does not apply to investment recommendations communicated to supervisors or internal investment committees.

46 A is correct. Brecksen does not consider the multi-factor analysis a critical component of the analysis or the resulting investment recommendation and thus, under Standards V(A) and (C), is not required to maintain a record of the analysis within the completed report.

Apfelbaum uses traditional "top-down" fundamental analysis in the investment process. The report followed the traditional format of previous reports on the same company. It contained a complete fundamental analysis and recommendation—indicating diligence and reasonable basis. The report also contained a multi-factor analysis—which is a quantitative analysis tool. If quantitative analysis were the basis of the investment recommendation, it would constitute a change in the general investment principles used by the firm. According to Standard V(B)–Communications with Clients and Prospective Clients, Brecksen and Grohl would be required to promptly disclose those changes to clients and prospective clients.

47 A is correct. Removing the multi-factor analysis from the research report does not constitute a violation. Grohl diligently prepared the internal document according to the firm's traditional format with a complete fundamental analysis and recommendation—indicating diligence and a reasonable basis for his recommendation. It would be wise for Grohl to retain records of the multi-factor analysis but he need not retain the analysis in the research report to comply with Standards V(A)–Diligence and Reasonable Basis or V(C)–Record Retention.

48 A is correct. According to Standard V(A)–Diligence and Reasonable Basis, research report conclusions or recommendations may represent the consensus of a group and not necessarily the views of the individual members listed. If the member believes that the consensus opinion has a reasonable basis, then he need not dissociate from the report.

49 B is correct. Zardt violated the Standard relating to Performance Presentation because he did not verify the accuracy of the return information before its distribution. According to Standard III(D), analysts may promote the success or accuracy of their recommendations, but they must make reasonable efforts to ensure that the information is fair, accurate, and complete. In addition to providing attribution, Zardt should take steps to ensure the accuracy of the data prior to distributing the material.

50 B is correct. Telline has a duty to preserve the confidentiality of current, former, and prospective clients. Telline violates Standard III(E)–Preservation of Confidentiality when he reveals information about the Leighton Family Trust.

51 A is correct. Brown conducts an independent and careful analysis of the plans' benefits for shareholders as well as the takeover offer. In doing so she puts the client's interests ahead of the firm's. Brown's actions are consistent with Standard III(A)–Loyalty, Prudence, and Care; Standard V(A)–Diligence and Reasonable Basis; and Standard III(B)–Fair Dealing.

52 B is correct. Telline is not likely to receive appropriate guidance on developing or revising investment policy statements from the Standard relating to Fair Dealing. Standard III(B) provides members with guidance on treating clients fairly when making investment recommendations, providing investment analysis, or taking investment action. Telline could obtain guidance from the Standards relating to Loyalty, Prudence, and Care and Suitability. Both Standard III(A) and (C) provide guidance for members in determining client objectives and the suitability of investments.

53 A is correct. Telline is careful to consider the investment's suitability for Caper's account. Telline's actions are consistent with CFA Institute Standards III(A)–Loyalty, Prudence, and Care and III(B)–Fair Dealing. Telline determines that the other client does not have the cash available in his account and selling existing holdings does not make sense.

54 B is correct. The firm violates Standard III(B)–Fair Dealing. Under Aiklin's policy, some clients for whom an IPO is suitable may not receive their pro-rata share of the issue. CFA Standards recommend that firms allocate IPOs on a pro-rata basis to clients, not to portfolio managers.

55 C is correct. Telline violates Standard III(B)–Fair Dealing by over-allocating shares to Caper. Telline carefully reviews both the proposed IPO and his client accounts to determine suitability. He fails to allocate the IPO shares on a pro-rata basis to all clients for whom the investment is suitable.

56 B is correct. Excerpt 2 is inconsistent with CFA Standards because it addresses only officers and only their direct reports, that is, employees whom they directly supervise. Standard IV (C) states that "any investment professionals who have employees subject to their control or influence" exercise supervisory responsibility. Excerpt 3 is consistent with CFA Standards. It is based on a quote from the *Standards of Practice Handbook* stating that "if a member or candidate determines that information is material, the member . . . should make reasonable efforts to achieve public dissemination." Members are not required to achieve public dissemination and those bound by a duty of loyalty or a duty to preserve confidentiality would refrain from doing so because it would breach their duty.

57 A is correct. According to Standard I(B)–Independence and Objectivity, members must use reasonable care and judgment to achieve and maintain independence and objectivity in their professional activities. Although it was sent to Voser's office, the gift basket is a private gift from Voser's sister and not likely to affect Voser's professional activities. According to Excerpt 4 of the Crawfood compliance manual and Standard IV(B)–Additional Compensation Arrangements, employees must obtain permission from their employer before accepting gifts, compensation, or other benefits that compete with, or might create a conflict of interest with, the employer's interests. The gift basket does not create a conflict or compete with the employer's interests.

58 A is correct. Voser most likely violated the Standard relating to loyalty to employer, Standard IV(A). While Voser used public information to develop the recommendation to purchase Greenhornfood shares, the company compliance guide states that she should not trade or cause others to trade in securities of companies that may be potential takeover targets. Voser's recommendation

caused her sister to trade in Greenhornfood, violating the company's compliance policies, and possibly harming her employer in its attempt to acquire Greenhornfood.

By advising others to invest in a food retailer that she considered an attractive acquisition target, Voser deprived her employer of the advantage of her skills and abilities and may have caused harm to her employer. Voser could have recommended Greenhornfood to Craw rather than her sister as an acquisition target. Although the sister's trade in Greenhornfood was small, a large trade might have moved the stock price and caused harm to Crawfood in terms of additional cost.

59 C is correct. Voser least likely violated the Standard relating to diligence and reasonable basis. Voser initially applied the mosaic theory and had a reasonable basis for the trade as required by Standard V(A). Eventually, she came into possession of material nonpublic information (corporate logo on a document, overheard conversation). According to Standard II(A), once in possession of material nonpublic information, she is prohibited from acting or causing others to act. Voser also violated her duty of loyalty to her employer, Standard IV(A), by encouraging others to trade in Greenhornfood and possibly harming Crawfood's attempts to acquire the smaller company at an attractive price.

60 C is correct. Craw did not adequately fulfill his responsibilities as a supervisor. As stated in the *Standards of Practice Handbook*, members and candidates with supervisory responsibility also must understand what constitutes an adequate compliance system for their firms and make reasonable efforts to see that appropriate compliance procedures are established, documented, communicated to covered personnel, and followed. "Adequate" procedures are those designed to meet industry standards, regulatory requirements, the requirements of the Code and Standards, and the circumstances of the firm. Once compliance procedures are established, the supervisor must also make reasonable efforts to ensure that the procedures are monitored and enforced. According to Standard IV(C)–Responsibilities of Supervisors, adequate compliance procedures require that once a violation is discovered, Craw conduct a thorough investigation to determine the scope of wrongdoing.

61 A is correct. As stated on page ix of the *Standards of Practice Handbook*, "Members and candidates must self disclose on the annual Professional Conduct Statement all matters that question their professional conduct, such as involvement in civil litigation, a criminal investigation, or being the subject of a written complaint." Standard VII(A)–Conduct as Members and Candidates in the CFA Program prohibits conduct that compromises the reputation of the CFA designation including misrepresenting information on the Professional Conduct Statement. Members are encouraged but not required to report violations of others. At a minimum, Craw should remind Voser of her duty to report the investigation.

READING

47

Application of the Code and Standards: Level II

LEARNING OUTCOMES

Mastery	The candidate should be able to:
☐	**a.** evaluate practices, policies, and conduct relative to the CFA Institute Code of Ethics and Standards of Professional Conduct;
☐	**b.** explain how the practices, policies, and conduct do or do not violate the CFA Institute Code of Ethics and Standards of Professional Conduct.

INTRODUCTION

1

This reading presents cases to illustrate how the CFA Institute Code of Ethics and Standards of Professional Conduct (Code and Standards) can be applied in situations requiring professional and ethical judgement. Exhibit 1 presents a useful framework to help guide individuals in their ethical decision-making process and application of the Code and Standards. By identifying where the Code and Standards might be relevant and considering actions and consequences within this framework, individuals can make more ethically sound decisions.

Although the framework's components do not need to be addressed in the sequence shown, a review of the outcome should conclude the process. This review provides insights for improved decision making in the future.

Exhibit 1 A Framework for Ethical Decision Making

- Identify: Relevant facts, stakeholders and duties owed, ethical principles, conflicts of interest
- Consider: Situational influences, additional guidance, alternative actions
- Decide and act
- Reflect: Was the outcome as anticipated? Why or why not?

© 2021 CFA Institute. All rights reserved.

This reading presents a number of scenarios involving individuals in private and institutional asset management. The first five cases focus on identifying whether violations of the Code and Standards occurred, with discussion and rationale as to why or why not a violation may have taken place. The last two cases focus on identifying violations of the Code and Standards, taking necessary corrective actions, and developing a policy statement to help prevent future violations by a firm's employees. As you read through these cases, consider how applying the framework might have helped each individual to make decisions.

2 SERENGETI ADVISORY SERVICES

a evaluate practices, policies, and conduct relative to the CFA Institute Code of Ethics and Standards of Professional Conduct

b explain how the practices, policies, and conduct do or do not violate the CFA Institute Code of Ethics and Standards of Professional Conduct

Serengeti Advisory Services (Serengeti) is an equity research firm based in Tanzania.[1] It was founded five years ago by three CFA charterholders: Fatima Bashar, Hasini Shah, and Amara Kariuki. Serengeti analysts conduct investment research on listed African companies and sell the research to institutional asset managers on an annual subscription basis. Their clients are predominantly asset managers based in North America and Europe. Each month, subscribers receive two or three company and/or industry research reports. In her role as CEO and head of research, Bashar has recently started a premium subscription service. For an additional fee, clients receive six additional research reports of their choosing per year. Because of the cost of the premium service, Bashar offers it only to clients that she believes can afford it. (Question 1)

Serengeti's research analysts provide investment recommendations for small- to mid-cap African companies. For companies not covered by Serengeti, Bashar has created an "approved list" of broker/dealers from across the continent whose research she believes meets Serengeti's standards. Bashar tells her clients, "We partner with these broker/dealers to provide research on companies not covered by our firm. In addition, these relationships provide Serengeti's analysts with insights and information from a wide variety of independent sources that they may use in their own research. Some of the procedures I have implemented to ensure proper use of third-party research include the following:

- **Procedure 1**: When a research report, ours or theirs, cites specific quotations as attributable to "leading analysts" and "leading experts," I require the analyst who wrote the report to name the specific references.
- **Procedure 2**: When a research report contains statistical estimates of forecasts prepared by others (such as economists) and identifies the source, I require the analyst who wrote the report to remove any qualifying statements or caveats that may have been used.
- **Procedure 3**: When a research report contains charts and graphs prepared by others, I require the analyst who wrote the report to cite their sources." (Question 2)

1 **Serenteti Advisory Services:** Renée K. Blasky, CFA, CIPM, and Michael G. McMillan, PhD, CFA. *Ethics Cases.* © 2020 CFA Institute. All rights reserved. Consistent with the 11th Edition of the *Standards of Practice Handbook*.

To be on Serengeti's approved list, broker/dealers must undergo a stringent due diligence process carried out by Bashar's team and comply with Procedures 1–3. If the firm passes the initial due diligence, it goes through a more formal process ending with approval by Serengeti's management committee. Thereafter, Serengeti conducts a follow-up due diligence process annually to ensure continuing compliance.

Over the years, Bashar has established the following referral arrangement with the firms on the approved list: Whenever Serengeti's analysts prepare a research report using information sourced from an approved broker/dealer, Bashar encourages the firm's clients to trade through that broker/dealer. Bashar believes this referral arrangement, which she discloses to clients and prospective clients, provides additional incentive for broker/dealers to share insights and information with her analysts at no cost. Because of Serengeti's large client base, broker/dealers seek to be placed on the firm's "approved list." (Question 3).

Bashar discusses this referral arrangement during a recent phone conversation with Jon Grant, CFA, of Grant Asset Management, one of Serengeti's largest clients.

> *Grant:* I have found the research partnerships that Serengeti has established with broker/dealers to be very beneficial. The due diligence that you conduct on the broker/dealers on your approved list saves me from investigating these firms on my own. As a result, I always follow the analysts' investment recommendations and execute my clients' trades through the broker/dealer that produced the research report or whose information was used in your research. I noticed that Olatunji Financial, one of the largest broker/dealers in Nigeria, is not on your approved list. Why is that?
>
> *Bashar:* Thank you so much for the compliment and your support. I am glad that Serengeti can provide this service, and I know the broker/dealers also appreciate your business. I have heard that Olatunji Financial conducts outstanding research on Nigerian-listed companies, but we have not had an opportunity to work with them. (Question 4)

A few days later, Bashar receives an email from Amope Olatunji, CFA, president and CEO of Olatunji Financial Services (OFS). In the email, Olatunji states, "Jon Grant has told me about the great equity research that your firm produces. I would like OFS to be added to Serengeti's approved list. To this end, OFS would like to invite you and two of your colleagues to our upcoming West Africa Investment Forum that we host at our headquarters in Lagos, Nigeria. We will pay all of your expenses related to attending the forum, including air travel and overnight hotel accommodations."

Bashar knows that attending this forum will be beneficial for Serengeti. First, it will give her an opportunity to meet and interview executives of publicly listed companies who are attending or whose offices are located in Lagos. Second, it will give Kariuki, the head of business development, the chance to meet with prospective clients. Third, it will give Shah, the chief compliance officer, an opportunity to start the due diligence process on OFS and follow up with the other approved Nigerian broker/dealers in the city. During a meeting with Kariuki and Shah to discuss how to proceed with Olatunji's invitation, Bashar states, "As I see it, there are three possible options we have:

Option 1	We allow OFS to pay for all of our expenses.
Option 2	We allow OFS to pay for our air travel and hotel accommodations, and we pay for our meals and entertainment.
Option 3	We allow OFS to pay for the forum's meals and entertainment, and we pay for our air travel and hotel accommodations." (Question 5)

At the forum, Bashar, Shah, and Kariuki separate to talk with the attendees. When Kariuki walks over to the recycling bin to dispose of her water bottle, she notices a document on top of the bin. Picking up the document, she realizes it is a listing of OFS clients with contact numbers, email addresses, and average trading volumes over the past three years. After taking a photo of the document with her phone, she places it back in the bin and proceeds to seek out those clients with the highest trading volumes. (Question 6)

The next morning, Bashar, Shah, and Olatunji meet to discuss the possibility of OFS being added to Serengeti's approved list. The following are excerpts from their meeting:

Olatunji: Bashar, I propose you open a discretionary personal account with OFS. I will use the cash that you deposit into the account to purchase a "model portfolio" based on the Nigerian equities our analysts are recommending. Any time our analysts change their recommendations or issue new ones, I will execute trades in your account when we do it for our other discretionary clients. I will also send you an instant message before I make the trade just to let you know, and then I will call you afterward to discuss the trade in more detail. With this arrangement, I can prove to you how well our analysts' recommendations perform, and you will be able to participate directly.

Bashar: That's a good idea. Could you send me portfolio performance data monthly as you do with your other clients? Please be aware that I will not disclose this model portfolio to Serengeti's clients until I see how well your analysts' recommendations perform. In the meantime, Shah will continue to conduct due diligence on OFS to determine whether it should be added to our approved list of broker/dealers. (Question 7).

The next day, Bashar opens a discretionary personal account with OFS. Within a few days, Olatunji is able to build the model portfolio in Bashar's account despite liquidity issues in the market. Shortly thereafter, Olatunji calls Bashar.

Olatunji: I have good news for you. At the end of next week, our team of analysts will be issuing a strong buy recommendation on Swann Bank (SWNB), a commercial bank that is headquartered in Lagos. Because its stock is thinly traded, I have already started building a position in SWNB in your portfolio. Three days before the research report is issued, I have instructed our traders to simultaneously buy and sell shares of SWNB to increase its liquidity in the market so it will be easier for our clients to purchase it for their portfolios. (Questions 8, 9, 10)

Bashar: Thank you for the information and update.

Later that same day, after Olatunji establishes a position in SWNB in Bashar's portfolio, Bashar calls Grant to inform him about Olatunji's upcoming recommendation. (Question 11)

Case Questions

1 Does Bashar *most likely* violate the CFA Institute Code and Standards with regard to the premium subscription service?

  A Yes

  B No, because she offers it only to clients who can afford it

  C No, because clients must pay an additional fee for this service

A is correct. Bashar violates Standard III(B): Fair Dealing because she offers the premium subscription service only to clients who she believes can afford it. According to Standard III(B), "Members and Candidates must deal fairly and objectively with all clients when providing investment analysis, making investment recommendations, taking investment action, or engaging in other professional activities." According to the guidance for Standard III(B), "members and candidates may differentiate their services to clients, but different levels of service must not disadvantage or negatively affect clients. In addition, the different service levels should be disclosed to clients and prospective clients and should be available to everyone (i.e., different service levels should not be offered selectively)."

B is incorrect. Bashar offers the premium service only to clients who she believes can afford it and does not make the service available to all clients. According to the guidance for Standard III(B), the premium subscription "should be disclosed to clients and prospective clients and should be available to everyone (i.e., different service levels should not be offered selectively)." Bashar's view on which clients can afford the premium service is irrelevant.

C is incorrect. Although clients pay an additional fee for the premium subscription service, Bashar offers this service only to clients who she believes can afford it. According to the guidance for Standard III(B), "members and candidates may provide more personal, specialized, or in-depth service to clients who are willing to pay for premium services through higher management fees or higher levels of brokerage. Members and candidates may differentiate their services to clients, but different levels of service must not advantage or negatively affect clients. In addition, different service levels should be disclosed to clients and prospective clients and should be available to everyone (i.e., different service levels should not be offered selectively)."

2 Which of Bashar's procedures regarding the use of third-party research *mostly likely* violates the CFA Institute Code and Standards?

A Procedure 1

B Procedure 2

C Procedure 3

B is correct. Bashar violated Standard I(C): Misrepresentation by asking the analyst who wrote the report to remove qualifying statements or caveats from reports that include the presentation of statistical estimates of forecasts prepared by others. According to the guidance for Standard I(C), when presenting statistical estimates of forecasts prepared by others, the source should be identified along with the qualifying statements or caveats that may have been included.

A is incorrect. According to the guidance for Standard I(C), to avoid plagiarism, Bashar does need to ensure that when reports cite specific quotations, the specific references are named instead of attributing them generally to "leading analysts" and "investment experts." The guidance further states that "members and candidates should disclose whether the research being presented to clients comes from another source—from either within or outside the member's or candidate's firm. This allows clients to understand who has the expertise for a research report and whether the work is being done by the analyst, other members of the firm, or an outside party."

C is incorrect. According to the guidance for Standard I(C), to avoid plagiarism, Bashar does need to ensure that when reports use charts or graphs prepared by others, the sources are cited.

3 Does Bashar's referral arrangement with the broker/dealers on the approved list *most likely* violate the CFA Institute Code and Standards?

A No

B Yes, because the broker/dealers may not provide best price and execution

C Yes, because she is placing Serengeti's interest before those of her clients

A is correct. There is nothing to indicate that Bashar's referral of clients to the broker/dealer producing the research report or whose information was used by Serengeti is a violation of the CFA Institute Code and Standards. Bashar discloses to clients the broker/dealer responsible for the report or information and suggests that clients trade through that broker/dealer. According to Standard VI(C): Referral Fees, "Members and Candidates must disclose to their employer, clients, and prospective clients, as appropriate, any compensation, consideration, or benefit received from or paid to others for the recommendation of products or services." There is no quid pro quo benefit being given to Serengeti or to the broker/dealers. In this case, Bashar is merely suggesting to the firm's clients that they trade through a particular broker/dealer because of the research or other information that the firm provided. The fact that Bashar thinks these suggestions provide additional incentive for broker/dealers to share insights and information with her analysts at no cost is irrelevant. It is the asset manager client that makes the final decision as to where to direct its trades. Therefore, Bashar has not violated Standard VI(C): Referral Fees.

B is incorrect. It is not Bashar's responsibility to determine whether the broker/dealer that she is referring her asset manager clients to provides best price and best execution. Instead, it is the duty of the asset managers to seek best price and best execution. Policies related to soft commissions as well as to best price and best execution relate to asset managers that have discretion over brokers executing transactions. Therefore, Bashar has not violated Standard III(A): Loyalty, Prudence, and Care.

C is incorrect. The referral arrangement does not place the interests of Serengeti above those of its clients. Bashar only suggests to her clients that they trade through the broker/dealer that provided the research report or other information to Serengeti. This action is not a violation of Standard III(A): Loyalty, Prudence, and Care.

4 With regard to Grant and Bashar's phone conversation about Serengeti's research partnerships, who *most likely* violated the CFA Institute Code and Standards?

- **A** Grant
- **B** Bashar
- **C** Both Bashar and Grant

A is correct. Grant violated Standard III(A): Loyalty, Prudence, and Care by solely relying on Bashar's suggestions about which broker/dealer to use to execute his clients' trades instead of seeking the broker/dealer that provides best price and best execution. Standard III(A) states, "Members and Candidates have a duty of loyalty to their clients and must act with reasonable care and exercise prudent judgment. Members and Candidates must act for the benefit of their clients and place their clients' interests before their employer's or their own interests." Grant has an obligation to deliver best price and best execution for every trade, which may entail splitting orders among several brokers if necessary to achieve full execution.

B is incorrect. Bashar did not violate the CFA Institute Code and Standards during the phone conversation with Grant about the research partnerships. She simply encouraged Grant to use the broker/dealers who produced the research report or whose information was used. According to the guidance for Standard III(A): Loyalty, Prudence, and Care, Grant is "obligated to seek 'best price' and 'best execution' and be assured by the client that the goods or services purchased from the brokerage will benefit the account beneficiaries. 'Best execution' refers to a trading process that seeks to maximize the value of the client's portfolio within the client's stated investment objectives and constraints."

C is incorrect. Only Grant violated the CFA Institute Code and Standards. Grant violated Standard III(A) by solely relying on Bashar's recommendation of a broker/dealer to execute his clients' trades instead of ensuring that the broker/dealer provides best price and best execution. Standard III(A) states, "Members and Candidates have a duty of loyalty to their clients and must act with reasonable care and exercise prudent judgment." Grant has an obligation to deliver best price and best execution for client trades, which may entail splitting orders among several brokers for full execution. Bashar does not violate the CFA Institute Code and Standards during the phone conversation with Grant because she discloses the referral arrangement that Serengeti has with the broker/dealers on its approved list.

5 To avoid violating the CFA Institute Code and Standards, which of Bashar's responses to Olatunji's forum invitation would be *most appropriate*?

A Option 1

B Option 2

C Option 3

C is correct. To prevent violating Standard I(B): Independence and Objectivity, Bashar's, Kariuki's, and Shah's air travel and overnight hotel accommodations should be paid for by Serengeti. Under Standard I(B), it would be permissible for OFS to pay for meals and entertainment provided at the forum. This action would not likely jeopardize their independence and objectivity when determining whether to add OFS to Serengeti's approved list of broker/dealers. Standard I(B) states, "Members and Candidates must use reasonable care and judgment to achieve and maintain independence and objectivity in their professional activities. Members and Candidates must not offer, solicit, or accept any gift, benefit, compensation, or consideration that reasonably could be expected to compromise their own or another's independence and objectivity." According to the guidance for Standard I(B), "To avoid the appearance of compromising their independence and objectivity, best practice dictates that members and candidates always use commercial transportation at their expense or at the expense of their firm rather than accept paid travel arrangements from an outside company."

A is incorrect. Allowing OFS to pay for all of their expenses could reasonably be expected to compromise the independence and objectivity of Bashar, Shah, and Kariuki when evaluating whether OFS should be included on Serengeti's approved list. Standard I(B): Independence and Objectivity states, "Members and Candidates must use reasonable care and judgment to achieve and maintain independence and objectivity in their professional activities. Members and Candidates must not offer, solicit, or accept any gift, benefit, compensation, or consideration that reasonably could be expected to compromise their own or another's independence and objectivity." According to the guidance for Standard I(B), "To avoid the appearance of compromising their independence and objectivity, best practice dictates that members and candidates always use commercial transportation at their expense or at the expense of their firm rather than accept paid travel arrangements from an outside company."

B is incorrect. Allowing OFS to pay for their air travel and overnight hotel accommodations could reasonably be expected to compromise their independence and objectivity when evaluating whether OFS should be included on Serengeti's approved list. Standard I(B): Independence and Objectivity requires "Members and Candidates to use reasonable care and judgment to achieve and maintain independence and objectivity in their professional activities. Members and Candidates must not offer, solicit, or accept any gift, benefit, compensation, or consideration that reasonably could be expected to compromise their own or another's independence and objectivity." Travel and accommodations are the largest costs associated with attending the event. According to the guidance for Standard I(B), "To avoid the appearance of compromising their independence and objectivity, best practice dictates that members and candidates always use commercial transportation at their expense or at the expense of their firm rather than accept paid travel arrangements from an outside company."

6 Did Kariuki *most likely* violate the CFA Institute Code and Standards by seeking out OFS clients on the list with the highest trading volumes at the forum?

A No

B Yes, with regard to Standard I(D): Misconduct

C Yes, with regard to Standard II(A): Material Nonpublic Information

B is correct. Kariuki violated Standard I(D): Misconduct by using OFS's confidential and proprietary information that she found on top of the recycling bin. Standard I(D) states, "Members and Candidates must not engage in any professional conduct involving dishonesty, fraud, or deceit or commit any act that reflects adversely on their professional reputation, integrity, or competence." Kariuki is effectively misappropriating confidential information that she knows does not belong to her and is not meant for public consumption, which is a form of dishonesty. The circumstances of her acquiring the information and the misconduct of the person who carelessly handled the information does not mitigate Kariuki's improper behavior in using the information after it is discovered. The person responsible for disposing of this confidential list likely violated Standard III(E): Preservation of Confidentiality by not ensuring the safe disposal of the list so it would not be found by others who did not have permission to access the information.

A is incorrect. Kariuki violated Standard I(D): Misconduct by using OFS's confidential and proprietary information that she found on top of the recycling bin. Standard I(D) states that "Members and Candidates must not engage in any professional conduct involving dishonesty, fraud, or deceit or commit any act that reflects adversely on their professional reputation, integrity, or competence." The person responsible for disposing of this confidential list likely violated Standard III(E): Preservation of Confidentiality by not ensuring the safe disposal of the list so it would not be found by others who did not have permission to access the information.

C is incorrect. Kariuki did not violate Standard II(A): Material Nonpublic Information because the confidential nonpublic information that she obtains would not have an impact on the price of a security or affect investment decision making. Instead, it is a list of OFS clients, their contact numbers, email addresses, and trading volumes. According to the guidance for Standard II(A), "Information is 'material' if its disclosure would probably have an impact on the price of a security or if reasonable investors would want to know the information before making an investment decision. In other words, information is material if it would significantly alter the total mix of information currently available about a security in such a way that the price of the security would be affected."

7 With respect to the creation and management of the model portfolio, have Bashar or Olatunji *most likely* violated the CFA Institute Code and Standards?

A Only Olatunji violated the CFA Institute Code and Standards.

B Both Bashar and Olatunji violated the CFA Institute Code and Standards.

C Neither Bashar nor Olatunji violated the CFA Institute Code and Standards.

C is correct. Neither Bashar nor Olatunji violated the CFA Institute Code and Standards.

- Bashar: It is not a violation of the CFA Institute Code and Standards for her to open a discretionary account at OFS and allow Olatunji to create and manage a model portfolio using his analysts' recommendations. It is also not a violation for her not to disclose this information to Serengeti's clients. These actions will not compromise her independence or objectivity [Standard I(B)] or create a conflict of interest [Standard VI(A)].

- Olatunji: It is not a violation for him to create and manage Bashar's model portfolio using his analysts' recommendations. Olatunji is not violating Standard III(B): Fair Dealing because he is executing trades in Bashar's account at the same time he executes trades in other client accounts. In addition, he is not violating Standard III(B) by instant messaging Bashar before he executes a trade in her account, by calling her afterward to provide more detail on the trade, or by providing her with monthly performance reports on her portfolio. There is no indication that these actions are any different from what Olatunji would do for any other client if they requested the same treatment.

A is incorrect. Olatunji did not violate the CFA Institute Code and Standards. It is not a violation for him to create and manage a portfolio model for Bashar based on his analysts' recommendations. Olatunji is not violating Standard III(B): Fair Dealing because he is executing trades in Bashar's account at the same time he executes trades in other client accounts. In addition, he is not violating Standard III(B) by instant messaging Bashar before he executes a trade in her account, by calling her afterward to provide more details on the recommendation, or by providing her with monthly performance reports on her portfolio. There is no indication that these actions are any different from what Olatunji would do for any other client, if they requested the same treatment.

B is incorrect. Neither Bashar nor Olatunji violated the CFA Institute Code and Standards.

- Bashar: It is not a violation for her to open a discretionary account at OFS and allow Olatunji to create and manage a model portfolio using his analysts' recommendations. In addition, it is also not a violation for her not to disclose this information to Serengeti's clients. These actions will not compromise her independence or objectivity [Standard I(B)] or create a conflict of interest [Standard VI(A)].
- Olatunji: It is not a violation for him to create and manage Bashar's model portfolio using his analysts' recommendations. Olatunji is not violating Standard III(B): Fair Dealing because he is executing trades in Bashar's account at the same time he executes trades in other client accounts. In addition, he is not violating Standard III(B) by instant messaging Bashar before he executes a trade in her account, by calling her afterward to provide more details on the trade, or by providing her with monthly performance reports on her portfolio. There is no indication that these actions are any different from what Olatunji would do for any other client if they requested the same treatment.

8 In telling Bashar about the upcoming recommendation on SWNB, Olatunji is *most likely* violating the CFA Institute Standard of Professional Conduct on:

A Fair Dealing.

B Preservation of Confidentiality.

C Communication with Clients and Prospective Clients.

A is correct. Olatunji violated Standard III(B): Fair Dealing by telling Bashar about the "strong buy" recommendation on SWNB before the research report is disseminated to all clients. According to the guidance for Standard III(B), the standard "addresses the manner in which investment recommendations or changes in prior recommendations are disseminated to clients. Each member or candidate is obligated to ensure that information is disseminated in such a way that all clients have a fair opportunity to act on every recommendation."

B is incorrect. Standard III(E): Preservation of Confidentiality addresses keeping information about current, former, and prospective clients confidential. According to the guidance for Standard III(E), the standard "requires that members and candidates preserve the confidentiality of information communicated to them by their clients, prospective clients, and former clients." Olatunji violated Standard III(B): Fair Dealing by telling Bashar about the strong buy recommendation before the research report is disseminated to all clients. Olatunji would need to communicate the strong buy recommendation to all clients at approximately the same time to meet Standard III(B).

C is incorrect. According to the guidance for Standard V(B): Communication with Clients and Prospective Clients, the standard "addresses member and candidate conduct with respect to communicating with clients. Developing and maintaining clear, frequent, and thorough communication practices is critical to providing high-quality financial services to clients." Olatunji violated Standard III(B): Fair Dealing by telling Bashar about the strong buy recommendation before the research report is disseminated to all clients.

9 In building a position in SWNB in Bashar's portfolio, who *most likely* violated the CFA Institute Code and Standards?

A Only Olatunji

B Only Bashar

C Both Bashar and Olatunji

C is correct. Both Olatunji and Bashar violated the CFA Institute Code and Standards.

- Olatunji violated Standard III(B): Fair Dealing by purchasing shares of SWNB in Bashar's model portfolio before the recommendation is disseminated to other clients. According to the guidance for Standard III(B), the standard "addresses the manner in which investment recommendations or changes to prior recommendations are disseminated to clients. Each member or candidate is obligated to ensure that the information is disseminated in such a manner that all clients have a fair opportunity to act on every recommendation." If Olatunji had disseminated the strong buy recommendation on SWNB to all clients at approximately the same time he disseminated it to Bashar, he would not be in violation of Standard III(B).
- Bashar violated Standard I(A): Knowledge of the Law. Bashar knows that Olatunji is building a position in SWNB in her portfolio before the research report on SWNB is issued. According to the guidance for Standard I(A), "members and candidates are responsible for violations in which they *knowingly* participate or assist." The guidance for Standard I(A) states that "if a member or candidate has reasonable grounds to believe that imminent or ongoing client or employer activities are illegal or unethical, the member or candidate must dissociate, or separate, from the activity." As head of Serengeti, Bashar needed to advise Olatunji that she rejects the advance purchase of SWNB, requiring an unwinding of the trade, and then communicate Olatunji's violations to the appropriate regulators.

A is incorrect. Bashar also violated the CFA Institute Standards of Professional Conduct. Bashar violated Standard I(A): Knowledge of the Law because she should know that Olatunji violated the Code and Standards by selectively disclosing the strong buy recommendation on SWNB to her and building a position in it before the report is issued to all clients. According to the guidance for Standard I(A), "Members and candidates are responsible for violations in which they *knowingly* participate or assist." The guidance for Standard I(A) states that "if a member or candidate has reasonable grounds to believe that imminent or ongoing client or employer activities are illegal or unethical, the member or candidate must dissociate, or separate, from the activity."

B is incorrect. Olatunji also violated the CFA Institute Standards of Professional Conduct. Olatunji violated Standard III(B): Fair Dealing by purchasing shares of SWNB in Bashar's model portfolio before the recommendation is disseminated to other clients. According to the guidance for Standard III(B), the standard "addresses the manner in which investment recommendations or changes to prior recommendations are disseminated to clients. Each member or candidate is obligated to ensure that the information is disseminated in such a manner that all clients have a fair opportunity to act on every recommendation."

10 In his instructions to his traders, Olatunji *most likely* violated all of the following CFA Institute Standards of Professional Conduct *except*:

A Market Manipulation.

B Priority of Transactions.

C Responsibilities of Supervisors.

B is correct. Olatunji did not violate Standard VI(B): Priority of Transactions. According to Standard VI(B), "Investment transactions for clients and employers must have priority over investment transactions in which a Member or Candidate is the beneficial owner." According to the guidance for the standard, it "is designed to prevent any potential conflict of interest or the appearance of a conflict of interest with respect to personal transactions." In this scenario, Olatunji did not trade for himself, beneficially or otherwise.

A is incorrect. Olatunji violated Standard II(B): Market Manipulation. By instructing his traders to simultaneously buy and sell shares of SWNB a few days before the research report is issued, Olatunji engaged in transaction-based manipulation. Olatunji asked his traders to engage in "wash trading," which is the simultaneous buying and selling of SWNB to create misleading and artificial activity in SWNB's stock.

C is incorrect. By instructing his traders to simultaneously buy and sell shares of SWNB (called "wash trading," which creates misleading and artificial activity in the stock) just before the research report is issued, Olatunji violated Standard II(B): Market Manipulation, as well as Standard IV(C): Responsibilities of Supervisors. According to Standard IV(C), as the traders' supervisor, Olatunji "must make reasonable efforts to ensure that anyone subject to their supervision or authority complies with applicable laws, rules, regulations, and the Code and Standards."

11 By informing Grant about the recommendation on SWNB, Bashar *most likely* violated all of the following CFA Institute Standards of Professional Conduct *except*:

A Fair Dealing.

B Priority of Transactions.

C Preservation of Confidentiality.

C is correct. Bashar did not violate Standard III(E): Preservation of Confidentiality, which addresses keeping information about current, former, and prospective clients confidential. According to the guidance for Standard III(E), the standard "requires that members and candidates preserve the confidentiality of information communicated to them by their clients, prospective clients, and former clients." Bashar violated Standard III(B): Fair Dealing by telling Grant about the strong buy recommendation before telling her other clients. In addition, Bashar violated Standard V(B): Priority of Transactions.

A is incorrect. Bashar told Grant about the new recommendation before telling her other clients; therefore, Bashar violated Standard III(B): Fair Dealing. According to the guidance for Standard III(B), the standard "addresses the manner in which investment recommendations or changes in prior recommendations are disseminated to clients. Each member or candidate is obligated to ensure that information is disseminated in such a manner that all clients have a fair opportunity to act on every recommendation."

B is incorrect. By waiting until the position in SWNB is established in her portfolio before telling Grant about the upcoming recommendation, Bashar violated Standard VI(B): Priority of Transactions. According to this standard, "Investment transactions for clients and employers must have priority over investment transactions in which a Member or Candidate is the beneficial owner."

3 BANCO LIBERTAD

a evaluate practices, policies, and conduct relative to the CFA Institute Code of Ethics and Standards of Professional Conduct

b explain how the practices, policies, and conduct do or do not violate the CFA Institute Code of Ethics and Standards of Professional Conduct

Banco Libertad (BL) is a private bank based in the country of Urutina.[2] Founded in 1957, BL provides investment management, securities research, real estate financing, and wealth management services to high-net-worth individuals.

Sofia Maduro, CFA, is the managing director of BL's investment management division. Prior to joining BL 15 months ago, she spent 20 years as a portfolio manager for a global wealth management firm that has a large operation in Urutina. This morning, Maduro is meeting with two of the portfolio managers that she supervises: Julio Ortiz, CFA, and Guadalupe Sanchez, CFA. Ortiz and Sanchez work as a team at BL's headquarters in Urutina, managing approximately USD5.2 billion in assets. All of Ortiz's clients reside in either Urutina or Chiladour, and all of Sanchez's clients reside in either Urutina or Panaguay. For compliance purposes, applicable law for them is the country where their clients reside.

Urutina ranks first among countries in the region based on its economic growth (GDP), prosperity, innovation, and infrastructure. It has fair and transparent capital markets and no personal income taxes.

Chiladour is a neighboring country with a developing and rapidly modernizing economy. Because it has a nascent capital market, its wealthy citizens prefer to invest their money in Urutina.

Panaguay shares a border with Urutina. It is a politically unstable country with an emerging economy that is highly dependent on commodities, such as petroleum and agricultural products. It has established capital markets with strict securities laws and regulations, but as result of the political instability, its wealthy citizens prefer to invest their money in Urutina.

Exhibit 2 summarizes each portfolio manager's client jurisdictions. Exhibit 3 contains a comparison of the securities regulations and laws in each country as well as BL's policies.

2 **Banco Libertad:** Barbara Mainzer, CFA, and Michael G. McMillan, PhD, CFA. *Ethics Cases.* © 2019 CFA Institute. All rights reserved. Consistent with the 11th Edition of the *Standards of Practice Handbook*.

Exhibit 2 Client Jurisdictions

	Urutina	Chiladour	Panaguay
J. Ortiz	Y	Y	
G. Sanchez	Y		Y

Exhibit 3 Securities Laws and Regulations and BL Policies

	Urutina	Chiladour	Panaguay	BL Policies
Use of social media to post investment information (performance reports, investment opinions, and recommendations)	Allowed	Allowed	Prohibited	No policy
Use of instant messaging for trade executions and confirmations	Prohibited	Allowed	Allowed	No policy
Disclosure of personal or confidential information to law enforcement officials	Allowed	Prohibited	Prohibited	No policy
Insider trading	Prohibited	Prohibited	Prohibited	Prohibited
Referral fees	No disclosure required	No disclosure required	No disclosure required	No disclosure required

After Maduro welcomes the two portfolio managers to her office and they are comfortably seated, the following conversation takes place.

Maduro: Thank you for taking the time out of your busy schedules to meet with me. It must be difficult to manage the portfolios of clients who live in different countries and in different time zones. I want to commend both of you on your use of technology to communicate with your clients. Your use of social media to post investment recommendations, performance reports, and investment opinions has set a great example for other managers at BL.

Ortiz: Thank you for the compliment and for allowing Guadalupe and I to use instant messaging (IM) platforms to manage our clients' portfolios. Now, all our clients have to do is IM us whenever they want to buy and sell securities, and we, in turn, IM them with their trade confirmations once the transactions are complete.

Maduro: On a more serious note, it has been brought to my attention that law enforcement officials from the governments of Chiladour and Panaguay have contacted both of you requesting confidential information on your clients. What are you doing about these inquiries?

Ortiz: I have a client who is the former Minister of Finance in Chiladour. Since establishing his own financial consulting business five years ago, he has deposited more than USD5 million with BL. Now he is under investigation for allegedly embezzling from the Chiladour Treasury. In response to the inquiry into his personal finances, I have given law enforcement officials from Chiladour all of the information they requested on this client.

Sanchez: I have a client who is a general in the Panaguay Army. His grandfather founded the country's second-largest petroleum processing facility. When the general opened his account last summer, he deposited approximately USD35 million and told me that he had inherited the money from his grandfather, who had recently passed away. According to law enforcement officials in Panaguay, this money allegedly came from bribes the general received from directing military contracts to specific corporations. Like Ortiz, I was so disgusted by the client's behavior that I gave the law enforcement officials all of the information they requested on this client. (Questions 1, 2, 3)

Maduro: Thank you for doing this. It is important that BL maintain good relations with law enforcement officials throughout the region. Speaking of good relations, I noticed that many of your clients have been referred by law firms in Chiladour and Panaguay. Do you have any referral arrangements with these firms?

Ortiz and Sanchez: Our referral arrangements are very informal, which is why we do not bother disclosing them to clients. Every year we purchase tickets for the principals of these firms to attend CONMEBOL COPA AMERICA, the men's international football tournament. In addition, we always take them out to dinner whenever we travel to Chiladour and Panaguay to meet with our clients or when the principals come to Urutina to meet with their clients. (Question 4)

Maduro: Do you have any other arrangements that I am not aware of?

Ortiz and Sanchez: Now that you mention it, yes we do. We have special arrangements with two of our largest clients, Juan Fabre and Caroline Zeissl. After our securities analysts post changes to their research recommendations on the BL client website, we immediately call these two to discuss these changes in detail and any impact on their portfolios. Clients can sign up for email or text alerts to notify them when analyst recommendations change, but we feel calling Fabre and Zeissl personally is important because of their asset size. In addition, whenever Fabre and Zeissl are interested in participating in an IPO, we invite them to the company roadshow presentations organized by the investment bank underwriting the issue. Because of capacity constraints, we cannot invite our other clients who are interested in these IPOs to the presentations, so we do not disclose the roadshow opportunity to them. Whether or not they receive an invitation, we make IPO opportunities available to all our clients for whom the investment is suitable. (Questions 5, 6)

Ortiz: I have known and worked with Fabre and his family for about 15 years. In any year that his portfolio outperforms its benchmark, he allows my family and I to use his beach house in Jamaica for a week of our choosing. This arrangement is acknowledged in writing by the client and documented as part of the client's investment policy statement (IPS). Three months ago, my family and I used his beach house because, for the first time since he made the offer, his portfolio outperformed its benchmark this past year.

Sanchez: I have known and worked with Zeissl and her family for about 10 years. The Zeissl family owns the Aneka, a luxury hotel and spa in the Bahamas. She has said that in any year that her portfolio outperforms its benchmark, my family and I can spend a week at the Aneka without charge. As Julio described, this arrangement is documented with the client. Her portfolio also outperformed its benchmark this past year, so my family and I spent a week at the Aneka this past summer.

Maduro: Not having reviewed your client files, I was unaware of these arrangements, so thank you for telling me about them. I do have another meeting shortly, so let's end for today. (Question 7)

Case Questions

1 Have Ortiz and/or Sanchez violated the CFA Institute Code and Standards in their use of social media?

- **A** Neither has violated Code and Standards.
- **B** Only Sanchez has violated Code and Standards.
- **C** Both Ortiz and Sanchez have violated Code and Standards.

B is correct. Sanchez has violated Standard I(A): Professionalism–Knowledge of the Law. According to the Guidance for Standard I(A): "Members and candidates who practice in multiple jurisdictions may be subject to varied securities laws and regulations. If applicable law is stricter than the requirements of the Code and Standards, members and candidates must adhere to applicable law; otherwise they must adhere to the Code and Standards." In this case, applicable law is determined by where the clients reside, which for Sanchez is Urutina and Panaguay. Applicable law in Panaguay is stricter than the Code and Standards because it prohibits the use of social media to post investment information (Exhibit 3), so Sanchez must follow the law in Panaguay. Sanchez is allowed to use social media in Urutina because applicable law is similar to the Code and Standards, which does not prohibit the use of social media in this way.

A is incorrect. Sanchez has violated Standard I(A) because the use of social media to post investment information is prohibited in Panaguay, where applicable law is stricter than the Code and Standards.

C is incorrect. The use of social media to post investment information is allowed in Chiladour and Urutina, and Ortiz has not violated the Code and Standards. Applicable laws in these countries are similar to the Code and Standards, which does not prohibit the use of social media in this way. Sanchez has violated Standard I(A) because the use of social media to post investment information is prohibited in Panaguay.

2 Has Maduro violated the CFA Institute Code and Standards by allowing Ortiz and Sanchez to use social media to post investment information?

- **A** No
- **B** Yes, by allowing Sanchez to use social media
- **C** Yes, by allowing both Ortiz and Sanchez to use social media

B is correct. Maduro has allowed, and in doing so approved of, Sanchez's social media use with her clients in Panaguay, despite its use for posting investment information being prohibited in Panaguay, where the applicable law is stricter than the Code and Standards. According to Standard IV(C): Duties to Employers–Responsibilities of Supervisors, "Members and Candidates must make reasonable efforts to ensure that anyone subject to their supervision or authority complies with the applicable laws, rules, regulations, and the Code and Standards."

In this instance, in her supervisory capacity and knowing that Sanchez invests on behalf of her clients in Panaguay, Maduro must ensure, through appropriate education, support, and training, that Sanchez is aware of and complies with applicable Panaguayan laws and regulations.

A is incorrect. Maduro has allowed, and in doing so approved of, Sanchez's use of social media with her clients in Panaguay, where it is prohibited for posting investment information (i.e., applicable law is stricter than the Code and Standards).

C is incorrect. Maduro has allowed, and in doing so approved of, Sanchez's use of social media with her clients in Panaguay, where it is prohibited for posting investment information (i.e., applicable law is stricter than the Code and Standards).

3 Have Ortiz and Sanchez violated the CFA Institute Code and Standards by using instant messaging and disclosing information on their clients to law enforcement officials?

A Yes, both are violations.

B It was only a violation to use instant messaging.

C It was only a violation to disclose information on their clients to law enforcement officials.

A is correct. Ortiz and Sanchez violated Standard I(A): Professionalism–Knowledge of the Law because they have used IM to communicate with their clients in Urutina where doing so is prohibited. According to the vignette, applicable law is the country where the clients reside. Both Ortiz and Sanchez have clients in Urutina, which prohibits the use of IM to communicate investment information. Because applicable law in Urutina is stricter than the Code and Standards (which allows the use of IM), Ortiz and Sanchez must follow applicable law. In addition, the disclosure of personal or confidential information is prohibited in Chiladour and Panaguay. Applicable law in these countries is stricter than the Code and Standards, which requires members and candidates to disclose client information when requested by law enforcement. In disclosing the confidential information, Ortiz and Sanchez violated applicable law, and in doing so violated Standard I(A): Professionalism–Knowledge of the Law. Guidance for this Standard states, "Members and candidates who practice in multiple jurisdictions may be subject to varied securities laws and regulations. If applicable law is stricter than the requirements of the Code and Standard, members and candidates must adhere to applicable law; otherwise they must adhere to the Code and Standards."

B is incorrect. Ortiz and Sanchez violated the Code and Standards by using IM to communicate with their clients and by disclosing information on their clients to law enforcement officials. Both are violations of applicable law and therefore violations of the Code and Standards.

C is incorrect. Ortiz and Sanchez violated the Code and Standards by disclosing information on their clients to law enforcement officials and by using IM to communicate with their clients. Both are violations of applicable law and therefore violations of the Code and Standards.

4 Have Ortiz and Sanchez violated the CFA Institute Code and Standards in their referral arrangement with lawyers in Chiladour and Panaguay?

A Yes

B No, because they disclosed the arrangement to Maduro

C No, because disclosure is not required in Urutina, Chiladour, and Panaguay

A is correct. Ortiz and Sanchez have violated Standard VI(C): Conflicts of Interest–Referral Fees by not disclosing to BL, clients, and prospective clients the benefit they receive and pay to the lawyers who recommend clients to them. According to Standard VI(C), "Members and Candidates must disclose to their employer, clients, and prospective clients, as appropriate, any compensation, consideration, or benefit received from or paid to others for the recommendation of products or services."

They have also violated Standard I(A): Professionalism–Knowledge of the Law. In this case, Ortiz and Sanchez must follow the Code and Standards because these are stricter than applicable country laws. Guidance for Standard I(A) states, "Members and candidates who practice in multiple jurisdictions may be subject to varied securities laws and regulations. If applicable law is stricter than the requirements of the Code and Standard, members and candidates must adhere to applicable law; otherwise they must adhere to the Code and Standards."

B is incorrect. Although Ortiz and Sanchez disclosed the fee arrangement to Maduro, they disclosed it after the fact, and they have not disclosed the arrangement to clients and prospective clients.

C is incorrect. Although disclosure is not required in Urutina, Chiladour, or Panaguay, Ortiz and Sanchez must follow the stricter of "applicable laws, rules, and regulations." In this case, the Code and Standards are stricter than applicable law because they require disclosure of referral arrangements.

5 Do Ortiz and Sanchez violate the CFA Institute Code and Standards by calling Fabre and Zeissl to discuss changes in research recommendations made by BL's analysts?

A No

B Yes, because they are not dealing with their other clients fairly

C Yes, because they are not treating their other clients with loyalty, prudence, and care

A is correct. Ortiz and Sanchez have not violated the Code and Standards. Ortiz and Sanchez call these clients immediately after BL analysts post changes to their research recommendations on the BL client website (i.e., they have been publicly posted or made publicly accessible). Because Fabre and Zeissl are the portfolio managers' largest clients, they receive additional personal service because they presumably pay higher fees than other clients or have a large amount of assets under management with the firm. Guidance for Standard III(B): Duties to Clients–Fair Dealing states, "Each member or candidate is obligated to ensure that information is disseminated in such a manner that all clients have a fair opportunity to act on every recommendation."

B is incorrect. Ortiz and Sanchez call Fabre and Zeissl after the changes in recommendations have been posted on BL's website (i.e., publicly posted or made publicly accessible). Therefore, they are not treating their other clients unfairly but rather are calling their largest clients to discuss the changes in more detail.

C is incorrect. By calling Fabre and Zeissl, their largest clients, after the analysts post their recommendation changes on BL's website (i.e., publicly posted or made publicly accessible), Ortiz and Sanchez are acting for the benefit of their clients and placing their clients' interests before BL's or their own interests.

6 Have Ortiz and Sanchez violated the CFA Institute Code and Standards by inviting Fabre and Zeissl to IPO roadshow presentations?

A No

B Yes, because they do not invite other clients to the roadshow

C Yes, because they do not disclose this opportunity to other clients

A is correct. Inviting clients to the IPO roadshow is not a violation of the Code and Standards. Fabre and Zeissl are their largest clients, and in addition, being invited to the IPO roadshow does not disadvantage other clients who are interested in the IPO. Therefore, they have not violated Standard III(B): Duties to Clients–Fair Dealing. The Guidance for Standard III(B) states, "Members and candidates may differentiate their services to clients, but different levels of service must not disadvantage or negatively affect clients." Attendance at the roadshow does not give Fabre and Zeissl any special advantage or priority over other clients in IPO participation.

B is incorrect. Not inviting to the roadshow other clients who are interested in the IPO does not disadvantage those clients.

C is incorrect. Ortiz and Sanchez are not obligated to disclose this opportunity to their other clients because it does not disadvantage those clients.

7 Have Ortiz and Sanchez violated the CFA Institute Code and Standards in not disclosing their respective arrangements with Fabre and Zeissl?

A Yes

B No, because Ortiz and Sanchez have documented the arrangements in the clients' IPS

C No, because these types of client arrangements do not have to be disclosed

A is correct. By not disclosing the arrangements to BL (Maduro) and obtaining her written consent, Ortiz and Sanchez are violating Standard IV(B): Duties to Employers–Additional Compensation Arrangements because the arrangements could result in them being partial to Fabre's and Zeissl's accounts at the expense of their other clients. Standard IV(B) states, "Members and Candidates must not accept gifts, benefits, compensation, or consideration that competes with or might reasonably be expected to create a conflict of interest with their employer's interest unless they obtain written consent from all parties involved."

B is incorrect. Although Ortiz and Sanchez obtained written acknowledgement from Fabre and Zeissl at the time each client made their offer, and included it as part of the IPS, they did not obtain their employer's written consent for the arrangement.

C is incorrect. According to Standard IV(B): Duties to Employers–Additional Compensation Arrangements, members and candidates are required to disclose these types of arrangements to their employer and obtain written consent from all parties. Ortiz and Sanchez did not disclose the arrangements to Maduro until after the arrangements had been accepted and used. In addition, they had only written consent from their clients.

4 QUANTHOUSE

a evaluate practices, policies, and conduct relative to the CFA Institute Code of Ethics and Standards of Professional Conduct

b explain how the practices, policies, and conduct do or do not violate the CFA Institute Code of Ethics and Standards of Professional Conduct

QuantHouse (QH) is a global investment firm that pioneered the use of quantitative techniques to implement investment strategies.[3] Three years ago, the firm hired Daniel Singh, PhD, CFA, a well-known finance professor. Singh created the Artificial Trading Model (ATM), a comprehensive model that captures and processes a substantial amount of publicly available information (company financial data, news, and industry information) and then makes investment decisions largely without human interaction. ATM has three components: an Alpha Model, a Risk Model, and an Optimizer. The Alpha Model evaluates public companies based on their earnings and valuation. The Risk Model identifies stock-specific risk and common factor risks (industry specific, country specific, and stock fundamental risks). The Optimizer takes the output from the Alpha and Risk Models, balances them against one other, and recommends an optimal client portfolio based on the client's chosen benchmark.

3 **QuantHouse:** Asjeet S. Lamba, PhD, CFA, and Michael G. McMillan, PhD, CFA. *Ethics Cases.* © 2018 CFA Institute. All rights reserved. Consistent with the 11th Edition of the *Standards of Practice Handbook*.

Singh uses the ATM model exclusively for QH's institutional clients and does not mention it when talking with his high-net-worth individual clients. Singh and his team of programmers update the Alpha and Risk components of the ATM on a quarterly basis. On an annual basis, consistent with QH's guidelines for all computer-based models, Singh reviews the Optimizer component and conducts extensive scenario tests with the overall model. (Questions 1, 2)

Recently, some of QH's institutional clients have been voicing concerns about their portfolios' underperformance. In particular, they have expressed dissatisfaction with the overexposure to certain industries in their portfolios, an element that is partly controlled by the ATM's ability to manage risk. In response to these complaints, QH's director of research and Singh's supervisor, Charlotte Ringfield, CFA, asks Singh to review the model. After doing so, Singh finds that the Optimizer is incorrectly reading the Risk Model's assessment of common risk factors and, as a result, is not weighting them appropriately.

Singh then meets with his team of analysts who helped create the model to determine the source of the error. They find that some of the Risk Model components are sending information to the Optimizer in decimal form while other components are sending information in percentages. This improper scaling has resulted in the Optimizer giving inappropriate weights to some of the common risk factors. After discovering the source of the error, Singh and his team meet with Ringfield to present their findings. Singh advocates that the error be fixed as soon as possible, but Ringfield disagrees and tells him to correct the error when the Risk Model is updated at the end of the quarter. She also asks Singh to temporarily disable the common risk factors in the Risk Model until the model is updated. Ringfield then asks Singh and his team not to mention the error to others and reminds them that they signed confidentiality and nondisclosure agreements when they were hired. She goes on to say that she and the investment committee will handle all disclosures to clients and senior management once the model is updated at the end of the quarter. (Questions 3, 4)

Two weeks later, after a very turbulent period in the financial markets, more clients complain about their portfolios' underperformance. Ringfield tells the portfolio managers about the error. When clients inquire about their portfolio's performance, Ringfield attributes the performance to market volatility and the functioning of the model's common risk factors. (Question 5)

Disturbed by the behavior of his colleagues and superiors, who have not yet revealed the error to clients, Singh decides to leave QH. He interviews with QH's largest competitor, Algos-R-Us (ARU). During his interview with ARU's hiring committee, Singh shows them a proprietary model that he has been developing for the past two years in his spare time (nights and weekends). His new model, which he calls StockStar, is based on years of academic research at his university, and Singh considers it to be his life's work. In addition to back-testing the model, Singh has used StockStar to manage his personal portfolio and the portfolios of his family in order to generate actual performance results. Impressed by Singh and his model, the hiring committee not only offers him a job but also offers to pay him a special licensing fee for the use of his StockStar model. Singh accepts the offer and returns to QH to tender his resignation. (Question 6)

On his first day at ARU, Singh presents the following information to the marketing department about the StockStar model's performance. This information will be incorporated into a new marketing brochure that will be mailed to current and prospective clients.

StockStar Model Performance: Actual and Back-Tested Returns*

	Model's Performance (%)	Benchmark Return (%)	Excess Return (%)
Back-tested returns			
2012	10.5	6.5	4.0
2013	−2.5	−6.0	3.5
2014	20.3	25.0	−4.7
2015	0.5	−6.8	7.3
Actual returns			
2016	8.5	2.2	6.3
2017	12.0	−0.5	12.5

* The benchmark return is based on the S&P/ASX 200 index. Note that past performance does not guarantee future results.

In the brochure, Singh states the following: "This model has been used to manage real portfolios over the past two years and has outperformed its benchmark in both years. In back-tests, the model has outperformed its benchmark in three out of four years." (Question 7)

Case Questions

1 Does Singh violate the CFA Institute Code and Standards by using the ATM model exclusively for QH's institutional clients?

A No

B Yes, because the model may be suitable for some non-institutional clients

C Yes, because he must at least mention the model when talking to high-net worth individuals

A is correct. It is not a violation of Standard III(B): Duties to Clients–Fair Dealing to use different investment models when working with different types of clients. The ATM model may be suitable only for institutional clients because of the size of their portfolios, among other factors. Standard III(B) requires members and candidates to treat all clients fairly when disseminating investment recommendations; making material changes to prior investment recommendations; or when taking investment action with regard to general purchases, new issues, or secondary offerings. Each client has unique needs, investment criteria, and investment objectives, so not all investment opportunities are suitable for all clients.

B is incorrect. It is not a violation of the Code and Standards to use different investment models when working with different types of clients, because some models may be suitable only for specific clients. There is no information in the case to indicate that the model is suitable for clients other than institutional investors.

C is incorrect. There is no duty under the Code and Standards to disclose all available investment models when working with different types of clients, because some models may be suitable only for specific clients. There is no information in the case to indicate that the model is suitable for clients other than institutional investors.

2 Prior to the performance concerns voiced by QH's institutional clients, did Singh violate the CFA Institute Code and Standards in the updating of the ATM model components?

A Yes

B No, because he updates the model's Alpha and Risk components on a quarterly basis

C No, because he followed the firm's guidelines and annually reviews the Optimizer and conducts scenario testing on the overall model

A is correct. Singh violated Standard V(A): Investment Analysis, Recommendations, and Actions–Diligence and Reasonable Basis. Members and candidates must understand the statistical significance of the model results they recommend and must be able to explain these results to clients. Singh did not take adequate care to ensure a thorough review of the model was taking place with appropriate frequency. Although the Alpha and Risk components are updated quarterly, he reviews the Optimizer, which links the two prior components, and the overall model itself on only an annual basis. Singh should have tested each of the model's components, and their combined interactions, with the same quarterly frequency.

B is incorrect. Although Singh updates the Alpha and Risk components quarterly, he reviews the Optimizer, which links these two components, and the overall model itself on only an annual basis. He should review all components as well as the overall model with the same frequency.

C is incorrect. Singh should review all components as well as the overall model on the timeframe that is appropriate (quarterly here) and at a minimum conforms to the firm's guidelines.

3 According to the CFA Institute Code and Standards, what is the next action (from those below) that Singh should take following his conversation with Ringfield about the model error?

A Dissociate from the firm.

B Contact the firm's clients.

C Contact senior management.

C is correct. Singh should contact senior management before dissociating himself from QH or contacting QH clients. Upon discovery of the error, Singh should try to fix the model immediately. By not fixing the model immediately, Singh is harming QH clients. In not gaining Ringfield's approval to fix the error immediately, the next action Singh should take is to contact Ringfield's boss or senior management to make them aware of the situation.

According to Standard I(A): Professionalism–Knowledge of the Law, "If a member or candidate has reasonable grounds to believe that imminent or ongoing client or employer activities are illegal or unethical, the member or candidate must dissociate, or separate, from the activity. In extreme cases, dissociation may require a member or candidate to leave his or her employment. Members and candidates may take the following intermediate steps to dissociate from ethical violations of others when direct discussions with the person or persons committing the violation are unsuccessful. The first step should be to attempt to stop the behavior by bringing it to the attention of the employer through a supervisor or the firm's compliance department. If this attempt is unsuccessful, then members and candidates have a responsibility to step away and dissociate from the activity."

A is incorrect. Dissociating from the firm is the final step in the process. Standard I(A) establishes next steps as follows: "The first step should be to attempt to stop the behavior by bringing it to the attention of the employer through a supervisor or the firm's compliance department. If this attempt is unsuccessful, then members and candidates have a responsibility to step away and dissociate from the activity."

B is incorrect. Contacting the firm's clients directly is not a permitted intermediate step under Standard I(A).

4 Did Singh violate the CFA Institute Code and Standards by not immediately fixing the error in the ATM model?

A Yes

B No, because the error will be fixed next quarter

C No, because Singh disabled the common risk factors in the Risk Model as ordered by Ringfield

A is correct. By not fixing the error in the ATM model immediately, Singh is violating Standard III(A): Duties to Clients–Loyalty, Prudence, and Care. Members and candidates have a duty of loyalty to their clients and must act with reasonable care and exercise prudent judgement. By not immediately fixing the error in the model, Singh is not acting for the benefit of clients, nor is he placing client interests before his employer's or his own interests.

B is incorrect. The error should have been fixed immediately. According to Standard III(A), "Investment actions must be carried out for the sole benefit of the client and in a manner the member or candidate believes, given the known facts and circumstances, to be in the best interest of the client." It was in the best interest of clients to fix the error in the model immediately.

C is incorrect. Disabling the common risk factors in the Risk Model did not address the underlying error. The error should have been fixed immediately regardless of Ringfield's order. According to Standard III(A), "investment actions must be carried out for the sole benefit of the client and in a manner the member or candidate believes, given the known facts and circumstances, to be in the best interest of the client." It was in the best interest of clients to fix the error in the model immediately.

5 Did Ringfield violate the CFA Institute Code and Standards when talking with clients about their portfolios' underperformance?

A Yes

B No, because the market was turbulent

C No, because the model's common risk factors were to blame

A is correct. When talking with clients about their portfolios' underperformance, Ringfield was in violation of Standard I(C): Professionalism–Misrepresentation. Members and candidates must not knowingly make any misrepresentations relating to investment analysis, recommendations, actions, or other professional activities. By attributing the underperformance of client portfolios to market volatility, she is not telling them the real reason for the underperformance. In addition, the common risk factors have been disabled, so they are not functioning as intended for the model.

B is incorrect. The reason for the underperformance of client portfolios is the error in the model, not market turbulence. Ringfield violated Standard I(C) regarding misrepresentations relating to investment analysis, recommendations, actions, or other professional activities.

C is incorrect. The reason for the underperformance of client portfolios is the error in the model, not the model's common risk factors. Ringfield violated Standard I(C) regarding misrepresentations relating to investment analysis, recommendations, actions, or other professional activities.

6 Did Singh violate the CFA Institute Code and Standards with respect to his duties to his employer, QuantHouse, in developing his StockStar model?

A No

B Yes, because the model was developed while he was working at QH

C Yes, because he invests his personal and family portfolios using the model

A is correct. Singh is not in violation of the CFA Institute Code and Standards, Standard IV(A): Duties to Employers–Loyalty. In this case, Singh developed the StockStar model in his spare time (on nights and weekends) and used the model to manage only his personal and family portfolios. In addition, he has not been compensated for the model.

B is incorrect. It is not a violation of Standard IV(A): Duties to Employers–Loyalty to develop a model in his spare time.

C is incorrect. It is not a violation of Standard IV(A): Duties to Employers–Loyalty, to use his personal and family portfolios to test or invest in the model.

7 In the table that Singh provides to the marketing department, does he violate the CFA Institute Code and Standards?

A No, because he presented the performance information in the manner required by the CFA Institute Code and Standards.

B Yes, because he should have included only the actual performance results of the model.

C Yes, because he should have disclosed that he used his personal and family portfolios to generate actual results.

A is correct. Singh has not violated Standard III(D): Duties to Clients–Performance Presentation related to performance presentation because he has presented both the actual and back-tested performance of the model and clearly distinguished between the two. He has also noted that past performance does not guarantee future results.

B is incorrect. Standard III(D): Duties to Clients–Performance Presentation encourages full disclosure of investment performance data. Both actual and simulated performance measures are allowed as long as they are clearly disclosed. Singh fully explained the performance results being reported, stating that results are simulated (back tested) when model results are used and indicating that the actual and back-tested results are gross of fees.

C is incorrect. Standard III(D): Duties to Clients–Performance Presentation does not prohibit showing past performance of funds managed as long as appropriate disclosures are made, including the person's role in generating that performance. Singh fully explained the performance results being reported, stating that results are simulated (back tested) when model results are used and indicating that the actual and back-tested results are gross of fees.

JR AND ASSOCIATES

5

a evaluate practices, policies, and conduct relative to the CFA Institute Code of Ethics and Standards of Professional Conduct

b explain how the practices, policies, and conduct do or do not violate the CFA Institute Code of Ethics and Standards of Professional Conduct

Jacobs, Riccio, and Associates (JRA) is a global investment advisory firm that primarily provides high-net-worth individuals and their families with personalized wealth management solutions such as wealth planning, retirement planning, investment management, and trust and fiduciary services.[4] In addition, the firm has a small number of institutional clients. JRA employs 25 investment advisers and portfolio managers.

Benjamin Jacobs, CFA, and Andrew Riccio, CFA, founded JRA 10 years ago. Prior to establishing the firm, Jacobs worked as a lawyer for Brightman Partners, a large and prestigious law firm that specializes in real estate, family law, and estate planning. Riccio worked as a Certified Public Accountant for Earnest & Olds (E&O), a multinational professional services firm that specializes in providing tax, consulting, and advisory services to corporations and individuals. Kathy Parker, CFA, joined the firm as the third senior partner two years after it was founded. Previously, she had worked for the Frontline Group, a broker/dealer. JRA acquires most of its clients through referral arrangements put in place by the three senior partners.

Jacobs has a fee-sharing arrangement with his former colleagues at Brightman Partners (BP) when they refer clients to JRA. The annual investment fee stated in JRA's marketing brochure is higher than the fee most of its clients pay because Jacobs offers a discount on the investment fee to clients who are referred by BP lawyers. This discount encourages the BP lawyers to market JRA's services to their clients. In return, JRA shares a portion of the client's' annual investment advisory fee with the referring lawyer. The lawyers at BP disclose this fee-sharing arrangement with the clients that they refer to JRA. JRA discloses all of this information in the JRA investment management agreement that individuals sign at the time they become clients. (Questions 1, 2)

Riccio offers a similar fee discount and sharing arrangement to accountants at his previous firm, Earnest & Olds (E&O), who refer their clients to JRA. Over time, however, Riccio has observed that many of JRA's clients are reluctant to tell their investment adviser about securities and real estate holdings that are managed at other firms. As a result, the adviser does not have a complete understanding of the client's overall financial position. To assist JRA advisers in developing more-realistic and accurate investment policy statements, the accountants at E&O provide a copy of their referred client's tax returns to the client's JRA adviser after they open an account at JRA. This step allows JRA advisers to "know their client" better and provides greater transparency into their client's financial condition. In return, JRA advisers provide their clients' quarterly account statements to their E&O accountants to help with their tax planning and year-end tax preparation. Client approval is not needed for this information sharing because clients sign confidentiality statements directly with their E&O accountants and JRA advisers, and because they often view their investment adviser and their accountant as a team. (Question 3)

Kathy Parker has a somewhat different referral arrangement in place with the Frontline Group. Frontline's brokerage unit refers all of its small institutional clients (pension plans, profit sharing plans, and endowments) that are looking for investment management to JRA. In return, all of the trading from these accounts continues to be executed through Frontline's broker/dealer. Because Frontline continues to provide "best price and best execution" to these clients, Parker believes no additional client disclosures are necessary because client trading is unaffected. (Question 4)

Since starting JRA, Jacobs and Riccio have developed a close relationship with Tim Carroll, an independent consultant they met at a networking event. Carroll is hired by pension funds to solicit and review proposals from investment advisers who wish to manage a portion of the pension fund's assets. Over the years, Carroll has been instrumental in JRA's success by referring several of his pension fund clients to the firm

4 **JR and Associates:** Marcus Allan Ingram, Ph.D., CFA, and Michael G. McMillan, PhD, CFA. *Ethics Cases*. © 2018 CFA Institute. All rights reserved. Consistent with the 11th Edition of the *Standards of Practice Handbook*.

because of the firm's outstanding performance record and superior client service. To thank Carroll for all of his hard work on JRA's behalf (regardless of whether Carroll's pension fund clients actually hire JRA), Jacobs and Riccio each make sizable annual donations to Carroll's Children's Charity, a non-profit organization Carroll created to benefit orphans. Because these donations are made annually, they are not disclosed to the pension funds referred by Carroll who become JRA clients. (Question 5)

Recently, JRA hired Mufid Othan, an investment adviser and CFA charterholder who previously worked at JRA's largest competitor, Sack International. To attract Othan and his large "book of clients," JRA offered him $500 for each client he "brought over" from Sack. While at Sack, Othan was allowed to connect with all of his clients through his personal social media platforms. This not only enabled him to build an electronic database containing the names, addresses, phone numbers, and email addresses of all his clients but also helped him to provide superior client service by "following" his clients' personal and professional lives. When Othan tendered his resignation from Sack, he was immediately escorted out of the building. Othan spent the following weekend contacting all of his clients via social media to tell them about his resignation and to encourage them to join him at JRA. He did not disclose to them, however, that he was being paid $500 for each client he brought over from Sack. (Questions 6, 7)

A few weeks after beginning work at JRA, Othan hired Zane Ode, a recent college graduate, who recently found out she had passed Level III of the CFA Program examination. After hearing the good news about her success with Level III, Ode posted the following comments in a CFA candidate chatroom:

Comment 1 "I can't believe I passed the exam; the ethics questions were super hard."

Comment 2 "Wow, I scored above the Minimum Performance Score (MPS) on derivatives. I still don't know what answer was right for the two-part contango–backwardation question."

Comment 3 "The graders must have been quite lenient in grading my answers to the constructed response questions."

Ode now has three and a half years of experience in the investment industry. Nevertheless, Othan has already made a habit of introducing her to current and prospective clients as the firm's "newest CFA," and Ode has said nothing to correct him. (Questions 8, 9)

Case Questions

1 Does Jacobs violate the CFA Institute Code and Standards by offering his referral clients a lower investment advisory fee than the one quoted in JRA's marketing brochure?

A No

B Yes, because JRA is misrepresenting its fees

C Yes, because JRA is not dealing with its clients fairly

A is correct. Jacobs is not in violation of the CFA Institute Code and Standards. According to Standard III (B): Duties to Clients–Fair Dealing, members and candidates may provide more personal, specialized, or in-depth service to clients who are willing to pay for premium services through higher management fees or higher levels of brokerage. The term "fair" implies that the member or candidate must take care not to discriminate against any clients when disseminating investment recommendations or taking investment action.

B is incorrect. JRA is not misrepresenting its fees, because some of its clients are paying the fees that are disclosed in its marketing brochure. In addition, the advertised fees represent the highest fees that clients would pay.

C is incorrect. Standard III (B): Duties to Clients–Fair Dealing focuses on investment recommendations and taking investment action. The case provides no evidence that non-referred clients are being discriminated against or that referred clients are receiving preferential treatment, with respect to the dissemination of investment recommendations or the taking of investment action. Referred clients are simply receiving discounted fees.

2 Does Jacobs violate the CFA Institute Code and Standards in his disclosure of referral arrangements to his clients?

A Yes

B No, because the lawyers disclose to their clients the discount that JRA offers

C No, because the discount and the fee-sharing arrangement is disclosed to individuals at the time they sign the investment management agreement

A is correct. Jacobs is in violation of Standard VI (C): Conflicts of Interest–Referral Fees, which states, "Members and candidates must disclose to their employers, clients, and prospective clients, as appropriate, any compensation, consideration, or benefit received from or paid to others for the recommendation of products or services…. Appropriate disclosure means that members and candidates must advise the client or prospective client, before entry into any formal agreement for services, of any benefit given or received from the recommendation of any services provided by the member or candidate." In this case, the disclosure does not occur until the time the individual signs the investment management agreement, which is too late.

B is incorrect. The case facts state that BP lawyers disclose the fee-sharing arrangement to the clients they refer to JRA. The case facts do not state whether the lawyers disclose the discount offered by JRA. The behavior of the BP lawyers, however, is not covered by the Code and Standards. Disclosures, or lack thereof, by BP lawyers do nothing to mitigate JRA's duties and responsibilities.

C is incorrect. The discount is disclosed to JRA clients at the time they sign the investment management agreement. According to the Standard VI (C), disclosure must occur before the client enters into a formal agreement.

3 Do JRA advisers violate the CFA Institute Code and Standards by sharing client information with the accountants at E&O?

A Yes

B No, because the client views representatives from both firms as a team

C No, because the client has signed confidentiality agreements with both firms

A is correct. JRA advisers have violated the confidentiality of their clients by not obtaining client approval (written approval is recommended) in advance of sharing their information between each firm. According to Standard III(E): Duties to Clients–Preservation of Confidentiality, members and candidates are required to preserve the confidentiality of information communicated to them by their clients, prospective clients, and former clients. This standard is applicable when (1) the member or candidate receives information because of his or her special ability to conduct a portion of the client's business or personal affairs and (2) the member or candidate receives information that arises from or is relevant to that portion of the client's business that is the subject of the special or confidential relationship.

B is incorrect. Although the client has signed confidentiality agreements with both firms, the client has not signed an agreement allowing the sharing of information between the firms.

C is incorrect. Although the client may view representatives from both firms as a team, neither team has received client approval in advance of sharing the client's information. As a practical matter, if JRA advisers request information from prospective clients regarding other investment income and assets and the prospect denies existence of such assets, the adviser is under no obligation to perform additional due diligence to ascertain the existence of other assets.

4 Has Parker violated the CFA Institute Code and Standards in her referral arrangement with Frontline Group?

A Yes

B No, because Frontline Group continues to provide "best price" and "best execution"

C No, because nothing has changed—all client trades are still executed by Frontline

A is correct. By not disclosing the referral arrangement to clients who were referred to her by Frontline Group, Parker has violated Standard VI (C): Conflicts of Interest–Referral Fees, which states, "Members and candidates must disclose to their employers, clients, and prospective clients, as appropriate, any compensation, consideration, or benefit received from or paid to others for the recommendation of products or services…. Appropriate disclosure means that members and candidates must advise the client or prospective client, before entry into any formal agreement for services, of any benefit given or received from the recommendation of any services provided by the member or candidate." In this case, there is no evidence to suggest Parker disclosed her referral arrangement with Frontline Group to prospective clients. By not doing so, Parker violated Standard VI (C).

B is incorrect. Regardless of whether Frontline provides "best price" and "best execution" or whether the execution of client trades remains unchanged by Frontline, Parker must still disclose the referral arrangement to her clients.

C is incorrect. Parker must still disclose the referral arrangement to her clients, regardless of the fact that all client trades continue to be executed by Frontline.

5 Did Jacobs and Riccio violate the CFA Institute Code and Standards by making annual donations to Carroll's Children's Charity?

A No

B Yes, because these donations create a conflict of interest

C Yes, because these donations represent additional compensation to Carroll

B is correct. The donations made by Jacobs and Riccio give Carroll an incentive to refer potential clients to JRA and at the very least give the perception that Carroll's objectivity and independence have been compromised. Jacobs and Riccio are in violation of Standard I(B): Professionalism–Independence and Objectivity, which states, "Members and candidates must use reasonable care and judgement to achieve and maintain independence and objectivity in their professional activities. Members and candidates must not offer, solicit, or accept any gift, benefit, compensation, or consideration that reasonably could be expected to compromise their own or another's independence and objectivity."

A is incorrect. As already noted, donations made by Jacobs and Riccio give Carroll an incentive to refer potential clients to JRA. This at the very least gives the perception that Carroll's objectivity and independence have been compromised, and so Jacobs and Riccio are in violation of the Code and Standards, specifically Standard I(B): Professionalism–Independence and Objectivity.

C is incorrect. The donations were made to Carroll's charity and do not represent additional compensation to Carroll. Additional compensation is defined in Standard IV(B): Duties to Employers–Additional Compensation Arrangements as "gifts, benefits, or compensation, or consideration that competes with or might reasonably be expected to create a conflict of interest with their employer's interest." An additional compensation arrangement is one that creates a conflict of interest between the member or candidate and her employer.

6 Did Othan violate the CFA Institute Code and Standards by contacting his Sack International clients via social media after leaving Sack?

A No

B Yes, because he is using client confidential information

C Yes, because the client information he is using belongs to Sack

A is correct. Othan is not in violation of the CFA Institute Code and Standards. According to Standard IV(A): Duties to Employers–Loyalty, "Members and candidates should understand and abide by all applicable firm policies and regulations as to the acceptable use of social media platforms to interact with clients and prospective clients. This is especially important when a member or candidate is planning to leave an employer." In this case, Sack allowed Othan to use his personal social media platforms to connect with clients. In addition, he did not contact his former clients via social media to inform them about his departure until after he resigned from Sack.

B is incorrect. Contacting his clients via social media after leaving Sack, does not require Othan to use confidential client information.

C is incorrect. Othan used his personal social media platforms to connect with clients. These platforms are not the property of Sack.

7 Did Othan violate the CFA Institute Code and Standards by not disclosing to clients that he was receiving $500 for each client that he brought over to JRA from Sack?

A No

B Yes, because this is a referral fee

C Yes, because this is additional compensation

A is correct. Othan is not in violation of the CFA Institute Code and Standards. The $500 does not have to be disclosed to clients because it is not a referral fee or additional compensation, and it does not create a conflict of interest with his employer, clients, or prospective clients.

B is incorrect. According to Standard VI(C): Conflicts of Interest–Referral Fees, referral fees are "any compensation, consideration, or benefit received from or paid to others for the recommendation of products or services." The $500 Othan received from JRA for each client he brought over from Sack is not a referral fee because the $500 is being paid by the employer (JRA) to the employee (Othan) for services provided. This amount is compensation paid by the firm, not a fee charged to clients.

C is incorrect. Additional compensation is defined in Standard IV(B): Duties to Employers–Additional Compensation Arrangements as "gifts, benefits, or compensation, or consideration that competes with or might reasonably be expected to create a conflict of interest with their employer's interest." The $500 is not additional compensation, and there is no conflict with the employer's interests. Although disclosure of all bonus arrangements may add clarity, the Code and Standards do not require members and candidates to disclose how they are compensated.

8 Which of the comments Ode posted in the CFA candidate chatroom violated the CFA Institute Code and Standards?

A Comment 1.

B Comment 2.

C Comment 3.

B is correct. Ode's comment 2 violated Standard VII(A): Responsibilities as a CFA Institute Member or CFA Candidate–Conduct as Participants in the CFA Institute Programs: "CFA Institute program rules, regulations, and policies prohibit candidates from disclosing confidential material gained during the exam process." Examples of information that cannot be disclosed by candidates sitting for an exam include but are not limited to the following:

- Specific detail of questions appearing on the exam (contango–backwardation).
- Broad topical areas and formulas tested or not tested on the exam (derivatives).

In this case, Ode disclosed specific details of questions appearing on the exam.

A is incorrect. In saying that the ethics questions were super hard, Ode did not disclose confidential information gained during the exam process.

C is incorrect. In saying that the graders must have been quite lenient in grading her answers to the constructed response questions, Ode did not disclose confidential information gained during the exam process.

9 Did Othan violate the CFA Institute Code and Standards in his description of Ode?

A Yes

B No, because Ode will be a CFA charterholder in another six months

C No, because Ode has successfully completed all three levels of the CFA Program

A is correct. Othan is in violation of the CFA Institute Code and Standards. Ode is not yet a CFA charterholder, and in referencing her as the firm's "newest CFA," Othan is misrepresenting Ode. Standard VII(B): Responsibilities as a CFA Institute Member or CFA Candidate–Reference to CFA Institute, the CFA Designation, and the CFA Program states that "'CFA Charterholders' are those individuals who have earned the right to use the CFA designation granted by CFA Institute. These people have satisfied certain requirements, including completion of the CFA Program, and required years of acceptable work experience." The recommended procedures for Ode's compliance with Standard VII(B) include educating others in the firm, including re-educating Othan, about her status.

B is incorrect. To be a CFA charterholder, Ode needs to have completed the required four years of work experience.

C is incorrect. The fact that she has completed all three levels of the CFA Program does not make Ode a CFA charterholder. To be a CFA charterholder, she must also have the required four years of work experience.

MAGADI ASSET MANAGEMENT 6

a **evaluate practices, policies, and conduct relative to the CFA Institute Code of Ethics and Standards of Professional Conduct**

b **explain how the practices, policies, and conduct do or do not violate the CFA Institute Code of Ethics and Standards of Professional Conduct**

Magadi Asset Management (Magadi) is a global investment management firm based in Nairobi, Kenya. Magadi manages dedicated equity, fixed income, and real estate funds, as well as other alternative investment vehicles.[5] The firm's clients include pension schemes, sovereign wealth funds, and high-net-worth individuals. Frederick Omondi, CFA, is Magadi's president and chief investment officer. Under Omondi, the CFA Code of Ethics and Standards of Professional Conduct has been adopted as the firm's Code of Conduct for Magadi's employees.

Last year, Omondi established a proprietary trading desk at Magadi. The role of the proprietary traders is to actively trade African securities for the firm's benefit. Proprietary traders do not execute orders for Magadi's institutional or retail clients; these orders are handled by traders on the main trading desk. To increase cooperation among traders and encourage the sharing of best execution practices, both trading desks are located on the same floor at Magadi's headquarters. This proximity has allowed proprietary traders to hear customer order flow and also see customer order information on the computer screens of the main traders. To encourage collaboration between the two trading desks, Omondi offers bonuses to proprietary traders who provide trading ideas to the main traders for the benefit of their clientele.

To allay client concerns about potential front-running, Omondi has told clients that information concerning their orders and business affairs is kept confidential. He further explains the firm has instituted a firm-wide policy that expressly states the following: "Employees may not discuss the business affairs of any client with any other person, except on a strict need-to-know basis. Trade orders made by the proprietary traders that may be similar to client orders must be executed after the clients' orders have been fully executed by the main dealing desk traders." (Questions 1, 2)

Omondi's biggest business success this year was a large mandate from a sovereign wealth fund to invest in Magadi's managed funds. To secure the mandate win, Omondi hired, as a "sub-adviser" to the managed funds, a business development agent with contacts at the highest level within the government responsible for the sovereign wealth fund. Despite having very limited experience as a financial consultant, the agent had a number of close relationships with senior managers at the sovereign wealth fund because of his connections to the government officials responsible for the fund. The payments made by Omondi, through the sub-adviser, included a "deal fee" and other expenses that facilitated the governmental support of the sovereign wealth fund investment. Omondi did not require the agent to provide details regarding its activities or the specific expenses covered by the fee. The agent's expenses are charged to Omondi's managed funds. As a thank you for being awarded the mandate, Omondi made donations to the favorite charities of the sovereign wealth fund's top management, as he had promised during the due diligence process. (Question 3)

Three years ago, Magadi launched the Pan Africa Frontier Fund (PAFF), a non-listed equity unit trust with an investment mandate that prohibits the use of leverage. The mandate requires the following:

- 80% of the companies in the portfolio to be traded on at least one of the 17 securities exchanges operating within Africa;
- the portfolio be invested in a minimum of eight countries at all times;
- no more than 30% of the portfolio's value can be invested in any single country;
- no more than 10% of the portfolio's value can be invested in cash and cash equivalents; and
- no single security can account for more than 15% of the portfolio's value.

5 **Magadi Asset Management:** Renée K. Blasky, CFA, CIPM, and Michael G. McMillan, PhD, CFA. *Ethics Cases*. © 2018 CFA Institute. All rights reserved. Consistent with the 11th Edition of the *Standards of Practice Handbook*.

Since its launch, PAFF has significantly underperformed its peers and has had several quarters of negative returns. As a result, it ranks in the bottom performance quadrant relative to its peers.

Omondi recently hired Bukenya Kirabo, CFA, to take over management of PAFF. Kirabo was hired to improve PAFF's performance and move the fund to the top performance quadrant in rankings based on his extensive experience and knowledge of African equities, as well as his reputation as an astute investment manager. Kirabo has more than two decades of experience analyzing and investing in public companies across Africa. After graduating from a top local university, he moved to London, where he worked for a global asset management firm. Five years ago, Kirabo was transferred to the firm's regional office in Africa to manage one of the firm's local funds. During the last five years, Kirabo has generated average annual returns of 23%. Since returning to Africa, Kirabo has witnessed notable improvements in African securities markets, particularly in the area of settlement risk. Many local markets remain relatively illiquid, however, and most public companies in Africa are under-researched compared with other emerging markets. As a result, systematic risk is considerably higher in African markets than in other emerging markets.

Three months after being hired at Magadi, Kirabo meets with Omondi to review PAFF's most recent quarterly performance. During the meeting, he states, "PAFF's solid performance this quarter is a result of three changes I made:

> Change 1: Because of strong cash inflows into PAFF, I have increased the maximum level of cash and cash equivalents to 15% of the portfolio. Given the illiquid nature of many markets in which we are investing, I believe it is more prudent, and less risky, to take sufficient time to find attractive investment opportunities and build position holdings.
>
> Change 2: I have increased the portfolio's geographic diversification from 11 countries (stock exchanges) to 13. Securities traded on 13 different African stock exchanges (up from 11 previously) are now represented in the portfolio. This higher level of diversification has improved the portfolio's Sharpe and information ratios.
>
> Change 3: To increase accountability for PAFF's performance, I am now making all buy and sell decisions for PAFF. Previously, when the team of analysts was making the investment decisions, it was difficult to attribute an individual's contribution to fund performance."

Kirabo next meets with the marketing department to discuss PAFF's new sales campaign. During the meeting, he states, "Please include all of the mandate changes I have made in PAFF in the new brochure that will be distributed to prospective clients. You can also include the five-year investment performance I achieved while managing a fund at my previous employer. Please do not state where the performance was earned, however, because my previous employer is a direct competitor of Magadi. Finally, because the mandate changes are relatively trivial, there is no need to inform existing clients." (Questions 4, 5, 6)

PAFF currently owns 9% of the common stock of Mtume's, a mining company listed on the Botswana Stock Exchange. Kirabo has been reducing the fund's holdings in Mtume because of the company's declining revenues and profits. This morning, Kirabo speaks with Olivia Moroka, Mtume's chief financial officer. During their conversation, Moroka tells Kirabo, "You may want to stop selling your shares of Mtume, because our board of directors just received a very attractive all-cash offer of BWP500 million (Botswana pula) to purchase one of our mining subsidiaries. Although nothing is definite, the board will be meeting next week to vote on the offer."

After getting off the phone with Moroka, Kirabo calls the Magadi analyst who follows Mtume and tells her about his conversation. The analyst then incorporates the expected subsidiary sales price into her financial model of Mtume. The output from her revised model indicates that the sale proceeds will significantly enhance Mtume's credit standing and its ability to reinstitute shareholder cash distributions on an earlier-than-expected schedule and in larger-than-expected amounts. When the analyst tells Kirabo about her findings, Kirabo immediately calls the proprietary and main traders to tell them to start buying "any and all" shares of Mtume. He then calls Omondi and tells him about his conversation with Moroka. After Omondi gets off the phone with Kirabo, Omondi calls his broker and purchases shares in Mtume for his personal account and the family accounts that he controls. (Questions 7, 8)

Case Questions

1 By allowing customer order information to be known to the traders on the proprietary desk, did traders on the main trading desk most likely violate the CFA Institute Code and Standards?

- A Yes
- B No, because this information was not shared outside of the firm
- C No, because proprietary traders were not allowed to act on this information until after client orders were executed

A is correct. Traders on the main trading desk are in violation of Standard III(E): Duties to Clients–Preservation of Confidentiality. This standard requires members and candidates to preserve the confidentiality of information communicated to them by their clients, prospective clients, and former clients. The sharing of office space such that the proprietary traders can see the screens of the main traders is inappropriate because it allows confidential client information to be disclosed to individuals (proprietary traders) who did not need to know the information. To avoid sharing confidential information and violating firm policy, the main traders should have taken necessary action to ensure the client information was not advertently or inadvertently shared with the proprietary desk traders.

B is incorrect. Although the information was not shared externally, the main traders still allowed its disclosure to individuals who did not meet the "need to know" requirement and, in doing so, violated Standard III(E).

C is incorrect. Whether or not the proprietary traders acted on the information is irrelevant in this case. Traders on the main trading desk needed to take the necessary action to prevent the disclosure of confidential information and, in not doing so, they violated Standard III(E).

2 Did Omondi most likely violate the CFA Institute Code and Standards in supervising the employees in the two trading desks?

- A Yes
- B No, because he implemented a policy to prevent front-running
- C No, because he encouraged collaboration between the two departments

A is correct. Omondi was in violation of Standard IV(C): Duties to Employers–Responsibilities of Supervisors. Members and candidates must promote actions by all employees under their supervision and authority to comply with applicable laws, rules, regulations, firm policies and the Code and Standards. Omondi failed to establish effective policies and procedures reasonably designed to prevent traders on the proprietary dealing desk from obtaining confidential customer information. Although the proprietary traders did not have direct access to the computer system used by the main traders to execute customer orders, by being co-located on the same floor, the proprietary traders could still view customer order information on the main traders' computer screens and hear them discuss customer orders. Omondi could have located one set of traders in a separate space or a different floor with security access restrictions. Omondi would also likely be in violation of Standard I(C): Professionalism–Misrepresentation, because his representation to customers was incorrect—that is, client information was made available to other employees outside of those operating on a "need-to-know" basis.

B is incorrect. Although Omondi announced a policy to mitigate front-running, the policy was ineffective and, as implemented, did not prevent or address the sharing of confidential client information (orders) to individuals who did not need to know this information (proprietary traders).

C is also incorrect. Encouraging collaboration between the two trading desks does not address the fact that the proprietary traders could see and hear confidential information about client orders from the main trading desk. Omondi failed to establish sufficient policies and procedures to ensure compliance with the Code and Standards as well as firm policy for the traders under his supervision.

3 Omondi most likely violated the CFA Institute Code and Standards when dealing with the sovereign wealth fund's top managers:

A Only by making charitable donations.

B Only by hiring a sub-adviser because of his high-level government contacts.

C By both A and B.

C is correct. Omondi was in violation of Standard I(B): Professionalism–Independence and Objectivity. "When working to earn a new investment allocation, members and candidates should not offer gifts, contributions, or other compensation to influence the decision of the hiring representative. The offering of these items with the intent to impair another person's independence and objectivity would not comply with Standard I(B). Such prohibited actions may include offering donations to a charitable organization."

To better serve clients, investment professionals may delegate to third parties work that requires particular specialization, knowledge, or expertise. For instance, an investment adviser may hire sub-advisers to handle a particular strategy or investment style outside the scope of the adviser's ability or experience. A global adviser may hire a sub-adviser to manage an asset allocation invested in a particular country or region, and the payments to the sub-adviser would be legitimate investment expenses that could properly be passed on to investors in the fund. It is clear from the facts of this case, however, that Omondi is not hiring a true sub-adviser but instead paying locally connected officials to secure access for the sovereign wealth fund's investment. The "sub-adviser" has limited financial experience but is close to the government officials, and the "deal fees" are not supported by any documentation that details legitimate investment expenses. The "sub-advising expenses" charged by Omondi to the fund could, in all likelihood, be funding corrupt transactions and bribes through local intermediaries. This practice violates multiple standards, including I(A) Knowledge of the Law (because the conduct would violate any type of anti-bribery laws); I(C) Misrepresentation (improperly labeling the expenditures as investment fees); V(A) Diligence and Reasonable Basis (no reasonable and adequate basis for the "investment" action); and V(C) Record Retention (No appropriate records to support the action).

4 According to the CFA Institute Code and Standards, which of the changes in the PAFF Fund does Kirabo *not* have to disclose?

A Change 1.

B Change 2.

C Change 3.

B is correct. Change 2 is not required to be disclosed because, by increasing the country exposure to 13 nations, Kirabo is still within the 80% stated mandate. The investment process has not fundamentally changed. Changes 1 and 3 are modifications to the investment process that, according to Standard V(B): Investment Analysis, Recommendations, and Actions–Communication with Clients and Prospective Clients, must be disclosed. According to Standard V(B), members and candidates must disclose to clients and prospective clients the basic format and general principles of the investment processes they use to analyze investments, select securities, and construct portfolios, and they must promptly disclose any changes that might materially affect those processes.

A is incorrect. Change 1 is a change in the fund's mandate because the maximum amount of cash that the fund can hold has been increased to 15% from 10%.

C is incorrect. Change 3 is a change to the investment process, because all purchase and sell decisions are now being made by Kirabo instead of the team of analysts.

5 Does Kirabo most likely violate the CFA Institute Code and Standards by including his prior performance in the PAFF marketing brochure?

A No

B Yes, because the brochure should have stated the name of the firm where he earned prior performance

C Yes, because the marketing brochure should not show fund performance earned at a prior firm as part of his performance track record

B is correct. Kirabo was in violation of Standard III(D): Duties to Clients–Performance Presentation. Standard III(D) does not prohibit showing past performance of funds managed at a prior firm, as long as showing that record is accompanied by appropriate disclosures about where the performance took place and the person's specific role in achieving that performance. Kirabo does not disclose the name of the prior firm or that he alone managed the fund and was solely responsible for its performance. Consequently, he is in violation of Standard III(D): Performance Presentation. Kirabo would also be required to receive permission in writing from his previous employer to take his performance records with him when he left the firm, because the performance record is an asset of the firm, not of the individual employee. If he did not receive prior written permission, he would also be in violation of Standard IV(A): Duties to Employers–Loyalty, which requires members and candidates to protect their employers' interests, even when leaving the firm.

C is incorrect. There is no prohibition on including past investment performance under Standard III(D) so long as there are disclosures that clearly indicate it was earned at a previous entity and what the role the manager played in achieving that performance.

6 According to the CFA Institute Code and Standards, whom must Kirabo most likely inform of the material changes related to the PAFF?

A Current clients only.

B Prospective clients only.

C Current and prospective clients.

C is correct. According to Standard V(B): Investment Analysis, Recommendations, and Actions–Communication with Clients and Prospective Clients, Kirabo must disclose to current and prospective clients both the fund mandate change and the change in the investment process.

7 Did Kirabo most likely violate the CFA Institute Code and Standards by purchasing additional shares of Mtume?

A Yes

B No, because the information that Kirabo learned from Moroka was not definite

C No, because his decision was based on the output from the analyst's revised model

A is correct. Kirabo was in violation of Standard II(A): Integrity of Capital Markets–Material Nonpublic Information. Members and candidates who possess material nonpublic information that could affect an investment's value must not act or cause others to act on the information. Information is "material" if its disclosure would probably affect the price of a security or if reasonable investors would want to know the information before making an investment decision. In addition to the substance and specificity of the information, the source or relative reliability of the information also determines materiality. In this case, factual information from a corporate insider regarding the purchase of a subsidiary is likely to be material. Although the offer is not definite or officially accepted by the board, its source, substance, and specificity are enough to make the information material. The output from the analyst's revised model was affected by the insider information.

B is incorrect. Information does not have to be definite to trigger the violation; it need only be considered both material and nonpublic. In this case, factual information from a corporate insider regarding the purchase of a subsidiary is both nonpublic and material. Trading on this information violates Standard II(A).

C is incorrect. The output from the analyst's revised model was affected by the insider information. Thus, his decision was based on material, nonpublic information, which violates Standard II(A): Material Nonpublic Information.

8 Did Omondi most likely violate the CFA Institute Code and Standards by purchasing shares for his personal and family accounts?

A Yes

B No, because the information is not definite

C No, because the board has not voted on the offer

A is correct. Omondi violated Standard II(A): Integrity of Capital Markets–Material Nonpublic Information. Members and candidates who possess material nonpublic information that could affect an investment's value must not act or cause others to act on the information. Information is "material" if its disclosure would probably affect the price of a security or if reasonable investors would want to know the information before making an investment decision. In addition to the substance and specificity of the information, the source or relative reliability of the information also determines materiality. In this case, factual information from a corporate insider regarding the purchase of a subsidiary is likely to be considered material. Although the offer is not definite or officially accepted by the board, its source, substance, and specificity are enough to make the information material. By purchasing shares informed by material nonpublic information, Omondi violated Standard II(A).

B is incorrect. Information does not have to be definite to trigger the violation; it need only be considered both material and nonpublic. In this case, factual information from a corporate insider regarding the purchase of a subsidiary is both nonpublic and material. Trading on this information violates Standard II(A).

C is incorrect. There is no requirement that the information must be about something that has actually occurred, such as the action having been taken. Under Standard II(A), "Information is considered material if its disclosure would probably have an impact on the price of a security or if reasonable investors would want to know the information before making an investment decision." Both statements are true here. Thus the information is material, and also nonpublic, so trading on this information violates Standard II(A): Material Nonpublic Information.

7 EDVARD STARK

a evaluate practices, policies, and conduct relative to the CFA Institute Code of Ethics and Standards of Professional Conduct

b explain how the practices, policies, and conduct do or do not violate the CFA Institute Code of Ethics and Standards of Professional Conduct

Edvard Stark, CFA, is a private client adviser for Eyearne Bank, a small private bank.[6] In his role, Stark constructs and manages globally diversified fixed-income and equity portfolios for his clients based on the clients' respective investment objectives, risk tolerance, and time horizon. As part of his service, Stark periodically reviews client assets held outside the bank and makes recommendations for those assets. Clients have often followed Stark's advice. In providing this service, Stark has been able to cultivate stronger relationships and build his client assets under management at the bank.

Stark has been following developments in digital currencies, also known as cryptocurrencies, for some time. When the national securities regulator announced, some months ago, its decision to regulate cryptocurrencies as securities and began issuing guidance on cryptocurrency best practices, Stark concluded it was time to consider digital currencies for himself and his clients. Intrigued by the rapid appreciation in value many cryptocurrencies have exhibited, he believes cryptocurrencies may offer clients the potential for higher returns as well as diversification benefits.

Stark spends two weekends researching the top cryptocurrencies. All are digital currencies created to facilitate different types of secure transactions over the internet. He learns that cryptocurrencies are "held" in online wallets set up by individual account holders and that individuals may earn additional cryptocurrency tokens by helping administer the cryptocurrency network through an activity called "mining." Stark has read that it is difficult for later entrants to a cryptocurrency network to make money through mining because competitive pressure tends to raise the required level of capital investment over time, so he decides to focus his efforts and research on the newer cryptocurrencies.

After considering several of the newer cryptocurrencies, Stark decides the best opportunity is with a digital currency called Meerine. To limit his risk of being wrong on the cryptocurrency's potential, Stark decides to give a buy recommendation to only a few of his smallest clients. He recommends a 1% position in Meerine to these clients. Each of these clients establishes an online wallet to hold his cryptocurrency tokens and buys the recommended position in Meerine.

As Stark monitors Meerine's price over the next month, he learns more about its trading patterns and its acceptance in the marketplace. Although Meerine's price exhibits significant volatility, Stark feels optimistic about its potential. From his research, he knows there may also be an opportunity in mining Meerine's currency.

6 **Edvard Stark:** David B. Stevens, CIMC, CFA. *Ethics Cases*. © 2017 CFA Institute. All rights reserved. Consistent with the 11th Edition of the *Standards of Practice Handbook*.

Mining would involve using his own computing resources to help process Meerine's digital transactions, but in return, he could earn additional Meerine tokens for his Meerine account.

To learn how to do this, Stark attends a local cryptocurrency conference and numerous workshops on mining. Stark believes mining Meerine's currency will give him a better understanding of cryptocurrencies and the technology supporting Meerine; this understanding, in turn, will help him make better cryptocurrency investment recommendations for his clients.

After mining Meerine's currency by running the mining software as a background process on his home computer for several months, Stark believes he is competent in his understanding of cryptocurrencies and their underlying technology. Mining has also provided him with a way to augment his salary from Eyearne Bank by adding Meerine tokens to his digital account. During this time, Meerine's price has continued to rise strongly. Stark decides to recommend a 3% Meerine position for all clients.

In his client review meetings, Stark highlights Meerine's cryptocurrency as an exciting opportunity. He illustrates the low correlation of cryptocurrencies with traditional assets and shows the strong performance of Meerine since his initial 1% buy recommendation. He shares with clients that he is mining the currency for Meerine and discusses his new 3% buy recommendation with each client. His clients, knowing little about cryptocurrencies, have few questions and no objections. Stark is pleased and feels his recommendation has been well received.

Because Meerine is a newer cryptocurrency, its daily trading volume is low and it will take his clients several days to establish their positions. As a miner, Stark receives a steady flow of Meerine tokens into his digital wallet from his mining activities. He offers his larger clients the opportunity to buy Meerine tokens directly from him so that they do not miss out on any potential appreciation of Meerine while trying to establish their positions.

Identify violations or possible violations of the Code and Standards by Stark. For each identified violation, state what actions Stark should have taken and make a short policy statement a firm could use to guide employees to help prevent similar violations in the future.

This case highlights ethical challenges individuals may face during their careers as markets evolve and innovative financial products are introduced. The violations or potential violations of the Code and Standards in this case relate to a member's duties to clients; duties to employer; duties regarding investment analysis, recommendations, and actions; and duty to disclose conflicts of interest to the employer and clients.

Duties to Clients

Standard III(B): Fair Dealing states that members and candidates must treat all clients fairly when taking investment action with regard to general purchases, new issues, or secondary offerings. Stark's offer to directly fill orders for his largest clients without making the same offer to all his clients is a breach of Standard III(B).

Standard III(B) does not state that all clients must be treated "equally." Members and candidates may differentiate their services to clients, but different levels of service must not disadvantage or negatively affect clients. When making investments in new offerings, however, members and candidates should distribute the issues to all customers for whom the investments are appropriate in a fair and equitable manner.

In this instance, Stark has clearly violated Standard III(B). Stark's offer to fill allocations from his Meerine account for only his largest clients puts his other clients at an economic disadvantage. Stark has a duty to all his clients to provide fair and impartial access to Meerine tokens.

Actions Required

Because Stark knows there is a limited market for Meerine tokens, he should either (1) offer each of his clients the opportunity to buy Meerine tokens directly from him, collect their orders, and then allocate his available tokens to each client in proportion to their planned investment or (2) not offer to sell any of his tokens to his clients.

Policy Statement for a Firm

"All client accounts participating in a new issue or security with limited liquidity will be executed as a block trade and shall receive the same execution price. All trade allocations to client accounts shall be made on a pro rata basis prior to or immediately following part or all of a block trade."

Standard III(C): Suitability obligates members and candidates who are in an investment advisory relationship with clients to consider carefully the needs, circumstances, and objectives of the clients when determining the appropriateness and suitability of a given investment. In judging the suitability of a potential investment, the member or candidate should review many aspects of the client's knowledge, experience related to investing, and financial situation. These aspects include, but are not limited to, the risk profile of the investment as compared with the constraints of the client, the impact of the investment on the diversity of the portfolio, and whether the client has the means or net worth to assume the associated risk. Although the national securities regulator is now regulating cryptocurrencies, they are still more suitable for speculation than as an investment, given that no clear consensus exists for determining future expected value for cryptocurrencies.

Although Stark has considered the potential risk reduction benefits from diversification, his recommendation that all his clients buy a 3% position in Meerine without specific regard to suitability regarding client circumstances or whether this investment is consistent with each client's written objectives, mandates, or constraints is a clear violation of Standard III(C). An additional violation of this standard is Stark's decision to initially recommend the Meerine investment only for his smallest accounts. Rather than being determined by his clients' investment objectives, including risk tolerance, his decision is driven by the desire to limit his personal and Eyearne Bank's risk of being wrong in his recommendation.

Actions Required

Although Stark is clearly excited about the possible benefits of cryptocurrencies, he needs to properly assess each client's circumstances and determine on the basis of her risk tolerance, goals, and objectives whether the client should invest in Meerine and, if so, what the appropriate level of exposure is for that client.

Policy Statement for a Firm

"When making any investment recommendations to clients, investment advisers must carefully consider the impact the proposed change will have on portfolio diversification, how the investment's risk parameters align with the client's assessed risk tolerance, and whether the proposed investment fits within the overall investment strategy, taking into account the client's time horizon, return objectives, and constraints, as well as the type and nature of the client."

Duties to Employers

Standard IV(B): Additional Compensation Arrangements requires members and candidates to obtain permission from their employer before accepting compensation or other benefits from third parties for any services that might create a conflict with their employer's interest.

Stark has begun mining Meerine for additional cryptocurrency compensation. Doing so creates a conflict of interest with Eyearne Bank, because mining Meerine's cryptocurrency involves activities that compete with Eyearne's services. Mining involves verifying transactions that occur outside of traditional banking channels. As part of normal operations, banks facilitate transactions through credit cards and checking accounts. In mining, Stark is supporting a service that is competitive with the bank, which creates a conflict. Earning outside compensation is not itself a violation of the Code and Standards, but Stark should disclose it to his employer for the consideration of conflicts.

Actions Required

Stark needs to disclose to his supervisor or the compliance department at Eyearne Bank his intention to mine Meerine and the potential earnings expected from this activity. He will need to receive written consent from Eyearne before beginning any mining activity.

Policy Statement for a Firm

"Employees must disclose any external employment or compensation arrangement to the firm and receive express written permission before undertaking any such arrangement. Failure to comply is a violation of company policy and is subject to disciplinary procedures up to and including termination."

Investment Analysis, Recommendations, and Actions

Under Standard V(A): Diligence and Reasonable Basis, members and candidates must exercise diligence, independence, and thoroughness in making investment recommendations. Although Stark had done some research before recommending that clients buy cryptocurrency, he was still in the learning process when he made the buy recommendation to his smallest clients; therefore, he is in violation of Standard V(A).

Standard V(A) does not require perfect knowledge but does require diligence and thoroughness from members and candidates in gathering as much information and knowledge as possible to inform their professional judgement before making an investment recommendation in order to have a reasonable and adequate basis for making the recommendation.

Actions Required

Stark should develop a written evaluation of cryptocurrencies and Meerine in particular, detailing the background information and decision framework that support his investment recommendation for cryptocurrencies and Meerine. Stark's report should consider risks as well as benefits.

Policy Statement for a Firm

"Purchases or recommendations to purchase are limited to securities on the 'Approved for Investment Purchase List' (Approved List). Securities can be added to the Approved List after review and approval by the Investment Committee. A written research report detailing risks and opportunities is required for evaluation by the Investment Committee. The report should also note whether the security is considered speculative or non-speculative."

Conflicts of Interest

Under Standard VI(A): Disclosure of Conflicts, members and candidates must make full and fair disclosure of all matters that could reasonably be expected to impair their independence and objectivity. Members and candidates must maintain their objectivity when rendering investment advice. Requiring members and candidates to disclose all matters that reasonably could be expected to impair the member's or candidate's objectivity allows clients to judge an adviser's motives and possible biases for themselves.

Stark's mining of Meerine and his recommendation that clients invest in Meerine is a conflict because he is advocating that his clients buy an investment with limited liquidity in which he has a personal holding. His clients' purchases would likely cause Meerine's price to rise, thereby directly benefiting Stark's position. His lack of full disclosure is a violation of Standard VI(A). Furthermore, his decision to sell some of his own cryptocurrency directly to his clients is a conflict that needs to be disclosed to all his clients who are considering his recommendation to buy Meerine, as well as to his employer, Eyearne Bank. Although he reveals his mining activity in client meetings held after his recommendation of a 3% position in Meerine for all clients, clients should be given an alternative cryptocurrency to invest in to avoid the direct conflict. Also, the information he shares in the client meetings does not fully disclose his conflicted position, because his clients have limited knowledge of cryptocurrencies and may not understand the conflict of Stark's mining activities and his investment recommendation to buy Meerine.

Because Stark's clients have a limited knowledge of cryptocurrencies, his duty to disclose his conflict of interest is of paramount importance so that his clients can fully evaluate his recommendation.

Actions Required

Stark should clearly disclose to his clients and Eyearne Bank his conflict of interest in mining Meerine and recommending Meerine for purchase to his clients. Because cryptocurrencies are relatively unfamiliar to most of his clients, he will need to make sure his clients fully understand his conflict. Before recommending Meerine to clients, Stark should also determine a suitable alternative cryptocurrency from those he researched for those clients who are uncomfortable with the conflict of interest.

Policy Statement for a Firm

"Employees shall not use their position, directly or indirectly, for private gain or financial benefit, to advance personal interests, or to obtain favors or benefits for themselves, their families, or any other person. Effective conflict management requires all employees to identify and disclose to the company's Compliance Officer all actual or potential conflicts of interest as they become aware of them. Because it is impossible to describe every conflict of interest, all employees are required to exercise sound judgment, seek advice when appropriate, escalate concerns, obtain review of certain activities as required by this policy and other applicable business and jurisdiction-specific policies and procedures, and adhere to the highest ethical standards."

SUBATH AGARWAY

8

a evaluate practices, policies, and conduct relative to the CFA Institute Code of Ethics and Standards of Professional Conduct

b explain how the practices, policies, and conduct do or do not violate the CFA Institute Code of Ethics and Standards of Professional Conduct

Subath Agarway, CFA, has recently joined CrowdWisdom as vice president, and he is in charge of due diligence.[7] Agarway is the ninth employee of CrowdWisdom, a young venture capital company that matches investors with startup companies in need of capital. His position at the online company is a newly created one. As head of the due diligence function, Agarway's role is to identify suitable companies for CrowdWisdom to offer to potential investors. Agarway is the only CFA charterholder on the team, which includes two co-founders, Craig Miller and Stephane Etienne. Both Miller and Etienne have substantial experience and strong networks from working at other industry startup companies.

Since its startup four years ago, CrowdWisdom has grown rapidly, funding 50 startup companies with almost $10 million from investors through its online matching platform. CrowdWisdom's business model markets to a wide range of startup companies seeking public capital. Startups in need of funds submit a listing application to CrowdWisdom. Application approval by CrowdWisdom's due diligence function allows companies to list on the platform for a fee, thereby becoming visible to platform investors as possible investments. Investors on the CrowdWisdom platform include both sophisticated and unsophisticated investors. Owing to a successful business model, Agarway is receiving an unprecedented number of applications from startups wishing to list on the company's platform.

The company's business plan calls for aggressive growth to maintain market share and secure CrowdWisdom's next round of funding. The founders' mandate is to list 100 companies on the CrowdWisdom platform in the next 18 months. In the longer term, the founders hope to do an initial public offering of CrowdWisdom's stock.

CrowdWisdom's early success has resulted in part from Miller's and Etienne's work in attracting platform investors who are willing to capitalize young startup companies. Leveraging their collective network, the founders created a large database of potential platform investors shortly after CrowdWisdom was created. As investors began investing on the platform, the founders pioneered an "Investor Club" whose members were the most active in providing capital through the CrowdWisdom platform. Investor Club members receive access to market intelligence research in addition to the research on CrowdWisdom listed companies that Agarway prepares and posts on the website.

To keep the database growing, Miller asks Agarway to consider companies whose customers appear to be a strong fit from a potential future investor standpoint. Agarway has experience marketing equity investments to customers of platform companies and knows that many companies have successfully raised funds by soliciting their own customers to become investors. Agarway also knows that CrowdWisdom's policies must comply with rules governing marketing over the internet, which include opt-in/opt-out preferences, age of person(s) marketed to, and required disclosures.

During the next two months, Agarway reviews the presentation materials for more than 100 companies that want to list on the CrowdWisdom platform. Agarway uses a process of due diligence he developed over several years, most recently as head of research for his previous employer, FunderWise, a lesser-known crowdfunding platform.

7 **Subath Agarway:** Cynthia Harrington, CFA. *Ethics Cases*. © 2017 CFA Institute. All rights reserved. Consistent with the 11th edition of the *Standards of Practice Handbook*.

Agarway's due diligence process consists of a two-step process he developed through trial and error at FunderWise. First, he reviews materials provided by companies to screen out those with a potential market for their product or service of less than $1 billion and those with perceived product or service viability concerns. Together, these criteria typically screen out 75% of applicant companies. Second, Agarway investigates the remaining companies by closely reviewing audited financial statements and interviewing company executives and customers. He is confident in his process and has personally invested in several FunderWise listed companies using this approach.

After considerable time and effort investigating the companies that made it past the first screen, Agarway's additional research leads him to reject almost all the remaining companies. The rejected companies appear to have issues with improper revenue recognition, questionable user claims, and regulatory litigation.

Of the few remaining applications, Agarway believes one of the most promising is that of a company called Deko, an information technology startup. Deko has impressive founders, attractive prospects, and a unique product. Additionally, Deko seems to have an enviable customer base that CrowdWisdom could approach for future investor funding activity. Deko is unique in that most of its users are preteens and teenagers who love Deko's software, which allows them to create digital collections of their possessions and then share these collections with their friends online. The company's strategy is to market its crowdfunded shares through email communications to the young users. The email contains an announcement on the company's crowdfunding offer and states the offer is available to adults over the age of 18. Companies with loyal users who often bring in friends and family have proven to be among the more successful at equity crowdfunding campaigns.

During this time, Agarway is asked to take on additional responsibilities. His days and evenings include speaking with founders of listed companies, answering investor questions, and working with attorneys to finalize listing transactions for new companies.

Several months later, after the founders present at two global startup conferences, Agarway's stack of applications for review grows to 300 companies. To meet CrowdWisdom's aggressive growth goals, Miller and Etienne suggest to Agarway that he target an application acceptance rate of 10%. They suggest Agarway research at least half of the applying companies in his second-stage process to meet the 10% acceptance rate.

Agarway patiently explains his process and his challenge in finding time to review applicants. In response, the founders suggest he find ways to reduce the time spent on each application. Miller and Etienne also recommend the acceptance of two companies whose founders Miller and Etienne met at the recent conferences.

Several activities in the case are or could be in violation of the Code and Standards. Identify violations or possible violations, state what actions Agarway and the firm should take to correct the violations, and make a short policy statement a firm could use to guide employees to help prevent similar violations from occurring in the future.

This case highlights challenges individuals may face during their careers when working for younger firms whose core business may not be traditional financial services or investment management, or when working for firms where they may be the only CFA charterholder or one of just a few CFA charterholders employed.

Professionalism

Standard I(A): Knowledge of the Law requires candidates and members to understand the applicable laws and regulations of the countries and jurisdictions where they engage in professional activities. Agarway should review the global rules governing online marketing to Deko's teen and preteen customers. The company strategy of offering equity to users' parents (or other adults in the household) through communications to its teen and preteen user base may put CrowdWisdom and Agarway at risk because it is illegal in many countries to collect information on such individuals over the internet without first obtaining parental permission. Unless Agarway can confirm that Deko is in compliance with this requirement, the use of CrowdWisdom's platform to solicit preteens could be against the law. Because Deko has cleared Agarway's due diligence process, if Deko were to be added to CrowdWisdom's platform, CrowdWisdom and Deko could be at risk for prosecution.

Actions Required

Although Agarway does not need to know the laws in every jurisdiction, he does need to stay informed about relevant legal limitations. Agarway and CrowdWisdom should establish a procedure whereby employees are regularly informed about changes in applicable laws and regulations. CrowdWisdom should also have legal counsel available to review planned additions to the platform to ensure that the company's strategy is not in conflict with relevant law.

Policy Statement for a Firm

"When determining whether a company should be included on the platform, careful consideration must be taken to determine whether the company's business strategy violates laws related to marketing and solicitation, particularly if the strategy targets minors or vulnerable adults (those with physical or mental disabilities). If the strategy targets minors or vulnerable adults, legal counsel will be consulted before listing the company on the platform."

Standard I(B): Independence and Objectivity obligates members and candidates to use reasonable care and judgment to achieve and maintain independence and objectivity in their professional activities. Pressure by CrowdWisdom's founders to modify Agarway's due diligence process to increase the number of listing approvals and shorten the review time frame is likely to put Agarway's independence and objectivity at risk.

Actions Required

Agarway and CrowdWisdom's senior leaders need to create and document a company-approved due diligence process, which will likely blend Agarway's past work with input from CrowdWisdom's founders. As a CFA charterholder, Agarway will need to be comfortable that the process has a reasonable basis and can be applied objectively.

Policy Statement for a Firm

"The selection of companies for inclusion on the platform will comply with the due diligence process approved by the firm's Board as detailed in the Selection Due Diligence Memorandum approved on 25 January 20XX."

Conflicts of Interest

Standard VI(A): Disclosure of Conflicts requires members and candidates to make full and fair disclosure of all matters that could reasonably be expected to impair their independence and objectivity. Identifying and managing conflicts is a reality of working in the investment industry, where conflicts are often present. When a conflict cannot be reasonably avoided, clear and complete disclosure of its existence is necessary. Some possible conflicts of interest exist in this scenario: CrowdWisdom's Investor Club selective access to additional market intelligence research and Agarway's personal investment in several companies that could be competitors of firms he is evaluating for the platform or future additions to the CrowdWisdom platform. Conflicts of interest may be inevitable and must be disclosed in a timely manner so that all parties involved can understand the circumstances and potential effects.

Actions Required

CrowdWisdom's Investor Club, which provides select investors with preferential access to additional market intelligence research, needs to be disclosed so that all investors can understand and evaluate the circumstances, the possible impact, and the potential disadvantage they may be placed at relative to Investor Club members.

In his personal portfolio, Agarway has invested in companies that could be competitors of firms he is reviewing in his due diligence work. His personal investments need to be disclosed to both his supervisor and CrowdWisdom's compliance officer. Additionally, if these firms also list on the CrowdWisdom platform, Agarway's personal investments would need to be disclosed to CrowdWisdom's users so that they can evaluate the independence and objectivity of each company's inclusion on the platform.

Policy Statement for a Firm

"Employees of the firm must disclose all personal investment holdings to the company's Compliance Officer, and that disclosure must be updated quarterly for public stocks and when invested for private holdings. All employee investments in companies that raise funds through the firm must be approved in advance by the Compliance Officer—or in the case of the founders, by the Board—before the transactions' closing and must be communicated to the firm's clients."

Glossary

Abandonment option The ability to terminate a project at some future time if the financial results are disappointing.

Abnormal earnings See *residual income*.

Abnormal return The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

Absolute convergence The idea that developing countries, regardless of their particular characteristics, will eventually catch up with the developed countries and match them in per capita output.

Absolute valuation model A model that specifies an asset's intrinsic value.

Absolute version of PPP An extension of the law of one price whereby the prices of goods and services will not differ internationally once exchange rates are considered.

Accounting estimates Estimates used in calculating the value of assets or liabilities and in the amount of revenue and expense to allocate to a period. Examples of accounting estimates include, among others, the useful lives of depreciable assets, the salvage value of depreciable assets, product returns, warranty costs, and the amount of uncollectible receivables.

Accumulated benefit obligation The actuarial present value of benefits (whether vested or non-vested) attributed, generally by the pension benefit formula, to employee service rendered before a specified date and based on employee service and compensation (if applicable) before that date. The accumulated benefit obligation differs from the projected benefit obligation in that it includes no assumption about future compensation levels.

Accuracy The percentage of correctly predicted classes out of total predictions. It is an overall performance metric in classification problems.

Acquirer The company in a merger or acquisition that is acquiring the target.

Acquiring company See *acquirer*.

Acquisition The purchase of some portion of one company by another; the purchase may be for assets, a definable segment of another entity, or the entire company.

Activation function A functional part of a neural network's node that transforms the total net input received into the final output of the node. The activation function operates like a light dimmer switch that decreases or increases the strength of the input.

Active factor risk The contribution to active risk squared resulting from the portfolio's different-than-benchmark exposures relative to factors specified in the risk model.

Active return The return on a portfolio minus the return on the portfolio's benchmark.

Active risk The standard deviation of active returns.

Active risk squared The variance of active returns; active risk raised to the second power.

Active share A measure of how similar a portfolio is to its benchmark. A manager who precisely replicates the benchmark will have an active share of zero; a manager with no holdings in common with the benchmark will have an active share of one.

Active specific risk The contribution to active risk squared resulting from the portfolio's active weights on individual assets as those weights interact with assets' residual risk.

Adjusted funds from operations (AFFO) Funds from operations adjusted to remove any non-cash rent reported under straight-line rent accounting and to subtract maintenance-type capital expenditures and leasing costs, including leasing agents' commissions and tenants' improvement allowances.

Adjusted present value As an approach to valuing a company, the sum of the value of the company, assuming no use of debt, and the net present value of any effects of debt on company value.

Adjusted R^2 A measure of goodness-of-fit of a regression that is adjusted for degrees of freedom and hence does not automatically increase when another independent variable is added to a regression.

Administrative regulations or administrative law Rules issued by government agencies or other regulators.

Advanced set An arrangement in which the reference interest rate is set at the time the money is deposited.

Advanced settled An arrangement in which a forward rate agreement (FRA) expires and settles at the same time, at the FRA expiration date.

Agency costs Costs associated with the conflict of interest present when a company is managed by non-owners. Agency costs result from the inherent conflicts of interest between managers and equity owners.

Agency costs of equity The smaller the stake managers have in the company, the less their share in bearing the cost of excessive perquisite consumption—consequently, the less their desire to give their best efforts in running the company.

Agency issues Conflicts of interest that arise when the agent in an agency relationship has goals and incentives that differ from the principal to whom the agent owes a fiduciary duty. Also called *agency problems* or *principal–agent problems*.

Agglomerative clustering A bottom-up hierarchical clustering method that begins with each observation being treated as its own cluster. The algorithm finds the two closest clusters, based on some measure of distance (similarity), and combines them into one new larger cluster. This process is repeated iteratively until all observations are clumped into a single large cluster.

Allowance for loan losses A balance sheet account; it is a contra asset account to loans.

Alpha The return on an asset in excess of the asset's required rate of return; the risk-adjusted return.

American Depositary Receipt A negotiable certificate issued by a depositary bank that represents ownership in a non-US company's deposited equity (i.e., equity held in custody by the depositary bank in the company's home market).

Analysis of variance (ANOVA) The analysis that breaks the total variability of a dataset (such as observations on the dependent variable in a regression) into components representing different sources of variation. With reference to regression, ANOVA provides the inputs for an *F*-test of

the significance of the regression as a whole, as well as the inputs for the coefficient of determination and the standard error of the estimate.

Application programming interface (API) A set of well-defined methods of communication between various software components and typically used for accessing external data.

Arbitrage (1) The simultaneous purchase of an undervalued asset or portfolio and sale of an overvalued but equivalent asset or portfolio in order to obtain a riskless profit on the price differential. Taking advantage of a market inefficiency in a risk-free manner. (2) The condition in a financial market in which equivalent assets or combinations of assets sell for two different prices, creating an opportunity to profit at no risk with no commitment of money. In a well-functioning financial market, few arbitrage opportunities are possible. (3) A risk-free operation that earns an expected positive net profit but requires no net investment of money.

Arbitrage-free models Term structure models that project future interest rate paths that emanate from the existing term structure. Resulting prices are based on a no-arbitrage condition.

Arbitrage-free valuation An approach to valuation that determines security values consistent with the absence of any opportunity to earn riskless profits without any net investment of money.

Arbitrage opportunity An opportunity to conduct an arbitrage; an opportunity to earn an expected positive net profit without risk and with no net investment of money.

Arbitrage portfolio The portfolio that exploits an arbitrage opportunity.

Ask price The price at which a trader will sell a specified quantity of a security. Also called *ask*, *offer price*, or *offer*.

Asset-based approach Approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities.

Asset-based valuation An approach to valuing natural resource companies that estimates company value on the basis of the market value of the natural resources the company controls.

Asset beta The unlevered beta; reflects the business risk of the assets; the asset's systematic risk.

Asset purchase An acquisition in which the acquirer purchases the target company's assets and payment is made directly to the target company.

Asymmetric information The differential of information between corporate insiders and outsiders regarding the company's performance and prospects. Managers typically have more information about the company's performance and prospects than owners and creditors.

At market contract When a forward contract is established, the forward price is negotiated so that the market value of the forward contract on the initiation date is zero.

Authorized participants (APs) A special group of institutional investors who are authorized by the ETF issuer to participate in the creation/redemption process. APs are large broker/dealers, often market makers.

Autocorrelations The correlations of a time series with its own past values.

Autoregressive model (AR) A time series regressed on its own past values in which the independent variable is a lagged value of the dependent variable.

Backtesting The process that approximates the real-life investment process, using historical data, to assess whether an investment strategy would have produced desirable results.

Backward integration A merger involving the purchase of a target ahead of the acquirer in the value or production chain; for example, to acquire a supplier.

Backward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving backward through the network's layers.

Backwardation A condition in futures markets in which the spot price exceeds the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is higher than the longer-term futures contract price.

Bag-of-words (BOW) A collection of a distinct set of tokens from all the texts in a sample dataset. BOW does not capture the position or sequence of words present in the text.

Bankruptcy A declaration provided for by a country's laws that typically involves the establishment of a legal procedure that forces creditors to defer their claims.

Barbell portfolio Fixed-income portfolio that combines short and long maturities.

Base error Model error due to randomness in the data.

Basic earnings per share (EPS) Net earnings available to common shareholders (i.e., net income minus preferred dividends) divided by the weighted average number of common shares outstanding during the period.

Basis The difference between the spot price and the futures price. As the maturity date of the futures contract nears, the basis converges toward zero.

Basis trade A trade based on the pricing of credit in the bond market versus the price of the same credit in the CDS market. To execute a basis trade, go long the "underpriced" credit and short the "overpriced" credit. A profit is realized as the implied credit prices converge.

Bear hug A tactic used by acquirers to circumvent target management's objections to a proposed merger by submitting the proposal directly to the target company's board of directors.

Bearish flattening Term structure shift in which short-term bond yields rise more than long-term bond yields, resulting in a flatter yield curve.

Benchmark value of the multiple In using the method of comparables, the value of a price multiple for the comparison asset; when we have comparison assets (a group), the mean or median value of the multiple for the group of assets.

Best ask The offer to sell with the lowest ask price. Also called *best offer* or *inside ask*.

Best bid The offer to buy with the highest bid price. Also called the *inside bid*.

Best offer See *best ask*.

Bias error Describes the degree to which a model fits the training data. Algorithms with erroneous assumptions produce high bias error with poor approximation, causing underfitting and high in-sample error.

Bid–ask spread The ask price minus the bid price.

Bid price The price at which a trader will buy a specified quantity of a security. Also called *bid.*

Bill-and-hold basis Sales on a bill-and-hold basis involve selling products but not delivering those products until a later date.

Blockage factor An illiquidity discount that occurs when an investor sells a large amount of stock relative to its trading volume (assuming it is not large enough to constitute a controlling ownership).

Bond indenture A legal contract specifying the terms of a bond issue.

Bond risk premium The expected excess return of a default-free long-term bond less that of an equivalent short-term bond.

Bond yield plus risk premium method An estimate of the cost of common equity that is produced by summing the before-tax cost of debt and a risk premium that captures the additional yield on a company's stock relative to its bonds. The additional yield is often estimated using historical spreads between bond yields and stock yields.

Bonding costs Costs borne by management to assure owners that they are working in the owners' best interest (e.g., implicit cost of non-compete agreements).

Bonus issue of shares *See stock dividend.*

Book value Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value of equity Shareholders' equity (total assets minus total liabilities) minus the value of preferred stock; common shareholders' equity.

Book value per share The amount of book value (also called carrying value) of common equity per share of common stock, calculated by dividing the book value of shareholders' equity by the number of shares of common stock outstanding.

Bootstrap aggregating (or bagging) A technique whereby the original training dataset is used to generate *n* new training datasets or bags of data. Each new bag of data is generated by random sampling with replacement from the initial training set.

Bootstrapping The use of a forward substitution process to determine zero-coupon rates by using the par yields and solving for the zero-coupon rates one by one, from the shortest to longest maturities.

Bottom-up approach With respect to forecasting, an approach that usually begins at the level of the individual company or a unit within the company.

Breakup value The value derived using a sum-of-the-parts valuation.

Breusch–Pagan test A test for conditional heteroskedasticity in the error term of a regression.

Bullet portfolio A fixed-income portfolio concentrated in a single maturity.

Bullish flattening Term structure change in which the yield curve flattens in response to a greater decline in long-term rates than short-term rates.

Bullish steepening Term structure change in which short-term rates fall by more than long-term yields, resulting in a steeper term structure.

Buy-side analysts Analysts who work for investment management firms, trusts, bank trust departments, and similar institutions.

Buyback *See share repurchase.*

Callable bond Bond that includes an embedded call option that gives the issuer the right to redeem the bond issue prior to maturity, typically when interest rates have fallen or when the issuer's credit quality has improved.

Canceled shares Shares that were issued, subsequently repurchased by the company, and then retired (cannot be reissued).

Cannibalization Cannibalization occurs when an investment takes customers and sales away from another part of the company.

Capital charge The company's total cost of capital in money terms.

Capital deepening An increase in the capital-to-labor ratio.

Capital rationing A capital rationing environment assumes that the company has a fixed amount of funds to invest.

Capital structure The mix of debt and equity that a company uses to finance its business; a company's specific mixture of long-term financing.

Capitalization of earnings method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capitalization rate The divisor in the expression for the value of perpetuity. In the context of real estate, it is the divisor in the direct capitalization method of estimating value. The cap rate equals net operating income divided by value.

Capitalized cash flow method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity. Also called *capitalized cash flow model.*

Capitalized income method In the context of private company valuation, a valuation model based on an assumption of a constant growth rate of free cash flow to the firm or a constant growth rate of free cash flow to equity.

Capped floater Floating-rate bond with a cap provision that prevents the coupon rate from increasing above a specified maximum rate. It protects the issuer against rising interest rates.

Carry arbitrage model A no-arbitrage approach in which the underlying instrument is either bought or sold along with an opposite position in a forward contract.

Carry benefits Benefits that arise from owning certain underlyings; for example, dividends, foreign interest, and bond coupon payments.

Carry costs Costs that arise from owning certain underlyings. They are generally a function of the physical characteristics of the underlying asset and also the interest forgone on the funds tied up in the asset.

Cash available for distribution See *adjusted funds from operations.*

Cash-generating unit The smallest identifiable group of assets that generates cash inflows that are largely independent of the cash inflows of other assets or groups of assets.

Cash offering A merger or acquisition that is to be paid for with cash; the cash for the merger might come from the acquiring company's existing assets or from a debt issue.

Cash settlement A procedure used in certain derivative transactions that specifies that the long and short parties settle the derivative's difference in value between them by making a cash payment.

Catalyst An event or piece of information that causes the marketplace to re-evaluate the prospects of a company.

Categorical dependent variables An alternative term for qualitative dependent variables.

CDS spread A periodic premium paid by the buyer to the seller that serves as a return over a market reference rate required to protect against credit risk.

Ceiling analysis A systematic process of evaluating different components in the pipeline of model building. It helps to understand what part of the pipeline can potentially improve in performance by further tuning.

Centroid The center of a cluster formed using the *k*-means clustering algorithm.

Chain rule of forecasting A forecasting process in which the next period's value as predicted by the forecasting equation is substituted into the right-hand side of the equation to give a predicted value two periods ahead.

Cheapest-to-deliver The debt instrument that can be purchased and delivered at the lowest cost yet has the same seniority as the reference obligation.

Classification and regression tree A supervised machine learning technique that can be applied to predict either a categorical target variable, producing a classification tree, or a continuous target variable, producing a regression tree. CART is commonly applied to binary classification or regression.

Clean surplus relation The relationship between earnings, dividends, and book value in which ending book value is equal to the beginning book value plus earnings less dividends, apart from ownership transactions.

Club convergence The idea that only rich and middle-income countries sharing a set of favorable attributes (i.e., are members of the "club") will converge to the income level of the richest countries.

Cluster A subset of observations from a dataset such that all the observations within the same cluster are deemed "similar."

Clustering The sorting of observations into groups (clusters) such that observations in the same cluster are more similar to each other than they are to observations in other clusters.

Cobb–Douglas production function A function of the form $Y = K^{\alpha} L^{1-\alpha}$ relating output (Y) to labor (L) and capital (K) inputs.

Coefficient of determination The percentage of the variation of the dependent variable that is explained by the independent variable. Also referred to as the "*R*-squared" or "R^2."

Cointegrated Describes two time series that have a long-term financial or economic relationship such that they do not diverge from each other without bound in the long run.

Collateral return The component of the total return on a commodity futures position attributable to the yield for the bonds or cash used to maintain the futures position. Also called *collateral yield*.

Collection frequency (CF) The number of times a given word appears in the whole corpus (i.e., collection of sentences) divided by the total number of words in the corpus.

Commercial real estate properties Income-producing real estate properties; properties purchased with the intent to let, lease, or rent (in other words, produce income).

Commodity swap A type of swap involving the exchange of payments over multiple dates as determined by specified reference prices or indexes relating to commodities.

Common size statements Financial statements in which all elements (accounts) are stated as a percentage of a key figure, such as revenue for an income statement or total assets for a balance sheet.

Company fundamental factors Factors related to the company's internal performance, such as factors relating to earnings growth, earnings variability, earnings momentum, and financial leverage.

Company share-related factors Valuation measures and other factors related to share price or the trading characteristics of the shares, such as earnings yield, dividend yield, and book-to-market value.

Comparables Assets used as benchmarks when applying the method of comparables to value an asset. Also called *comps*, *guideline assets*, or *guideline companies*.

Competition laws A law that promotes or maintains market competition by regulating anti-competitive conduct. Known as "antitrust law" in the United States, "anti-monopoly law" in China and Russia, and often referred to as "trade practices law" in the United Kingdom and Australia.

Compiled financial statements Financial statements that are not accompanied by an auditor's opinion letter.

Complexity A term referring to the number of features, parameters, or branches in a model and to whether the model is linear or non-linear (non-linear is more complex).

Composite variable A variable that combines two or more variables that are statistically strongly related to each other.

Comprehensive income All changes in equity other than contributions by, and distributions to, owners; income under clean surplus accounting; includes all changes in equity during a period except those resulting from investments by owners and distributions to owners. Comprehensive income equals net income plus other comprehensive income.

Comps Assets used as benchmarks when applying the method of comparables to value an asset.

Concentrated ownership Ownership structure consisting of an individual shareholder or a group (controlling shareholders) with the ability to exercise control over the corporation.

Conditional convergence The idea that convergence of per capita income is conditional on the countries having the same savings rate, population growth rate, and production function.

Conditional heteroskedasticity Heteroskedasticity in the error variance that is correlated with the values of the independent variable(s) in the regression.

Conditional VaR (CVaR) The weighted average of all loss outcomes in the statistical (i.e., return) distribution that exceed the VaR loss. Thus, CVaR is a more comprehensive measure of tail loss than VaR is. Sometimes referred to as the *expected tail loss* or *expected shortfall*.

Confusion matrix A grid used for error analysis in classification problems, it presents values for four evaluation metrics including true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Conglomerate discount The discount possibly applied by the market to the stock of a company operating in multiple, unrelated businesses.

Conglomerate merger A merger involving companies that are in unrelated businesses.

Consolidation The combining of the results of operations of subsidiaries with the parent company to present financial statements as if they were a single economic unit. The assets, liabilities, revenues, and expenses of the subsidiaries are combined with those of the parent company, eliminating intercompany transactions.

Constant dividend payout ratio policy A policy in which a constant percentage of net income is paid out in dividends.

Constant returns to scale The condition that if all inputs into the production process are increased by a given percentage, then output rises by that same percentage.

Contango A condition in futures markets in which the spot price is lower than the futures price; also, the condition in which the near-term (closer to expiration) futures contract price is lower than the longer-term futures contract price.

Contingent consideration Potential future payments to the seller that are contingent on the achievement of certain agreed-on occurrences.

Continuing earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *persistent earnings*, or *underlying earnings*.

Continuing residual income Residual income after the forecast horizon.

Continuing value The analyst's estimate of a stock's value at a particular point in the future.

Control premium An increment or premium to value associated with a controlling ownership interest in a company.

Convergence The property by which as expiration approaches, the price of a newly created forward or futures contract will approach the price of a spot transaction. At expiration, a forward or futures contract is equivalent to a spot transaction in the underlying.

Conversion period For a convertible bond, the period during which bondholders have the right to convert their bonds into shares.

Conversion price For a convertible bond, the price per share at which the bond can be converted into shares.

Conversion rate (or ratio) For a convertible bond, the number of shares of common stock that a bondholder receives from converting the bond into shares.

Conversion value For a convertible bond, the value of the bond if it is converted at the market price of the shares. Also called *parity value*.

Convertible bond Bond with an embedded conversion option that gives bondholders the right to convert their bonds into the issuer's common stock during a pre-determined period at a pre-determined price.

Convexity A measure of how interest rate sensitivity changes with a change in interest rates.

Core earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *persistent earnings*, or *underlying earnings*.

Core real estate investment style Investing in high-quality, well-leased, core property types with low leverage (no more than 30% of asset value) in the largest markets with strong, diversified economies. It is a conservative strategy designed to avoid real estate–specific risks, including leasing, development, and speculation in favor of steady returns. Hotel properties are excluded from the core categories because of the higher cash flow volatility resulting from single-night leases and the greater importance of property operations, brand, and marketing.

Corpus A collection of text data in any form, including list, matrix, or data table forms.

Cost approach An approach that values a private company based on the values of the underlying assets of the entity less the value of any related liabilities. In the context of real estate, this approach estimates the value of a property based on what it would cost to buy the land and construct a new property on the site that has the same utility or functionality as the property being appraised.

Cost of carry model A model that relates the forward price of an asset to the spot price by considering the cost of carry (also referred to as future-spot parity model).

Cost of debt The cost of debt financing to a company, such as when it issues a bond or takes out a bank loan.

Cost of equity The required rate of return on common stock.

Covariance stationary Describes a time series when its expected value and variance are constant and finite in all periods and when its covariance with itself for a fixed number of periods in the past or future is constant and finite in all periods.

Covered bonds A senior debt obligation of a financial institution that gives recourse to the originator/issuer and a predetermined underlying collateral pool.

Covered interest rate parity The relationship among the spot exchange rate, the forward exchange rate, and the interest rates in two currencies that ensures that the return on a hedged (i.e., covered) foreign risk-free investment is the same as the return on a domestic risk-free investment. Also called *interest rate parity*.

Cox-Ingersoll-Ross model A general equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is directly related to the level of interest rates.

Creation basket The list of securities (and share amounts) the authorized participant (AP) must deliver to the ETF manager in exchange for ETF shares. The creation basket is published each business day.

Creation/redemption The process in which ETF shares are created or redeemed by authorized participants transacting with the ETF issuer.

Creation units Large blocks of ETF shares transacted between the authorized participant (AP) and the ETF manager that are usually but not always equal to 50,000 shares of the ETF.

Credit correlation The correlation of credit (or default) risks of the underlying single-name CDS contained in an index CDS.

Credit curve The credit spreads for a range of maturities of a company's debt.

Credit default swap A derivative contract between two parties in which the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit derivative A derivative instrument in which the underlying is a measure of the credit quality of a borrower.

Credit event The event that triggers a payment from the credit protection seller to the credit protection buyer.

Credit protection buyer One party to a credit default swap; the buyer makes a series of cash payments to the seller and receives a promise of compensation for credit losses resulting from the default.

Credit protection seller One party to a credit default swap; the seller makes a promise to pay compensation for credit losses resulting from the default.

Credit risk The risk that the borrower will not repay principal and interest. Also called *default risk*.

Credit valuation adjustment The value of the credit risk of a bond in present value terms.

Cross-validation A technique for estimating out-of-sample error directly by determining the error in validation samples.

Current exchange rate For accounting purposes, the spot exchange rate on the balance sheet date.

Current rate method Approach to translating foreign currency financial statements for consolidation in which all assets and liabilities are translated at the current exchange rate. The current rate method is the prevalent method of translation.

Curvature One of the three factors (the other two are level and steepness) that empirically explain most of the changes in the shape of the yield curve. A shock to the curvature factor affects mid-maturity interest rates, resulting in the term structure becoming either more or less hump-shaped.

Curve trade Buying a CDS of one maturity and selling a CDS on the same reference entity with a different maturity.

Cyclical businesses Businesses with high sensitivity to business- or industry-cycle influences.

Data mining The practice of determining a model by extensive searching through a dataset for statistically significant patterns.

Data preparation (cleansing) The process of examining, identifying, and mitigating (i.e., cleansing) errors in raw data.

Data snooping The subconscious or conscious manipulation of data in a way that produces a statistically significant result (i.e., the *p*-value is sufficiently small or the *t*-statistic sufficiently large to indicate statistical significance), such as by running multiple simulations and naively accepting the best result. Also known as p-hacking.

Data wrangling (preprocessing) This task performs transformations and critical processing steps on cleansed data to make the data ready for ML model training (i.e., preprocessing), and includes dealing with outliers, extracting useful variables from existing data points, and scaling the data.

"Dead-hand" provision A poison pill provision that allows for the redemption or cancellation of a poison pill provision only by a vote of continuing directors (generally directors who were on the target company's board prior to the takeover attempt).

Debt rating An objective measure of the quality and safety of a company's debt based upon an analysis of the company's ability to pay the promised cash flows. It includes an analysis of any indentures.

Deep learning Algorithms based on deep neural networks, ones with many hidden layers (more than two), that address highly complex tasks, such as image classification, face recognition, speech recognition, and natural language processing.

Deep neural networks Neural networks with many hidden layers—at least 2 but potentially more than 20—that have proven successful across a wide range of artificial intelligence applications.

Default risk See *credit risk*.

Defined benefit pension plans Plan in which the company promises to pay a certain annual amount (defined benefit) to the employee after retirement. The company bears the investment risk of the plan assets.

Defined contribution pension plans Individual accounts to which an employee and typically the employer makes contributions, generally on a tax-advantaged basis. The amounts of contributions are defined at the outset, but the future value of the benefit is unknown. The employee bears the investment risk of the plan assets.

Definitive merger agreement A contract signed by both parties to a merger that clarifies the details of the transaction, including the terms, warranties, conditions, termination details, and the rights of all parties.

Delay costs Implicit trading costs that arise from the inability to complete desired trades immediately. Also called *slippage.*

Delta The relationship between the option price and the underlying price, which reflects the sensitivity of the price of the option to changes in the price of the underlying. Delta is a good approximation of how an option price will change for a small change in the stock.

Dendrogram A type of tree diagram used for visualizing a hierarchical cluster analysis; it highlights the hierarchical relationships among the clusters.

Dependent variable The variable whose variation about its mean is to be explained by the regression; the left-side variable in a regression equation. Also referred to as the *explained variable*.

Depository Trust and Clearinghouse Corporation A US-headquartered entity providing post-trade clearing, settlement, and information services.

Descriptive statistics The study of how data can be summarized effectively.

Diluted earnings per share (Diluted EPS)Net income, minus preferred dividends, divided by the weighted average number of common shares outstanding considering all dilutive securities (e.g., convertible debt and options); the EPS that would result if all dilutive securities were converted into common shares.

Dilution A reduction in proportional ownership interest as a result of the issuance of new shares.

Dimension reduction A set of techniques for reducing the number of features in a dataset while retaining variation across observations to preserve the information contained in that variation.

Diminishing marginal productivity When each additional unit of an input, keeping the other inputs unchanged, increases output by a smaller increment.

Direct capitalization method In the context of real estate, this method estimates the value of an income-producing property based on the level and quality of its net operating income.

Discount To reduce the value of a future payment in allowance for how far away it is in time; to calculate the present value of some future amount. Also, the amount by which an instrument is priced below its face value.

Discount factor The present value or price of a risk-free single-unit payment when discounted using the appropriate spot rate.

Discount for lack of control An amount or percentage deducted from the pro rata share of 100% of the value of an equity interest in a business to reflect the absence of some or all of the powers of control.

Discount for lack of marketability An amount of percentage deducted from the value of an ownership interest to reflect the relative absence of marketability.

Discount function Discount factors for the range of all possible maturities. The spot curve can be derived from the discount function and vice versa.

Discount rate Any rate used in finding the present value of a future cash flow.

Discounted abnormal earnings model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Discounted cash flow (DCF) analysis In the context of merger analysis, an estimate of a target company's value found by discounting the company's expected future free cash flows to the present.

Discounted cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows. In the context of real estate, this method estimates the value of an income-producing property based on discounting future projected cash flows.

Discounted cash flow model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Discriminant analysis A multivariate classification technique used to discriminate between groups, such as companies that either will or will not become bankrupt during some time frame.

Dispersed ownership Ownership structure consisting of many shareholders, none of which has the ability to individually exercise control over the corporation.

Divestiture The sale, liquidation, or spin-off of a division or subsidiary.

Dividend A distribution paid to shareholders based on the number of shares owned.

Dividend coverage ratio The ratio of net income to dividends.

Dividend discount model (DDM) A present value model of stock value that views the intrinsic value of a stock as present value of the stock's expected future dividends.

Dividend displacement of earnings The concept that dividends paid now displace earnings in all future periods.

Dividend imputation tax system A taxation system that effectively assures corporate profits distributed as dividends are taxed just once and at the shareholder's tax rate.

Dividend index point A measure of the quantity of dividends attributable to a particular index.

Dividend payout ratio The ratio of cash dividends paid to earnings for a period.

Dividend policy The strategy a company follows with regard to the amount and timing of dividend payments.

Dividend rate The annualized amount of the most recent dividend.

Dividend yield Annual dividends per share divided by share price.

Divisive clustering A top-down hierarchical clustering method that starts with all observations belonging to a single large cluster. The observations are then divided into two clusters based on some measure of distance (similarity). The algorithm then progressively partitions the intermediate clusters into smaller ones until each cluster contains only one observation.

Document frequency (DF) The number of documents (texts) that contain a particular token divided by the total number of documents. It is the simplest feature selection method and often performs well when many thousands of tokens are present.

Document term matrix (DTM) A matrix where each row belongs to a document (or text file), and each column represents a token (or term). The number of rows is equal to the number of documents (or text files) in a sample text dataset. The number of columns is equal to the number of tokens from the BOW built using all the documents in the sample dataset. The cells typically contain the counts of the number of times a token is present in each document.

Dominance An arbitrage opportunity when a financial asset with a risk-free payoff in the future must have a positive price today.

Double taxation system Corporate earnings are taxed twice when paid out as dividends. First, corporate pretax earnings are taxed regardless of whether they will be distributed as dividends or retained at the corporate level. Second, dividends are taxed again at the individual shareholder level.

Downstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary) such that the investor company records a profit on its income statement. An example is a sale of inventory by the investor company to the associate or by a parent to a subsidiary company.

Dual-class shares Shares that grant one share class superior or even sole voting rights, whereas the other share class has inferior or no voting rights.

Due diligence Investigation and analysis in support of a recommendation; the failure to exercise due diligence may sometimes result in liability according to various securities laws.

Dummy variable A type of qualitative variable that takes on a value of 1 if a particular condition is true and 0 if that condition is false.

Duration A measure of the approximate sensitivity of a security to a change in interest rates (i.e., a measure of interest rate risk).

Dutch disease A situation in which currency appreciation driven by strong export demand for resources makes other segments of the economy (particularly manufacturing) globally uncompetitive.

Earnings surprise The difference between reported EPS and expected EPS. Also referred to as *unexpected earnings*.

Earnings yield EPS divided by price; the reciprocal of the P/E.

Economic profit See *residual income*.

Economic sectors Large industry groupings.

Economic value added (EVA®) A commercial implementation of the residual income concept; the computation of EVA® is the net operating profit after taxes minus the cost of capital, where these inputs are adjusted for a number of items.

Economies of scale A situation in which average costs per unit of good or service produced fall as volume rises. In reference to mergers, the savings achieved through the consolidation of operations and elimination of duplicate resources.

Edwards–Bell–Ohlson model A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share.

Effective convexity Sensitivity of duration to changes in interest rates.

Effective duration Sensitivity of the bond's price to a 100 bps parallel shift of the benchmark yield curve, assuming no change in the bond's credit spread.

Effective spread Two times the difference between the execution price and the midpoint of the market quote at the time an order is entered.

Eigenvalue A measure that gives the proportion of total variance in the initial dataset that is explained by each eigenvector.

Eigenvector A vector that defines new mutually uncorrelated composite variables that are linear combinations of the original features.

Embedded options Contingency provisions found in a bond's indenture or offering circular representing rights that enable their holders to take advantage of interest rate movements. They can be exercised by the issuer, by the bondholder, or automatically depending on the course of interest rates.

Ensemble learning A technique of combining the predictions from a collection of models to achieve a more accurate prediction.

Ensemble method The method of combining multiple learning algorithms, as in ensemble learning.

Enterprise value Total company value (the market value of debt, common equity, and preferred equity) minus the value of cash and investments.

Enterprise value multiple A valuation multiple that relates the total market value of all sources of a company's capital (net of cash) to a measure of fundamental value for the entire company (such as a pre-interest earnings measure).

Equilibrium The condition in which supply equals demand.

Equity carve-out A form of restructuring that involves the creation of a new legal entity and the sale of equity in it to outsiders.

Equity charge The estimated cost of equity capital in money terms.

Equity REITs REITs that own, operate, and/or selectively develop income-producing real estate.

Equity swap A swap transaction in which at least one cash flow is tied to the return on an equity portfolio position, often an equity index.

Error autocorrelations The autocorrelations of the error term.

Error term The difference between an observation and its expected value, where the expected value is based on the true underlying population relation between the dependent and independent variables. Also known simply as the *error*.

ESG integration An ESG investment approach that focuses on systematic consideration of material ESG factors in asset allocation, security selection, and portfolio construction decisions for the purpose of achieving the product's stated investment objectives.

Estimated parameters With reference to a regression analysis, the estimated values of the population intercept and population slope coefficients in a regression.

***Ex ante* tracking error** A measure of the degree to which the performance of a given investment portfolio might be expected to deviate from its benchmark; also known as *relative VaR*.

***Ex ante* version of PPP** The hypothesis that expected changes in the spot exchange rate are equal to expected differences in national inflation rates. An extension of relative purchasing power parity to expected future changes in the exchange rate.

Ex-dividend Trading ex-dividend refers to shares that no longer carry the right to the next dividend payment.

Ex-dividend date The first date that a share trades without (i.e., "ex") the right to receive the declared dividend for the period.

Excess earnings method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Exchange ratio The number of shares that target stockholders are to receive in exchange for each of their shares in the target company.

Exercise date The date when employees actually exercise stock options and convert them to stock.

Exercise value The value of an option if it were exercised. Also sometimes called *intrinsic value*.

Expanded CAPM An adaptation of the CAPM that adds to the CAPM a premium for small size and company-specific risk.

Expectations approach A procedure for obtaining the value of an option derived from discounting at the risk-free rate its expected future payoff based on risk neutral probabilities.

Expected exposure The projected amount of money an investor could lose if an event of default occurs, before factoring in possible recovery.

Expected holding-period return The expected total return on an asset over a stated holding period; for stocks, the sum of the expected dividend yield and the expected price appreciation over the holding period.

Expected shortfall See *conditional VaR*.

Expected tail loss See *conditional VaR*.

Exploratory data analysis (EDA) The preliminary step in data exploration, where graphs, charts, and other visualizations (heat maps and word clouds) as well as quantitative methods (descriptive statistics and central tendency measures) are used to observe and summarize data.

Exposure to foreign exchange risk The risk of a change in value of an asset or liability denominated in a foreign currency due to a change in exchange rates.

Extendible bond Bond with an embedded option that gives the bondholder the right to keep the bond for a number of years after maturity, possibly with a different coupon.

External growth Company growth in output or sales that is achieved by buying the necessary resources externally (i.e., achieved through mergers and acquisitions).

Extra dividend *See special dividend*.

F1 score The harmonic mean of precision and recall. F1 score is a more appropriate overall performance metric (than accuracy) when there is unequal class distribution in the dataset and it is necessary to measure the equilibrium of precision and recall.

Factor A common or underlying element with which several variables are correlated.

Factor betas An asset's sensitivity to a particular factor; a measure of the response of return to each unit of increase in a factor, holding all other factors constant.

Factor portfolio See *pure factor portfolio*.

Factor price The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors.

Factor risk premium The expected return in excess of the risk-free rate for a portfolio with a sensitivity of 1 to one factor and a sensitivity of 0 to all other factors. Also called *factor price*.

Factor sensitivity See *factor betas*.

Failure to pay When a borrower does not make a scheduled payment of principal or interest on any outstanding obligations after a grace period.

Fair market value The market price of an asset or liability that trades regularly.

Fair value The amount at which an asset (or liability) could be bought (or incurred) or sold (or settled) in a current transaction between willing parties, that is, other than in a forced or liquidation sale. As defined in IFRS and US GAAP, it is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date.

Feature engineering A process of creating new features by changing or transforming existing features.

Feature selection A process whereby only pertinent features from the dataset are selected for model training. Selecting fewer features decreases model complexity and training time.

Features The independent variables (X's) in a labeled dataset.

Financial contagion A situation in which financial shocks spread from their place of origin to other locales. In essence, a faltering economy infects other, healthier economies.

Financial distress Heightened uncertainty regarding a company's ability to meet its various obligations because of lower or negative earnings.

Financial transaction A purchase involving a buyer having essentially no material synergies with the target (e.g., the purchase of a private company by a company in an unrelated industry or by a private equity firm would typically be a financial transaction).

First-differencing A transformation that subtracts the value of the time series in period $t - 1$ from its value in period t.

First-order serial correlation Correlation between adjacent observations in a time series.

Fitting curve A curve which shows in- and out-of-sample error rates (E_{in} and E_{out}) on the y-axis plotted against model complexity on the x-axis.

Fixed price tender offer Offer made by a company to repurchase a specific number of shares at a fixed price that is typically at a premium to the current market price.

Fixed-rate perpetual preferred stock Non-convertible, non-callable preferred stock with a specified dividend rate that has a claim on earnings senior to the claim of common stock, and no maturity date.

Flight to quality During times of market stress, investors sell higher-risk asset classes such as stocks and commodities in favor of default-risk-free government bonds.

Flip-in pill A poison pill takeover defense that dilutes an acquirer's ownership in a target by giving other existing target company shareholders the right to buy additional target company shares at a discount.

Flip-over pill A poison pill takeover defense that gives target company shareholders the right to purchase shares of the acquirer at a significant discount to the market price, which has the effect of causing dilution to all existing acquiring company shareholders.

Float Amounts collected as premium and not yet paid out as benefits.

Floored floater Floating-rate bond with a floor provision that prevents the coupon rate from decreasing below a specified minimum rate. It protects the investor against declining interest rates.

Flotation cost Fees charged to companies by investment bankers and other costs associated with raising new capital.

Forced conversion For a convertible bond, when the issuer calls the bond and forces bondholders to convert their bonds into shares, which typically happens when the underlying share price increases above the conversion price.

Foreign currency transactions Transactions that are denominated in a currency other than a company's functional currency.

Forward curve The term structure of forward rates for loans made on a specific initiation date.

Forward dividend yield A dividend yield based on the anticipated dividend during the next 12 months.

Forward integration A merger involving the purchase of a target that is farther along the value or production chain; for example, to acquire a distributor.

Forward P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Forward price The fixed price or rate at which the transaction, scheduled to occur at the expiration of a forward contract, will take place. This price is agreed to at the initiation date of the forward contract.

Forward pricing model The model that describes the valuation of forward contracts.

Forward propagation The process of adjusting weights in a neural network, to reduce total error of the network, by moving forward through the network's layers.

Forward rate An interest rate determined today for a loan that will be initiated in a future period.

Forward rate agreement An over-the-counter forward contract in which the underlying is an interest rate on a deposit. A forward rate agreement (FRA) calls for one party to make a fixed interest payment and the other to make an interest payment at a rate to be determined at contract expiration.

Forward rate model The forward pricing model expressed in terms of spot and forward interest rates.

Forward rate parity The proposition that the forward exchange rate is an unbiased predictor of the future spot exchange rate.

Forward value The monetary value of an existing forward contract.

Franking credit A tax credit received by shareholders for the taxes that a corporation paid on its distributed earnings.

Free cash flow The actual cash that would be available to the company's investors after making all investments necessary to maintain the company as an ongoing enterprise (also referred to as free cash flow to the firm); the internally generated funds that can be distributed to the company's investors (e.g., shareholders and bondholders) without impairing the value of the company.

Free cash flow hypothesis The hypothesis that higher debt levels discipline managers by forcing them to make fixed debt service payments and by reducing the company's free cash flow.

Free cash flow method Income approach that values an asset based on estimates of future cash flows discounted to present value by using a discount rate reflective of the risks associated with the cash flows.

Free cash flow to equity The cash flow available to a company's common shareholders after all operating expenses, interest, and principal payments have been made and necessary investments in working and fixed capital have been made.

Free cash flow to equity model A model of stock valuation that views a stock's intrinsic value as the present value of expected future free cash flows to equity.

Free cash flow to the firm The cash flow available to the company's suppliers of capital after all operating expenses (including taxes) have been paid and necessary investments in working and fixed capital have been made.

Free cash flow to the firm model A model of stock valuation that views the value of a firm as the present value of expected future free cash flows to the firm.

Frequency analysis The process of quantifying how important tokens are in a sentence and in the corpus as a whole. It helps in filtering unnecessary tokens (or features).

Friendly transaction A potential business combination that is endorsed by the managers of both companies.

Functional currency The currency of the primary economic environment in which an entity operates.

Fundamental factor models A multifactor model in which the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices.

Fundamentals Economic characteristics of a business, such as profitability, financial strength, and risk.

Funds available for distribution (FAD) See *adjusted funds from operations.*

Funds from operations (FFO) Net income (computed in accordance with generally accepted accounting principles) *plus* (1) gains and losses from sales of properties and (2) depreciation and amortization.

Futures price The price at which the parties to a futures contract agree to exchange the underlying (or cash). In commodity markets, the price agreed on to deliver or receive a defined quantity (and often quality) of a commodity at a future date.

Futures value The monetary value of an existing futures contract.

FX carry trade An investment strategy that involves taking long positions in high-yield currencies and short positions in low-yield currencies.

Gamma A measure of how sensitive an option's delta is to a change in the underlying. The change in a given instrument's delta for a given small change in the underlying's value, holding everything else constant.

Generalize When a model retains its explanatory power when predicting out-of-sample (i.e., using new data).

Generalized least squares A regression estimation technique that addresses heteroskedasticity of the error term.

Going-concern assumption The assumption that the business will maintain its business activities into the foreseeable future.

Going-concern value A business's value under a going-concern assumption.

Goodwill An intangible asset that represents the excess of the purchase price of an acquired company over the value of the net identifiable assets acquired.

Grant date The day that stock options are granted to employees.

Green bond Bonds in which the proceeds are designated by issuers to fund a specific project or portfolio of projects that have environmental or climate benefits.

Greenmail The purchase of the accumulated shares of a hostile investor by a company that is targeted for takeover by that investor, usually at a substantial premium over market price.

Greenwashing The risk that a green bond's proceeds are not actually used for a beneficial environmental or climate-related project.

Grid search A method of systematically training a model by using various combinations of hyperparameter values, cross validating each model, and determining which combination of hyperparameter values ensures the best model performance.

Gross domestic product A money measure of the goods and services produced within a country's borders over a stated period.

Gross lease A lease under which the tenant pays a gross rent to the landlord, who is responsible for all operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Ground truth The known outcome (i.e., target variable) of each observation in a labelled dataset.

Growth accounting equation The production function written in the form of growth rates. For the basic Cobb–Douglas production function, it states that the growth rate of output equals the rate of technological change plus α multiplied by the growth rate of capital plus $(1 - \alpha)$ multiplied by the growth rate of labor.

Growth capital expenditures Capital expenditures needed for expansion.

Growth option The ability to make additional investments in a project at some future time if the financial results are strong. Also called *expansion option.*

Guideline assets Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline companies Assets used as benchmarks when applying the method of comparables to value an asset.

Guideline public companies Public-company comparables for the company being valued.

Guideline public company method A variation of the market approach; establishes a value estimate based on the observed multiples from trading activity in the shares of public companies viewed as reasonably comparable to the subject private company.

Guideline transactions method A variation of the market approach; establishes a value estimate based on pricing multiples derived from the acquisition of control of entire public or private companies that were acquired.

Harmonic mean A type of weighted mean computed by averaging the reciprocals of the observations and then taking the reciprocal of that average.

Hazard rate The probability that an event will occur, given that it has not already occurred.

Hedonic index Unlike a repeat-sales index, a hedonic index does not require repeat sales of the same property. It requires only one sale. The way it controls for the fact that different properties are selling each quarter is to include variables in the regression that control for differences in the characteristics of the property, such as size, age, quality of construction, and location.

Heteroskedastic With reference to the error term of regression, having a variance that differs across observations.

Heteroskedasticity The property of having a nonconstant variance; refers to an error term with the property that its variance differs across observations.

Heteroskedasticity-consistent standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Hierarchical clustering An iterative unsupervised learning procedure used for building a hierarchy of clusters.

Highest and best use The concept that the best use of a vacant site is the use that would result in the highest value for the land. Presumably, the developer that could earn the highest risk-adjusted profit based on time, effort, construction and development cost, leasing, and exit value would be the one to pay the highest price for the land.

Historical exchange rates For accounting purposes, the exchange rates that existed when the assets and liabilities were initially recorded.

Historical scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.
Historical simulation A simulation method that uses past return data and a random number generator that picks observations from the historical series to simulate an asset's future returns.
Historical simulation method The application of historical price changes to the current portfolio.
Historical stress testing The process that tests how investment strategies would perform under some of the most negative (i.e., adverse) combinations of events and scenarios.
Ho–Lee model The first arbitrage-free term structure model. The model is calibrated to market data and uses a binomial lattice approach to generate a distribution of possible future interest rates.
Holding period return The return that an investor earns during a specified holding period; a synonym for total return.
Holdout samples Data samples that are not used to train a model.
Homoskedasticity The property of having a constant variance; refers to an error term that is constant across observations.
Horizontal merger A merger involving companies in the same line of business, usually as competitors.
Horizontal ownership Companies with mutual business interests (e.g., key customers or suppliers) that have cross-holding share arrangements with each other.
Hostile transaction An attempt to acquire a company against the wishes of the target's managers.
Human capital The accumulated knowledge and skill that workers acquire from education, training, or life experience.
Hybrid approach With respect to forecasting, an approach that combines elements of both top-down and bottom-up analyses.
Hyperparameter A parameter whose value must be set by the researcher before learning begins.
I-spreads Shortened form of "interpolated spreads" and a reference to a linearly interpolated yield.
Illiquidity discount A reduction or discount to value that reflects the lack of depth of trading or liquidity in that asset's market.
Impairment Diminishment in value as a result of carrying (book) value exceeding fair value and/or recoverable value.
Impairment of capital rule A legal restriction that dividends cannot exceed retained earnings.
Implementation shortfall The difference between the money return (or value) on a notional or paper portfolio and the actual portfolio return (or value).
Implied volatility The standard deviation that causes an option pricing model to give the current option price.
In-sample forecast errors The residuals from a fitted time-series model within the sample period used to fit the model.
iNAVs "Indicated" net asset values are intraday "fair value" estimates of an ETF share based on its creation basket.
Income approach A valuation approach that values an asset as the present discounted value of the income expected from it. In the context of real estate, this approach estimates the value of a property based on an expected rate of return. The estimated value is the present value of the expected future income from the property, including proceeds from resale at the end of a typical investment holding period.
Incremental VaR (IVaR) A measure of the incremental effect of an asset on the VaR of a portfolio by measuring the difference between the portfolio's VaR while including a specified asset and the portfolio's VaR with that asset eliminated.
Indenture A written contract between a lender and borrower that specifies the terms of the loan, such as interest rate, interest payment schedule, or maturity.
Independent board directors Directors with no material relationship with the company with regard to employment, ownership, or remuneration.
Independent regulators Regulators recognized and granted authority by a government body or agency. They are not government agencies per se and typically do not rely on government funding.
Independent variable A variable used to explain the dependent variable in a regression; a right-side variable in a regression equation. Also referred to as the *explanatory variable*.
Index CDS A type of credit default swap that involves a combination of borrowers.
Indicator variable A variable that takes on only one of two values, 0 or 1, based on a condition. In simple linear regression, the slope is the difference in the dependent variable for the two conditions. Also referred to as a *dummy variable*.
Industry structure An industry's underlying economic and technical characteristics.
Information gain A metric which quantifies the amount of information that the feature holds about the response. Information gain can be regarded as a form of non-linear correlation between Y and X.
Information ratio (IR) Mean active return divided by active risk; or alpha divided by the standard deviation of diversifiable risk.
Informational frictions Forces that restrict availability, quality, and/or flow of information and its use.
Inside ask See *best ask*.
Inside bid See *best bid*.
Inside spread The spread between the best bid price and the best ask price. Also called the *market bid-ask spread*, *inside bid-ask spread*, or *market spread*.
Insiders Corporate managers and board directors who are also shareholders of a company.
Inter-temporal rate of substitution The ratio of the marginal utility of consumption s periods in the future (the numerator) to the marginal utility of consumption today (the denominator).
Intercept The expected value of the dependent variable when the independent variable in a simple linear regression is equal to zero.
Interest rate risk The risk that interest rates will rise and therefore the market value of current portfolio holdings will fall so that their current yields to maturity then match comparable instruments in the marketplace.
Interlocking directorates Corporate structure in which individuals serve on the board of directors of multiple corporations.
Internal rate of return Abbreviated as IRR. Rate of return that discounts future cash flows from an investment to the exact amount of the investment; the discount rate that makes the present value of an investment's costs (outflows) equal to the present value of the investment's benefits (inflows).

International Fisher effect The proposition that nominal interest rate differentials across currencies are determined by expected inflation differentials.

Intrinsic value The value of an asset given a hypothetically complete understanding of the asset's investment characteristics; the value obtained if an option is exercised based on current conditions. The difference between the spot exchange rate and the strike price of a currency.

Inverse price ratio The reciprocal of a price multiple—for example, in the case of a P/E, the "earnings yield" E/P (where P is share price and E is earnings per share).

Investment value The value to a specific buyer, taking account of potential synergies based on the investor's requirements and expectations.

ISDA Master Agreement A standard or "master" agreement published by the International Swaps and Derivatives Association. The master agreement establishes the terms for each party involved in the transaction.

Judicial law Interpretations of courts.

Justified (fundamental) P/E The price-to-earnings ratio that is fair, warranted, or justified on the basis of forecasted fundamentals.

Justified price multiple The estimated fair value of the price multiple, usually based on forecasted fundamentals or comparables.

***K*-fold cross-validation** A technique in which data (excluding test sample and fresh data) are shuffled randomly and then are divided into k equal sub-samples, with $k - 1$ samples used as training samples and one sample, the kth, used as a validation sample.

***K*-means** A clustering algorithm that repeatedly partitions observations into a fixed number, k, of non-overlapping clusters.

***K*-nearest neighbor** A supervised learning technique that classifies a new observation by finding similarities ("nearness") between this new observation and the existing data.

Kalotay–Williams–Fabozzi (KWF) model An arbitrage-free term structure model that describes the dynamics of the log of the short rate and assumes constant drift, no mean reversion, and constant volatility.

Key rate durations Sensitivity of a bond's price to changes in specific maturities on the benchmark yield curve. Also called *partial durations*.

***k*th-order autocorrelation** The correlation between observations in a time series separated by k periods.

Labeled dataset A dataset that contains matched sets of observed inputs or features (X's) and the associated output or target (Y).

Labor force Everyone of working age (ages 16 to 64) who either is employed or is available for work but not working.

Labor force participation rate The percentage of the working age population that is in the labor force.

Labor productivity The quantity of real GDP produced by an hour of labor. More generally, output per unit of labor input.

Labor productivity growth accounting equation States that potential GDP growth equals the growth rate of the labor input plus the growth rate of labor productivity.

Lack of marketability discount An extra return to investors to compensate for lack of a public market or lack of marketability.

LASSO Least absolute shrinkage and selection operator is a type of penalized regression which involves minimizing the sum of the absolute values of the regression coefficients. LASSO can also be used for regularization in neural networks.

Latency The elapsed time between the occurrence of an event and a subsequent action that depends on that event.

Law of one price A principle that states that if two investments have the same or equivalent future cash flows regardless of what will happen in the future, then these two investments should have the same current price.

Leading dividend yield Forecasted dividends per share over the next year divided by current stock price.

Leading P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Learning curve A curve that plots the accuracy rate (= 1 – error rate) in the validation or test samples (i.e., out-of-sample) against the amount of data in the training sample, which is thus useful for describing under- and overfitting as a function of bias and variance errors.

Learning rate A parameter that affects the magnitude of adjustments in the weights in a neural network.

Level One of the three factors (the other two are steepness and curvature) that empirically explain most yield curve shape changes. A shock to the level factor changes the yield for all maturities by an almost identical amount.

Leveraged buyout A transaction whereby the target company management team converts the target to a privately held company by using heavy borrowing to finance the purchase of the target company's outstanding shares.

Leveraged recapitalization A post-offer takeover defense mechanism that involves the assumption of a large amount of debt that is then used to finance share repurchases. The effect is to dramatically change the company's capital structure while attempting to deliver a value to target shareholders in excess of a hostile bid.

Libor–OIS spread The difference between Libor and the overnight indexed swap rate.

Limit order book The book or list of limit orders to buy and sell that pertains to a security.

Lin-log model A regression model in which the independent variable is in logarithmic form.

Linear classifier A binary classifier that makes its classification decision based on a linear combination of the features of each data point.

Linear regression Regression that models the straight-line relationship between the dependent and independent variables. Also known as *least squares regression* and *ordinary least squares regression*.

Linear trend A trend in which the dependent variable changes at a constant rate with time.

Liquidating dividend A dividend that is a return of capital rather than a distribution from earnings or retained earnings.

Liquidation To sell the assets of a company, division, or subsidiary piecemeal, typically because of bankruptcy; the form of bankruptcy that allows for the orderly satisfaction of creditors' claims after which the company ceases to exist.

Liquidation value The value of a company if the company were dissolved and its assets sold individually.

Liquidity preference theory A term structure theory that asserts liquidity premiums exist to compensate investors for the added interest rate risk they face when lending long term.

Liquidity premium The premium or incrementally higher yield that investors demand for lending long term.

Local currency The currency of the country where a company is located.

Local expectations theory A term structure theory that contends the return for all bonds over short periods is the risk-free rate.

Log-lin model A regression model in which the dependent variable is in logarithmic form.

Log-linear model With reference to time-series models, a model in which the growth rate of the time series as a function of time is constant.

Log-log model A regression model in which both the dependent and independent variables are in logarithmic form. Also known as the *double-log model*.

Log-log regression model A regression that expresses the dependent and independent variables as natural logarithms.

Logistic regression (logit model) A qualitative-dependent-variable multiple regression model based on the logistic probability distribution.

Long/short credit trade A credit protection seller with respect to one entity combined with a credit protection buyer with respect to another entity.

Look-ahead bias The bias created by using information that was unknown or unavailable in the time periods over which backtesting is conducted, such as company earnings and macroeconomic indicator values.

Lookback period The time period used to gather a historical data set.

Loss given default The amount that will be lost if a default occurs.

Macroeconomic factor model A multifactor model in which the factors are surprises in macroeconomic variables that significantly explain equity returns.

Macroeconomic factors Factors related to the economy, such as the inflation rate, industrial production, or economic sector membership.

Maintenance capital expenditures Capital expenditures needed to maintain operations at the current level.

Majority shareholders Shareholders that own more than 50% of a corporation's shares.

Majority-vote classifier A classifier that assigns to a new data point the predicted label with the most votes (i.e., occurrences).

Managerialism theories Theories that posit that corporate executives are motivated to engage in mergers to maximize the size of their company rather than shareholder value (a form of agency cost).

Marginal VaR (MVaR) A measure of the effect of a small change in a position size on portfolio VaR.

Market approach Valuation approach that values an asset based on pricing multiples from sales of assets viewed as similar to the subject asset.

Market conversion premium per share For a convertible bond, the difference between the market conversion price and the underlying share price, which allows investors to identify the premium or discount payable when buying a convertible bond rather than the underlying common stock.

Market conversion premium ratio For a convertible bond, the market conversion premium per share expressed as a percentage of the current market price of the shares.

Market efficiency A finance perspective on capital markets that deals with the relationship of price to intrinsic value. The **traditional efficient markets formulation** asserts that an asset's price is the best available estimate of its intrinsic value. The **rational efficient markets formulation** asserts that investors should expect to be rewarded for the costs of information gathering and analysis by higher gross returns.

Market fragmentation Trading the same instrument in multiple venues.

Market impact The effect of the trade on transaction prices. Also called *price impact.*

Market timing Asset allocation in which the investment in the market is increased if one forecasts that the market will outperform T-bills.

Market value of invested capital The market value of debt and equity.

Mature growth rate The earnings growth rate in a company's mature phase; an earnings growth rate that can be sustained long term.

Maximum drawdown The worst cumulative loss ever sustained by an asset or portfolio. More specifically, maximum drawdown is the difference between an asset's or a portfolio's maximum cumulative return and its subsequent lowest cumulative return.

Mean reversion The tendency of a time series to fall when its level is above its mean and rise when its level is below its mean; a mean-reverting time series tends to return to its long-term mean.

Mean square error (MSE) The sum of squares error divided by the degrees of freedom, $n - k - 1$; in a simple linear regression, $n - k - 1 = n - 2$.

Mean square regression (MSR) The sum of squares regression divided by the number of independent variables k; in a simple linear regression, $k = 1$.

Merger The absorption of one company by another; two companies become one entity and one or both of the pre-merger companies ceases to exist as a separate entity.

Metadata Data that describes and gives information about other data.

Method based on forecasted fundamentals An approach to using price multiples that relates a price multiple to forecasts of fundamentals through a discounted cash flow model.

Method of comparables An approach to valuation that involves using a price multiple to evaluate whether an asset is relatively fairly valued, relatively undervalued, or relatively overvalued when compared to a benchmark value of the multiple.

Midquote price The average, or midpoint, of the prevailing bid and ask prices.

Minority interest The proportion of the ownership of a subsidiary not held by the parent (controlling) company.

Minority shareholders Shareholders that own less than 50% of a corporation's shares.

Mispricing Any departure of the market price of an asset from the asset's estimated intrinsic value.

Mixed offering A merger or acquisition that is to be paid for with cash, securities, or some combination of the two.

Model specification With reference to regression, the set of variables included in the regression and the regression equation's functional form.

Molodovsky effect The observation that P/Es tend to be high on depressed EPS at the bottom of a business cycle and tend to be low on unusually high EPS at the top of a business cycle.

Momentum indicators Valuation indicators that relate either price or a fundamental (such as earnings) to the time series of their own past values (or in some cases to their expected value).

Monetary assets and liabilities Assets and liabilities with value equal to the amount of currency contracted for, a fixed amount of currency. Examples are cash, accounts receivable, accounts payable, bonds payable, and mortgages payable. Inventory is not a monetary asset. Most liabilities are monetary.

Monetary/non-monetary method Approach to translating foreign currency financial statements for consolidation in which monetary assets and liabilities are translated at the current exchange rate. Non-monetary assets and liabilities are translated at historical exchange rates (the exchange rates that existed when the assets and liabilities were acquired).

Monetizing Unwinding a position to either capture a gain or realize a loss.

Monitoring costs Costs borne by owners to monitor the management of the company (e.g., board of director expenses).

Monte Carlo simulation A technique that uses the inverse transformation method for converting a randomly generated uniformly distributed number into a simulated value of a random variable of a desired distribution. Each key decision variable in a Monte Carlo simulation requires an assumed statistical distribution; this assumption facilitates incorporating non-normality, fat tails, and tail dependence as well as solving high-dimensionality problems.

Mortgages Loans with real estate serving as collateral for the loans.

Multicollinearity A regression assumption violation that occurs when two or more independent variables (or combinations of independent variables) are highly but not perfectly correlated with each other.

Multiple linear regression Linear regression involving two or more independent variables.

Multiple linear regression model A linear regression model with two or more independent variables.

Mutual information Measures how much information is contributed by a token to a class of texts. MI will be 0 if the token's distribution in all text classes is the same. MI approaches 1 as the token in any one class tends to occur more often in only that particular class of text.

Mutually exclusive projects Mutually exclusive projects compete directly with each other. For example, if Projects A and B are mutually exclusive, you can choose A or B, but you cannot choose both.

N-grams A representation of word sequences. The length of a sequence varies from 1 to *n*. When one word is used, it is a unigram; a two-word sequence is a bigram; and a 3-word sequence is a trigram; and so on.

***n*-Period moving average** The average of the current and immediately prior $n - 1$ values of a time series.

Naked credit default swap A position where the owner of the CDS does not have a position in the underlying credit.

Name entity recognition An algorithm that analyzes individual tokens and their surrounding semantics while referring to its dictionary to tag an object class to the token.

Negative serial correlation Serial correlation in which a positive error for one observation increases the chance of a negative error for another observation, and vice versa.

Net asset balance sheet exposure When assets translated at the current exchange rate are greater in amount than liabilities translated at the current exchange rate. Assets exposed to translation gains or losses exceed the exposed liabilities.

Net asset value The difference between assets and liabilities, all taken at current market values instead of accounting book values.

Net asset value per share Net asset value divided by the number of shares outstanding.

Net lease A lease under which the tenant pays a net rent to the landlord and an additional amount based on the tenant's pro rata share of the operating costs, utilities, maintenance expenses, and real estate taxes relating to the property.

Net liability balance sheet exposure When liabilities translated at the current exchange rate are greater assets translated at the current exchange rate. Liabilities exposed to translation gains or losses exceed the exposed assets.

Net operating income Gross rental revenue minus operating costs but before deducting depreciation, corporate overhead, and interest expense. In the context of real estate, a measure of the income from the property after deducting operating expenses for such items as property taxes, insurance, maintenance, utilities, repairs, and insurance but before deducting any costs associated with financing and before deducting federal income taxes. It is similar to EBITDA in a financial reporting context.

Net regulatory burden The private costs of regulation less the private benefits of regulation.

Network externalities The impact that users of a good, a service, or a technology have on other users of that product; it can be positive (e.g., a critical mass of users makes a product more useful) or negative (e.g., congestion makes the product less useful).

Neural networks Highly flexible machine learning algorithms that have been successfully applied to a variety of supervised and unsupervised tasks characterized by nonlinearities and interactions among features.

No-arbitrage approach A procedure for obtaining the value of an option based on the creation of a portfolio that replicates the payoffs of the option and deriving the option value from the value of the replicating portfolio.

No-growth company A company without positive expected net present value projects.

No-growth value per share The value per share of a no-growth company, equal to the expected level amount of earnings divided by the stock's required rate of return.

Non-cash rent An amount equal to the difference between the average contractual rent over a lease term (the straight-line rent) and the cash rent actually paid during a period. This figure is one of the deductions made from FFO to calculate AFFO.

Non-convergence trap A situation in which a country remains relatively poor, or even falls further behind, because it fails to implement necessary institutional reforms and/or adopt leading technologies.

Non-monetary assets and liabilities Assets and liabilities that are not monetary assets and liabilities. Non-monetary assets include inventory, fixed assets, and intangibles, and non-monetary liabilities include deferred revenue.

Non-renewable resources Finite resources that are depleted once they are consumed; oil and coal are examples.
Non-residential properties Commercial real estate properties other than multi-family properties, farmland, and timberland.
Nonearning assets Cash and investments (specifically cash, cash equivalents, and short-term investments).
Nonstationarity With reference to a random variable, the property of having characteristics, such as mean and variance, that are not constant through time.
Normal EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normalized EPS*.
Normalized earnings The expected level of mid-cycle earnings for a company in the absence of any unusual or temporary factors that affect profitability (either positively or negatively).
Normalized EPS The EPS that a business could achieve currently under mid-cyclical conditions. Also called *normal EPS*.
Normalized P/E P/E based on normalized EPS data.
Notional amount The amount of protection being purchased in a CDS.
NTM P/E Next 12-month P/E: current market price divided by an estimated next 12-month EPS.
Off-the-run A series of securities or indexes that were issued/created prior to the most recently issued/created series.
On-the-run The most recently issued/created series of securities or indexes.
One hot encoding The process by which categorical variables are converted into binary form (0 or 1) for machine reading. It is one of the most common methods for handling categorical features in text data.
One-sided durations Effective durations when interest rates go up or down, which are better at capturing the interest rate sensitivity of bonds with embedded options that do not react symmetrically to positive and negative changes in interest rates of the same magnitude.
One-tier board Board structure consisting of a single board of directors, composed of executive (internal) and non-executive (external) directors.
Opportunity cost The value that investors forgo by choosing a particular course of action; the value of something in its best alternative use.
Optimal capital structure The capital structure at which the value of the company is maximized.
Option-adjusted spread (OAS) Constant spread that, when added to all the one-period forward rates on the interest rate tree, makes the arbitrage-free value of the bond equal to its market price.
Orderly liquidation value The estimated gross amount of money that could be realized from the liquidation sale of an asset or assets, given a reasonable amount of time to find a purchaser or purchasers.
Organic growth Company growth in output or sales that is achieved by making investments internally (i.e., excludes growth achieved through mergers and acquisitions).
Other comprehensive income Changes to equity that bypass (are not reported in) the income statement; the difference between comprehensive income and net income.
Other post-employment benefits Promises by the company to pay benefits in the future, such as life insurance premiums and all or part of health care insurance for its retirees.
Out-of-sample forecast errors The differences between actual and predicted values of time series outside the sample period used to fit the model.
Overfitting When a model fits the training data too well and so does not generalize well to new data.
Overnight indexed swap (OIS) rate An interest rate swap in which the periodic floating rate of the swap equals the geometric average of a daily unsecured overnight rate (or overnight index rate).
Pairs trading An approach to trading that uses pairs of closely related stocks, buying the relatively undervalued stock and selling short the relatively overvalued stock.
Par curve A hypothetical yield curve for coupon-paying Treasury securities that assumes all securities are priced at par.
Par swap A swap in which the fixed rate is set so that no money is exchanged at contract initiation.
Parametric method A method of estimating VaR that uses the historical mean, standard deviation, and correlation of security price movements to estimate the portfolio VaR. Generally assumes a normal distribution but can be adapted to non-normal distributions with the addition of skewness and kurtosis. Sometimes called the *variance–covariance method* or the *analytical method*.
Partial regression coefficients The slope coefficients in a multiple regression. Also called *partial slope coefficients*.
Partial slope coefficients The slope coefficients in a multiple regression. Also called *partial regression coefficients*.
Parts of speech An algorithm that uses language structure and dictionaries to tag every token in the text with a corresponding part of speech (i.e., noun, verb, adjective, proper noun, etc.).
Payout amount The loss given default times the notional.
Payout policy The principles by which a company distributes cash to common shareholders by means of cash dividends and/or share repurchases.
Payouts Cash dividends and the value of shares repurchased in any given year.
Pecking order theory The theory that managers consider how their actions might be interpreted by outsiders and thus order their preferences for various forms of corporate financing. Forms of financing that are least visible to outsiders (e.g., internally generated funds) are most preferable to managers and those that are most visible (e.g., equity) are least preferable.
PEG ratio The P/E-to-growth ratio, calculated as the stock's P/E divided by the expected earnings growth rate.
Penalized regression A regression that includes a constraint such that the regression coefficients are chosen to minimize the sum of squared residuals *plus* a penalty term that increases in size with the number of included features.
Pension obligation The present value of future benefits earned by employees for service provided to date.
Perfect capital markets Markets in which, by assumption, there are no taxes, transaction costs, or bankruptcy costs and in which all investors have equal ("symmetric") information.
Perpetuity A perpetual annuity, or a set of never-ending level sequential cash flows, with the first cash flow occurring one period from now.
Persistent earnings Earnings excluding nonrecurring components. Also referred to as *core earnings*, *continuing earnings*, or *underlying earnings*.

Pet projects Projects in which influential managers want the corporation to invest. Often, unfortunately, pet projects are selected without undergoing normal capital budgeting analysis.

Physical settlement Involves actual delivery of the debt instrument in exchange for a payment by the credit protection seller of the notional amount of the contract.

Point-in-time data Data consisting of the exact information available to market participants as of a given point in time. Point-in-time data is used to address look-ahead bias.

Poison pill A pre-offer takeover defense mechanism that makes it prohibitively costly for an acquirer to take control of a target without the prior approval of the target's board of directors.

Poison puts A pre-offer takeover defense mechanism that gives target company bondholders the right to sell their bonds back to the target at a pre-specified redemption price, typically at or above par value; this defense increases the need for cash and raises the cost of the acquisition.

Portfolio balance approach A theory of exchange rate determination that emphasizes the portfolio investment decisions of global investors and the requirement that global investors willingly hold all outstanding securities denominated in each currency at prevailing prices and exchange rates.

Positive serial correlation Serial correlation in which a positive error for one observation increases the chance of a positive error for another observation; a negative error for one observation increases the chance of a negative error for another observation.

Potential GDP The maximum amount of output an economy can sustainably produce without inducing an increase in the inflation rate. The output level that corresponds to full employment with consistent wage and price expectations.

Precision In error analysis for classification problems it is ratio of correctly predicted positive classes to all predicted positive classes. Precision is useful in situations where the cost of false positives (FP), or Type I error, is high.

Preferred habitat theory A term structure theory that contends that investors have maturity preferences and require yield incentives before they will buy bonds outside of their preferred maturities.

Premise of value The status of a company in the sense of whether it is assumed to be a going concern or not.

Premium leg The series of payments the credit protection buyer promises to make to the credit protection seller.

Premiums Amounts paid by the purchaser of insurance products.

Present value model A model of intrinsic value that views the value of an asset as the present value of the asset's expected future cash flows.

Present value of growth opportunities The difference between the actual value per share and the no-growth value per share. Also called *value of growth*.

Presentation currency The currency in which financial statement amounts are presented.

Price improvement When trade execution prices are better than quoted prices.

Price momentum A valuation indicator based on past price movement.

Price multiples The ratio of a stock's market price to some measure of value per share.

Price-setting option The operational flexibility to adjust prices when demand varies from what is forecast. For example, when demand exceeds capacity, the company could benefit from the excess demand by increasing prices.

Price-to-earnings ratio (P/E) The ratio of share price to earnings per share.

Priced risk Risk for which investors demand compensation for bearing (e.g., equity risk, company-specific factors, macroeconomic factors).

Principal components analysis (PCA) An unsupervised ML technique used to transform highly correlated features of data into a few main, uncorrelated composite variables.

Principle of no arbitrage In well-functioning markets, prices will adjust until there are no arbitrage opportunities.

Prior transaction method A variation of the market approach; considers actual transactions in the stock of the subject private company.

Private market value The value derived using a sum-of-the-parts valuation.

Probability of default The probability that a bond issuer will not meet its contractual obligations on schedule.

Probability of survival The probability that a bond issuer will meet its contractual obligations on schedule.

Procedural law The body of law that focuses on the protection and enforcement of the substantive laws.

Production-flexibility option The operational flexibility to alter production when demand varies from forecast. For example, if demand is strong, a company may profit from employees working overtime or from adding additional shifts.

Project sequencing To defer the decision to invest in a future project until the outcome of some or all of a current project is known. Projects are sequenced through time, so that investing in a project creates the option to invest in future projects.

Projection error The vertical (perpendicular) distance between a data point and a given principal component.

Prospective P/E A P/E calculated on the basis of a forecast of EPS; a stock's current price divided by next year's expected earnings.

Protection leg The contingent payment that the credit protection seller may have to make to the credit protection buyer.

Protection period Period during which a bond's issuer cannot call the bond.

Provision for loan losses An income statement expense account that increases the amount of the allowance for loan losses.

Proxy fight An attempt to take control of a company through a shareholder vote.

Proxy statement A public document that provides the material facts concerning matters on which shareholders will vote.

Prudential supervision Regulation and monitoring of the safety and soundness of financial institutions to promote financial stability, reduce system-wide risks, and protect customers of financial institutions.

Pruning A regularization technique used in CART to reduce the size of the classification or regression tree—by pruning, or removing, sections of the tree that provide little classifying power.

Purchasing power gain A gain in value caused by changes in price levels. Monetary liabilities experience purchasing power gains during periods of inflation.

Purchasing power loss A loss in value caused by changes in price levels. Monetary assets experience purchasing power loss during periods of inflation.

Purchasing power parity (PPP) The idea that exchange rates move to equalize the purchasing power of different currencies.

Pure expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *unbiased expectations theory*.

Pure factor portfolio A portfolio with sensitivity of 1 to the factor in question and a sensitivity of 0 to all other factors.

Putable bond Bond that includes an embedded put option, which gives the bondholder the right to put back the bonds to the issuer prior to maturity, typically when interest rates have risen and higher-yielding bonds are available.

Qualitative dependent variables Dummy variables used as dependent variables rather than as independent variables.

Quality of earnings analysis The investigation of issues relating to the accuracy of reported accounting results as reflections of economic performance. Quality of earnings analysis is broadly understood to include not only earnings management but also balance sheet management.

Random forest classifier A collection of a large number of decision trees trained via a bagging method.

Random walk A time series in which the value of the series in one period is the value of the series in the previous period plus an unpredictable random error.

Rational efficient markets formulation See *market efficiency.*

Readme files Text files provided with raw data that contain information related to a data file. They are useful for understanding the data and how they can be interpreted correctly.

Real estate investment trusts Tax-advantaged entities (companies or trusts) that own, operate, and—to a limited extent—develop income-producing real estate property.

Real estate operating companies Regular taxable real estate ownership companies that operate in the real estate industry in countries that do not have a tax-advantaged REIT regime in place or that are engage in real estate activities of a kind and to an extent that do not fit in their country's REIT framework.

Real interest rate parity The proposition that real interest rates will converge to the same level across different markets.

Real options Options that relate to investment decisions such as the option to time the start of a project, the option to adjust its scale, or the option to abandon a project that has begun.

Rebalance return A return from rebalancing the component weights of an index.

Recall Also known as *sensitivity*, in error analysis for classification problems it is the ratio of correctly predicted positive classes to all actual positive classes. Recall is useful in situations where the cost of false negatives (FN), or Type II error, is high.

Reconstitution When dealers recombine appropriate individual zero-coupon securities and reproduce an underlying coupon Treasury.

Recovery rate The percentage of the loss recovered.

Redemption basket The list of securities (and share amounts) the authorized participant (AP) receives when it redeems ETF shares back to the ETF manager. The redemption basket is published each business day.

Reference entity The borrower (debt issuer) covered by a single-name CDS.

Reference obligation A particular debt instrument issued by the borrower that is the designated instrument being covered.

Regime With reference to a time series, the underlying model generating the times series.

Regression analysis A tool for examining whether a variable is useful for explaining another variable.

Regression coefficients The intercept and slope coefficient(s) of a regression.

Regular expression (regex) A series of texts that contains characters in a particular order. Regex is used to search for patterns of interest in a given text.

Regularization A term that describes methods for reducing statistical variability in high-dimensional data estimation problems.

Regulatory arbitrage Entities identify and use some aspect of regulations that allows them to exploit differences in economic substance and regulatory interpretation or in foreign and domestic regulatory regimes to their (the entities') advantage.

Regulatory burden The costs of regulation for the regulated entity.

Regulatory capture Theory that regulation often arises to enhance the interests of the regulated.

Regulatory competition Regulators may compete to provide a regulatory environment designed to attract certain entities.

Reinforcement learning Machine learning in which a computer learns from interacting with itself or data generated by the same algorithm.

Relative-strength indicators Valuation indicators that compare a stock's performance during a period either to its own past performance or to the performance of some group of stocks.

Relative valuation models A model that specifies an asset's value relative to the value of another asset.

Relative VaR See *ex ante tracking error*.

Relative version of PPP The hypothesis that changes in (nominal) exchange rates over time are equal to national inflation rate differentials.

Renewable resources Resources that can be replenished, such as a forest.

Rental price of capital The cost per unit of time to rent a unit of capital.

Repeat sales index As the name implies, this type of index relies on repeat sales of the same property. In general, the idea supporting this type of index is that because it is the same property that sold twice, the change in value between the two sale dates indicates how market conditions have changed over time.

Replacement cost In the context of real estate, the value of a building assuming it was built today using current construction costs and standards.

Reporting unit For financial reporting under US GAAP, an operating segment or one level below an operating segment (referred to as a component).

Required rate of return The minimum rate of return required by an investor to invest in an asset, given the asset's riskiness.

Residential properties Properties that provide housing for individuals or families. Single-family properties may be owner-occupied or rental properties, whereas multi-family properties are rental properties even if the owner or manager occupies one of the units.

Residual The difference between an observation and its predicted value, where the predicted value is based on the estimated linear relation between the dependent and independent variables using sample data.

Residual autocorrelations The sample autocorrelations of the residuals.

Residual income Earnings for a given period, minus a deduction for common shareholders' opportunity cost in generating the earnings. Also called *economic profit* or *abnormal earnings*.

Residual income method Income approach that estimates the value of all intangible assets of the business by capitalizing future earnings in excess of the estimated return requirements associated with working capital and fixed assets.

Residual income model (RIM) A model of stock valuation that views intrinsic value of stock as the sum of book value per share plus the present value of the stock's expected future residual income per share. Also called *discounted abnormal earnings model* or *Edwards–Bell–Ohlson model*.

Residual loss Agency costs that are incurred despite adequate monitoring and bonding of management.

Restructuring Reorganizing the capital structure of a firm.

Return on capital employed Operating profit divided by capital employed (debt and equity capital).

Return on invested capital A measure of the after-tax profitability of the capital invested by the company's shareholders and debtholders.

Reverse carry arbitrage A strategy involving the short sale of the underlying and an offsetting opposite position in the derivative.

Reverse stock split A reduction in the number of shares outstanding with a corresponding increase in share price but no change to the company's underlying fundamentals.

Reverse stress testing A risk management approach in which the user identifies key risk exposures in the portfolio and subjects those exposures to extreme market movements.

Reviewed financial statements A type of non-audited financial statements; typically provide an opinion letter with representations and assurances by the reviewing accountant that are less than those in audited financial statements.

Rho The change in a given derivative instrument for a given small change in the risk-free interest rate, holding everything else constant. Rho measures the sensitivity of the option to the risk-free interest rate.

Risk budgeting The allocation of an asset owner's total risk appetite among groups or divisions (in the case of a trading organization) or among strategies and managers (in the case of an institutional or individual investor).

Risk decomposition The process of converting a set of holdings in a portfolio into a set of exposures to risk factors.

Risk factors Variables or characteristics with which individual asset returns are correlated. Sometimes referred to simply as *factors*.

Risk parity A portfolio allocation scheme that weights stocks or factors based on an equal risk contribution.

Robust standard errors Standard errors of the estimated parameters of a regression that correct for the presence of heteroskedasticity in the regression's error term.

Roll When an investor moves its investment position from an older series to the most current series.

Roll return The component of the return on a commodity futures contract attributable to rolling long futures positions forward through time. Also called *roll yield*.

Rolling down the yield curve A maturity trading strategy that involves buying bonds with a maturity longer than the intended investment horizon. Also called *riding the yield curve*.

Rolling windows A backtesting method that uses a rolling-window (or walk-forward) framework, rebalances the portfolio after each period, and then tracks performance over time. As new information arrives each period, the investment manager optimizes (revises and tunes) the model and readjusts stock positions.

Root mean squared error (RMSE) The square root of the average squared forecast error; used to compare the out-of-sample forecasting performance of forecasting models.

Sale-leaseback A situation in which a company sells the building it owns and occupies to a real estate investor and the company then signs a long-term lease with the buyer to continue to occupy the building. At the end of the lease, use of the property reverts to the landlord.

Sales comparison approach In the context of real estate, this approach estimates value based on what similar or comparable properties (comparables) transacted for in the current market.

Scaled earnings surprise Unexpected earnings divided by the standard deviation of analysts' earnings forecasts.

Scaling The process of adjusting the range of a feature by shifting and changing the scale of the data. Two of the most common ways of scaling are normalization and standardization.

Scatter plot A chart in which two variables are plotted along the axis and points on the chart represent pairs of the two variables. In regression, the dependent variable is plotted on the vertical axis and the independent variable is plotted along the horizontal axis. Also known as a scattergram and a *scatter diagram*.

Scenario analysis A technique for exploring the performance and risk of investment strategies in different structural regimes.

Scree plots A plot that shows the proportion of total variance in the data explained by each principal component.

Screening The application of a set of criteria to reduce a set of potential investments to a smaller set having certain desired characteristics.

Seasonality A characteristic of a time series in which the data experience regular and predictable periodic changes; for example, fan sales are highest during the summer months.

Secured overnight financing rate (SOFR) A daily volume-weighted index of rates on qualified cash borrowings collateralized by US Treasuries that is expected to replace Libor as a floating reference rate for swaps.

Securities offering A merger or acquisition in which target shareholders are to receive shares of the acquirer's common stock as compensation.

Security selection risk See *active specific risk*.

Segmented markets theory A term structure theory that contends yields are solely a function of the supply and demand for funds of a particular maturity.

Self-regulating organizations (SROs) Self-regulating bodies that are given recognition and authority, including enforcement power, by a government body or agency.

Self-regulatory bodies Private, non-governmental organizations that both represent and regulate their members. Some self-regulating organizations are also independent regulators.

Sell-side analysts Analysts who work at brokerages.

Sensitivity analysis A technique for exploring how a target variable (e.g., portfolio returns) and risk profiles are affected by changes in input variables (e.g., the distribution of asset or factor returns).

Sentence length The number of characters, including spaces, in a sentence.

Serially correlated With reference to regression errors, errors that are correlated across observations.

Service period For employee stock options, usually the period between the grant date and the vesting date.

Settled in arrears An arrangement in which the interest payment is made (i.e., settlement occurs) at the maturity of the underlying instrument.

Settlement In the case of a credit event, the process by which the two parties to a CDS contract satisfy their respective obligations.

Shaping risk The sensitivity of a bond's price to the changing shape of the yield curve.

Share repurchase A transaction in which a company buys back its own shares. Unlike stock dividends and stock splits, share repurchases use corporate cash.

Shareholder activism Strategies used by shareholders to attempt to compel a company to act in a desired manner.

Shareholders' equity Total assets minus total liabilities.

Shark repellents A pre-offer takeover defense mechanism involving the corporate charter (e.g., staggered boards of directors and supermajority provisions).

Simple linear regression (SLR) A regression that summarizes the relation between the dependent variable and a single independent variable.

Simulation A technique for exploring how a target variable (e.g. portfolio returns) would perform in a hypothetical environment specified by the user, rather than a historical setting.

Single-name CDS Credit default swap on one specific borrower.

Sinking fund bond A bond that requires the issuer to set aside funds over time to retire the bond issue, thus reducing credit risk.

Slope coefficient The coefficient of an independent variable that represents the average change in the dependent variable for a one-unit change in the independent variable.

Soft margin classification An adaptation in the support vector machine algorithm which adds a penalty to the objective function for observations in the training set that are misclassified.

Special dividend A dividend paid by a company that does not pay dividends on a regular schedule or a dividend that supplements regular cash dividends with an extra payment.

Spin-off A form of restructuring in which shareholders of a parent company receive a proportional number of shares in a new, separate entity; shareholders end up owning stock in two different companies where there used to be one.

Split-off A form of restructuring in which shareholders of the parent company are given shares in a newly created entity in exchange for their shares of the parent company.

Split-rate tax system In reference to corporate taxes, a split-rate system taxes earnings to be distributed as dividends at a different rate than earnings to be retained. Corporate profits distributed as dividends are taxed at a lower rate than those retained in the business.

Spot curve The term structure of spot rates for loans made today.

Spot price The current price of an asset or security. For commodities, the current price to deliver a physical commodity to a specific location or purchase and transport it away from a designated location.

Spot rate The interest rate that is determined today for a risk-free, single-unit payment at a specified future date.

Spot yield curve The term structure of spot rates for loans made today.

Stabilized NOI In the context of real estate, the expected NOI when a renovation is complete.

Stable dividend policy A policy in which regular dividends are paid that reflect long-run expected earnings. In contrast to a constant dividend payout ratio policy, a stable dividend policy does not reflect short-term volatility in earnings.

Standard error of the estimate A measure of the fit of a regression line, calculated as the square root of the mean square error. Also known as the *standard error of the regression* and the *root mean square error*.

Standard error of the forecast A measure of the uncertainty associated with a forecasted value of the dependent variable that depends on the standard error of the estimate, the variability of the independent variable, the deviation of the forecasted independent variable from the mean in the regression, and the number of observations.

Standard error of the slope coefficient The standard error of the slope, which in a simple linear regression is the ratio of the model's standard error of the estimate (s_e) to the square root of the variation of the independent variable.

Standardized beta With reference to fundamental factor models, the value of the attribute for an asset minus the average value of the attribute across all stocks, divided by the standard deviation of the attribute across all stocks.

Standardized unexpected earnings Unexpected earnings per share divided by the standard deviation of unexpected earnings per share over a specified prior time period.

Static trade-off theory of capital structure A theory pertaining to a company's optimal capital structure. The optimal level of debt is found at the point where additional debt would cause the costs of financial distress to increase by a greater amount than the benefit of the additional tax shield.

Statistical factor model A multifactor model in which statistical methods are applied to a set of historical returns to determine portfolios that best explain either historical return covariances or variances.

Statutes Laws enacted by legislative bodies.

Statutory merger A merger in which one company ceases to exist as an identifiable entity and all its assets and liabilities become part of a purchasing company.

Steady-state rate of growth The constant growth rate of output (or output per capita) that can or will be sustained indefinitely once it is reached. Key ratios, such as the capital–output ratio, are constant on the steady-state growth path.

Steepness The difference between long-term and short-term yields that constitutes one of the three factors (the other two are level and curvature) that empirically explain most of the changes in the shape of the yield curve.

Stock dividend A type of dividend in which a company distributes additional shares of its common stock to shareholders instead of cash.

Stock purchase An acquisition in which the acquirer gives the target company's shareholders some combination of cash and securities in exchange for shares of the target company's stock.

Stop-loss limit Constraint used in risk management that requires a reduction in the size of a portfolio, or its complete liquidation, when a loss of a particular size occurs in a specified period.

Straight bond An underlying option-free bond with a specified issuer, issue date, maturity date, principal amount and repayment structure, coupon rate and payment structure, and currency denomination.

Straight-line rent The average annual rent under a multi-year lease agreement that contains contractual increases in rent during the life of the lease.

Straight-line rent adjustment See *non-cash rent.*

Straight voting Voting structure in which shareholders are granted the right of one vote for each share owned.

Stranded assets Assets that are obsolete or not economically viable.

Strategic transaction A purchase involving a buyer that would benefit from certain synergies associated with owning the target firm.

Stress tests A risk management technique that assesses the portfolio's response to extreme market movements.

Stripping A dealer's ability to separate a bond's individual cash flows and trade them as zero-coupon securities.

Subsidiary merger A merger in which the company being purchased becomes a subsidiary of the purchaser.

Substantive law The body of law that focuses on the rights and responsibilities of entities and relationships among entities.

Succession event A change of corporate structure of the reference entity, such as through a merger, a divestiture, a spinoff, or any similar action, in which ultimate responsibility for the debt in question is unclear.

Sum of squares error (SSE) The sum of the squared deviations of (1) the value of the dependent variable and (2) the value of the dependent variable based on the estimated regression line. Also referred to as the *residual sum of squares.*

Sum of squares regression (SSR) The sum of the squared deviations of (1) the value of the dependent variable based on the estimated regression line and (2) the mean of the dependent variable.

Sum of squares total (SST) The sum of the squared deviations of the dependent variable from its mean; the variation of the dependent variable. Also referred to as the *total sum of squares.*

Sum-of-the-parts valuation A valuation that sums the estimated values of each of a company's businesses as if each business were an independent going concern.

Summation operator A functional part of a neural network's node that multiplies each input value received by a weight and sums the weighted values to form the total net input, which is then passed to the activation function.

Supernormal growth Above-average or abnormally high growth rate in earnings per share.

Supervised learning Machine learning where algorithms infer patterns between a set of inputs (the *X*'s) and the desired output (*Y*). The inferred pattern is then used to map a given input set into a predicted output.

Support vector machine A linear classifier that determines the hyperplane that optimally separates the observations into two sets of data points.

Survivorship bias The bias that results when data as of a given date reflects only those entities that have survived to that date. Entities can include any element of an index or list that is constituted through time: stocks, investment funds, etc. Survivorship bias is a form of look-ahead bias.

Sustainable growth rate The rate of dividend (and earnings) growth that can be sustained over time for a given level of return on equity, keeping the capital structure constant and without issuing additional common stock.

Swap curve The term structure of swap rates.

Swap rate The "price" that swap traders quote among one another. It is the rate at which the present value of all the expected floating-rate payments received over the life of the floating-rate bond equal the present value of all the expected fixed-rate payments made over the life of the fixed-rate bond.

Swap rate curve The term structure of swap rates.

Swap spread The difference between the fixed rate on an interest rate swap and the rate on a Treasury note with equivalent maturity; it reflects the general level of credit risk in the market.

Systematic risk Risk that affects the entire market or economy; it cannot be avoided and is inherent in the overall market. Systematic risk is also known as non-diversifiable or market risk.

Systemic risk The risk of failure of the financial system.

Tail risk The risk that losses in extreme events could be greater than would be expected for a portfolio of assets with a normal distribution.

Takeover A merger; the term may be applied to any transaction but is often used in reference to hostile transactions.

Takeover premium The amount by which the takeover price for each share of stock must exceed the current stock price in order to entice shareholders to relinquish control of the company to an acquirer.

Tangible book value per share Common shareholders' equity minus intangible assets reported on the balance sheet, divided by the number of shares outstanding.

Target In machine learning, the dependent variable (*Y*) in a labeled dataset; the company in a merger or acquisition that is being acquired.

Target capital structure A company's chosen proportions of debt and equity.

Target company See *target.*

Target payout ratio A strategic corporate goal representing the long-term proportion of earnings that the company intends to distribute to shareholders as dividends.

Taxable REIT subsidiaries Subsidiaries that pay income taxes on earnings from non-REIT-qualifying activities like merchant development or third-party property management.

Technical indicators Momentum indicators based on price.

TED spread A measure of perceived credit risk determined as the difference between Libor and the T-bill yield of matching maturity.

Temporal method A variation of the monetary/non-monetary translation method that requires not only monetary assets and liabilities, but also non-monetary assets and liabilities that are measured at their current value on the balance sheet date to be translated at the current exchange rate. Assets and liabilities are translated at rates consistent with the timing of their measurement value. This method is typically used when the functional currency is other than the local currency.

Tender offer A public offer whereby the acquirer invites target shareholders to submit ("tender") their shares in return for the proposed payment.

Term frequency (TF) Ratio of the number of times a given token occurs in all the texts in the dataset to the total number of tokens in the dataset.

Term premium The additional return required by lenders to invest in a bond to maturity net of the expected return from continually reinvesting at the short-term rate over that same time horizon.

Terminal price multiples The price multiple for a stock assumed to hold at a stated future time.

Terminal share price The share price at a particular point in the future.

Terminal value of the stock The analyst's estimate of a stock's value at a particular point in the future. Also called *continuing value of the stock*.

Test sample A data sample that is used to test a model's ability to predict well on new data.

Theta The change in a derivative instrument for a given small change in calendar time, holding everything else constant. Specifically, the theta calculation assumes nothing changes except calendar time. Theta also reflects the rate at which an option's time value decays.

Time series A set of observations on a variable's outcomes in different time periods.

Tobin's *q* The ratio of the market value of debt and equity to the replacement cost of total assets.

Token The equivalent of a word (or sometimes a character).

Tokenization The process of splitting a given text into separate tokens. Tokenization can be performed at the word or character level but is most commonly performed at word level.

Top-down approach With respect to forecasting, an approach that usually begins at the level of the overall economy. Forecasts are then made at more narrowly defined levels, such as sector, industry, and market for a specific product.

Total factor productivity (TFP) A multiplicative scale factor that reflects the general level of productivity or technology in the economy. Changes in total factor productivity generate proportional changes in output for any input combination.

Total invested capital The sum of market value of common equity, book value of preferred equity, and face value of debt.

Tracking error The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking risk*.

Tracking risk The standard deviation of the differences between a portfolio's returns and its benchmark's returns; a synonym of active risk. Also called *tracking error*.

Trailing dividend yield The reciprocal of current market price divided by the most recent annualized dividend.

Trailing P/E A stock's current market price divided by the most recent four quarters of EPS (or the most recent two semi-annual periods for companies that report interim data semi-annually). Also called *current P/E*.

Training sample A data sample that is used to train a model.

Tranche CDS A type of credit default swap that covers a combination of borrowers but only up to pre-specified levels of losses.

Transaction exposure The risk of a change in value between the transaction date and the settlement date of an asset of liability denominated in a foreign currency.

Treasury shares/stock Shares that were issued and subsequently repurchased by the company.

Trend A long-term pattern of movement in a particular direction.

Triangular arbitrage An arbitrage transaction involving three currencies that attempts to exploit inconsistencies among pairwise exchange rates.

Trimming Also called truncation, it is the process of removing extreme values and outliers from a dataset.

Triple-net leases Common leases in the United States and Canada that require each tenant to pay its share of the following three operating expenses: common area maintenance and repair expenses; property taxes; and building insurance costs. Also known as *NNN leases*.

Two-tier board Board structure consisting of a supervisory board that oversees a management board.

Unbiased expectations theory A term structure theory that contends the forward rate is an unbiased predictor of the future spot rate. Also called the *pure expectations theory*.

Unconditional heteroskedasticity Heteroskedasticity of the error term that is not correlated with the values of the independent variable(s) in the regression.

Uncovered interest rate parity The proposition that the expected return on an uncovered (i.e., unhedged) foreign currency (risk-free) investment should equal the return on a comparable domestic currency investment.

Underlying earnings Earnings excluding nonrecurring components. Also referred to as *continuing earnings*, *core earnings*, or *persistent earnings*.

Unexpected earnings The difference between reported EPS and expected EPS. Also referred to as an *earnings surprise*.

Unit root A time series that is not covariance stationary is said to have a unit root.

Unsupervised learning Machine learning that does not make use of labeled data.

Upfront payment The difference between the credit spread and the standard rate paid by the protection buyer if the standard rate is insufficient to compensate the protection seller. Also called *upfront premium*.

Upfront premium See *upfront payment*.

Upstream A transaction between two related companies, an investor company (or a parent company) and an associate company (or a subsidiary company) such that the associate company records a profit on its income statement. An example is a sale of inventory by the associate to the investor company or by a subsidiary to a parent company.

Validation sample A data sample that is used to validate and tune a model.

Valuation The process of determining the value of an asset or service either on the basis of variables perceived to be related to future investment returns or on the basis of comparisons with closely similar assets.

Value additivity An arbitrage opportunity when the value of the whole equals the sum of the values of the parts.

Value at risk (VaR) The minimum loss that would be expected a certain percentage of the time over a certain period of time given the assumed market conditions.

Value of growth The difference between the actual value per share and the no-growth value per share.

Variance error Describes how much a model's results change in response to new data from validation and test samples. Unstable models pick up noise and produce high variance error, causing overfitting and high out-of-sample error.

Vasicek model A partial equilibrium term structure model that assumes interest rates are mean reverting and interest rate volatility is constant.

Vega The change in a given derivative instrument for a given small change in volatility, holding everything else constant. A sensitivity measure for options that reflects the effect of volatility.

Venture capital investors Private equity investors in development-stage companies.

Vertical merger A merger involving companies at different positions of the same production chain; for example, a supplier or a distributor.

Vertical ownership Ownership structure in which a company or group that has a controlling interest in two or more holding companies, which in turn have controlling interests in various operating companies.

Vested benefit obligation The actuarial present value of vested benefits.

Vesting date The date that employees can first exercise stock options.

Visibility The extent to which a company's operations are predictable with substantial confidence.

Voting caps Legal restrictions on the voting rights of large share positions.

Web spidering (scraping or crawling) programs Programs that extract raw content from a source, typically web pages.

Weighted average cost of capital (WACC) A weighted average of the after-tax required rates of return on a company's common stock, preferred stock, and long-term debt, where the weights are the fraction of each source of financing in the company's target capital structure.

Weighted harmonic mean See *harmonic mean*.

White-corrected standard errors A synonym for robust standard errors.

White knight A third party that is sought out by the target company's board to purchase the target in lieu of a hostile bidder.

White squire A third party that is sought out by the target company's board to purchase a substantial minority stake in the target—enough to block a hostile takeover without selling the entire company.

Winner's curse The tendency for the winner in certain competitive bidding situations to overpay, whether because of overestimation of intrinsic value, emotion, or information asymmetries.

Winsorization The process of replacing extreme values and outliers in a dataset with the maximum (for large value outliers) and minimum (for small value outliers) values of data points that are not outliers.

Write-down A reduction in the value of an asset as stated in the balance sheet.

Yield curve factor model A model or a description of yield curve movements that can be considered realistic when compared with historical data.

Zero A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

Zero-coupon bond A bond that does not pay a coupon but is priced at a discount and pays its full face value at maturity.

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.