

*Jerzy Marzec*

**Katedra Ekonometrii**

# Badanie niewypłacalności kredytobiorcy na podstawie modeli logitowych i probitowych\*

## 1. Wprowadzenie

Jednym z głównych obszarów działalności banku komercyjnego, a zarazem źródłem przychodów jest udzielanie kredytów. Zatem nieodłącznym elementem tej sfery działalności banku jest ryzyko kredytowe, związane z udzieleniem pojedynczego kredytu. Realizuje się ono w chwili, gdy kredytobiorca – dłużnik nie wywiązuje się z umowy kredytowej, tzn. nieterminowo spłaca raty kapitałowe lub odsetki albo, w ostateczności, odstępuje od spłaty swych zobowiązań wobec banku. Zatem niedotrzymanie umowy przez klienta oznacza dla banku częściową, a niekiedy całkowitą utratę środków pieniężnych udostępnionych niesolidnemu kredytobiorcy. Dodatkowo bank ponosi koszty utraconych korzyści z tytułu nie zapłaconych przez dłużnika odsetek od kapitału. Bank ma obowiązek dokonywać klasyfikacji należności pozostających do spłaty przez klienta. Zgodnie z treścią uchwały Komisji Nadzoru Bankowego w sprawie zasad tworzenia rezerw na ryzyko związane z działalnością banków, należności, w przypadku których opóźnienie w spłacie kapitału lub odsetek wynosi więcej niż 1 miesiąc, są klasyfikowane jako należności zagrożone. W szczególności, jeżeli opóźnienie to jest nie większe niż 3 miesiące, to należności te zalicza się do należności poniżej standardu, jeżeli opóźnienie wynosi powyżej 3 miesięcy i nie dłużej niż 6 miesięcy, klasyfikuje się je jako należności wątpliwe, natomiast gdy opóźnienie wynosi powyżej 6 miesięcy – jako należności stracone<sup>1</sup>. Drugim kryterium ustalania ryzyka bankowego w odniesieniu do należności,

---

\* Praca wykonana w ramach projektu badawczego nr 1-H02B-022-18, finansowanego przez Komitet Badań Naukowych. Autor pragnie wyrazić podziękowanie prof. dr. hab. Jackowi Osiewalskiemu za cenne uwagi w trakcie przygotowywania niniejszego opracowania.

<sup>1</sup> Uchwała nr 8/1999 Komisji Nadzoru Bankowego z 22 grudnia 1999 r.

obok terminowości spłat kapitału i odsetek, jest sytuacja ekonomiczno-finansowa dłużnika. W związku z tym ryzykiem bank jest zobowiązany do tworzenia rezerw celowych m.in. na należności należące do grupy należności zagrożonych. Tworzone rezerwy nie stanowią kapitału pracującego i nie przynoszą przychodów odsetkowych. Przeciwnie, generują one koszty utraconych korzyści z tytułu przychodów odsetkowych, które bank uzyskałby, gdyby zamrożone w postaci rezerw celowych środki pieniężne zainwestował np. w wolne od ryzyka obligacje skarbu państwa. Ponadto bank musi ponieść koszty odsetkowe z tytułu pozyskania środków (od deponentów), które następnie są przeznaczone na rezerwy. Wysokość tworzonych rezerw na należności zagrożone zależy od kategorii należności, a obecnie tworzone są one w wysokości:

- 20% kwoty należności zakwalifikowanych do kategorii „poniżej standardu”,
- 50% kwoty należności zakwalifikowanych do kategorii „wątpliwe”,
- 100% kwoty należności zakwalifikowanych do kategorii „stracone”<sup>2</sup>.

Wobec powyższego posiadanie przez bank w portfelu kredytowym należności zagrożonych, czyli „złych” kredytów, ma wiele negatywnych konsekwencji, m.in. powoduje pogorszenie płynności finansowej, wpływa ujemnie na wynik finansowy, a w ostateczności może doprowadzić do bankructwa banku. Zatem analiza i monitorowanie ryzyka kredytowego, czyli m.in. podejmowanie działań w celu jego minimalizacji, należą do fundamentalnych zadań towarzyszących działalności kredytowej banku.

Podstawowym narzędziem zarządzania ryzykiem kredytowym, związanym z udzieleniem pojedynczego kredytu, jest analiza wniosku kredytowego. Ma ona na celu ocenę w sensie prawnym i ekonomicznym wiarygodność kredytową potencjalnego kredytobiorcy. Jest sporządzana przez analityka kredytowego, który opierając się na wewnętrznych procedurach decyduje, czy konkretny wniosek klienta zostanie zweryfikowany pozytywnie, czy negatywnie. Do obowiązków analityka należy m.in. podjęcie działań w celu sprawdzenia, czy klient ma zdolność prawną do zaciągania zobowiązania, czy działa legalnie (dotyczy firm), czy posiada odpowiednie zabezpieczenie kredytu. Musi on także zbadać obecną oraz przyszłą zdolność klienta do spłaty rat kapitałowo-odsetkowych. W interesie banku jest, aby analityk kredytowy przy formułowaniu oceny wniosku kredytowego używał ściśle określonych i obiektywnych metod. Oczywiście im kwota kredytu jest niższa, tym procedura weryfikacji klienta jest prostsza, bardziej automatyczna. Procedury udzielania kredytu, w tym badanie zdolności kredytowej, są więc ściśle zestandaryzowane. Narzędziem wspomagającym pracę analityka kredytowego w tym zakresie są tzw. modele scoringowe (punktacyjne). Służą one do obliczenia wskaźnika informującego o stopniu ryzyka kredytowego (stopniu wypłacalności klienta) dla konkretnego wniosku kredytowego przy wykorzystaniu mierzalnych i obiektywnych cech klienta. Na tej podstawie klient – potencjalny kredytobiorca – jest kwalifikowany do jednej z dwóch rozłącznych grup, tj. grupy klientów podwyższo-

<sup>2</sup> Zob. Uchwała nr 8/1999 Komisji Nadzoru Bankowego z 22 grudnia 1999 r.

nego bądź obniżonego ryzyka. Bank uzależnia decyzję o udzieleniu kredytu od tego, do której grupy ryzyka klient został przydzielony. Klienci z pierwszej grupy nie kwalifikują się do uzyskania kredytu ze względu na wysokie ryzyko kredytowe dla banku.

Modele scoringowe mają zastosowanie przede wszystkim do weryfikacji wniosków kredytowych klientów indywidualnych. Ta grupa klientów jest najliczniejszą i w miarę jednorodną grupą klientów banku. Średnia wartość przyznanych kredytów detalicznych jest niewielka w stosunku do kredytów komercyjnych, ale ich udział ilościowy w portfelu banku jest dominujący<sup>3</sup>. Z tego względu korzystanie z tych modeli na etapie weryfikacji wniosków kredytowych klientów indywidualnych jest szczególnie wskazane. Głównymi zaletami modeli scoringowych jest m.in. ujednoczenie i obiektywizacja oceny wszystkich wniosków kredytowych oraz obniżenie kosztów obsługi klienta poprzez usprawnienie procesu rozpatrywania wniosków kredytowych.

Istotą każdego modelu scoringowego jest mechanizm, który umożliwia na podstawie tylko wybranych, ale najistotniejszych cech klienta predykcję ryzyka związanego z udzieleniem mu kredytu. Najczęściej w tym celu adaptuje się modele statystyczno-matematyczne, które wykorzystują historyczne dane o udzielonych kredytach, przebiegu ich spłaty i cechach klientów – kredytobiorców, aby na ich podstawie prognozować prawdopodobieństwo niewypłacalności dowolnego kredytobiorcy. Najbardziej znanymi metodami statystycznymi wykorzystywanymi do konstrukcji modeli scoringowych są m.in. analiza dyskryminacyjna oraz modele logitowe i probitowe. Dobór zmiennych wyjaśniających ryzyko pojedynczego kredytu, konstrukcja modelu statystycznego i sposób jego estymacji decydują o jakości modelu scoringowego, czyli o jego zdolności prognostycznej. O możliwościach zastosowania modeli scoringowych i innych metod statystycznych w analizie ryzyka kredytowego pisali w ostatnich latach m.in. M. Gruszczyński [2001] i W. Kuryłek [2000].

## 2. Model logitowy i probitowy

Niech  $Y$  oznacza dyskretną zmienną losową przyjmującą jedynie dwie wartości, zero albo jeden. Funkcja prawdopodobieństwa tej zmiennej losowej dana jest wzorem:

$$\begin{aligned} \Pr(Y = 1) &= F(x \cdot \beta), \\ \Pr(Y = 0) &= 1 - F(x \cdot \beta), \end{aligned} \tag{1}$$

gdzie  $x$  to wektor-wiersz wartości zmiennych egzogenicznych<sup>4</sup>, które mają potencjalny wpływ na zmienną  $Y$ , a  $\beta$  to wektor-kolumna nieznanych parametrów, które odzwierciedlają kierunek i siłę oddziaływania zmian wartości

<sup>3</sup> Na podstawie obserwacji dotyczących analizowanego banku komercyjnego.

<sup>4</sup> Dla uproszczenia zapisu pominięto indeks numeru obserwacji.

zmiennych  $x$  na  $Y$  (zob. [Greene 1993]). Rozkład zmiennej  $Y$  zależy poprzez funkcję  $F(\cdot)$  od wskaźnika  $x \cdot \beta$ , czyli liniowej funkcji parametrów i wartości zmiennych egzogenicznych. Kluczowym założeniem jest przyjęcie odpowiedniej postaci analitycznej dla funkcji  $F(\cdot)$ , która określa rodzaj modelu. Z punktu widzenia wykorzystania modelu (1) do predykcji zmiennej  $Y$  wymaga się, aby postać  $F(\cdot)$  została tak dobrana, by:

$$\begin{cases} \lim_{x \cdot \beta \rightarrow -\infty} \Pr(Y = 1) = 0, \\ \lim_{x \cdot \beta \rightarrow +\infty} \Pr(Y = 1) = 1. \end{cases} \quad (2)$$

Wynika stąd, że funkcja  $F(\cdot)$  jest rosnącą funkcją wskaźnika  $x \cdot \beta$ , o zbiorze wartości ograniczonym do przedziału  $(0;1)$ .

W literaturze klasę modeli, w których zmienną objaśnianą jest zmienna zero-jedynkowa (binarna), nazywa się modelami dwumianowymi (dychotomicznymi) lub modelami dyskretnego wyboru (ang. *binary choice models*). Szerzej są one omówione np. w pracach [Amemiya 1981, 1985] oraz [Aldrich, Nelson 1984]. Z polskich pozycji literaturowych warto wspomnieć monografię M. Gruszczyńskiego [2001], która jest poświęcona wykorzystaniu modeli i prognoz zmiennych jakościowych w finansach i bankowości. Analiza modeli dwumianowych dotyczy przede wszystkim specyfikacji zmiennych egzogenicznych, metody estymacji parametrów  $\beta$  w ramach przyjętej klasy modelu oraz ich wykorzystania do prognozowania zmiennej  $Y$ .

Najprostszym przypadkiem modeli dwumianowych jest liniowy model prawdopodobieństwa (ang. *linear probability model*), który otrzymamy przyjmując, że  $F(x \cdot \beta) = x \cdot \beta$ . Oznacza to, że zależności między objaśniającą zmienną dychotomiczną a zmiennymi  $x$  przedstawia się w postaci regresji liniowej. Skoro wartość oczekiwana zmiennej  $Y$  wynosi  $F(x \cdot \beta)$ , to próbkowy model regresji dla pojedynczej obserwacji przybiera postać:

$$y = E(y) + (y - E(y)) = F(x \cdot \beta) + \varepsilon = x \cdot \beta + \varepsilon, \quad (3)$$

gdzie  $\varepsilon$  to składnik losowy, który przyjmuje wartość  $1 - x \cdot \beta$  lub  $-x \cdot \beta$  z prawdopodobieństwem odpowiednio  $x \cdot \beta$  lub  $1 - x \cdot \beta$ . Zatem jego wartość oczekiwana  $\varepsilon$  wynosi zero, a wariancja  $x \cdot \beta \cdot (1 - x \cdot \beta)$ , więc wariancja zmienia się systematycznie wraz ze zmianami  $x$ , co wskazuje na jego heteroskedastyczność. Zatem stosowanie zwykłej metody najmniejszych kwadratów (MNK) może prowadzić do błędnych wyników. Możliwe jest natomiast użycie uogólnionej (szacowanej) metody najmniejszych kwadratów EGLS, FGLS – zob. [Greene 1993]), jednakże należy dodatkowo narzucić nierównościowe ograniczenia na prawdopodobieństwo. W przeciwnym razie można uzyskać, jak w przypadku MNK, nieinterpretowalne wyniki, tj. ujemne lub większe od jedności prawdopodobieństwo, że  $Y = 1$ . Z tego też powodu stosowanie modelu (3) nie jest wskazane, aczkolwiek wyniki uzyskane MNK można wykorzystać

jako oceny wstępne do iteracyjnych metod estymacji modeli dwumianowych, np. metody największej wiarygodności.

Przypomnijmy, że warunek (2) spełnia każda dystrybuanta zmiennej losowej ciągłej. W literaturze najczęściej stosuje się modele probitowe lub logitowe. Jeżeli za  $F(\cdot)$  przyjmiemy dystrybuantę zmiennej losowej o rozkładzie standaryzowanym normalnym  $\Phi(\cdot)$ , to mówimy o modelu probitowym, gdy zaś dystrybuantę rozkładu logistycznego – o modelu logitowym<sup>5</sup>. Analityczna postać dystrybuanty zmiennej losowej o rozkładzie logistycznym jest następująca:

$$\Pr(Y < x \cdot \beta) = \frac{e^{x \cdot \beta}}{1 + e^{x \cdot \beta}} = \Lambda(x \cdot \beta). \quad (4)$$

Estymacja modelu (1), gdzie  $F(\cdot)$  jest dystrybuantą jednego z rozkładów prawdopodobieństwa, oparta jest najczęściej na metodzie największej wiarygodności. Obserwację o numerze  $t$ , czyli  $y_t$  (realizację jednowymiarowej zmiennej losowej  $Y_t$ ) traktujemy jako pojedyncze losowanie z rozkładu Bernoulliego. Prawdopodobieństwo sukcesu wynosi  $F(x_t \cdot \beta)$ , więc funkcja prawdopodobieństwa  $Y_t$  ma postać  $[F(x_t \cdot \beta)]^{y_t} \cdot [1 - F(x_t \cdot \beta)]^{1 - y_t}$ , gdzie  $x_t$  jest wektorem-wierszem wartości  $k$ -zmiennych objaśniających dla obserwacji o numerze  $t$ , a wektor nieznanymi parametrów  $\beta$  ma wymiar  $k \times 1$ . Zatem łączną funkcję prawdopodobieństwa w przypadku  $T$  niezależnych obserwacji można zapisać jako iloczyn funkcji prawdopodobieństwa, otrzymując:

$$\Pr(Y_1 = y_1, \dots, Y_T = y_T) = \prod_{t=1}^T F(x_t \cdot \beta)^{y_t} \cdot (1 - F(x_t \cdot \beta))^{1 - y_t}. \quad (5)$$

W celu zastosowania MNW logarytmujemy funkcję wiarygodności, uzyskując:

$$\ln L(\beta | y, X) = \sum_t [y_t \cdot \ln F(x_t \cdot \beta) + (1 - y_t) \cdot \ln(1 - F(x_t \cdot \beta))], \quad (6)$$

gdzie  $y$  oznacza (inaczej niż wcześniej) wektor  $T \times 1$  obserwacji zmiennej objaśnianej, a  $X$  macierz  $T \times k$  wartości zmiennych objaśniających. Warunek konieczny dla uzyskania rozwiązania maksymalizującego funkcję wiarygodności ma postać układu  $k$  równań:

$$\frac{\partial \ln L(\beta | y, X)}{\partial \beta} = \sum_t x_t' \cdot \left[ \frac{y_t \cdot f(x_t \cdot \beta)}{F(x_t \cdot \beta)} - \frac{(1 - y_t) \cdot f(x_t \cdot \beta)}{1 - F(x_t \cdot \beta)} \right] = \mathbf{0}_{k \times 1}, \quad (7)$$

gdzie  $f(x_t \cdot \beta)$  jest pochodną  $F(x_t \cdot \beta)$  względem wskaźnika  $x_t \cdot \beta$ , czyli wartością funkcji gęstości w punkcie  $x_t \cdot \beta$ . Powyższy układ równań jest nieliniowy ze względu na  $\beta$ , więc jego rozwiązanie wymaga odpowiednich technik numerycznych. W przypadku modelu logitowego układ równań (7) ma postać:

<sup>5</sup> Wykorzystuje się również dystrybuanty innych rozkładów, np. Weibulla.

$$\frac{\partial \ln L(\beta | y, X)}{\partial \beta} = \sum_i x'_i \cdot [y_i - \Lambda(x_i \cdot \beta)] = \mathbf{0}, \quad (8)$$

natomiast dla modelu probitowego równania te można zapisać następująco:

$$\frac{\partial \ln L(\beta | y, X)}{\partial \beta} = \sum_i x'_i \cdot \left[ y_i \cdot \frac{\phi(x_i \cdot \beta)}{\Phi(x_i \cdot \beta)} - \frac{(1 - y_i) \cdot \phi(x_i \cdot \beta)}{(1 - \Phi(x_i \cdot \beta))} \right] = \mathbf{0}, \quad (9)$$

gdzie  $\Phi(\cdot)$  i  $\phi(\cdot)$  są odpowiednio dystrybuantą i gęstością zmiennej losowej o standaryzowanym rozkładzie normalnym.

Macierz drugich pochodnych cząstkowych w przypadku modelu logitowego i probitowego jest macierzą ujemnie określoną dla dowolnego  $\beta$ ; zob. [Amemiya 1985]. Zatem logarytm funkcji wiarygodności jest funkcją globalnie wklęsłą, więc estymator MNW otrzymujemy rozwiązując układ równań (8) lub (9). Hesjan zostanie wykorzystany do obliczenia asymptotycznych błędów średnich szacunku, a jego postać w przypadku modelu logitowego jest następująca:

$$H_{\text{logit}} = \frac{\partial^2 \ln L(\beta | y, X)}{\partial \beta \partial \beta'} = - \sum_i \Lambda(x_i \cdot \beta) \cdot [1 - \Lambda(x_i \cdot \beta)] \cdot x'_i \cdot x_i, \quad (10)$$

a w modelu probitowym:

$$H_{\text{probit}} = \frac{\partial^2 \ln L(\beta | y, X)}{\partial \beta \partial \beta'} = - \sum_i \lambda_i \cdot [\lambda_i + x_i \beta] \cdot x'_i \cdot x_i,$$

gdzie:

$$\lambda_i = \lambda_{0i} = \frac{-\phi(x_i \cdot \beta)}{1 - \Phi(x_i \cdot \beta)} \quad \text{dla } y_i = 0 \quad (11)$$

i

$$\lambda_i = \lambda_{1i} = \frac{\phi(x_i \cdot \beta)}{\Phi(x_i \cdot \beta)} \quad \text{dla } y_i = 1.$$

Korzystając z własności MNW, błędy średnie szacunku poszczególnych parametrów otrzymujemy jako pierwiastki elementów diagonalnych asymptotycznej macierzy kowariancji postaci:

$$V_{MNW} = -E \left[ \frac{\partial^2 \ln L(\beta | y, X)}{\partial \beta \partial \beta'} \right]^{-1}, \quad (12)$$

gdzie  $E[\cdot]$  oznacza wartość oczekiwaną w rozkładzie próbkowym wektora obserwacji  $y$ . Estymator tej macierzy,  $\hat{V}_{MNW}$ , uzyskuje się przyjmując w równaniu (12) dla  $\beta = \hat{\beta}_{MNW}$ , gdzie  $\hat{\beta}_{MNW}$  jest estymatorem MNW. W przypadku modelu logitowego  $E[H_{\text{logit}}] = H_{\text{logit}}$ , wystarczy więc jedynie odwrócić macierz

$-H_{\text{logit}}$ , a następnie obliczyć jej wartość dla ocen MNW. Natomiast w przypadku modelu probitowego T. Amemiya [1981, 1985] pokazał, że:

$$E \left[ \frac{\partial^2 \ln L(\beta | y, X)}{\partial \beta \partial \beta'} \right]_{\text{probit}} = \sum_i \lambda_{0i} \cdot \lambda_{1i} \cdot x_i' \cdot x_i \quad (13)$$

Związki między parametrami modeli probitowych i logitowych oraz inne metody estymacji modeli dwumianowych przedstawia T. Amemiya [1985].

W następnej części przedstawiono wyniki empiryczne estymacji obu modeli z wykorzystaniem metody MNW. Następnie dokonano predykcji ryzyka niewypłacalności w przypadku różnych typów klientów banku.

### 3. Wyniki empiryczne

Do estymacji modelu logitowego i probitowego wykorzystano dane pochodzące z polskiego banku komercyjnego. Dane te obejmują informacje o kredytach detalicznych, tj. kredytach konsumpcyjnych i hipotecznych, udzielonych w okresie 1.01.2000–30.09.2001 r. Przyjmijmy, że zmienna objaśniana  $Y$  przyjmuje dwie wartości:

1)  $Y = 1$ , w przypadku gdy kredytobiorca na dzień 30.09.2001 r. ma zaległości w spłacie rat kapitałowo-odsetkowych<sup>6</sup>, tzn. opóźnienie w spłacie ostatniej raty wynosi więcej niż 1 miesiąc,

2)  $Y = 0$ , w przypadku gdy kredytobiorca na dzień 30.09.2001 r. w terminie spłaca raty kapitałowo-odsetkowe od zaciągniętego kredytu.

Powyższa definicja zmiennej  $Y$  wynika z uregulowań prawnych dotyczących tworzenia rezerw na należności zagrożone, opisanych na wstępie niniejszego opracowania. Mianowicie, gdy zaległości w spłacie rat kapitałowo-odsetkowych przekroczą okres 1 miesiąca, to należność zostaje automatycznie przekwalifikowana do kategorii „poniżej standardu”, co dla banku skutkuje tworzeniem rezerwy w wysokości 20%, po 3 miesiącach 50%, a po 6 miesiącach 100% wartości zadłużenia<sup>7</sup>. Możemy zatem przyjąć dla uproszczenia, że gdy  $Y = 1$ , wówczas kredytobiorca jest „złym” klientem banku, a w przeciwnym wypadku „dobrym”. W przypadku gdybyśmy rozważali więcej niż dwie kategorie zmiennej  $Y$  (np. analizowali należności normalne, poniżej standardu, wątpliwe, stracone), to wówczas wykorzystamy modele wielomianowe (ang. *multinomial models*); zob. [Amemiya 1985]).

Jeżeli chodzi o dobór zmiennych objaśniających ryzyko niewypłacalności pojedynczego kredytobiorcy, to w literaturze z tego zakresu proponuje się, aby uwzględnić m.in.:

<sup>6</sup> W przypadku kredytów z odsetkami płatnymi z góry klient ma obowiązek spłacać jedynie raty kapitałowe.

<sup>7</sup> Rezerwy są tworzone od kwoty należności zagrożonych (bez narosłych odsetek karnych, kosztów upomnień klienta itp.) pomniejszonych o stosowne zabezpieczenia; zob. Uchwała nr 8/1999 Komisji Nadzoru Bankowego z 22 grudnia 1999 r.

- zmienne charakteryzujące cechy osobowe i demograficzne kredytobiorcy, np. płeć, wiek, stan cywilny, miejsce zamieszkania, wykształcenie, liczba osób na utrzymaniu itp.,
- zmienne charakteryzujące zatrudnienie, m.in. zawód, miejsce pracy,
- zmienne ekonomiczne opisujące zamożność: posiadanie własnego domu (mieszkania), posiadanie samochodu itp.,
- zmienne finansowe przedstawiające dotychczasowe relacje klienta z bankiem: posiadanie kart płatniczych, rachunków depozytowych, zaciąganie i przebieg spłaty dotychczasowych kredytów; zob. [Gruszczyński 2000].

W niniejszej pracy przy doborze zmiennych objaśniających uwzględniono powyższe wskazówki, jednakże w przypadku analizowanego banku część informacji o klientach była niekompletna, co wpłynęło zarówno na liczbę zmiennych, jak i liczebność zbioru obserwacji poddanych modelowaniu. W analizie tej wykorzystano kilkadziesiąt tysięcy rachunków kredytowych<sup>8</sup>, a jako potencjalne zmienne wyjaśniające ryzyko pojedynczej umowy kredytowej przyjęto:

- płeć (zmienna przyjmuje wartość 1, jeżeli klientem jest mężczyzna, 0 w przypadku kobiety),
- wiek klienta (w latach),
- wpływy, tzn. wielkość kwartalnych wpływów w latach 2000–2001 (w tys. zł) na rachunki *a vista* kredytobiorcy w badanym banku (przede wszystkim rachunki oszczędnościowo-rozliczeniowe ROR); jeżeli kredytobiorca nie posiada rachunku ROR w tym banku, przyjęto, że wpływy wynoszą zero,
- posiadanie przez kredytobiorcę rachunku ROR w analizowanym banku (1 – posiada, 0 – nie posiada),
- informację o tym, czy kredytobiorca posiada karty płatnicze wydane przez bank (1 – posiada przynajmniej jedną kartę płatniczą, 0 – nie posiada),
- sposób pozyskania klienta jako kredytobiorcy (1 – poprzez pośrednika kredytowego, 0 – bezpośrednio przez bank),
- typ kredytu (1 – kredyt konsumpcyjny, 0 – kredyt hipoteczny),
- podstawowe źródło dochodu uzyskiwanego przez kredytobiorcę (zmienna *zrdoch*), tj. umowa o pracę, albo renta lub emerytura, albo własna działalność, umowa o dzieło lub umowa zlecenie, albo inne źródło, np. stypendium.

Ostatnia zmienna może przyjmować cztery różne wartości. Chcąc ją uwzględnić w równaniu regresji z wyrazem wolnym, za referencyjną wartość tej zmiennej przyjęto „umowę o pracę” (dla 75% kredytobiorców stanowi podstawowe źródło dochodu). Zatem w modelu uwzględniono typ źródła dochodu, wprowadzając trzy zmienne zero-jedynkowe:

- *zrdoch1* = 1, gdy źródłem dochodu kredytobiorcy jest renta lub emerytura, *zrdoch1* = 0 w przeciwnym wypadku,

<sup>8</sup> Nie podano dokładnej liczby analizowanych rachunków, ponieważ mogłoby to posłużyć do identyfikacji banku, który wyraził zgodę na ich wykorzystanie w niniejszym opracowaniu pod warunkiem niewskazywania źródła danych.



–  $zrdoch2 = 1$ , gdy źródłem dochodu kredytobiorcy jest własna działalność, umowa o dzieło lub umowa zlecenie,  $zrdoch2 = 0$  w przeciwnym wypadku,

–  $zrdoch3 = 1$ , gdy źródłem dochodu jest np. stypendium,  $zrdoch3 = 0$  w przeciwnym wypadku.

Z powyższego wynika, że gdy  $zrdoch1 = 0$ ,  $zrdoch2 = 0$  i  $zrdoch3 = 0$ , to źródłem dochodu kredytobiorcy jest umowa o pracę.

Wprowadzenie do modelu zmiennej informującej, czy kredytobiorca posiada ROR lub przynajmniej jedną kartę płatniczą, ma odzwierciedlać jego dotychczasowe relacje z bankiem (przywiązanie, lojalność wobec banku). Szczególnego wyjaśnienia wymaga zmienna, która informuje o sposobie pozyskania przez bank klienta jako kredytobiorcy. Z uwagi na fakt, że obsługa klientów detalicznych w zakresie sprzedaży produktów kredytowych jest pracochłonna i często nie przynosi bezpośrednich efektów, badany bank podpisał umowę z zewnętrznymi firmami, które zajmują się pozyskaniem i weryfikacją klientów – potencjalnych kredytobiorców. Za pośrednictwo firmy te otrzymują prowizję jako procent od wartości udzielonych kredytów. Pozyskani w ten sposób kredytobiorcy w przeważającej większości nie posiadali rachunku ROR w badanym banku i nie korzystali wcześniej z innych usług. Zmienna ta, jak pokażą wyniki empiryczne, będzie miała decydujący wpływ na określenie prawdopodobieństwa wypłacalności klienta. Podział rachunków kredytowych na konsumpcyjne i hipoteczne jest spowodowany różną formą ich zabezpieczeń. W przypadku kredytów konsumpcyjnych podstawową formą zabezpieczenia jest weksel in blanco lub poręczenie osób trzecich. Natomiast kredyty hipoteczne są zabezpieczone przede wszystkim hipoteką na nieruchomości będącej własnością kredytobiorcy, własnościowym spółdzielczym prawie do lokalu mieszkalnego lub domu jednorodzinnego, udziale we współwłasności lub prawie użytkowania wieczystego. Zabezpieczenia kredytów hipotecznych dają bankowi pewniejszą i szybszą możliwość wyegzekwowania długu od niesolidnego klienta; przeciętne prawdopodobieństwo spłacenia tego kredytu powinno więc być istotnie wyższe niż kredytu konsumpcyjnego.

Do estymacji parametrów modelu probitowego i logitowego wykorzystano dane o rachunkach kredytowych klientów detalicznych. Podstawowe charakterystyki tego zbioru danych przedstawia tabela 1. Udział „złych” kredytów, w przypadku których klienci zalegali ze spłatą rat kapitałowo-odsetkowych, wynosi 20%. Kobiety były równie często kredytobiorcami jak mężczyźni. Większość z kredytobiorców była dotychczasowymi klientami banku, posiadała rachunki ROR, a co trzeci korzystał z kart płatniczych. Kredyty hipoteczne stanowiły tylko 6% całkowitej liczby kredytów, a 38% umów kredytowych podpisano przy udziale pośrednika. W uzupełnieniu można dodać, że średni wiek kredytobiorcy wynosił 40 lat, a dla 75% z nich głównym źródłem uzyskiwanego dochodu była umowa o pracę, dla 17% – renta lub emerytura, a dochody z własnej działalności uzyskiwało 6% kredytobiorców. Wielkość miesięcz-

nych wpływów na rachunki ROR właścicieli rachunków kredytowych nie przekraczała kwoty 10,2 tys. zł<sup>9</sup>.

Tabela 1. Podstawowe ilościowe informacje o rachunkach kredytowych i ich właścicielach

Wyszczególnienie	Struktura (%)
Udział „złych” kredytów	20
Struktura według płci (udział mężczyzn)	53
Posiadający ROR	56
Posiadający karty płatnicze	33
Klient/kredyt pozyskany przez pośrednika	38
Struktura według typu kredytu (konsumpcyjny)	94

Źródło: opracowanie własne.

W przypadku analizowanych danych jakościowych istniała obawa, że występuje silna współliniowość między zmiennymi objaśniającymi. W tym celu zastosowano miernik uwarunkowania macierzy  $X'X$  (zob. [Osiewalski 1992])<sup>10</sup>. Jednakże uzyskane wyniki nie potwierdziły występowania współliniowości. Obliczono także podstawową miarę zależności między zmiennymi, tj. współczynniki korelacji liniowej Pearsona, co przedstawia tabela 2.

Tabela 2. Współczynniki korelacji liniowej Pearsona między zmiennymi

Zmienna	Y	Płeć	Wiek	Wpływy	ROR	Karty	Pośrednik	Typ kredytu	Źródło dochodu
Y	1								
Płeć	0,041	1							
Wiek	-0,002	-0,126	1						
Wpływy	-0,020	0,012	-0,008	1					
ROR	-0,448	-0,030	-0,170	0,038	1				
Karty	-0,291	-0,043	-0,136	0,042	0,619	1			
Pośrednik	0,506	0,038	0,140	-0,034	-0,834	-0,518	1		
Typ kredytu	0,116	0,033	0,046	-0,018	-0,197	-0,172	0,203	1	
Źródło dochodu	-0,043	0,022	0,248	0,034	0,030	-0,035	-0,035	-0,027	1

Źródło: opracowanie własne.

<sup>9</sup> 19% z nich ma wpływy wyższe niż 10 tys. zł, a mediana wynosi 2,6 tys. zł. Co ósmy klient (przeważnie o wysokich dochodach) ma zaciągnięty więcej niż jeden kredyt.

<sup>10</sup> Za miarę współliniowości przyjęto pierwiastek ilorazu maksymalnej i minimalnej wartości własnej macierzy  $R_n$ , przy czym  $R_n = W^{-1}(X'X)W^{-1}$ , gdzie  $W$  jest macierzą diagonalną stopnia  $k$  zawierającą na przekątnej długości kolumn macierzy  $X$ .

Na podstawie danych z tabeli 2 możemy stwierdzić, że występuje silna korelacja między sposobem pozyskania klienta i posiadaniem rachunku ROR a niewypłacalnością kredytobiorcy. Ponadto kredytobiorcy pozyskani przez pośredników w przeważającej większości nie posiadali rachunku ROR w analizowanym banku.

Do estymacji modelu logitowego i probitowego wykorzystano opisaną wcześniej metodę największej wiarygodności, dostępną w pakiecie Gauss 3.2 firmy Aptech Systems. W celu zbadania numerycznej zbieżności wykorzystywanego algorytmu optymalizacyjnego zbadano wrażliwość uzyskiwanych wyników na dobór punktów startowych. Za oceny wstępne wektora  $\beta$  przyjęto najpierw oceny estymatora MNK dla liniowego modelu prawdopodobieństwa (zob. tabela 3), następnie współczynniki korelacji liniowej Pearsona (z pierwszej kolumny tabeli 2) oraz założono, że wstępną oceną  $\beta$  jest wektor składający się z zer.

Tabela 3. Oceny MNK parametrów liniowego modelu prawdopodobieństwa

Zmienna	Oceny	Błędy średnie szacunku	Statystyka $t$
Stała	0,165	0,012	14,302
Płeć	0,007	0,004	2,091
Wiek	-0,002	0,00017	-11,450
Wpływy	0,000005	0,000007	-0,689
ROR	-0,067	0,007	-9,749
Karty	-0,022	0,005	-4,660
Pośrednik	0,355	0,006	55,006
Typ kredytu	0,022	0,007	3,062
Zrdoch1	-0,028	0,006	-4,612
Zrdoch2	0,023	0,007	3,101
Zrdoch3	-0,035	0,012	-2,938

Źródło: opracowanie własne.

W obu modelach (logitowym i probitowym) uzyskiwaliśmy te same rozwiązania bez względu na wybór jednego z trzech punktów startowych. Oceny MNW i błędy średnie szacunku dla obu modeli przedstawia tabela 4. Oczywiście oceny parametrów obu modeli są różne, ale z uwagi na zależności między parametrami tych modeli, o których mówi T. Amemiya [1985], wartości statystyk  $t$ -Studenta przyjmują zbliżone wartości. Oceny wszystkich parametrów są statystycznie istotne, przy czym jedynie w przypadku płci i typu kredytu błędy średnie szacunku są relatywnie duże w stosunku do ocen.

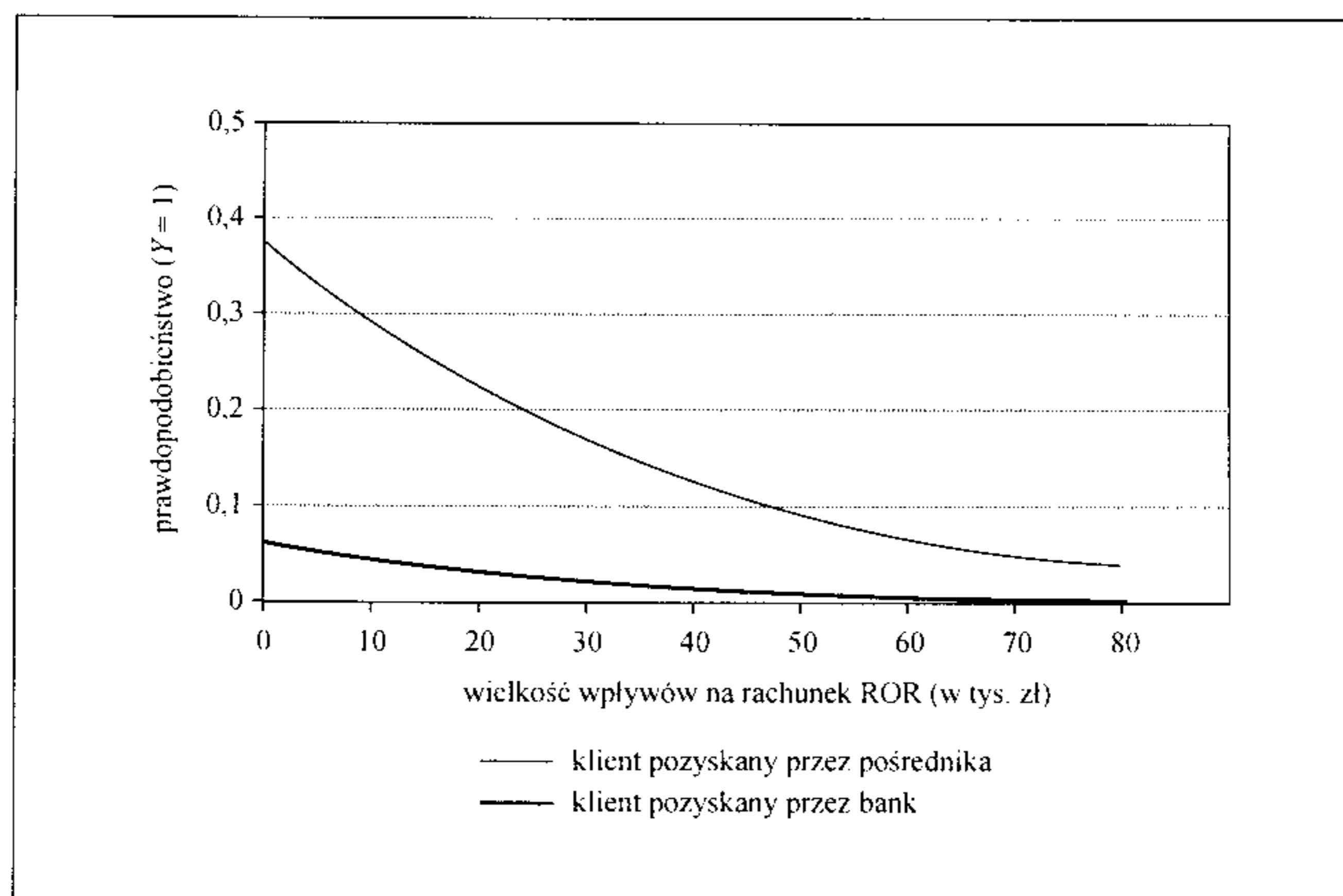
Do celów predykcji posłużymy się modelem probitowym, który daje prognozy niewiele różniące się od uzyskanych w modelu logitowym. Ewentualne różnice wynikają z faktu, że rozkład logitowy ma większą masę prawdopodobieństwa w ogonach niż rozkład normalny. W konsekwencji dla ustalonych

wartości wskaźnika  $x_i \cdot \beta > 0$  ( $x_i \cdot \beta < 0$ ) w modelu probitowym otrzymujemy wyższe (niższe) wartości prognozy zmiennej  $Y$  niż w modelu logitowym.

Tabela 4. Oceny MNW parametrów modelu probitowego i logitowego

Zmienna	Model probitowy			Model logitowy		
	oceny	błędy średnie szacunku	statystyka $t$	oceny	błędy średnie szacunku	statystyka $t$
Stała	-1,199	0,077	-15,521	-2,146	0,163	-13,201
Płeć	0,043	0,018	2,439	0,053	0,024	2,260
Wiek	-0,009	0,001	-10,017	-0,015	0,001	-13,220
Wpływy	-0,017	0,002	-9,394	-0,068	0,007	-10,223
ROR	-0,285	0,038	-7,588	-0,357	0,070	-5,092
Karty	-0,174	0,033	-5,236	-0,303	0,069	-4,365
Pośrednik	1,269	0,031	40,789	2,230	0,055	40,782
Typ kredytu	0,181	0,065	2,780	0,377	0,153	2,473
Zrdoch1	-0,089	0,029	-3,081	-0,138	0,039	-3,591
Zrdoch2	0,311	0,040	7,721	0,490	0,059	8,327
Zrdoch3	-0,227	0,075	-3,039	-0,429	0,125	-3,445

Źródło: opracowanie własne.



Rys. 1. Prognoza prawdopodobieństwa niewypłacalności klienta w zależności od zmiennej *pośrednik* i wielkości wpływów na ROR

Źródło: opracowanie własne.

Tabela 5. Prognoza prawdopodobieństwa niewypłacalności dla różnych typów klientów

Zmienna	Najczęstszy klient		Młody biznesmen	Starsza kobieta
	pośrednik = 1	pośrednik = 0		
Stała	1	1	1	1
Płeć	1	1	1	0
Wiek	40,2	40,2	18	60
Wpływy (w tys. zł)	10,2	10,2	10,2	1
ROR	1	1	0	1
Karty płatnicze	0	0	0	1
Pośrednik	1	0	1	0
Typ kredytu: konsumpcyjny	1	1	1	0
Zrdoch1	0	0	0	1
Zrdoch2	0	0	1	0
Zrdoch3	0	0	0	0
Pr( $Y = 1$ )	0,299	0,036	0,606	0,013

Źródło: opracowanie własne.

Znaki i wartości ocen parametrów przy zmiennych objaśniających informują o kierunku i sile wpływu tych zmiennych na prawdopodobieństwo niewypłacalności kredytobiorcy. Z powyższych danych wynika m.in., że prawdopodobieństwo niespłacenia kredytu w przypadku klienta będącego mężczyzną jest nieznacznie większe niż w przypadku kobiety. Prawdopodobieństwo to jest niższe, jeżeli klient posiada rachunek ROR lub korzysta z kart płatniczych, i maleje ono wraz z wiekiem klienta oraz wielkością wpływów na rachunek ROR. Forma zabezpieczenia ma wpływ na terminowość i rzetelność spłaty kredytu przez klienta, więc mniejsze ryzyko związane jest z kredytem hipotecznym niż konsumpcyjnym. Innym czynnikiem zmniejszającym niewypłacalność klienta jest źródło dochodu, przy czym studenci korzystający z kredytu studenckiego ( $zrdoch3 = 1$ ) oraz emeryci i renciści ( $zrdoch1 = 1$ ) są lepszymi (obciążonymi mniejszym ryzykiem) kredytobiorcami niż klienci utrzymujący się z umowy o pracę. Natomiast osoby prowadzące własną działalność gospodarczą ( $zrdoch2 = 1$ ) są najmniej pożądaną grupą klientów z uwagi na związane z nimi duże ryzyko kredytowe. Jednakże głównym czynnikiem, który ma wpływ na wielkość ryzyka związanego z pojedynczym wnioskiem kredytowym, jest sposób pozyskania klienta jako kredytobiorcy (zmienna *pośrednik*). W przypadku badanego banku korzystanie z usług pośredników, którzy mieli pozyskać klientów spełniających wymagania stawiane kredytobiorcom, doprowadziło do pogorszenia jakości posiadanego portfela kredytowego. Współpraca ta spowodowała dla banku utratę przychodów odsetkowych oraz naraziła go na dodatkowe koszty i często bezpowrotną utratę pożyczonego kapitału. Zmienna ta odgrywa wyraźnie decydującą rolę w kształtowaniu się prawdopodobieństwa niewypłacalności kredytobiorcy. Jej znaczenie ilustruje rys. 1, który przedstawia, w jaki sposób zmienia się prawdopodobieństwo

w zależności od zmiennych *pośrednik* i *wielkość wpływów na rachunek ROR* (przy wartościach pozostałych zmiennych zero-jedynkowych ustalonych jako wartości najczęstsze w próbie, a w przypadku *wiek klienta* na poziomie przeciętnym). W przypadku gdy wpływy klienta pozyskanego bezpośrednio przez bank i poprzez pośrednika są nie większe niż 10 tys. zł, prawdopodobieństwo zaniechania spłaty kredytu przez pierwszego klienta wynosi od 0,04 do 0,05, natomiast drugiego aż od 0,3 do 0,36. W przypadku klientów pozyskanych przez bank wielkość wpływów, w przeciwieństwie do klientów pozyskanych przez pośredników, nie ma zasadniczego wpływu na zmianę prawdopodobieństwa niewywiązywania się z umowy kredytowej.

Na podstawie modelu możemy dokonać oszacowania prawdopodobieństwa „złego” kredytu w przypadku różnych typów klientów. Szczegółową specyfikację wybranych sylwetek kredytobiorców, w tym najczęściej spotykanego, przedstawia tabela 5. Prawdopodobieństwo, że dotychczasowa klientka banku – starsza kobieta – korzystająca już z usług banku, nie będzie spłacać kredytu hipotecznego, jest bliskie zero. Natomiast w przypadku innego hipotetycznego klienta, którego bank pozyskał przez pośrednika – młodego mężczyzny uzyskującego dochód z własnej działalności na poziomie średnim w próbie (10 tys. zł), prawdopodobieństwo to wynosi aż 0,6. Oszacowane prawdopodobieństwo dla klienta pozyskanego przez bank o cechach najczęstszych w próbie (dotyczy zmiennych jakościowych) kształtuje się na poziomie 0,04, a dla klienta pozyskanego przez pośrednika prawdopodobieństwo to wynosi aż 0,3. Powyższe wyniki wskazują, że rola zmiennej *pośrednik* ma decydujący wpływ na to, czy kredyt można traktować jako „zły”.

## 5. Podsumowanie

Wyniki analizy empirycznej obejmującej rachunki kredytowe klientów detalicznych wskazują, że głównym czynnikiem zwiększającym prawdopodobieństwo niewypłacalności klienta jest sposób jego pozyskania. Bezpośredni kontakt pracownika banku z potencjalnym kredytobiorcą i weryfikacja przedłożonego wniosku kredytowego obniża w znaczący sposób ryzyko kredytowe. Natomiast korzystanie z usług firm zewnętrznych pośredniczących w pozyskaniu i udzielaniu kredytów zwiększa prawdopodobieństwo niewypłacalności klienta. Czynnikiem zmniejszającym ryzyko jest przywiązanie klienta do banku, które przejawia się stopniem korzystania przez niego z usług bankowych. Posiadanie rachunku ROR jest czynnikiem obniżającym ryzyko kredytowe, podobnie jak korzystanie z kart płatniczych.

Oczywiście występuje wiele innych czynników nie uwzględnionych w modelu, które być może istotnie przyczyniają się do kształtowania ryzyka kredytowego. Jednakże brak danych (bądź ich niekompletność) uniemożliwia przeprowadzenie szerszej analizy empirycznej i weryfikacji statystycznej, która prawdopodobnie przyniosłaby ciekawsze i pełniejsze wnioski.

Intencją autora jest przedstawienie, że w przypadku badanego banku możliwe jest zbudowanie sprawnego mechanizmu – modelu scoringowego, którego wykorzystanie w praktyce pozwoliłoby m.in. na ujednoczenie, obiektywizację i częściową automatyzację oceny wniosków kredytowych. W wymiarze finansowym przyczyniłoby się to do obniżenia kosztów związanych z tworzonymi rezerwami poprzez poprawę jakości portfela kredytowego oraz do obniżenia kosztów obsługi klienta poprzez usprawnienie procesu rozpatrywania wniosków kredytowych.

## Literatura

- Aldrich J., Nelson F. [1984], *Linear Probability, Logit, and Probit Models*, Sage, Beverly Hills, California.
- Amemiya T. [1981], *Qualitative Response Models: A Survey*, „Journal of Economic Literature”, vol. 19.
- Amemiya T. [1985], *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.
- Greene W.H. [1993], *Econometric Analysis*, Macmillan, New York.
- Gruszczyński M. [2001], *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Monografie i Opracowania SGH, Warszawa, nr 6.
- Kuryłek W. [2000], *Credit scoring – podejście statystyczne*, „Bank i Kredyt”, nr 6.
- Osiewalski J. [1992], *Uogólnione niescentrowane współczynniki zwiększenia wariacji*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, Kraków, nr 374.

## Test of Borrower Insolvency Using Logit and Probit Models

The purpose of this article is to use binary choice models in the analysis of credit agreement risk. The author presents construction and estimation of logit and probit models using maximum likelihood as a method of estimation. The paper describes the results of empirical research using data on consumer loans (loans to individuals), coming from one of Polish commercial banks. The results show that the probability that bank's client won't pay off instalments depends in particular on the way bank finds a client, on the type of the loan, and on the source of client's income. The author also calculates credit agreement risk in the case of various clients of the bank.