

Jerzy Marzec¹

Katedra Ekonometrii i Badań Operacyjnych
Uniwersytet Ekonomiczny w Krakowie

POMIAR KORZYŚCI Z ZASTOSOWANIA MODELU WIELOMIANOWEGO W OGRANICZANIU RYZYKA KREDYTOWEGO

Wprowadzenie

Jednym z głównych założeń, leżących u podstaw modeli ekonomicznych opisujących działalność banków, jest przyjęcie, że działają one na rynku niedoskonałym, który charakteryzuje się niepełną informacją i niepewnością. Na podstawową (tradycyjną) działalność banków składa się przyjmowanie depozytów i transformowanie je w kredyty z wykorzystaniem pracy ludzkiej i kapitału fizycznego; zob. Sealey i Lindley (1977). Zatem z ekonomicznego punktu widzenia zarządzanie bankiem sprowadza się w uproszczeniu do zarządzania aktywami i pasywami. Ryzyko kredytowe, podobnie jak ryzyko płynności, jest związane z bilansem banku i „oznacza potencjalną niemożność odzyskania pełnej wartości księgowej aktywów”; zob. Matthews i Thomson [2007] str. 69. Sytuacja ta może być wynikiem niespłacenia przez pożyczkobiorców rat kredytowych lub należnych odsetek. Bank jest przedsiębiorstwem maksymalizującym zysk, jednakże problem decyzyjny różni się od tego, który charakteryzuje typowe przedsiębiorstwo. Ryzyko, które towarzyszy udzielaniu kredytów, czyli sprzedaży podstawowego produktu bankowego, powoduje występowanie zjawiska ich racjonowania. To z kolei jest w sprzeczności z założeniami modeli popytu i podaży, w których zakłada się istnienie ceny równowagi, czyli takiego oprocentowania kredytów, po którym wszyscy potencjalni kredytobiorcy mogą zakupić kredyty. Teorie endogenicznego racjonowania kredytów należą do tych rozwiązań, które są spójne z zasadą maksymalizacji zysku i opisują racjonalne zachowanie się banków w sytuacji niedoskonałej informacji; zob. np. Hodgman (1960), Jaffee i Russell (1976) i pozycje bibliograficzne zawarte w książkach Matthews i Thomsona (2007) oraz Freixasa i Rocheta (2007). Modele te zakładają występowanie we wzajemnych relacjach między bankiem a kredytobiorcą takich zjawisk jak: asymetria informacji, negatywna selekcja, negatywne bodźce i pokusa nadużycia. Bank jest zainteresowany osłabieniem skutków powyższych czynników, szczególnie pierwszych trzech. Modele punktowej oceny zdolności kredytowej (ang. *credit scoring*) są właśnie tymi narzędziami, które umożliwiają *ex ante* określenie stopnia ryzyka związanego ze spłatą kredytu, czyli rozróżnienie na podstawie wcześniej zebranych infor-

¹ Uniwersytet Ekonomiczny w Krakowie, ul. Rakowicka 27, 31-510 Kraków, e-mail: marzecj@uek.krakow.pl. Artykuł powstał w ramach badań statutowych finansowanych przez Uniwersytet Ekonomiczny w Krakowie.

macji o kredytobiorcach bezpiecznych i ryzykownych, co przyczynia się do ograniczenia ryzyka kredytowego.

W niniejszym artykule prezentuje się pomiar korzyści finansowych z zastosowania modeli wielomianowych dla kategorii uporządkowanych w procesie podejmowania decyzji kredytowych. W tym celu wykorzystano elementy statystycznej teorii decyzji, która w warstwie normatywnej pozwala określić decyzje, które są optymalne ze względu na cele decydenta przy dostępnej, często ograniczonej informacji. Jeżeli bank pragnie maksymalizować swój oczekiwany zysk z działalności kredytowej, to powinien postępować według określonych zasad. W tym artykule proponuje się, aby w ramach systemu scoringowego stosować reguły udzielania kredytu, które zależą od marż kredytowych i prawdopodobieństw niespłacenia kredytów przez potencjalnych kredytobiorców. Ekonometryczny model wielomianowy dla kategorii uporządkowanych jest opisem zachowania się kredytobiorcy wobec spłaty długu. Jego estymacja na podstawie danych historycznych umożliwia wyznaczenie ocen a posteriori parametrów rozkładu prawdopodobieństwa zmiennej charakteryzującej niespłacalność kredytobiorcy. Niniejsze badania stanowią kontynuację tematyki dotyczącej wykorzystania modeli danych jakościowych w podejmowaniu decyzji kredytowych i oceny korzyści finansowych z tego tytułu dla banku; zob. Marzec (2009, 2010).

Proponowane podejście różni się istotnie od prezentowanego w literaturze polskiej. Wyniki badań empirycznych dotyczących oceny wiarygodności polskich kredytobiorców przedstawiają m.in. Chrzanowska i Witkowska (2005), Witkowska i Chrzanowska (2005, 2006), Misztal (2006). Wykorzystują oni przede wszystkim sieci neuronowe, drzewa klasyfikacyjne bądź analizę dyskryminacyjną. Prace te mają charakter czysto empiryczny, główny nacisk jest położony na porównywanie wyników klasyfikacji otrzymanych za pomocą wspomnianych metod. W literaturze światowej, takie podejście można znaleźć m.in. w artykułach: Fritz i Hosemann (2000), Bensic i inni (2005), Trinkle i Baldwin (2007). Natomiast niniejsze badania nawiązują do dyskusji prowadzonej m.in. przez Thomasa (2000) oraz Thomasa, Olivera i Handa (2005), którzy uważają, że w przyszłości istotną rolę dla banku będzie pełnić scoring zysku, gdyż umożliwia on na poziomie pojedynczego klienta szacowanie zysku ze sprzedaży różnych produktów. Wobec powyższego, interdyscyplinarne ujęcie oceny przydatności modeli statystycznych w ograniczaniu ryzyka kredytowego – prezentowane w niniejszym artykule – wydaje się być ciekawym od strony praktycznej i poprawnym od strony metodycznej spojrzeniem na jakże ważny element działalności w banku – podejmowanie decyzji kredytowych.

1. Modele wielomianowe – definicja

W literaturze ekonometrycznej modele dla jakościowych zmiennych endogenicznych lub inaczej modele dyskretnego wyboru (ang. *quantal response* lub *discrete choice models*) przedstawiają

zależność między wynikiem dokonywanych wyborów a egzogenicznymi zmiennymi objaśniającymi, które opisują cechy możliwych alternatyw lub indywidualne charakterystyki podmiotów podejmujących decyzje.

Podstawowa definicja tych modeli opiera się na zaproponowanej przez McFaddena koncepcji stochastycznej funkcji użyteczności (ang. *random utility function*); zob. *Structural Analysis ...* [1981].² Funkcja użyteczności związana z decyzją o numerze j , opisująca preferencje konsumenta t , reprezentowana jest przez nieobserwowalną (ukrytą) zmienną losową z_{tj} . Podmiot dokonuje takiego wyboru, który przynosi mu najwięcej korzyści, czyli maksymalizuje funkcję użyteczności. Zmienna z_{tj} zależy nie tylko od charakterystyk możliwych wyborów i cech decydenta, ale także od składnika losowego, który reprezentuje wpływ na podjęte decyzje zakłóceń losowych, błędów pomiaru, nieodpowiedniej specyfikacji postaci funkcji użyteczności oraz błędnego postrzegania wielkości kosztów jednostkowych charakteryzujących poszczególne wybory. Podjęta decyzja jest reprezentowana przez endogeniczną zmienną losową, przyjmującą skończoną liczbę wartości.

Najprostszym przypadkiem modelu dyskretnego wyboru jest model dychotomiczny, gdy $J = 2$. Jeżeli zmienna objaśniana mierzona jest na skali porządkowej, to otrzymujemy wielomianowy model dla kategorii uporządkowanych.

W niniejszym artykule przedmiotem analizy jest wielomianowy model dla kategorii uporządkowanych przy założeniu jednakowej liczby alternatyw oraz posiadania danych charakteryzujących jedynie podmiot dokonujący wybór. Wprowadzając ciągłe zmienne z_t , których wartości określają obserwowaną kategorię zmiennej y_t – podjętą decyzję, otrzymujemy model o następującej postaci (zob. McKelvey i Zavoina (1975)):

$$\begin{cases} z_t = \mathbf{x}_t \cdot \boldsymbol{\beta} + \varepsilon_t \\ y_{tj} = 1 & \text{gdy } \alpha_{j-1} < z_t < \alpha_j \text{ dla } t = 1, \dots, T \quad j = 1, \dots, J \\ y_{tj} = 0 & \text{w przeciwnym przypadku,} \end{cases} \quad (1)$$

gdzie α_j spełniające warunek $\alpha_{j-1} < \alpha_j < \alpha_{j+1}$ są tzw. punktami granicznymi (ucięcia) zmiennej z_t , zaś \mathbf{x}_t jest wektorem zmiennych egzogenicznych (lub ich znanych funkcji) charakteryzujących jednostkę podejmującą wybór. Pomocnicza zmienna y_{tj} przyjmuje wartość jeden, gdy $y_t = j$ albo zero w pozostałych przypadkach. Zmienna y_t jest zmienną skalarną, której wartości odpowiadają umownym numerom kategorii, czyli $1, 2, \dots, J$. O składnikach ε_t najczęściej zakłada się, jak w modelach dwumianowych, że są niezależnymi zmiennymi losowymi i posiadają identyczne rozkłady o wartości oczekiwanej równej zero i ustalonej wariancji, co jest jednym ze sposobów narzucenia identyfi-

² Retrospektywne ujęcie podstaw teoretycznych modeli danych dyskretnych przedstawia np. Kapłon [2005].

kowalności parametrów. Przyjmuje się zatem, że $\alpha_0 = -\infty$ i $\alpha_J = +\infty$ oraz $\alpha_1 = 0$, jeżeli w równaniu dla zmiennej z_t występuje wyraz wolny β_1 .

Prawdopodobieństwa zaobserwowania kategorii o numerach $j = 1, \dots, J$, czyli parametry rozkładu zmiennej dyskretnej y_t , są równe różnicy dystrybuant:

$$p_{ij} \equiv \Pr(y_{ij} = 1) = \Pr(\alpha_{j-1} < z_t < \alpha_j) = F(\alpha_j - \mathbf{x}_t \boldsymbol{\beta}) - F(\alpha_{j-1} - \mathbf{x}_t \boldsymbol{\beta}), \quad (2)$$

gdzie $F(a)$ jest wartością dystrybuanty zmiennej losowej ε_t w punkcie a . Najbardziej znanymi przypadkami modelu (1) są modele probitowy i logitowy, które otrzymujemy, gdy $F(\cdot)$ jest dystrybuantą standaryzowanej zmiennej losowej o rozkładzie normalnym albo zmiennej o standardowym rozkładzie logistycznym.

W przypadku tych modeli podstawową i najczęściej stosowaną metodą estymacji, gdy wykorzystuje się dane indywidualne, jest metoda największej wiarygodności, która ma charakter asymptotyczny; zob. Pratt (1981), Maddala (1983), Amemiya (1985). W przypadku niestandardowych modeli – np. z rozkładem t Studenta – w pełni probabilistycznym podejściu do zagadnienia konstrukcji modelu, estymacji jego parametrów i testowania przyjętych założeń na podstawie zarówno dla małej jak i dużej próby, jest wnioskowanie bayesowskie; zob. Albert i Chib (1993), Marzec (2006, 2008a).

2. Definicja modeli zastosowanych w badaniach

W przykładzie empirycznym zostaną przedstawione wyniki porównania zyskowności modeli wielomianowych, zastosowanych w scoringu kredytowym. W tym celu wykorzystano, obok modeli probitowego i logitowego, także specyfikację z rozkładem t Studenta.³

Rozszerzenie standardowych modeli, czyli logitowego i probitowego, może postępować w dwóch kierunkach. Pierwsza proponowana modyfikacja polega na przyjęciu innej postaci zależności między zmienną ukrytą z_t a zmiennymi objaśniającymi. W niniejszych badaniach zostanie wykorzystana formuła wielomianu drugiego stopnia względem zmiennych egzogenicznych w_{th}

$$\mathbf{x}_t \cdot \boldsymbol{\beta} = \beta_1 + \sum_h \beta_h \cdot w_{th} + \sum_h \sum_{i \geq h} \beta_{hi} \cdot w_{th} \cdot w_{ti}. \quad (3)$$

Osiewalski i Marzec (2004a) zaproponowali nazywać specyfikację (3) modelem II rzędu (z aproksymacją liniową), w odróżnieniu do przypadku, w którym $\mathbf{x}_t \cdot \boldsymbol{\beta} = \beta_1 + \sum_h \beta_h \cdot w_{th}$, czyli modelu I rzędu. Takie rozszerzenie spotyka się zwłaszcza w badaniach empirycznych, zob. np. *Discrimination ...* (1998), Hosmer i Lemeshow (2000), Cramer (2004), Moffatt (2005). O korzyściach z konstrukcji opisanej wzorem (3) pisał m.in. Marzec (2006).

³ Rezultaty uzyskane na podstawie modeli dwumianowych przedstawiono w publikacjach: Marzec (2009, 2010).

Druga modyfikacja – z punktu widzenia konstrukcji modeli danych jakościowych jest oryginalna i rzadko stosowana – polega na przyjęciu dla składnika ε_t rozkładu z szerszej klasy. W przypadku tej rodziny modeli Albert i Chib (1993) zaproponowali rozkład t Studenta o nieznannej liczbie stopni swobody $v \in (0; +\infty)$; zob. także Marzec (2003, 2008a), Osiewalski i Marzec (2004). Jednym z motywów zastosowania rozkładu t Studenta jest spostrzeżenie, że rozkład logistyczny może być aproksymowany powyższym rozkładem o stopniach swobody między 7 a 9; zob. Mudholkara i George'a (1978). Zatem model z rozkładem t Studenta stanowi proste uogólnienie modeli probitowego ($v \rightarrow +\infty$) i logitowego ($v \in (7; 9)$).

W niniejszym artykule rozważamy dwa modele: model z rozkładem t Studenta z aproksymacją II rzędu (M_1) oraz model z aproksymacją liniową: probitowy (M_2). Wyniki zaprezentowane w pracach Marzec (2006, 2008b) pokazały, że w modelu M_1 ocena parametru v wynosi około 5,7 a odchylenie standardowe 0,44, otrzymano więc model zbliżony do logitowego. W konsekwencji dla uproszczenia prezentowanej analizy pominięto model logitowy, gdyż uzyskane na jego podstawie wyniki są bardzo zbliżone do tych z modelu M_1 . Warto wspomnieć, że w ujęciu statystycznym model M_1 zdecydowanie lepiej opisuje zjawisko niespłacalności kredytów niż pozostałe dwa prostsze, które jednakże w zagadnieniach scoringu kredytowego są stosowane powszechnie.

3. Model wielomianowy w podejmowaniu decyzji kredytowych – pomiar korzyści

3.1. Opis problemu decyzyjnego

Model wielomianowy może stanowić rdzeń systemu scoringowego, gdy rozważa się kilka typów zachowań potencjalnego kredytobiorcy w odniesieniu do spłaty rat kapitałowo–odsetkowych. Przyjęto, że bank jest zainteresowany podjęciem decyzji o przyznaniu bądź odmowie kredytu w zależności od stanu natury, czyli podejścia kredytobiorcy do spłaty kredytu. W warunkach niepełnej informacji podjęcie decyzji o udzieleniu albo odmowie kredytu jest uzależnione od prawdopodobieństw zaobserwowania poszczególnych zachowań pożyczkobiorców oraz częściowych wypłat pieniężnych będących konsekwencjami decyzji banku i postaw klientów wobec spłaty zaciągniętych długów. Warto przypomnieć, że w modelu dychotomicznym rozróżnia się wyłącznie kredyty niespłacane i spłacane, co najczęściej oznacza, że w pierwszym przypadku bank traci w całości zarówno kapitał, jak i marżę kredytową, a tylko te drugie są źródłem przychodów odsetkowych. W omawianym przypadku dopuszcza się – w odróżnieniu do powyższego modelu – częściową utratę kapitału i odsetek. Istotną kwestią staje się więc zdefiniowanie zmiennej wielomianowej y_t , która reprezentuje niespłacalność kredytów, czyli stany natury.

Z uwagi na dostępność danych proponuje się, aby w modelu wielomianowym definicje zmiennej y_t oprzeć na klasyfikacji należności, która jest podstawą do tworzenia rezerw celowych. Posiadając dane o kredytach detalicznych z lat 2000–2001 rozróżniono cztery kategorie należności: normalne, poniżej standardu, wątpliwe i stracone. Przyjmuje się, iż bank traci cały kapitał i odsetki tylko w przypadku ostatniej kategorii, więc stopa odzysku wynosi zero. Należności o kategorii poniżej standardu i wątpliwe przynoszą tylko część oczekiwanych zysków. Dla uproszczenia zakłada się⁴, że stopa odzysku dla należności o kategorii poniżej standardu wynosi 0,8, w przypadku zaś należności wątpliwych kształtuje się na poziomie 0,5. Gdy kredyty są spłacane w terminie, to bank otrzymuje marżę kredytową równą m (w punktach procentowych), w krańcowym przypadku zaś traci marżę i kapitał.⁵ Decyzję banku odzwierciedla dwupunktowa zmienna d_t . Bank udziela kredytu ($d_t = 1$), gdy oczekiwana wypłata z tytułu jego udzielenia jest większa od zera albo większa od wypłaty otrzymanej, gdy odmawia kredytu ($d_t = 0$). Wybór wariantu zależy od szczegółowej konstrukcji funkcji wypłat, od tego czy rozważamy wyłącznie wypłaty pieniężne czy także koszty utraconych korzyści. Funkcję nagród, wynikających z potencjalnych decyzji banku w zależności od stanów natury (kategorii należności), czyli opis rzeczywistego problemu decyzyjnego, prezentuje tabela 1. Podstawy teoretyczne statystycznej teorii decyzji, wykorzystywanej w tej części badań, prezentowane są m.in. w monografiach: DeGroot (1981) oraz Coombs, Dawes i Tversky (1977). W kontekście statystycznej teorii decyzji problem udzielania kredytu w przypadku dwóch stanów natury omawia m.in. Osiewalski [2007].

Tabela 1.

Funkcja wypłat w przypadku czterech kategorii należności (m – marża kredytowa)⁶

Decyzje	Kategorie (stany natury)			
	$j=1$ normalne	$j=2$ poniżej standardu	$j=3$ wątpliwe	$j=4$ stracone
udzielić $d_t = 1$	m	$0,8 m - 0,2 (1+m)$	$0,5 m - 0,5 (1+m)$	$-(1+m)$
odmówić $d_t = 0$	$-m$	$-0,8 m$	$-0,5 m$	0
prawdopodobieństwo	p_{t1}	p_{t2}	p_{t3}	p_{t4}

Źródło: opracowanie własne.

Niech prawdopodobieństwa zdarzeń, że potencjalny kredyt o numerze t należy do pierwszej, drugiej, trzeciej albo czwartej kategorii należności, wynoszą odpowiednio p_{t1} , p_{t2} , p_{t3} albo p_{t4} , przy czym $p_{t1} + p_{t2} + p_{t3} + p_{t4} = 1$. Przy podejmowaniu decyzji kredytowych w warunkach niepewności za-

⁴ Stopy odzysku zostały ustalone arbitralnie, ale mogą być szacowane na podstawie informacji o przebiegu dotychczasowych spłat rat i odsetek od kredytów zagrożonych.

⁵ Przez marżę kredytową ($m > 0$) rozumie się różnicę między realnym oprocentowaniem kredytu, a kosztem pozyskania finansujących go środków, tj. oprocentowaniem depozytów. Dla uproszczenia przyjęto, że marża jest identyczna dla wszystkich kredytów.

⁶ Uwzględnienie kosztów alternatywnych jest umotywowane specyfiką działalności banku jako pośrednika między klientami mającymi nadwyżkę środków finansowych oraz tymi, którzy w danym momencie odczuwają ich brak. Deponenti lokują w banku swoje oszczędności, za które bank płaci im odsetki. Depozyty mają charakter czynnika produkcji, który generuje koszt, kredyty zaś są aktywami generującymi przychód. Bank nie może odmówić przyjęcia depozytu, co najwyżej może zniechęcić potencjalnego deponenta proponując mu niskie oprocentowanie. Zatem podejmując decyzję o odmowie kredytu bank traci możliwość uzyskania środków pieniężnych, które pokryłyby koszt pozyskania depozytów.

kłada się, że bank jest w stanie ocenić szanse wystąpienia różnych stanów natury. Informacja ta może być wyrażona w postaci rozkładu prawdopodobieństwa dla y_t – rozkładu a priori lub a posteriori. Ten pierwszy wyrażają subiektywną wiedzę posiadaną przez bank na ten temat. Jednakże parametry p_{ij} ($j = 1, \dots, 4$) nie są znane, więc w niniejszym artykule proponuje się je szacować według wzoru (2) na podstawie danych o udzielonych kredytach i modelu (1). W tym celu wykorzystano podejście bayesowskie, które umożliwia w oparciu o informację z próby aktualizację wstępnej wiedzy reprezentowanej przez rozkład a priori, która ostatecznie znajduje odzwierciedlenie w formie rozkładu posteriori.

Przy ustalonej funkcji wypłat i znanych wielkościach p_{ij} oczekiwana wypłata z tytułu udzielenia kredytu wynosi:

$$EW(d_t = 1) = m \cdot p_{t1} + (0,6m - 0,2)p_{t2} + 0,5 - (1 + m)p_{t4}, \quad (4)$$

w przypadku zaś odmowy:

$$EW(d_t = 0) = -m \cdot p_{t1} - 0,8m \cdot p_{t2} - 0,5m \cdot p_{t3}. \quad (5)$$

Optymalną decyzją kredytową jest ta, która przynosi największą wypłatę (najmniejszą stratę). Zatem bank udzieli kredytu klientowi ($d_t = 1$), gdy oczekiwany zysk z tego tytułu jest wyższy od kosztów utraconych korzyści, co zachodzi, gdy

$$EW(d_t = 1) > EW(d_t = 0). \quad (6)$$

W przeciwnym przypadku klient spotka się z odmową przyznania kredytu. Przyjmijmy, że zyskiem z tytułu przyznaniu kredytu o jednostkowej wartości jest oczekiwana wypłata $EW(d_t = 1)$, w przypadku zaś odmowy $EW(d_t = 0)$.

Wypłata $EW(d_t = 1)$ jest realną pieniężną korzyścią z działalności kredytowej, więc z punktu widzenia adekwatności kapitałowej może ona być podstawą do obliczenia minimalnego poziomu kapitału regulacyjnego, gdy bank stosuje metodę wewnętrznych ratingów. Wówczas $EW(d_t = 1)$ wzięta z przeciwnym znakiem i pomnożona przez wartość kredytu stanowi oczekiwaną stratę z tytułu niespłacenia pojedynczego kredytu.

Z ekonomicznego punktu widzenia przewaga proponowanego podejścia wynika z możliwości dokonania oceny przydatności rozważanych modeli w kontekście korzyści finansowych, które można osiągnąć w ramach scoringu kredytowego. Wielkość zysku można określić dla pojedynczego wniosku kredytowego lub portfela. Proponowana reguła pozwala na podejmowanie optymalnych decyzji kredytowych w oparciu o czynniki ekonomiczne i probabilistyczne, tj. marżę kredytową, prawdopodobieństwo częściowego i całkowitego niespłacenia długu przez kredytobiorcę. Zastosowanie tego narzędzia wspomagającego podejmowanie decyzji w warunkach niepewności prowadzi do zmniejszenia asymetrii informacji między pożyczkobiorcą a pożyczkodawcą, dodatkowo zaś

może osłabić napływ nieuczciwych klientów, czyli ograniczyć niepożądane skutki negatywnej selekcji. W ramach tego podejścia możliwe jest indywidualne ustalenie – w oparciu o wzory (4) i (5) – ceny kredytu dla danego kredytobiorcy, która jest adekwatna do poziomu ryzyka tegoż klienta czy produktu. Przeciwdziała to zjawisku rezygnacji wiarygodnych kredytobiorców z ubiegania się o kredyty z powodu negatywnych bodźców, gdy jednakowa, wysoka jego cena zawiera koszty ryzyka wszystkich potencjalnych pożyczkobiorców.

3.2. Prezentacja wyników

Przedmiotem analizy jest portfel kredytów detalicznych o łącznej wartości 422,6 mln zł, udzielonych w latach 2000–2001. Wartość kredytów należących do kategorii należności normalnych wynosi 379,7 mln zł, do kategorii poniżej standardu 13,6 mln zł, kredyty wątpliwe i stracone zaś kształtują się na poziomie 13,9 i 15,4 mln zł. Marża kredytowa (m) wynosi 10,7 punktu procentowego⁷, a macierz wypłat prezentuje Tabela 2. Zauważmy, że wyłącznie w przypadku kredytów normalnych wypłata z tytułu przyznania kredytu jest większa od kosztów utraconych korzyści w przypadku odmowy. Dla pozostałych kategorii należności decyzją korzystniejszą jest rezygnacja z udzielenia kredytu.

Tabela 2.

Macierz wypłat dla marży kredytowej równej 10,7 punktu procentowego.

Decyzje	Kategorie (stany natury)			
	$j=1$ normalne	$j=2$ poniżej standardu	$j=3$ wątpliwe	$j=4$ stracone
udzielić $d_t = 1$	0,107	-0,136	-0,5	-1,107
odmówić $d_t = 0$	-0,107	-0,086	-0,054	0

Źródło: opracowanie własne.

W wariancie optymistycznym, gdyby wszystkie kredyty zostały spłacone w całości, to przy marży równej 10,7 punktu procentowego bank osiągnąłby zysk maksymalny w kwocie 45,2 mln zł. Faktyczna wypłata (zysk, Z_{Fakt}) z tytułu udzielenia tych kredytów liczona według stanów natury wynosi 14,8 mln zł. Składa się na nią: 40,6 mln zł zysku z udzielenia kredytów z kategorii należności normalnych i straty w kwocie 1,8 mln zł, 7 mln zł i 17 mln zł w przypadku kredytów poniżej standardu, wątpliwych i straconych. Faktyczna wartość utraconych korzyści z tytułu odmowy udzielenia kredytów wynosi zero, gdyż badano wyłącznie wnioski kredytowe, które zostały zaakceptowane przez bank.

⁷ Średnią wartość marży kredytowej obliczono w oparciu o dane dla gospodarstw domowych z 2001 r. W tym okresie średnie ważone oprocentowanie kredytów detalicznych wynosiło 22,2 punktu procentowego, cena depozytów gospodarstw domowych zaś wynosiła 11,5 punktu procentowego. Obliczeń dokonano na podstawie danych pochodzących z NBP.

Oczekiwane wypłaty

Tabela 3 przedstawia oczekiwane wypłaty z podjętych decyzji kredytowych. Oczekiwana wypłata z tytułu udzielenia kredytów kształtuje się na poziomie 13,5 mln zł w modelu M_1 i 13,9 mln zł w M_2 , gdy wypłata faktyczna wynosi 14,8 mln zł. Wypłata uzyskana na podstawie modelu M_2 jest nieznacznie wyższa, o 0,5 mln zł. W trzech modelach oczekiwana wartość utraconych korzyści z tytułu odmowy kształtuje się na poziomie 41,2 mln zł. Przyjęcie za p_{t1} , p_{t2} , p_{t3} i p_{t4} częstości występowania poszczególnych kategorii w próbie, tj. 80,3%, 6%, 6,3% i 7,4%, odpowiada modelowi (1), w którym występuje jedynie wyraz wolny. W tym szczególnym modelu oczekiwana wypłata z tytułu udzielenia kredytów byłaby stratą równą 15 mln zł. W odniesieniu do rezultatów otrzymanych na podstawie proponowanych modeli (M_1 i M_2) oraz faktycznej wypłaty, równej 14,8 mln zł, model ten jest całkowicie nieskutecznym narzędziem prognozowania.

Tabela 3.

Oczekiwane wypłaty z podjętych decyzji kredytowych na podstawie modeli wielomianowych (w mln zł)

Decyzje	Model	
	M_1	M_2
udzielić kredytu	13,5	13,9
odmówić	-41,2	-41,2

Źródło: obliczenia własne.

Pomiar korzyści *ex post*

W celu obliczenia korzyści wynikających z zastosowania modelu wielomianowego do podejmowania decyzji kredytowych można wykorzystać rzeczywiste wartości zmiennych y_t . Uwzględnienie informacji o kategoriach należności poszczególnych kredytów powoduje, że analiza ta ma charakter *ex post*. W pierwszej kolejności wyznaczono optymalny podział portfela kredytów, co ilustruje Tabela 4. Jak wcześniej zauważono udzielone powinny być wyłącznie kredyty, które w przyszłości należałyby do kategorii należności normalnych. Bank zna postawy klientów wobec spłaty kredytów (stany natury), więc może określić najlepsze decyzje *ex post*. Na podstawie macierzy wypłat (z tabeli 2) i informacji o optymalnym podziale portfela otrzymuje się maksymalny zysk (Z_{Max}), który wynosi 38,71 mln zł.

Różnica między zyskiem maksymalnym a faktycznym ($Z_{Max} - Z_{Fakt}$) stanowi wielkość dodatkowego zysku, który bank mógłby osiągnąć, gdyby w odniesieniu do badanego portfela mógł powtórnie zastosować model scoringowy do podjęcia decyzji o udzieleniu albo odmowie kredytu. Różnica ta wynosi 23,91 mln zł.

Tabela 4.

Optymalny (*ex post*) podział portfela kredytów i maksymalne wypłaty (w mln zł)

Decyzje	Stany natury				Suma
	$j=1$	$j=2$, poniżej standardu	$j=3$	$j=4$	
	normalne	standardu	wątpliwe	stracone	
Podział portfela					

udzielić	379,7	0	0	0	379,7
odmówić	0	13,6	13,9	15,4	42,9
Maksymalne wypłaty					
udzielić	40,63	0	0	0	40,63
odmówić	0	-1,17	-0,75	0	-1,92

Źródło: obliczenia własne.

Następnie w oparciu o wyniki modeli M_1 i M_2 dokonano podziału rozważanych rachunków kredytowych na dwie grupy w zależności od tego, czy powinny zostać otwarte, czy nie. Wykorzystano w tym celu formułę (6). W kolejnym kroku, korzystając z macierzy nagród, zaprezentowanej w tabeli 2, dla każdej z czterech grup rachunków kredytowych obliczono korzyści finansowe wynikające z podjęcia decyzji o przyznaniu albo odmowie kredytu.

Tabela 5 przedstawia w ujęciu wartościowym podział badanych rachunków kredytowych w zależności od tego czy powinny być otwarte czy nie. Gdyby decyzje kredytowe były podejmowane na podstawie modelu M_1 , to całkowita wartość udzielonych kredytów byłaby niższa o 19%, czyli o 79,2 mln zł. Wartość udzielonych kredytów w poszczególnych kategoriach zmniejszyłaby się o 13% w pierwszej grupie ryzyka ($j=1$), 63% w drugiej oraz 79% w trzeciej i czwartej. Wyniki otrzymane dla modelu probitowego (M_2) są lepsze w odniesieniu do kategorii należności normalnych, ale gorsze w przypadku kredytów zagrożonych.

Tabela 5.

Podział i portfela kredytowego w zależności od decyzji kredytowej (w mln zł)

Decyzje	Stany natury				Suma
	$j=1$ normalne	$j=2$, poniżej standardu	$j=3$ wątpliwe	$j=4$ stracone	
Model M_1					
udzielić	332,2 (87%)	5,0 (37%)	2,9 (21%)	3,2 (21%)	343,3 (81%)
odmówić	47,5 (13%)	8,6 (63%)	11,0 (79%)	12,1 (79%)	79,2 (19%)
Suma	379,7	13,6	13,9	15,4	422,6
Model M_2					
udzielić	334,3 (88%)	5,1 (38%)	3,5 (25%)	4,0 (26%)	346,9 (82%)
odmówić	45,4 (12%)	8,5 (62%)	10,4 (75%)	11,4 (74%)	75,7 (18%)
Suma	379,7	13,6	13,9	15,4	422,6

Źródło: obliczenia własne.

Tabela 6 prezentuje korzyści finansowe wynikające z podjętych decyzji kredytowych w zależności od kategorii należności (stanów natury). Stanowią one iloczyn elementów tabel 2 i 5. Stosując model t Studenta (M_1), bank osiągnie zysk w kwocie 29,8 mln zł z tytułu udzielenia kredytów i jednocześnie poniesie stratę 6,4 mln zł z tytułu odmowy. Łączny zysk wyniesie 23,4 mln zł i będzie o 0,6 mln zł wyższy niż w przypadku modelu probitowego. Stanowi to 599 zł w przeliczeniu na jeden kredyt o średniej wartości równej około 11 tys. zł. Faktyczny zysk z tytułu udzielenia badanego portfela kredytów wynosi 14,8 mln zł. Zastosowanie modelu M_1 pozwoliłoby zatem na zwiększenie zysku o dodatkowe 8,6 mln zł, co stanowi 58% przyrost. Jeżeli odniesie się kwotę 8,6 mln zł do różnicy $Z_{Max} - Z_{Fakt} = 23,91$ mln zł, to otrzyma się względny miernik efektywności modelu

w scoringu. W modelu tym wynosi on 36%. W przypadku modelu probitowego (M_2) łączne korzyści wynikające z klasyfikacji są o 0,6 mln zł mniejsze, czyli o 2,5% w stosunku do specyfikacji M_1 . Skuteczność modelu M_2 jest więc niższa i wynosi 33,5%.

Tabela 6.

Korzyści finansowe wynikające z zastosowania modeli wielomianowych w scoringu (w mln zł)

Decyzje	Stany natury				Suma
	$j=1$ normalne	$j=2$, poniżej standardu	$j=3$ wątpliwe	$j=4$ stracone	
Model M_1					
udzielić	35,5	-0,7	-1,5	-3,6	29,8
odmówić	-5,1	-0,7	-0,6	0,0	-6,4
Suma	30,5	-1,4	-2,0	-3,6	23,4
Model M_2					
udzielić	35,8	-0,7	-1,8	-4,4	28,9
odmówić	-4,9	-0,7	-0,6	0,0	-6,1
Suma	30,9	-1,4	-2,3	-4,4	22,8

Źródło: obliczenia własne.

Wnioski

W niniejszych badaniach przedstawiono wykorzystanie modelu wielomianowego dla kategorii uporządkowanych w procesie udzielania kredytów w warunkach niepewności. Zastosowanie to przyniosło wiele korzyści z punktu widzenia banku. Umożliwiło ocenę *ex ante* i *ex post* skuteczności systemu scoringowego w kategoriach zysków i utraconych korzyści wynikających z podjętych decyzji kredytowych. Użycie modelu o rozkładzie t Studenta przyniosło o 2,5% wyższe korzyści finansowe w odniesieniu do modelu probitowego. Efektywność tego pierwszego modelu wyniosła 36%, drugiego zaś 33,5%.

Spojrzenie na problem konstrukcji systemu scoringowego w kategoriach statystycznej teorii decyzji umożliwia obiektywne porównanie badanych modeli ograniczających ryzyka pojedynczego kredytu. Wówczas pomiar korzyści z ich zastosowania opiera się na przesłankach ekonomicznych, co proponujemy w niniejszym artykule.

Literatura

- Albert J., S. Chib (1993), Bayesian Analysis of Binary and Polychotomous Response Data, „Journal of the American Statistical Association”, vol. 88, s. 669-679.
- Amemiya T. (1985), Advanced Econometrics, Harvard University Press, Cambridge (Massachusetts).
- Bensic M., N. Sarlija, M. Zekic-Susac (2005), Modelling Small-business Credit Scoring by Using Logistic Regression, Neural Networks and Decision Trees, “International Journal of Intelligent Systems in Accounting and Finance Management”, vol. 13 (3), s. 133-150.
- Chrzanowska M., D. Witkowska (2005), Zastosowanie sztucznych sieci neuronowych do klasyfikacji ryzyka kredytowego podmiotów gospodarczych, „Metody ilościowe w badaniach ekonomicznych V”, Wydawnictwo SGGW, Warszawa, s. 80-89.
- Coombs C.H., R.M. Dawes, A. Tversky (1977), Wprowadzenie do psychologii matematycznej, PWN, Warszawa.
- Cramer J.S. (2004), Scoring Bank Loans That May go Wrong: a Case Study, „Statistica Neerlandica”, vol. 58, nr 3, s. 365-380.

- DeGroot M. (1981), *Optymalne decyzje statystyczne*, PWN, Warszawa.
- Discrimination, Competition, and Loan Performance in FHA Mortgage Lending (1998), Berkovec J.A., G.B. Canner, S.A. Gabriel, T.H. Hannan, „*Review of Economics and Statistics*”, vol. 80 (2), s. 241-250.
- Freixas X., J.C. Rochet (2007), *Mikroekonomia bankowa*, CeDeWU, Warszawa.
- Fritz S., D. Hosemann (2000), Restructuring the Credit Process: Behaviour Scoring for German Corporates, „*International Journal of Intelligent Systems in Accounting, Finance and Management*”, vol. 9 (1), s. 9-21.
- Hodgman D. (1960), Credit risk and credit rationing, „*Quarterly Journal of Economics*”, vol. 74, str. 258-278.
- Hosmer D., S. Lemeshow (2000), *Applied Logistic Regression*, Wiley, New York.
- Jaffee D.M., T. Russell (1976), Imperfect Information, Uncertainty and Credit Rationing, „*Quarterly Journal of Economics*”, vol. 4, str. 651-666.
- Kapłon R. (2006), Analiza danych dyskretnych w ujęciu retrospektywnym, „*Badania Operacyjne i Decyzje*”, nr 1, s.55–72.
- Maddala G.S. (1983), *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
- Marzec J. (2003), Bayesowska analiza modeli dyskretnego wyboru (dwumianowych), „*Przegląd Statystyczny*”, t. 50, s. 129-146.
- Marzec J. (2006), Bayesowski model wielomianowy z rozkładem t Studenta dla kategorii uporządkowanych, „*Metody ilościowe w naukach ekonomicznych, Szóste Warsztaty Doktorskie z Zakresu Ekonometrii i Statystyki*” (red. A. Welfe), Wydawnictwo SGH w Warszawie, s. 123-144.
- Marzec J. (2008a), Bayesowska analiza i testowanie modeli dwumianowych z rozkładem t Studenta, „*Przegląd Statystyczny*”, t. 55, nr 2, s. 129-146.
- Marzec J. (2008b), Bayesowskie modele zmiennych jakościowych i ograniczonych w badaniach niespłacalności kredytów, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- Marzec J. (2009), Zdolności dyskryminacyjne modelu dwumianowego ze skończoną mieszkanką rozkładów normalnych w ocenie niespłacalności kredytów, [w:] „*Współczesne problemy statystyki, ekonometrii i matematyki stosowanej*”, *Studia i Prace Uniwersytetu Ekonomicznego w Krakowie* nr 3, s. 129-144.
- Marzec J. (2010), Modele dwumianowe w scoringu kredytowym – pomiar korzyści z ich zastosowania, „*Badania Operacyjne i Decyzje*”, nr 1, w druku.
- Matthews K., J. Thomson (2007), *Ekonomika bankowości*, PWE, Warszawa.
- McKelvey R.D., W. Zavoina (1975), A Statistical Model for the Analysis of Ordinary Level Dependent Variables, „*Journal of Mathematical Sociology*”, nr 4, s. 103-120.
- Misztal M. (2006), O zastosowaniu metody rekurencyjnego podziału w analizie ryzyka kredytowego, „*Modelowanie preferencji a ryzyko '05*”, Wydawnictwo AE w Katowicach, s. 453-468.
- Moffatt P.G. (2005), Hurdle Models of Loan Default, „*Journal of the Operational Research Society*”, nr 56, s. 1063–1071.
- Mudholkar G., E. George (1978), A Remark on the Shape of the Logistic Distribution, „*Biometrika*”, nr 65, s. 667-668.
- Osiewalski J. (2007), Bayesowska statystyka i teoria decyzji w analizie ryzyka kredytu detalicznego, [w:] „*Finansowe warunkowania decyzji ekonomicznych*” (red. D. Fatuła), Wydawnictwo Krakowskiej Szkoły Wyższej w Krakowie, Kraków, s. 15–28.
- Osiewalski J., J. Marzec (2004), Model dwumianowy II rzędu i skośny rozkład Studenta w analizie ryzyka kredytowego, „*Folia Oeconomica Cracoviensia*”, vol. 45. s. 63-84.
- Pratt J.W. (1981), Concavity of the Log Likelihood, „*Journal of the American Statistical Association*”, vol. 76, nr 373, s. 103-106.
- Sealey C.W., J.T. Lindley (1977), Inputs, outputs, and a theory of production and cost at depository financial institutions, „*The Journal of Finance*”, nr 3, str. 1251-1277.
- Structural Analysis of Discrete Data with Econometric Applications* (1981), ed. C. Manski, D. McFadden, MIT Press, Cambridge.
- Thomas L.C. (2000), A Survey of Credit and Behavioural Scoring: Forecasting Financial Risk of Lending to Consumers, „*International Journal of Forecasting*”, 16, s. 149-172.
- Thomas L.C., R.W. Oliver, D.J. Hand (2005), A Survey of the Issues in Consumer Credit Modelling Research, „*Journal of the Operational Research Society*”, 56 (9), s. 1006-1015.
- Trinkle B. A. Baldwin (2007), Interpretable credit model development via artificial neural networks, „*International Journal of Intelligent Systems in Accounting and Finance Management*”, vol. 15 (3-4), s. 123-147.
- Witkowska D., M. Chrzanowska (2005), Wybrane metody klasyfikacji kredytobiorców: modele logitowe i sieci neuronowe, „*Modelowanie preferencji a ryzyko '04*”, Wydawnictwo AE w Katowicach, s. 531-540.
- Witkowska D., M. Chrzanowska (2006), Drzewa klasyfikacyjne jako metoda grupowania klientów banku, „*Modelowanie preferencji a ryzyko '05*”, Wydawnictwo AE w Katowicach, s. 485-496.