

Jerzy Marzec
Akademia Ekonomiczna Krakowie

Modele wielomianowe dla kategorii uporządkowanych w badaniu niespłacalności kredytów konsumpcyjnych¹

Wprowadzenie

W literaturze ekonometrycznej modele dla jakościowych zmiennych endogenicznych określa się terminem modele dyskretnego wyboru (ang. *quantal response or discrete choice models*). Jeżeli zmienna objaśniana przyjmuje skończoną liczbę wartości i mierzona jest na skali porządkowej, wówczas otrzymujemy wielomianowy model dla kategorii uporządkowanych, który po raz pierwszy zaproponowali Aitchison i Silvey (1957). Przykładem zmiennych mierzonych na skali porządkowej jest np. informacja o zatrudnieniu (bezrobotny, zatrudniony w niepełnym wymiarze godzin, na pełny etat) lub skala odpowiedzi respondentów w badaniach ankietowych (np. skala Likerta: zdecydowanie tak, raczej tak, nie mam zdania, raczej nie i zdecydowanie nie). Klasyfikację modeli dyskretnych ze względu na rodzaj wartości jakie przyjmuje zmienna endogeniczna wraz z zastosowaniami w ekonomii prezentuje m.in. Maddala (1983). Zasadniczym celem niniejszego opracowania jest prezentacja wyników badań empirycznych, uzyskanych na podstawie modelu probitowego i logitowego. a dotyczących niespłacalności kredytów konsumpcyjnych w polskim banku komercyjnym.

1. Model wielomianowy dla kategorii uporządkowanych

Wielomianowy model dla kategorii uporządkowanych przyjmuje następującą postać (zob. McKelvey i Zavoina (1975)):

$$\begin{cases} z_t = x_t \beta + \varepsilon_t \\ y_{ij} = 1 \text{ gdy } \alpha_{j-1} < z_t < \alpha_j \text{ dla } t = 1, \dots, T \quad j = 1, \dots, J \\ y_{ij} = 0 \text{ w przeciwnym przypadku,} \end{cases} \quad (1)$$

gdzie z_t jest nieobserwowalną (ukrytą) zmienną ciągłą, której wartości określają obserwowaną kategorię zmiennej y_t , natomiast x_t jest wektorem zmiennych egzogenicznych (lub ich znanych funkcji), których wpływ na kategorie zmiennej y_t określa wektor parametrów β , zaś parametry α_j są tzw. punktami granicznymi. O zmiennej losowej ε_t najczęściej zakłada się, że posiada standaryzowany rozkład normalny lub logistyczny oraz przyjmuje się brak korelacji między ε_t i ε_s dla każdego $t \neq s$.² Warto przypomnieć, że szczególnym i najprostszym przypadkiem modeli wielomianowych jest model dychotomiczny (dla binarnej zmiennej y_t), który uzyskujemy dla $J = 2$. Z modelu (1) wynika, że prawdopodobieństwo zaobserwowania kategorii j wynosi

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

gdzie $F(\cdot)$ jest dystrybuantą zmiennej losowej ε_t . Identyfikowalność parametrów wymaga, aby $\alpha_0 = -\infty$ i $\alpha_J = +\infty$ (oraz $\alpha_1 = 0$, jeżeli w równaniu regresji dla zmiennej z_t występuje wyraz wolny, co zakładamy w tej pracy). W efekcie otrzymujemy

¹ Praca wykonana w ramach badań statutowych finansowanych przez Akademię Ekonomiczną w Krakowie w roku 2003 r. Autor pragnie wyrazić podziękowania Profesorowi Jackowi Osiewalskiemu i Mgr. Pawłowi Siarce za dyskusję oraz cenne i użyteczne komentarze.

² Zmienna z ma rozkład logistyczny, jeżeli jej dystrybuanta ma postać $F(z) = (1 + e^{-z})^{-1}$.

$$\begin{aligned}
 p_{t1} &= F(-x_t\beta) \\
 p_{t2} &= F(\alpha_2 - x_t\beta) - F(-x_t\beta) \\
 p_{t3} &= F(\alpha_3 - x_t\beta) - F(\alpha_2 - x_t\beta) \\
 &\dots \\
 p_{t,j-1} &= F(\alpha_{j-1} - x_t\beta) - F(\alpha_{j-2} - x_t\beta) \\
 p_{tj} &= 1 - F(\alpha_{j-1} - x_t\beta)
 \end{aligned} \tag{3}$$

W przypadku modeli wielomianowych podstawową i najczęściej stosowaną metodą estymacji, gdy obserwacje y_t są niezależne, jest metoda największej wiarygodności (MNW), przy czym funkcja wiarygodności ma postać

$$L(\alpha, \beta|y) = \prod_{t=1}^T \prod_{j=1}^J (p_{tj})^{y_{tj}}, \tag{4}$$

Jeżeli przez $f(\varepsilon_t)$ oznaczymy funkcję gęstości zmiennej ε_t , czyli $\partial F(\varepsilon_t)/\partial \varepsilon_t = f(\varepsilon_t)$, to warunki konieczne dla uzyskania rozwiązania maksymalizującego funkcję wiarygodności mają postać układu równań:

$$\begin{aligned}
 \frac{\partial \ln L(\alpha, \beta|y)}{\partial \beta} &= -\sum_{t=1}^T \sum_{j=1}^J y_{tj} \frac{f(\alpha_j - x_t\beta) - f(\alpha_{j-1} - x_t\beta)}{p_{tj}} x'_t = \mathbf{0}, \\
 \frac{\partial \ln L(\alpha, \beta|y)}{\partial \alpha_j} &= \sum_{t=1}^T \sum_{j=1}^J y_{tj} \frac{\delta_{j,h} \cdot f(\alpha_j - x_t\beta) - \delta_{j-1,h} \cdot f(\alpha_{j-1} - x_t\beta)}{p_{tj}} = 0
 \end{aligned} \tag{5}$$

gdzie $\delta_{j,h} = 1$ gdy $j=h$, natomiast $\delta_{j,h} = 0$ w przeciwnym przypadku. Powyższy układ równań musimy rozwiązać za pomocą metod numerycznych, np. metodą Newtona-Raphsona. Występowanie dużej liczby wyborów (J) i związanej z tym dużej liczby parametrów może prowadzić do komplikacji obliczeniowych.

Pojedyncze parametry β_k ($k=1, \dots, K$) w modelu (1) nie mają bezpośredniej interpretacji ekonomicznej. Wnioskowanie o sile i kierunku wpływu zmiennych egzogenicznych x_t może opierać się na efektach krańcowych zdefiniowanych dla zmiennych ciągłych jako³

$$\eta_{ijk}(\alpha, \beta) = \frac{\partial \Pr(y_{tj} = 1)}{\partial x_{tk}} = -(f(\alpha_j - x_t\beta) - f(\alpha_{j-1} - x_t\beta))\beta_k, \tag{6}$$

natomiast dla zmiennych zero-jedynkowych efekt krańcowy obliczamy wg wzoru

$$\eta_{ijk}(\alpha, \beta) = \Pr(y_{tj} = 1|x_{tk} = 1) - \Pr(y_{tj} = 1|x_{tk} = 0), \tag{7}$$

Przybliżone błędy średnie szacunku dla efektów krańcowych jako nieliniowej (ciągłej) funkcji wektora θ oblicza się, korzystając z własności MNW, w oparciu o następującą macierz

$$V(\eta_{ij}) = \left(\frac{\partial \eta_{ij}(\theta)}{\partial \theta} \right)' V_{MNW} \frac{\partial \eta_{ij}(\theta)}{\partial \theta}, \tag{8}$$

gdzie V_{MNW} jest asymptotyczną macierzą kowariancji estymatora MNW dla wektora θ , zaś $\partial \eta_{ij}(\theta)/\partial \theta$ jest macierzą pochodnych cząstkowych efektów krańcowych η_{ij} względem wektora parametrów $\theta = (\alpha' \beta)'$. Estymator macierzy $V(\eta_{ij})$ uzyskuje się przyjmując w (8) $\theta = \hat{\theta}_{MNW}$, gdzie $\hat{\theta}_{MNW}$ jest estymatorem MNW⁴.

O podstawowych własnościach metody MNW (zgodność i asymptotyczna normalność) pisze m.in. Amemiya (1985), natomiast Pratt (1981) prezentuje warunki konieczne i wystarczające na jednoznaczne istnienie estymatora MNW w modelu (1) w zależności od przyjętego rozkładu dla ε_t . Alternatywne bayesowskie podejście do estymacji modelu probitowego prezentowane jest przez Alberta i Chiba (1993), a także w pracy Marzec (2003d).

³ Wzór (6) jest dla przypadku, gdy w modelu (1) zmienna z_t jest liniową funkcją x_t .

⁴ Por. np. Greene (1993).

2. Wyniki empiryczne

Omówione powyżej podejście zostało wykorzystane do badania niespłacalności kredytów konsumpcyjnych udzielonych klientom detalicznym w polskim banku komercyjnym. W celu estymacji modelu probitowego i logitowego wykorzystaliśmy zbiór danych obejmujący 37 tysięcy rachunków. Wcześniej w pracach Marzec (2003a,b,c,d) przedstawiamy wyniki estymacji modeli dychotomicznych, tj. modelu probitowego, logitowego oraz ich uogólnienia, czyli modelu z rozkładem t-Studenta o nieznannej liczbie stopni swobody oraz wielomianowego modelu kategorii uporządkowanych, w oparciu o dane zawierające także kredyty hipoteczne. Konstrukcję zmiennej endogenicznej y_i oparliśmy na klasyfikacji należności wg uchwały Komisji Nadzoru Bankowego w sprawie zasad tworzenia rezerw na ryzyko związane z działalnością banków⁵. W przypadku kredytów detalicznych wyróżnia się cztery kategorie należności, tj. należności normalne, poniżej standardu, wątpliwe i stracone.

Przyjęliśmy zatem, iż zmienna objaśniana y_i przyjmuje cztery wartości ($J=4$), które jednocześnie oznaczają kategorie należności lub równoważnie okres opóźnienia w spłacie przez kredytobiorcę rat kapitałowo-odsetkowych:

j	Kategoria należności	Okres opóźnienia w spłacie	Obserwowany udział w próbie
1	Normalne	Do 1 miesiąca	79,1%
2	Poniżej standardu	Od 1 do 3 miesięcy	6,3%
3	Wątpliwe	Od 3 do 6 miesięcy	6,7%
4	Stracone	Powyżej 6 miesięcy	7,9%

Jako potencjalne zmienne egzogeniczne wyjaśniające ryzyko pojedynczej umowy kredytowej przyjęliśmy (jak we wcześniejszych pracach):⁶

- płeć (zmienna przyjmuje wartość 1, jeżeli klientem jest mężczyzna, 0 w przypadku kobiety),
- wiek kredytobiorcy (w setkach lat),
- wpływy, tzn. wielkość miesięcznych wpływów w latach 2000-2001 (w setkach tys. zł) na rachunki typu ROR kredytobiorcy w badanym banku,
- posiadanie ROR w analizowanym banku (1 – posiada, 0 – nie posiada),
- informację o tym, czy kredytobiorca posiada karty płatnicze lub kredytowe wydane przez bank (1 – posiada choć jedną kartę płatniczą, 0 – nie posiada),
- sposób udzielenia kredytu (1 – poprzez pośrednika kredytowego, 0 – bezpośrednio przez bank),
- okres trwania umowy kredytowej (w dziesiątkach lat),
- podstawowe źródło dochodu uzyskiwanego przez kredytobiorcę (zmienna *zrdoch*).

Ostatnia zmienna może przyjmować cztery różne wartości. Chcąc ją uwzględnić w równaniu regresji z wyrazem wolnym wprowadziliśmy trzy zmienne zerojedynkowe, przy czym za sytuację referencyjną przyjęliśmy „umowę o pracę” ($zrdoch1 = 0$ i $zrdoch2 = 0$, i $zrdoch3 = 0$), a w pozostałych przypadkach:

- $zrdoch1 = 1$, gdy źródłem dochodu kredytobiorcy jest renta lub emerytura,
- $zrdoch2 = 1$, gdy źródłem dochodu kredytobiorcy jest własna działalność, umowa o dzieło lub umowa zlecenie,
- $zrdoch3 = 1$ w przypadku innego źródła dochodu, np. stypendium.

W niniejszej pracy – podobnie jak w pracy Marzec (2003c,2003d) - założyliśmy, że w modelu (1) prawdopodobieństwo zakwalifikowania kredytu do jednej z kategorii należności może zależeć liniowo nie

⁵ Uchwała nr 8/1999 Komisji Nadzoru Bankowego z 22 grudnia 1999 r. stanowi zasady tworzenia przez banki rezerw celowych od należności zagrożonych.

⁶ Więcej informacji o zbiorze danych prezentujemy w pracy Marzec (2003a).

tylko od zmiennych egzogenicznych w_{th} , ale także od iloczynów tych zmiennych oraz kwadratów zmiennych ciągłych (wiek, wpływy, okres), co prowadzi do następującego, bardziej ogólnego modelu:

$$\begin{cases} z_t = \beta_1 + \sum_h w_{th}\beta_h + \sum_h \sum_{i \geq h} w_{th}w_{ti}\beta_{hi} + \varepsilon_t \\ y_{ij} = 1 \quad \text{gdy } \alpha_{j-1} < z_t < \alpha_j \quad \text{dla } t = 1, \dots, T \quad j = 1, \dots, 4 \\ y_{ij} = 0 \quad \text{w przeciwnym przypadku,} \end{cases} \quad (9)$$

gdzie $\alpha_0 = -\infty$, $\alpha_1 = 0$, $\alpha_4 = +\infty$ oraz $\alpha_2 < \alpha_3$.

Z punktu widzenia omówionych wcześniej metod wnioskowania, z_t jest nadal liniowo zależne od parametrów β , zatem taki nieliniowy sposób wprowadzenia zmiennych egzogenicznych nie wnosi żadnych komplikacji na etapie estymacji. Powyższa modyfikacja może przyczynić się do lepszego oszacowania p_{ij} , ponieważ jeżeli prawdziwa zależność z_t od w_{th} jest nieliniowa, to wielomian stopnia drugiego (względem w_{th}) jest lepszą aproksymacją (wyższego rzędu) niż wielomian stopnia pierwszego (funkcja liniowa względem w_{th}). Zasadność tego rozszerzenia potwierdza klasyczny test ilorazu wiarygodności (ang. *likelihood ratio test*; *LR*).⁷ Wartość logarytmu wiarygodności dla modelu probitowego postaci (9) wynosi -21606 , zaś dla modelu z restrykcjami ($\beta_{hi}=0$) kształtuje się na poziomie -21953 , zatem wartość *LR* jest wysoka i wynosi 694. Ponadto współczynnik determinacji McFaddena R^2 w modelu (9) wynosi 0,23, natomiast w modelu z restrykcją kształtuje się na poziomie 0,19.⁸ Podobnie w pełnym modelu logitowym *LR* wynosi 835 i $R^2=0,21$, zaś w modelu zredukowanym $R^2=0,19$. Tak niskie wartości R^2 są typowe z uwagi na polichotomiczny charakter zmiennej y_t . Z punktu widzenia wnioskowania statystycznego nieliniowy sposób wprowadzenia zmiennych egzogenicznych jest zarówno w modelu probitowym, jak i w logitowym uzasadniony.

Tabela 1 zawierają uśrednione po wszystkich obserwacjach wartości efektów krańcowych w przypadku modelu probitowego. Zauważmy, że jeżeli wiek kredytobiorcy wzrośnie o jeden rok (*ceteris paribus*), to prawdopodobieństwo terminowego spłacenia kredytu konsumpcyjnego (p_{t1}) wzrośnie średnio o prawie 0,018, a prawdopodobieństwo wystąpienia opóźnienia w spłacie od 1 do 3 miesięcy (p_{t2}) obniży się o 0,002, opóźnienia w spłacie od 3 do 6 miesięcy (p_{t3}) spadnie prawie o 0,005, zaś prawdopodobieństwo opóźnienia dłuższego niż 6 miesięcy (p_{t4}) obniży się 0,01. Natomiast wzrost wpływów o 1 tysiąc złotych spowoduje wzrost p_{t1} o 0,043 oraz spadek prawdopodobieństwa zakwalifikowania kredytu konsumpcyjnego do czwartej kategorii należności o 0,026. Ponadto wraz z wydłużeniem o jeden rok okresu kredytowania prawdopodobieństwo terminowej spłaty rat wzrośnie o 0,02. Jeżeli hipotetycznym klientem jest mężczyzna, to ryzyko niedotrzymania przez niego umowy kredytowej jest nieznacznie wyższe niż w przypadku kobiety. Udzielenie kredytu poprzez pośrednika także istotnie zwiększa ryzyko kredytowe. Studenci korzystający z kredytu studenckiego (*zrdoch3*) oraz emeryci i renciści (*zrdoch1*) są mniej ryzykownymi kredytobiorcami niż klienci zatrudnieni na umowę o pracę. Natomiast klienci prowadzący własną działalność gospodarczą (*zrdoch2*) charakteryzują się największym ryzykiem kredytowym spośród rozważanych typów klientów ze względu na źródło ich dochodów. Efekty krańcowe względem zmiennych *ROR* i *karty* charakteryzują się relatywnie dużymi błędami średnimi szacunku, co świadczy o braku wpływu tych zmiennych na wielkość p_{ij} . Wyniki dla modelu logitowego były prawie identyczne jak te dla modelu probitowego.

⁷ $LR = -2[\ln L(\theta_R|y) - \ln L(\theta|y)]$, gdzie $\ln L(\theta_R|y)$ jest wartością logarytmu wiarygodności dla modelu (9) z restrykcją $\beta_{hi}=0$. Jeżeli $T \rightarrow \infty$ statystyka *LR* ma rozkład χ^2 o liczbie stopni swobody równej liczbie restrykcji, zob. np. Greene (1993).

Tabela 1. Wartości uśrednionych efektów krańcowych $\eta_{jh} = T^{-1} \sum_t \partial \Pr(y_{ij} = 1) / \partial w_{th}$ dla modelu probitowego (błędy średnie szacunku w nawiasach).

zmienna	$j = 1$	$j = 2$	$j = 3$	$j = 4$
pleć	-0,008 (0,004)	0,003 (0,001)	0,002 (0,001)	0,003 (0,002)
wiek	0,179 (0,017)	-0,022 (0,004)	-0,045 (0,005)	-0,112 (0,010)
wpływy	4,280 (1,968)	-0,581 (0,219)	-1,085 (0,475)	-2,613 (1,276)
ROR	-0,048 (0,034)	0,018 (0,014)	0,014 (0,011)	0,016 (0,013)
karty	0,100 (0,093)	-0,011 (0,010)	-0,024 (0,023)	-0,065 (0,060)
pośrednik	-0,243 (0,007)	0,042 (0,002)	0,063 (0,002)	0,137 (0,004)
okres	0,199 (0,015)	-0,015 (0,004)	-0,043 (0,004)	-0,141 (0,009)
zrdoch1	0,020 (0,011)	-0,010 (0,003)	-0,008 (0,003)	-0,003 (0,006)
zrdoch2	-0,002 (0,011)	0,008 (0,002)	0,003 (0,003)	-0,008 (0,007)
zrdoch3	0,034 (0,021)	-0,009 (0,004)	-0,012 (0,005)	-0,013 (0,014)

Źródło: obliczenia własne.

Natomiast Tabela 2 zawiera wartości efektów krańcowych w przypadku modelu probitowego przy założeniu $\beta_{hi}=0$, który najczęściej wykorzystuje się w badaniach empirycznych. Wyniki te zbliżone są do rezultatów z Tabeli 1, a istotne różnice w wartościach efektów krańcowych dotyczą przede wszystkim kierunku wpływu zmiennej ROR oraz siły oddziaływania zmiennej *wpływy* na prawdopodobieństwo terminowej spłaty kredytu.

Tabela 2. Wartości uśrednionych efektów krańcowych η_{jh} dla modelu probitowego ($\beta_{hi}=0$).

zmienna	$j=1$	$j=2$	$j=3$	$j=4$
pleć	-0,010	0,002	0,003	0,006
wiek	0,189	-0,034	-0,050	-0,105
wpływy	0,608	-0,108	-0,160	-0,340
ROR	0,057	-0,010	-0,015	-0,032
karty	0,033	-0,006	-0,009	-0,018
pośrednik	-0,265	0,047	0,070	0,148
okres	0,188	-0,033	-0,049	-0,105
zrdoch1	0,018	-0,003	-0,005	-0,010
zrdoch2	-0,064	0,011	0,017	0,036
zrdoch3	0,024	-0,004	-0,006	-0,013

Źródło: obliczenia własne.

Oszacowany model wykorzystaliśmy do prognozowania okresu opóźnienia w spłacie kredytu dla wybranych klientów. W modelu probitowym, w przypadku „klienta najczęstszego”⁹ prawdopodobieństwo terminowej spłaty kredytu, jeżeli kredyt został mu udzielony bezpośrednio przez bank, jest wysokie i wynosi prawie 0,98. Jeśli zaś otrzymał go poprzez pośrednika, to jest ono znacznie niższe i wynosi 0,81. Prawdopodobieństwo, że opóźnienie spłaty kredytu przekroczy 6 miesięcy, w pierwszym przypadku jest praktycznie zerowe, natomiast w drugim kształtuje się na poziomie około 0,05. Wyniki te istotnie różnią się pomiędzy modelami, co przedstawia Tabela 3.

⁸ $R^2 = 1 - \ln L(\theta|y) / \ln L(\theta_0|y)$, gdzie $\ln L(\theta_0|y)$ jest wartością logarytmu wiarygodności dla modelu (9) z restrykcją $\beta_{hi}=0$ i $\beta_{hi}=0$, czyli modelu regresji dla z_t jedynie z wyrazem wolnym; zob. np. Amemiya (1981).

⁹ Tzn. o cechach najczęstszych w próbie (dotyczy zmiennych jakościowych) i przeciętnych (dla zmiennych ciągłych) w badanej zbiorowości; zob. Marzec (2003a,b,c,d).

Tabela 3. Prawdopodobieństwa zakwalifikowania kredytu „najczęstszego klienta” do poszczególnych kategorii należności w zależności od sposobu udzielenia mu kredytu.

Sposób udzielenia kredytu	Model probitowy							
	$j = 1$		$j = 2$		$j = 3$		$j = 4$	
Przez pośrednika	0,814	(0,043)	0,075	(0,031)	0,065	(0,016)	0,046	(0,016)
Przez bank	0,977	(0,003)	0,013	(0,002)	0,007	(0,001)	0,003	(0,001)
	Model logitowy							
Przez pośrednika	0,900	(0,029)	0,041	(0,018)	0,032	(0,009)	0,027	(0,009)
Przez bank	0,989	(0,002)	0,005	(0,001)	0,004	(0,000)	0,003	(0,001)

Źródło: obliczenia własne.

W celu oceny zdolności progностycznej oszacowanych powyżej modeli obliczyliśmy przeciętne prawdopodobieństwa zakwalifikowania badanych rachunków kredytowych do każdej z czterech kategorii należności. Jeżeli model dokonałby bezbłędnej klasyfikacji, to na przekątnej w Tabeli 4 znajdowałyby się jedynki, a poza nią były zera. Uzyskane wyniki, które dla wszystkich modeli są prawie identyczne, wskazują na niezbyt dobre ich własności progностyczne.

Prawdopodobieństwo poprawnego zakwalifikowania rachunków kredytowych z pierwszej kategorii należności jest wysokie i wynosi prawie 0,85. Niestety, model niezbyt trafnie klasyfikuje kredyty z pozostałych grup ryzyka. Zgodnie z intuicją dotyczy to zwłaszcza rachunków o kategorii „poniżej standardu” i „wątpliwe”. Odsetek poprawnie zakwalifikowanych rachunków w grupie drugiej wynosi tylko 11%, w trzeciej 13%, w czwartej prawie 21%, przy czym odpowiednio aż 62, 60 i 54 procent kredytów z tych kategorii zostało uznanych za należności normalne. Z punktu widzenia banku ważnym elementem polityki kredytowej jest minimalizowanie wartości kredytów zakwalifikowanych do grupy należności zagrożonych (druga, trzecia i czwarta kategoria należności). Straty spowodowane udzieleniem kredytu, który nigdy nie zostanie spłacony są o wiele większe niż potencjalny zysk jaki bank uzyskałaby, gdyby ten kredyt został spłacony wraz z odsetkami. Model zbyt optymistycznie zatem ocenia ryzyko związane z rachunkami kredytowymi z grupy zagrożonych, co z punktu widzenia zarządzania ryzykiem banku nie jest cechą pożądaną. Podobne wyniki uzyskano dla modelu probitowego na podstawie większego zbioru danych zawierającego kredyty hipoteczne, zob. Marzec (2003d). Być może jedną z przyczyn jest sposób definiowania zmiennej y_{it} , który wynika z uregulowań prawnych i nie odzwierciedla w pełni zachowań kredytobiorców. Ponadto udział w zbiorze danych kredytów z drugiej, trzeciej i czwartej kategorii należności jest relatywnie niewielki i wynosi odpowiednio 6,3%, 6,7% i 7,9%, co zapewne negatywnie wpływa na wyniki estymacji modelu i jakość prognoz.¹⁰

Tabela 4. Średnie prawdopodobieństwo zakwalifikowania rachunku kredytowego do danej kategorii należności T $\frac{1}{T} \sum_t \Pr(y_{it}=1 | y_{ij}=1)$ dla $i,j=1,\dots,4$ – model probitowy.

	$j=1$	$j=2$	$j=3$	$j=4$
$\Pr(y_{i1}=1 y_{ij}=1)$	0,848	0,616	0,598	0,539
$\Pr(y_{i2}=1 y_{ij}=1)$	0,049	0,110	0,114	0,116
$\Pr(y_{i3}=1 y_{ij}=1)$	0,049	0,123	0,129	0,140
$\Pr(y_{i4}=1 y_{ij}=1)$	0,054	0,151	0,160	0,205
Suma	1,000	1,000	1,000	1,000

Źródło: obliczenia własne.

¹⁰ Trudno oczekiwać, aby posiadane dane miały charakter próby zbilansowanej (udział kredytów każdej kategorii wynosi około 25%), ponieważ oznaczałoby to bankructwo banku.

Podsumowanie

W niniejszym opracowaniu zaprezentowano specyfikację i estymację modeli wielomianowych dla kategorii uporządkowanych. Z punktu widzenia zarządzania ryzykiem kredytowym, uzyskano interesujące wyniki informujące o sile i kierunku wpływu badanych zmiennych objaśniających na prawdopodobieństwa zakwalifikowania rachunków kredytowych do poszczególnych grup ryzyka kredytowego. Jednakże oszacowane modele nie wykazują zbyt dobrych zdolności prognostycznych. Istnieje więc potrzeba uogólnienia modeli probitowego i logitowego poprzez przyjęcie szerszej klasy rozkładów dla składnika losowego ε . Wyniki poprzednich badań wskazują na konieczność wykorzystania rozkładów o tzw. „grubych ogonach”, np. rozkładu t-Studenta; por. Marzec (2003c). Kolejną propozycją jest przyjęcie innej konstrukcji zmiennej endogenicznej, która lepiej odzwierciedlałaby rzeczywiste postawy kredytobiorców wobec ich zobowiązań wynikających z umowy kredytowej. Alternatywą dla badanych czterech kategorii zmiennej y_i byłoby wprowadzenie jedynie trzech, które informowałyby, czy kredyt spłacany jest terminowo, nieregularnie czy też w ogóle.

Literatura

- Aitchison J., S. Silvey, 1957, *The Generalization of Probit Analysis to the Case of Multiple Responses*, Biometrika, 44, s. 253-262.
- Albert J. Chib S., 1993, *Bayesian Analysis of Binary and Polychotomous Response Data*, Journal of the American Statistical Association, 88, s. 669-679.
- Amemiya T., 1981, *Qualitative Response Models: A Survey*, Journal of Economic Literature, 19.
- Amemiya T., 1985, *Advanced Econometrics*, Harvard University Press, Cambridge Massachusetts.
- Greene W.H., 1993, *Econometric Analysis*, Macmillan Publishing Company, New York.
- Gruszczyński M., 2001, *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Monografie i Opracowania SGH, Warszawa, nr 6.
- Maddala G.S., 1983, *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
- Marzec J., 2003a, *Badanie niewypłacalności kredytobiorcy na podstawie modeli logitowych i probitowych*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie nr 628 (w druku).
- Marzec J., 2003b, *Badanie niespłacalności kredytów za pomocą bayesowskich modeli dychotomicznych - założenia i wyniki*, Metody ilościowe w naukach ekonomicznych (red. A. Welfe), Wydawnictwo SGH w Warszawie.
- Marzec J., 2003c, *Bayesowska analiza modeli dyskretnego wyboru (dwumianowych)*, maszynopis.
- Marzec J., 2003d, *Bayesowska analiza wielomianowego modelu probitowego dla kategorii uporządkowanych*, Folia Oeconomica Cracoviensia, w druku.
- McKelvey R.D., W. Zavoina, 1975, *A Statistical Model for the Analysis of Ordinary Level Dependent Variables*, Journal of Mathematical Sociology, 4, s. 103-120.
- Pratt J.W., 1981, *Concavity of the Log Likelihood*, Journal of the American Statistical Association, vol. 76, nr. 373, s. 103-106.