

BAYESOWSKA ANALIZA WIELOMIANOWEGO MODELU PROBITOWEGO DLA KATEGORII UPORZĄDKOWANYCH¹

Praca przedstawiona na posiedzeniu Komisji Nauk Ekonomicznych i komisji Statystyczno-Demograficznej Oddziału PAN w Krakowie dniu 8 kwietnia 2003 r. przez Jerzego Marca.

ABSTRACT

J. Marzec. 2003, *Bayesian analysis of ordered multinomial probit model*. Folia Oeconomica Cracoviensia.

This paper presents the use of multinomial probit model for ordered data in the estimation of consumer and mortgage loans risk. Bayesian approach proposed by Albert and Chib (1993) for modeling ordered-response data are presented. To compute the posterior distribution of any desired quantity of interest, we apply the Gibbs sampling algorithm. We show results of empirical analysis of 4 categories of loan i.e. normal loans, overdue loans, slack loans and bad loan. We predict the probability of each categories, which depends on nine explanatory variables such as: sex, age, size and source of client's income, a type of loan, a period of loan, information about using a cheque deposit account and credit or payments (ATM) cards by debtors and the way a loan is given (by an agent or not).

KEY WORDS: ordered multinomial probit model, bayesian approach, Gibbs sampling, credit scoring model.

SŁOWA KLUCZOWE: wielomianowego modelu probitowego dla kategorii uporządkowanych, podejście bayesowskie, losowanie Gibbsa, kredytowy model scoringowy.

1. WSTĘP

W literaturze ekonometrycznej modele dla jakościowych zmiennych endogenicznych określa się mianem modeli dyskretnego wyboru (ang. *quantal response or discrete choice models*). Opisują one zależność między wynikiem dokonywanych wyborów a egzogenicznymi zmiennymi objaśniającymi, które mogą opisywać cechy możliwych alternatyw (wyborów) lub indywidualne charakterystyki podmiotów podejmujących decyzję. W celu kwantyfikacji zmiennej endogenicznej najczęściej przyjmuje się dla niej liczby naturalne, które reprezentują kody poszczególnych możliwych alternatyw (w przypadku kategorii nieuporządkowanych) bądź rangi (dla kategorii uporządkowanych) albo dane reprezentujące liczebność zbioru (wynikające ze zliczenia; ang. *count data*²). Stąd zbiór wartości zmiennej endogenicznej jest zbiorem skończonym lub przeliczalnym. Jeżeli liczba alternatyw jest większa niż dwa i skończona, to mamy do czynienia z modelami wielomianowymi (ang. *multinomial response, polychotomous or multiple choice models*), przy czym najczęściej stosowanymi są modele probitowe i logitowe. Przykład zmiennych mierzonych na skali porządkowej jest np. informacja o zatrudnieniu (bezrobotny, zatrudniony w niepełnym

¹ Autor pragnie wyrazić podziękowania Profesorowi Jackowi Osiewalskiemu za cenne uwagi i konstruktywną krytykę na kolejnych etapach powstawania niniejszego opracowania. Praca wykonana w ramach badań statutowych finansowanych przez Akademię Ekonomiczną w Krakowie w roku 2003 r.

² Gruszczyński (2001) używa określenia dane licznikowe dla *count data*.

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). Natomiast zmienna określająca np. wybór środka transportu do pracy (własne auto, autobus, tramwaj) lub wykonywany zawód (urzędnik, inżynier, polityk, przedsiębiorca) ma charakter kategorii nieuporządkowanej. Danymi reprezentującymi liczebność zbioru są np. ilość szkód komunikacyjnym popełnionych przez właściciela auta lub liczba rat kredytu, ze spłatą których zalega kredytobiorca. Te ostatnie są najczęściej wykorzystywane w ubezpieczeniach i wykorzystują rozkład Poissona lub jego liczne uogólnienia. 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 nowego – bayesowskiego – podejścia do estymacji wielomianowego modelu probitowego dla kategorii uporządkowanych. W szczególności przedstawimy propozycję Alberta i Chiba (1993), wykorzystującą losowanie Gibbsa. Następnie zastosujemy to podejście do badania ryzyka kredytowego pojedynczej umowy kredytowej dla klientów detalicznych pewnego polskiego banku komercyjnego. Niniejsze opracowanie stanowi kontynuację badań empiryczno-metodologicznych, prezentowanych wcześniej w pracach: Marzec (2003a,b,c). Podejście bayesowskie w przypadku modeli wielomianowych zastosowali m.in. Zellner (1983), Zellner i Rossi (1984), Abert i Chib (1993), Koop i Poirier (1993), McCulloch i Rossi (1994), McCulloch, Polson i Rossi (2000).

2. MODEL WIELOMIANOWOWY – DEFINICJA

Najprostsze definicje modeli wielomianowych opierają się na rozszerzeniu koncepcji modelu dla dychotomicznej zmiennej endogenicznej. Przyjmijmy, że zmienne endogeniczne y_t ($t=1, \dots, T$) przyjmują z ustalonym prawdopodobieństwem p_{ij} wartości $1, 2, \dots, J_t$, co zapisujemy³

$$\Pr(y_t = j) = F_{ij}(x, \theta) \text{ dla } t=1, \dots, T \text{ i } j=1, \dots, J_t \quad (1)$$

gdzie $F_{ij}(\cdot)$ jest znaną funkcją wiążącą prawdopodobieństwo p_{ij} z wielkościami θ i x oraz określającą klasę modelu, zaś θ i x są odpowiednio wektorem parametrów i wektorem zmiennych objaśniających (zmiennych egzogenicznych lub ich znanych funkcji); zob. Amemiya (1985). Najczęściej za funkcję $F_{ij}(\cdot)$ przyjmuje się dystrybuantę zmiennej losowej o rozkładzie normalnym lub logistycznym. Szczegółowa specyfikacja modelu statystycznego zależy od typu skali, na której dokonano pomiaru zmiennej endogenicznej oraz od zakresu posiadanych informacji, które niosą zmienne egzogeniczne; zob. Amemiya (1981, 1985), Maddala (1983). Inny sposób definicji modeli

³ Dla uproszczenia pominięto dodatkowy indeks t przy j .

wielomianowych, zaproponowany przez McFaddena, wykorzystuje koncepcję funkcji użyteczności, co ma szczególnie zastosowanie w przypadku kategorii nieuporządkowanych; zob. Maddala (1983) i literatura tam cytowana. Przypomnijmy, że szczególnym i najprostszym przypadkiem modeli wielomianowych jest model dychotomiczny (dla binarnej zmiennej y_t), który uzyskujemy, gdy $J_t = 2$ dla $t=1, \dots, T$. W przypadku modeli wielomianowych podstawową i najczęściej stosowaną metodą estymacji, gdy obserwacje y_t są niezależne, jest metoda największej wiarygodności (MNW). Funkcja wiarygodności dla modelu (1) ma postać

$$L(\theta|y) = \prod_{t=1}^T \prod_{j=1}^{J_t} (F_{ij})^{y_{ij}}, \quad (2)$$

gdzie y_{ij} jest zmienną zero-jedynkową, która przyjmuje wartość jeden, gdy obserwujemy $y_t = j$, natomiast zero – w przeciwnym przypadku. O podstawowych własnościach metody MNW (zgodność i asymptotyczna normalność) pisze m.in. Amemiya (1985). Zastosowanie tej metody wymaga wykorzystania numerycznych metod optymalizacji (np. metoda Newtona-Raphsona), co np. w przypadku dużej liczby wyborów (J_t) i związanej z tym dużej liczby parametrów może prowadzić do komplikacji obliczeniowych.

W dalszej części przedmiotem analizy będzie wyłącznie wielomianowy model probitowy dla kategorii uporządkowanych przy założeniu jednakowej liczby alternatyw ($J_t \equiv J$) oraz posiadania danych charakteryzujących jedynie podmiot dokonujący wybór. W literaturze podaje się, że model wielomianowy dla kategorii uporządkowanych jako pierwsi zaproponowali Aitchison i Silvey (1957). Wprowadzając ciągłe, nieobserwowalne (ukryte) zmienne z_t , których wartości określają obserwowaną kategorię zmiennej y_t , otrzymujemy model o następującej postaci (zob. McKelvey i Zavoina (1975))

$$\begin{cases} z_t = x_t \beta + \varepsilon_t \\ y_{ij} = 1 & \text{gdy } \alpha_{j-1} < z_t < \alpha_j \text{ dla } t = 1, \dots, T \quad j = 1, \dots, J \\ y_{ij} = 0 & \text{w przeciwnym przypadku,} \end{cases} \quad (3)$$

gdzie x_t jest wektorem zmiennych egzogenicznych, a parametry α_j są tzw. punktami granicznymi. Z uwagi na identyfikowalność parametrów zakłada się, że $\alpha_0 = -\infty$ i $\alpha_J = +\infty$ oraz $\alpha_j = 0$, jeżeli w równaniu dla zmiennej z_t występuje wyraz wolny, co zakładamy w tej pracy. Wówczas prawdopodobieństwo przyjęcia przez zmienną y_t kategorii j wynosi

$$p_{ij} \equiv \Pr(y_{ij} = 1) = \Pr(\alpha_{j-1} < z_t < \alpha_j) = F(\alpha_j - x_t \beta) - F(\alpha_{j-1} - x_t \beta), \quad (4)$$

gdzie $F(\cdot)$ jest dystrybuantą standaryzowanej zmiennej losowej ε_t , dla której przyjmuje się najczęściej rozkład logistyczny lub normalny. Pratt (1981) prezentuje warunki konieczne i wystarczające na jednoznaczne istnienie estymatora MNW w modelu (3) w zależności od przyjętego rozkładu dla ε_t .

W ekonomii, w przeciwieństwie do nauk przyrodniczych, modele wielomianowe dla kategorii uporządkowanych są rzadziej stosowane niż modele dla kategorii nieuporządkowanych. Spowodowane jest to tym, iż po pierwsze, zjawiska ekonomiczne cechują się dużą złożonością, więc przyjęcie a priori mocniejszej skali dla zmiennej objaśnianej wydaje się często być założeniem zbyt silnym. Po drugie, użycie modelu dla kategorii uporządkowanych, gdy w rzeczywistości zmienna objaśniana mierzona jest na skali nominalnej, choć poprawia efektywność estymacji, to prowadzi do poważnego obciążenia estymatora. W niniejszym opracowaniu konstrukcja zmiennej endogenicznej y_t (por. część 4) upoważnia nas do zastosowania modelu dla kategorii uporządkowanych.

3. BAYESOWKI WIELOMIANOWY MODEL PROBITOWY

Na gruncie bayesowskim model statystyczny jest zdefiniowany poprzez łączny rozkład prawdopodobieństwa dla obserwacji i parametrów lub innych wielkości nieobserwowalnych. W modelu hierarchicznym (3) łączna funkcja gęstości $p(y, z, \alpha, \beta)$, którą w przypadku rozkładu mieszanego (dyskretno-ciągłego) nazywa się uogólnioną funkcją gęstości, ma postać

$$p(y, z, \alpha, \beta) = p(y|z, \alpha) \cdot p(z|\beta) \cdot p(\alpha, \beta), \quad (5)$$

gdzie $p(\alpha, \beta)$ to rozkład a priori, który reprezentuje wstępną wiedzę badacza o parametrach, przy czym zakłada się niezależność parametrów α i β , więc $p(\alpha, \beta) = p(\alpha)p(\beta)$. W formule (5) zdegenerowany (jednopunktowy) rozkład próbkowy dla zaobserwowanego wektora y (warunkowy względem z i α) ma postać

$$p(y|z, \alpha) = \prod_{t=1}^T \sum_{j=1}^J [y_{tj} \cdot I(\alpha_{j-1} < z_t < \alpha_j)], \quad (6)$$

gdzie $I(w \in \Omega) = 1$, gdy $w \in \Omega$ i $I(w \in \Omega) = 0$, jeżeli $w \notin \Omega$. W modelu probitowym rozkład próbkowy dla wektora zmiennych ukrytych z , $p(z|\beta)$, jest wielowymiarowym rozkładem normalnym o wartości oczekiwanej $X\beta$ i jednostkowej macierzy kowariancji, gdzie macierz X o wymiarze $T \times k$ zawiera w kolumnach obserwacje poszczególnych zmiennych objaśniających. Przyjęte w pracy założenie o braku korelacji między ε_t i ε_s dla każdego $t \neq s$ jest powszechnie spotykane w literaturze.

Podstawowa zasada estymacji bayesowskiej sprowadza się do wyznaczenia z łącznego rozkładu prawdopodobieństwa (5) brzegowej gęstości dla wektora parametrów (α lub β) bądź wektora zmiennych ukrytych z , warunkowej względem wektora obserwacji y , czyli tzw. funkcji gęstości rozkładu a posteriori. Korzystając ze wzoru Bayesa otrzymujemy łączną funkcję gęstości rozkładu a posteriori dla α , β i z przy danym y , która przy założeniu niewłaściwego rozkładu jednostajnego a priori dla α i β , jak w pracy Aberta i Chiba (1993), ma postać

$$p(z, \alpha, \beta | y) = \frac{p(y, z, \alpha, \beta)}{p(y)} \propto \prod_{t=1}^T \left[(2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(z_t - x_t \beta)^2\right) \cdot \sum_{j=1}^J (y_{tj} \cdot I(\alpha_{j-1} < z_t < \alpha_j)) \right] \quad (7)$$

W celu uzyskania z formuły (7) brzegowych rozkładów a posteriori powinniśmy dokonać wielokrotnego całkowania, przy czym za sprawą niestandardowej postaci tego rozkładu całkowanie analityczne jest wykluczone. Model (3) charakteryzuje się wysokim wymiarem przestrzeni parametrów i wielkości nieobserwowalnych, więc zastosowanie tradycyjnych technik numerycznej aproksymacji całek jest nieskuteczne. Liczba wszystkich parametrów i wielkości nieobserwowalnych jest większa niż liczba obserwacji. Albert i Chib (1993) zaproponowali wykorzystanie w tym przypadku losowania Gibbsa (ang. *Gibbs Sampling*) jako metody numerycznej aproksymacji brzegowych rozkładów a posteriori. W celu zastosowania próbnika Gibbsa należy z gęstości (7) wyznaczyć pełne rozkłady warunkowe dla α , β i z , aby następnie na ich podstawie uzyskiwać poprzez wielokrotne generowanie liczb pseudolosowych próbki z rozkładu a posteriori (choć tylko w sensie asymptotycznym), zob. Casella i George (1992), Tierney (1994) lub Osiewalski (2001).

Warunkowy względem α , β i obserwowanej kategorii j zmiennej y_t rozkład a posteriori dla z_t jest uciętym rozkładem normalnym:

$$z_t | (\alpha, \beta, y_{tj} = 1) \sim N(x_t \beta, 1) \text{ ucięty na lewo (na prawo) od } \alpha_{j-1} \text{ (} \alpha_j \text{)}. \quad (8)$$

Natomiast warunkowy rozkład a posteriori dla $\beta = (\beta_1, \dots, \beta_k)'$, $p(\beta | \alpha, z, y)$, jest także wielowymiarowym rozkładem normalnym o wektorze wartości oczekiwanych β_z i macierzy kowariancji $(X'X)^{-1}$, co zapisujemy

$$\beta | (\alpha, z, y) \sim N^{(k)}(\beta_z, (X'X)^{-1}), \text{ gdzie } \beta_z = (X'X)^{-1} X'z. \quad (9)$$

Zauważmy, że gęstość rozkładu dla α_j warunkowego względem β , z , y i wszystkich α_i ($i \neq j$) jest proporcjonalna do wyrażenia

$$p\left(\alpha_j \middle| \alpha_{i \neq j}, \beta, z, y\right) \propto \prod_{t=1}^T [y_{t,j} \cdot I(\alpha_{j-1} < z_t < \alpha_j) + y_{t,j+1} \cdot I(\alpha_j < z_t < \alpha_{j+1})].$$

W efekcie pełny rozkład a posteriori dla α_j , warunkowy względem β , z , y i α_i ($i \neq j$), jest rozkładem jednostajnym

$$\alpha_j \middle| \alpha_{i \neq j}, \beta, z, y \sim U\left(\max(\alpha_{j-1}, (z_t | y_{t,j} = 1)), \min(\alpha_{j+1}, (z_t | y_{t,j+1} = 1))\right). \quad (10)$$

Próbkowanie Gibbsa polega na sekwencyjnym losowaniu z rozkładów (8), (9) i (10). Wcześniej należy przyjąć wartości początkowe dla α i β wykorzystując np. oceny MNW lub oceny estymatora metody najmniejszych kwadratów dla β w liniowym modelu prawdopodobieństwa,

czyli korzystając z formuły $(X'X)^{-1}X'y$, natomiast elementy wektora α przyjąć arbitralnie. Po wykonaniu dostatecznej liczby cykli wstępnych (n_0) otrzymujemy próbę z interesujących nas rozkładów a posteriori $\theta^{(n_0+1)}, \dots, \theta^{(n_0+n)}$, na podstawie której liczymy charakterystyki tych rozkładów jak wartości oczekiwane i odchylenia standardowe a posteriori dla poszczególnych parametrów (θ) lub ich funkcji $g(\theta)$, wg formuł $E[g(\theta)|y] = n^{-1} \sum_{h=n_0+1}^{n_0+n} g(\theta^{(h)}) = \bar{\mu}_n$ zaś

$$D[g(\theta)|y] = E[(g(\theta))^2|y] - (\bar{\mu}_n)^2.$$

Zasadniczym problemem pojawiającym się przy zastosowaniu metod typu Monte Carlo opartych na łańcuchach Markowa jest ocena zbieżności tego algorytmu. Możemy spodziewać się, że stabilizacja w kolejnych cyklach wartości oczekiwanych i odchyłeń standardowych a posteriori po odrzuceniu odpowiedniej liczby początkowych losowań, tzw. cykli spalonych, oznacza osiągnięcie zbieżności tej metody. Jedną z graficznych metod badania zbieżności algorytmu jest metoda CuSum zaproponowana przez Yu i Mykland, która wraz z innymi jest opisana w pracy Cowles i Carlin (1996). W niniejszym opracowaniu monitorujemy wartości standaryzowanej statystyki $CuSum_i$:

$$CuSum_i = \frac{\bar{\mu}_i - \bar{\mu}_n}{\sigma_n} \quad \text{dla } i = n_0 + 1, \dots, n_0 + n, \quad \text{gdzie } \bar{\mu}_i = (i - n_0)^{-1} \sum_{h=n_0+1}^i g(\theta^{(h)}). \quad (11)$$

Jeżeli zakres zmienności wielkości $CuSum_i$ maleje wraz z ilością cykli Gibbsa do zera lub kształtuje się na poziomie rzędu np. $\pm 5\%$ lub mniej, to możemy przyjąć, iż zbieżność algorytmu została osiągnięta. Rysunek 1 przedstawia kształtowanie się $CuSum_i$ dla parametrów α i β , po odrzuceniu 300 tys. cykli wstępnych. W tym momencie warto wspomnieć, że na gruncie bayesowskim przy dość ogólnych założeniach, oceny MNW można traktować jako przybliżone wartości oczekiwane rozkładu a posteriori parametrów, który w przypadku dużej liczby obserwacji można przybliżać rozkładem normalnym, zob. Osiewalski (1991). W dalszej części opracowania prezentujemy wyniki empiryczne uzyskane dla bardzo dużej liczby obserwacji, a zatem zastosowanie MNW w tym przypadku z punktu widzenia teorii powinno okazać się równoważne wnioskowaniu bayesowskiemu.

Rysunek 1

4. WYNIKI EMPIRYCZNE

Przedstawiony powyżej wielomianowy model probitowy oraz podejście bayesowskie wykorzystaliśmy do badania niewypłacalności kredytobiorców detalicznych. W tym celu posłużyliśmy się zbiorem danych, na podstawie których wcześniej dokonaliśmy estymacji modeli

dychotomicznych, tj. modelu probitowego, logitowego oraz ich uogólnienia, czyli modelu z rozkładem t-Studenta o nieznannej liczbie stopni swobody; zob. Marzec (2003a,b,c). Dane pochodzące z dużego, polskiego banku komercyjnego obejmowały dostępne informacje o kredytobiorcach indywidualnych i udzielonych im kredytach detalicznych, tj. kredytach konsumpcyjnych i hipotecznych. Liczyły one prawie 40 tysięcy rachunków i obejmowały okres od 01.01.2000 do 30.09.2001 r. W celu określenia zmiennej endogenicznej y_t wykorzystaliśmy klasyfikację należności, do której banki są zobowiązane treścią uchwały Komisji Nadzoru Bankowego w sprawie zasad tworzenia rezerw na ryzyko związane z działalnością banków⁴. W przypadku kredytów detalicznych wyróżnia się cztery kategorie należności, tj. należności normalne, poniżej standardu, wątpliwe i stracone. Sposób kwalifikacji poszczególnych kredytów zależy od terminowości spłat przez dłużnika kapitału i odsetek oraz jego sytuacji ekonomiczno-finansowej. Wobec powyższego przyjęliśmy, iż zmienna objaśniana y_t 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	80.3%
2	Poniżej standardu	Od 1 do 3 miesięcy	6.0%
3	Wątpliwe	Od 3 do 6 miesięcy	6.3%
4	Stracone	Powyżej 6 miesięcy	7.4%

Niech $y_{i1} = 1$, gdy kredyt zakwalifikowany został do kategorii należności normalnych, natomiast $y_{i2} = 1$, gdy kredyt jest oznaczony jako „poniżej standardu” itd.

Z punktu widzenia zarządzania ryzykiem kredytowym bank stara się minimalizować w globalnym portfelu kredytowym udział kredytów z drugiej, trzeciej i czwartej kategorii, czyli należności z grupy zagrożonych, wprowadzając m.in. metody oceny solidności kredytowej kredytobiorcy zwane metodami credit-scoring. Wyniki empiryczne przedstawione poniżej mogą być podstawą do budowy praktycznego modelu scoringowego.

Jako potencjalne zmienne egzogeniczne wyjaśniające ryzyko pojedynczej umowy kredytowej przyjęliśmy (jak we wcześniejszych pracach):⁵

- 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,

⁴ Uchwała nr 8/1999 Komisji Nadzoru Bankowego z 22 grudnia 1999 r. stanowi zasady tworzenia przez banki rezerw celowych od należności zagrożonych.

⁵ Dodatkowo wprowadziliśmy zmienną informującą o okresie trwania umowy kredytowej. Więcej informacji o zbiorze danych prezentujemy w pracy Marzec (2003a).

- 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),
- typ kredytu (1 – kredyt konsumpcyjny, 0 – kredyt hipoteczny),
- okres trwania umowy kredytowej (w dziesiątkach lat),
- 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 wprowadziliśmy trzy zmienne zerojedynkowe, a za punkt odniesienia 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, $zrdoch1 = 0$ w przeciwnym przypadku,
- $zrdoch2 = 1$, gdy źródłem dochodu kredytobiorcy jest własna działalność, umowa o dzieło lub umowa zlecenie, $zrdoch2 = 0$ w przeciwnym przypadku,
- $zrdoch3 = 1$ w przypadku innego źródła dochodu, np. stypendium, $zrdoch3 = 0$ w przeciwnym przypadku.

W niniejszej pracy – podobnie jak w pracy Marzec (2003c) - założyliśmy, że w modelu (3) prawdopodobieństwo zakwalifikowania kredytu do jednej z kategorii należności, może zależeć liniowo nie 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:

$$\begin{cases} z_t = \beta_1 + \sum_h w_{th}\beta_h + \sum_h \sum_{i \geq h} w_{th}w_{ti}\beta_{hi} + \varepsilon_t \\ y_{ij} = 1 & \text{gdy } \alpha_{j-1} < z_t < \alpha_j \text{ dla } t = 1, \dots, T \quad j = 1, \dots, J \\ y_{ij} = 0 & \text{w przeciwnym przypadku,} \end{cases} \quad (12)$$

Z punktu widzenia omówionych wcześniej metod wnioskowania, z_t jest nadal liniowo zależne od parametrów β , więc taki nieliniowy sposób wprowadzenia zmiennych egzogenicznych nie wnosi żadnych komplikacji na etapie estymacji. Powyższa modyfikacja może przyczynić się do lepszego oszacowania p_{ij} . 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}). Ponadto w modelu (12) iloraz pochodnych cząstkowych prawdopodobieństwa p_{ij} względem zmiennych w_{th} i w_{tg} , tzw. efektów krańcowych zależy od

wszystkich zmiennych egzogenicznych, czyli dla każdej obserwacji może być inny. Natomiast, jeżeli założymy $\beta_{hi}=0$, to iloraz ten jest równy ilorazowi parametrów, tj. β_h/β_g , a zatem nie zależy od wartości tych zmiennych. Powyższe rozszerzenie liczby czynników wyjaśniających p_{ij} może być przedmiotem statystycznej weryfikacji, której wyniki pokazujemy w dalszej części.

Uzyskane wyniki empiryczne wskazały (zgodnie z teorią), że w przypadku tak dużej liczby obserwacji, wyniki MNW są identyczne z rezultatami otrzymanymi za pomocą podejścia bayesowskiego. Uzyskanie zbieżności algorytmu Gibbsa wymagało wykonania dużej liczby cykli spalonych, ponieważ wymiar przestrzeni parametrów i wielkości nieobserwowalnych liczył prawie 40 tysięcy. Drugim czynnikiem negatywnie wpływającym na zbieżność próbkowania Gibbsa mogła być silna współliniowość spowodowana uwzględnieniem w modelu (12) iloczynów i kwadratów zmiennych w_{th} . Tabela 1 zawiera wartości oczekiwane i odchylenia standardowe a posteriori dla parametrów modelu (12), czyli dla wektorów α i β .

Tabela 1

W celu statystycznej weryfikacji zasadności wprowadzenia w modelu (12) iloczynów i kwadratów zmiennych w_{th} wykorzystaliśmy bayesowski odpowiednik klasycznego testu F na redukcję modelu, posługując się formą kwadratową:

$$u(\beta_{(2)}) = (\beta_{(2)} - b_{(2)})^T H_{22}^{-1} (\beta_{(2)} - b_{(2)}) / k_2, \text{ gdzie } \beta_{(2)} = [\beta_{13} \dots \beta_{66}]'. \quad (13)$$

Jeżeli brzegowy rozkład a posteriori wektora $\beta_{(2)}$ jest k_2 -wymiarowym rozkładem t-Studenta o $T-k$ stopniach swobody ($1 \leq k_2 \leq k$), wektorze niecentralności $b_{(2)}$ i macierzy precyzji H_{22}^{-1} , to rozkład a posteriori wielkości $u(\beta_{(2)})$ jest rozkładem F-Snedecora o $(k_2, T-k)$ stopniach swobody; zob. Zellner (1971). Dla restrykcji $\beta_{(2)} = [0 \dots 0]'$, która odpowiada redukcji modelu (12) do modelu liniowego względem w_{th} , wartość $u([0 \dots 0]')$ wynosi ponad 17. Otrzymana wartość znajduje się w prawym ogonie rozkładu $F(k_2, T-k)$, który dla tak dużej liczby obserwacji jest dobrą aproksymacją rozkładu a posteriori tej formy kwadratowej, gdy nie znamy dokładnej postaci brzegowego rozkładu a posteriori dla $\beta_{(2)}$. Ponieważ wybrany element podprzestrzeni parametrów ($\beta_{(2)} = [0 \dots 0]'$) znajduje się w podzbiorze wartości parametrów mało prawdopodobnych a posteriori, więc nie jesteśmy skłonni przyjąć, że $\beta_{(2)} = [0 \dots 0]'$ i nie dokonujemy redukcji modelu⁶.

Pojedyncze parametry β_h lub β_{hi} modelu wielomianowego nie mają bezpośredniej interpretacji, nawet w przypadku najprostszym, gdy $\beta_{hi} = 0$. Wpływ marginalnych zmian zmiennych objaśniających w_{th} na zmianę prawdopodobieństwa p_{ij} wyrażają efekty krańcowe bądź

⁶ Podobne wyniki otrzymaliśmy także w przypadku modeli dwumianowych, por. Marzec (2003c).

elastyczności. Tabela 2 zawiera efekty krańcowe - $\eta_{jh} = T^{-1} \sum_t \partial \Pr(y_{ij} = 1) / \partial w_{th}$ - obliczone jako pochodna cząstkowa p_{ij} względem w_{th} i uśrednione po wszystkich obserwacjach⁷.

Tabela 2

Interpretując efekty krańcowe można zauważyć, że dla ustalonej zmiennej objaśnianej suma η_{jh} wynosi zero, co wynika z własności sumowania się prawdopodobieństwa do jedności. W konsekwencji efekty krańcowe dla dwóch skrajnych alternatyw, tj. η_{1h} i η_{4h} , charakteryzują się przeciwnymi znakami. Z uwagi na przyjętą różniczkową definicję efektów krańcowych szczegółowo omówimy efekty krańcowe jedynie względem zmiennych ciągłych, natomiast w przypadku zmiennych zero-jedynkowych ograniczymy się do oceny jakościowej, zwracając uwagę jedynie na znak η_{jh} .

Jeżeli wiek kredytobiorcy zwiększy się o 1 rok, to przy założeniu *ceteris paribus* prawdopodobieństwo zakwalifikowania jego kredytu do kategorii należności normalnych (p_{11}) wzrośnie średnio o 0.00227 jednostki z błędem ± 0.00105 , równocześnie prawdopodobieństwo zakwalifikowania do kategorii poniżej standardu spadnie przeciętnie o 0.00032 (± 0.00020), zakwalifikowania do trzeciej kategorii obniży się o 0.00058 (± 0.00029), zaś do czwartej (p_{14}) o 0.00136 (± 0.00057). Natomiast wzrost wpływów na bieżący rachunek depozytowy kredytobiorcy o tysiąc złotych spowoduje przeciętny wzrost p_{11} o 0.02969 (± 0.00982), a więc o tyle samo obniży się prawdopodobieństwa zakwalifikowania kredytu do pozostałych kategorii należności, przy czym prawdopodobieństwo opóźnienia spłaty kredytu dłuższego niż 6 miesięcy (p_{14}) ulegnie obniżeniu średnio o 0.01780 (± 0.00628). Wraz z wydłużeniem okresu kredytowania klienta, prolongatą kredytu o rok, wielkość p_{11} i p_{12} przeciętnie wzrośnie odpowiednio o 0.0112 (± 0.0017) i 0.0006 (± 0.0003), natomiast p_{13} i p_{14} obniży się o 0.0020 (± 0.0005) i 0.0098 (± 0.0010). Udzielenie kredytu poprzez pośrednika, podobnie jak sprzedaż kredytu konsumpcyjnego zamiast hipotecznego, zwiększa ryzyko kredytowe, natomiast posiadanie karty płatniczej lub kredytowej zmniejsza je (powoduje wzrost p_{11}). Analogicznie, jeżeli hipotetycznym klientem jest mężczyzna, to ryzyko niedotrzymania przez niego umowy (zaniechania spłaty rat i odsetek) jest wyższe niż w przypadku kobiety. Studenci korzystający z kredytu studenckiego (*zrdoch3*) oraz emeryci i renciści (*zrdoch1*) są mniej ryzykownymi kredytobiorcami niż klienci zatrudnieni na umowę o pracę. Spośród czterech źródeł dochodów największe ryzyko kredytowe związane jest z udzieleniem kredytu klientom prowadzącym własną działalność gospodarczą (*zrdoch2*). Efekty krańcowe względem zmiennej ROR charakteryzują się relatywnie dużymi odchyleniami standardowe a posteriori, co świadczy o

⁷ Formuła oparta o pochodnych ma uzasadnienie w przypadku zmiennych ciągłych, natomiast w przypadku zmiennych zero-jedynkowych powinno się stosować formułę różnicy prawdopodobieństw, tzn. $\Pr(y_{ij}=1|w_{th}=1) - \Pr(y_{ij}=1|w_{th}=0)$, która ponadto gwarantuje spełnienie warunku $-1 < \eta_{jh} < 1$.

braku wpływu tej zmiennej na wielkość p_{ij} . Powyższe wyniki są jakościowo zgodne z wynikami uzyskanymi dla modeli dwumianowych w pracach Marzec (2003a,b,c).

Oszacowany model możemy wykorzystać do celów prognostycznych, czyli prognozowania okresu opóźnienia w spłacie kredytu, co jest główną podstawą do określenia kategorii należności. Dla uproszczenia rozważmy, jak w pracach Marzec (2003a,b,c), cztery hipotetyczne sylwetki klientów starających się o kredyt, które przedstawia Tabela 3.

Tabela 3

Spśród czterech potencjalnych kredytobiorców najmniejsze ryzyko kredytowe związane jest ze starszą panią utrzymującą się z emerytury w kwocie 1 tysiąca złotych netto, której udzielono kredyt hipoteczny. Prawdopodobieństwo terminowej spłaty przez nią rat kapitałowo-odsetkowych jest praktycznie równe jedności. Największe ryzyko kredytowe związane jest z kredytem, który został udzielony poprzez pośrednika młodemu mężczyźnie prowadzącemu własną działalność gospodarczą, który dotąd nie korzystał z innych usług badanego banku. Prawdopodobieństwo, że będzie on terminowo spłacał kredyt wynosi tylko 0.45, a prawdopodobieństwo opóźnienia spłaty od jednego do 3 miesięcy (od 3 do 6 miesięcy) wynosi 0.13 (0.16). Natomiast prawdopodobieństwo opóźnienia dłuższego niż 6 miesięcy (czwarta kategoria należności), które powoduje obowiązek tworzenia 100% rezerw celowych, wynosi aż 0.26. Warto wspomnieć, że kredyty udzielane przez pośredników przeznaczone były przede wszystkim na ratalne zakupy sprzętu AGD, mebli itp., więc dotyczyły one niewielkich kwot, lecz ich udział ilościowy w portfelu kredytowym był znaczący, a więc koszty ewentualnej windykacji należności byłyby duże. Identyfikacja zatem grupy klientów obciążonej dużym ryzykiem kredytowym jest podstawowym krokiem zmierzającym do minimalizacji ryzyka kredytowego portfela banku. Negatywny wpływ korzystania przez bank z usług pośredników kredytowych na ryzyko kredytowe potwierdza analiza „najczęstszego klienta”, tzn. o cechach najczęstszych w próbie (dotyczy zmiennych jakościowych) i przeciętnych (dla zmiennych ciągłych) w badanej zbiorowości. Jeżeli udzielono mu kredytu bezpośrednio przez bank, a zatem jego zdolność kredytowa została zweryfikowana przez pracownika banku, wówczas prawdopodobieństwo dotrzymania przez niego umowy jest bardzo wysokie i wynosi ponad 0.96, natomiast prawdopodobieństwo opóźnienia w spłacie dłuższego niż 6 miesięcy jest znikome – poniżej 0.01. Gdyby udzielono mu kredytu poprzez pośrednika, wówczas prawdopodobieństwo zakwalifikowania tego kredytu do pierwszej kategorii ryzyka obniżyłoby się do poziomu 0.8, zakwalifikowania do drugiej wynosiłoby 0.08, do trzeciej prawie 0.07, a prawdopodobieństwo całkowitego zaniechania spłaty kształtowałoby się na poziomie 0.05. Warto zauważyć, że oszacowane wielkości p_{ij} dla tego potencjalnego kredytobiorcy odpowiadają w przybliżeniu empirycznym udziałom ilości poszczególnych kategorii należności w badanym portfelu kredytów detalicznych.

Tabela 4

Następnie zbadaliśmy zdolność prognostyczną oszacowanego powyżej modelu. W tym celu obliczyliśmy przeciętne teoretyczne prawdopodobieństwa zakwalifikowania rachunków kredytowych do każdej z czterech kategorii należności; zob. Tabela 5. Prawdopodobieństwo poprawnego zakwalifikowania przez model kredytów z pierwszej kategorii należności jest wysokie i wynosi 0.86. 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 20%, przy czym odpowiednio aż 62, 60 i 55 procent kredytów z tych kategorii zostało uznanych za należności normalne. Model więc zbyt optymistycznie ocenia ryzyko związane z rachunkami kredytowymi, co z punktu widzenia zarządzania ryzykiem banku nie jest cechą pożądaną.

Tabela 5

5. PODSUMOWANIE

W niniejszym opracowaniu zaprezentowaliśmy, odwołując się do literatury przedmiotu, specyfikację i estymację bayesowskiego wielomianowego modelu probitowego dla kategorii uporządkowanych. Przedstawiliśmy losowanie Gibbsa jako automatyczną metodę numerycznej aproksymacji brzegowych rozkładów a posteriori i charakterystyk tych rozkładów. Warto wspomnieć, iż dla małej próby podejście bayesowskie jest polecane z uwagi na nieasymptotyczne (małopróbkowe) własności, na co zwrócił uwagę Zellner (1983), pierwszy stosując to podejście w przypadku logitowego modelu wielomianowego.

Z punktu widzenia zarządzania ryzykiem kredytowym, na podstawie zaprezentowanego modelu wielomianowego, uzyskaliśmy nowe wyniki empiryczne w stosunku do rezultatów otrzymanych dla modelu dwumianowego. Jednakże oszacowany model nie wykazuje zbyt dobrych zdolności prognostycznych, co może być spowodowane przyjęciem zbyt silnego założenia o rozkładzie normalnym dla składnika losowego w równaniu (3). Ponadto wprowadzenie dodatkowych zmiennych egzogenicznych prawdopodobnie poprawiłoby dopasowanie modelu do danych empirycznych. Wyniki dotychczasowych badań empirycznych uzasadniają potrzebę stosowania modelu wykorzystującego rozkład o „grubych ogonach”, tj. rozkładu t-Studenta, który w przypadku modelu dwumianowego zdecydowanie lepiej opisywał dane niż model probitowy czy logitowy, zob. Marzec (2003c). Budowa i estymacja modelu opartego na tym rozkładzie wydaje się być naturalnym kierunkiem dalszych pogłębionych badań metodologiczno-empirycznych.

BIBLIOGRAFIA

- 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*, vol. 19, s. 1483-1536.
- Amemiya T., 1985, *Advanced Econometrics*, Harvard University Press, Cambridge (Massachusetts).
- Casella G., E. George, 1992, *Explaining the Gibbs Sampler*, *The American Statistician*, 46.
- Cowles M.K., B.P. Carlin, 1996, *Markov Chain Monte Carlo Convergence Diagnostic: A Comparative Review*, *Journal of the American Statistical Association*, 91, s. 883-904.
- 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.
- Koop, G., D. Poirier, 1993, *Bayesian Analysis of Logit Models using Natural Conjugate Priors*, *Journal of Econometrics*, 56, s. 323-340.
- 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, 103-117.
- 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)*, *Przegląd Statystyczny*, tom 50, nr 4, s. 129-146.
- McCulloch R.E., N.G. Polson, P. E. Rossi, 2000, *A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters*, *Journal of Econometrics*, 99, s. 173-193.
- McCulloch R.E., P. E. Rossi, 1994, *An exact Likelihood Analysis of the Multinomial Probit Model*, *Journal of Econometrics*, 64, s. 207-240.
- 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.
- Osiewalski J., 1991, *Bayesowska estymacja i predykcja dla jednorównaniowych modeli ekonometrycznych*, *Akademia Ekonomiczna w Krakowie, Zeszyty Naukowe, Seria specjalna: Monografie*, nr 100, Kraków.
- Osiewalski J., 2001, *Ekonometria bayesowska w zastosowaniach*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Pratt J.W., 1981, *Concavity of the Log Likelihood*, *Journal of the American Statistical Association*, vol. 76, nr. 373, s. 103-106.
- Tierney L., 1994, *Markov Chains for Exploring Posterior Distributions (with discussion)*, *Annals of Statistics*, 22, s. 1701-1762.
- Wiśniewski J., 1986, *Ekonometryczne badanie zjawisk jakościowych (studium metodologiczne)*, Uniwersytet M. Kopernika, Toruń.
- Zellner A., 1971, *An Introduction to Bayesian Inference in Econometrics*, J. Wiley, New York 1971.
- Zellner A., 1983, *Bayesian Analysis of Simple Multinomial Logit Model*, *Economics Letters*, 11, s. 133-136.
- Zellner A., P. Rossi, 1984, *Bayesian Analysis of Dichotomous Quantal Response Models*, *Journal of Econometrics*, 25, s. 365-393.

Tabela 1.

Wartości oczekiwane i odchylenia standardowe a posteriori parametrów bayesowskiego wielomianowego modelu probitowego.

Zmienna	Parametr	E($\cdot y$)	D($\cdot y$)	Zmienna	Parametr	E($\cdot y$)	D($\cdot y$)
	α_2	0.329	0.005	$(w_3)^2$	β_{33}	-0.011	0.002
	α_3	0.787	0.013	$w_3 \cdot w_4$	β_{34}	6.269	6.155
Stała	β_1	-1.811	0.600	$w_3 \cdot w_5$	β_{35}	0.184	0.313
Płeć (w_1)	β_2	0.577	0.212	$w_3 \cdot w_6$	β_{36}	-1.872	1.692
Wiek (w_2)	β_3	-0.730	1.117	$w_3 \cdot w_7$	β_{37}	-1.086	0.445
Wpływy (w_3)	β_4	-12.552	6.277	$w_3 \cdot w_8$	β_{38}	0.026	0.328
ROR (w_4)	β_5	1.340	0.452	$w_3 \cdot w_9$	β_{39}	-14.501	2.927
Karty (w_5)	β_6	-1.261	0.632	$w_3 \cdot w_{10}$	β_{40}	4.762	0.495
Pośrednik (w_6)	β_7	1.955	0.134	$w_3 \cdot w_{11}$	β_{41}	-0.953	2.574
Typ kredytu (w_7)	β_8	0.438	0.554	$w_4 \cdot w_5$	β_{42}	0.491	0.548
Okres (w_8)	β_9	-1.596	0.487	$w_4 \cdot w_6$	β_{43}	0.003	0.113
Zrdoch1 (w_9)	β_{10}	-27.934	8.294	$w_4 \cdot w_7$	β_{44}	-1.233	0.403
Zrdoch2 (w_{10})	β_{11}	0.907	0.365	$w_4 \cdot w_8$	β_{45}	-0.257	0.227
Zrdoch3 (w_{11})	β_{12}	2.441	0.798	$w_4 \cdot w_9$	β_{46}	0.332	0.156
$w_1 \cdot w_2$	β_{13}	-0.382	0.163	$w_4 \cdot w_{10}$	β_{47}	-0.176	0.144
$w_1 \cdot w_3$	β_{14}	-0.158	0.327	$w_4 \cdot w_{11}$	β_{48}	0.626	0.500
$w_1 \cdot w_4$	β_{15}	0.209	0.075	$w_5 \cdot w_6$	β_{49}	0.050	0.124
$w_1 \cdot w_5$	β_{16}	-0.077	0.071	$w_5 \cdot w_7$	β_{50}	0.257	0.223
$w_1 \cdot w_6$	β_{17}	0.009	0.061	$w_5 \cdot w_8$	β_{51}	0.275	0.174
$w_1 \cdot w_7$	β_{18}	-0.334	0.181	$w_5 \cdot w_9$	β_{52}	0.112	0.147
$w_1 \cdot w_8$	β_{19}	-0.304	0.109	$w_5 \cdot w_{10}$	β_{53}	0.052	0.104
$w_1 \cdot w_9$	β_{20}	0.003	0.055	$w_5 \cdot w_{11}$	β_{54}	-0.124	0.259
$w_1 \cdot w_{10}$	β_{21}	-0.158	0.080	$w_6 \cdot w_8$	β_{55}	-2.222	0.205
$w_1 \cdot w_{11}$	β_{22}	-0.199	0.163	$w_6 \cdot w_9$	β_{56}	0.081	0.104
$(w_2)^2$	β_{23}	-0.326	0.650	$w_6 \cdot w_{10}$	β_{57}	-0.431	0.133
$w_2 \cdot w_3$	β_{24}	3.141	1.517	$w_6 \cdot w_{11}$	β_{58}	0.773	0.496
$w_2 \cdot w_4$	β_{25}	-0.554	0.358	$w_7 \cdot w_8$	β_{59}	1.964	0.320
$w_2 \cdot w_5$	β_{26}	0.950	0.346	$w_7 \cdot w_9$	β_{60}	28.250	8.294
$w_2 \cdot w_6$	β_{27}	-0.589	0.292	$w_7 \cdot w_{10}$	β_{61}	-0.091	0.253
$w_2 \cdot w_7$	β_{28}	0.571	0.899	$w_7 \cdot w_{11}$	β_{62}	-3.084	0.527
$w_2 \cdot w_8$	β_{29}	1.390	0.547	$(w_8)^2$	β_{63}	0.450	0.132
$w_2 \cdot w_9$	β_{30}	-0.710	0.311	$w_8 \cdot w_9$	β_{64}	-0.329	0.205
$w_2 \cdot w_{10}$	β_{31}	-0.874	0.392	$w_8 \cdot w_{10}$	β_{65}	-0.156	0.187
$w_2 \cdot w_{11}$	β_{32}	1.127	0.722	$w_8 \cdot w_{11}$	β_{66}	-2.270	0.375

Źródło: obliczenia własne.

Tabela 2.

Wartości oczekiwane i odchylenia standardowe a posteriori uśrednionych efektów krańcowych

$$T^{-1} \sum_t \partial \Pr(y_{ij} = 1) / \partial w_{th}$$

zmienna	j = 1		j = 2		j = 3		j = 4	
	E()	D()	E()	D()	E()	D()	E()	D()
pleć	-0.008	(0.003)	0.003	(0.001)	0.002	(0.001)	0.003	(0.002)
wiek	0.227	(0.105)	-0.032	(0.020)	-0.058	(0.029)	-0.136	(0.057)
wpływy	2.969	(0.982)	-0.426	(0.109)	-0.763	(0.247)	-1.780	(0.628)
ROR	-0.031	(0.026)	0.013	(0.010)	0.010	(0.008)	0.008	(0.011)
karty	0.082	(0.084)	-0.010	(0.009)	-0.020	(0.021)	-0.053	(0.055)
pośrednik	-0.220	(0.007)	0.037	(0.002)	0.056	(0.002)	0.127	(0.004)
typ kredytu	-1.277	(0.340)	0.213	(0.060)	0.355	(0.096)	0.709	(0.185)
okres kredytowania	0.112	(0.017)	0.006	(0.003)	-0.020	(0.005)	-0.098	(0.010)
zrdoch1	0.100	(0.026)	-0.046	(0.011)	-0.035	(0.009)	-0.020	(0.008)
zrdoch2	-0.009	(0.010)	0.009	(0.002)	0.004	(0.002)	-0.004	(0.006)
zrdoch3	0.030	(0.019)	-0.008	(0.004)	-0.009	(0.005)	-0.013	(0.012)

Źródło: obliczenia własne.

Tabela 3.

Opis wybranych sylwetek kredytobiorców.

zmienna	Najczęstszy pośrednik=1	klient pośrednik=0	Młody Biznesmen	Starsza pani
Płeć	1	1	1	0
Wiek (w latach)	40.2	40.2	21	60
Wpływy (w tys. zł/miesiąc)	10.2	10.2	0	1
ROR	1	1	0	1
Karty płatnicze	0	0	0	1
Pośrednik	1	0	1	0
Typ kredytu: konsumpcyjny	1	1	1	0
Okres kredytu (w latach)	2.6	2.6	2.6	2.6
Zrdoch1	0	0	0	1
Zrdoch2	0	0	1	0
Zrdoch3	0	0	0	0

Źródło: obliczenia własne.

Tabela 4.

Wartości oczekiwane i odchylenia standardowe a posteriori prawdopodobieństwa zakwalifikowania kredytu hipotetycznego klienta do poszczególnych kategorii należności.

		$j=1$	$j=2$	$j=3$	$j=4$
Najczęstszy (pośrednik=1)	E()	0.803	0.078	0.068	0.052
	D()	(0.044)	(0.012)	(0.015)	(0.016)
Najczęstszy (pośrednik=0)	E()	0.965	0.019	0.011	0.005
	D()	(0.004)	(0.002)	(0.001)	(0.001)
Młody Biznesmen	E()	0.450	0.130	0.164	0.256
	D()	(0.039)	(0.002)	(0.008)	(0.032)
Starsza Pani	E()	0.999	1.61E-56	1.11E-58	1.08E-61
	D()	(1.97E-53)	(2.55E-54)	(1.78E-56)	(1.76E-59)

Źródło: obliczenia własne.

Tabela 5.

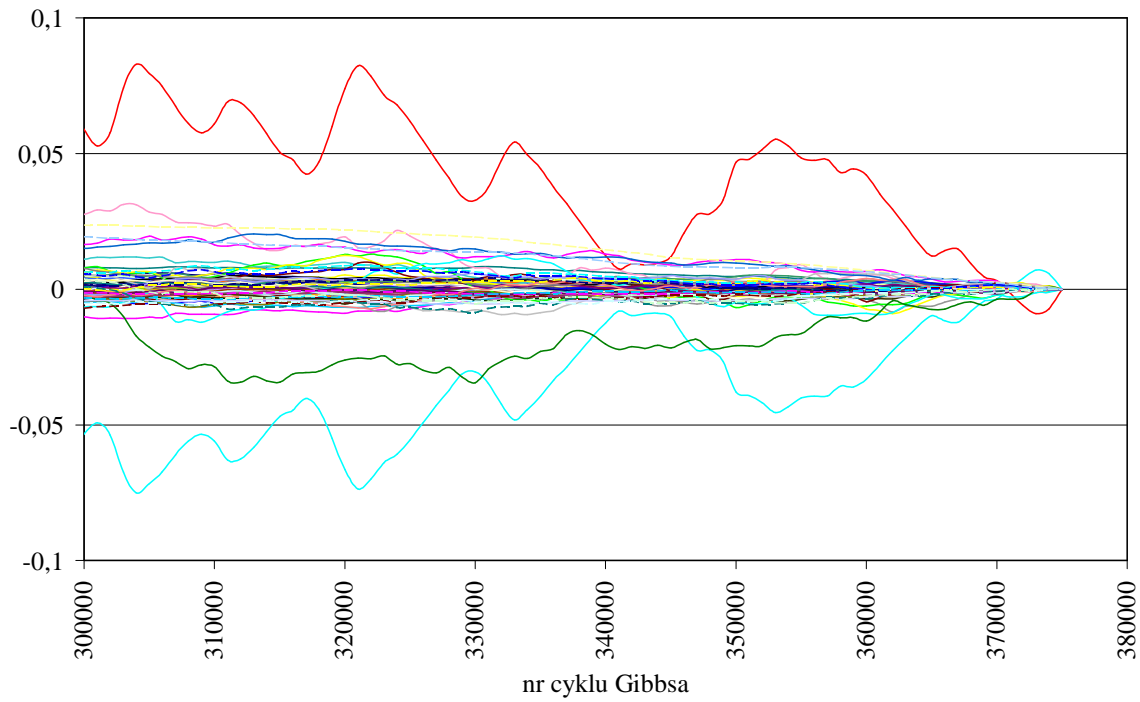
Średnie prawdopodobieństwo zakwalifikowania rachunku kredytowego do danej kategorii należności $T^{-1} \sum_i \Pr(y_{it} = 1 | y_{ij} = 1)$ dla $i, j = 1, \dots, 4$.

	$j=1$	$j=2$	$j=3$	$j=4$
	$y_{t1} = 1$	$y_{t2} = 1$	$y_{t3} = 1$	$y_{t4} = 1$
$\Pr(y_{t1} = 1 y_{ij} = 1)$	0.86	0.62	0.60	0.55
$\Pr(y_{t2} = 1 y_{ij} = 1)$	0.05	0.11	0.11	0.11
$\Pr(y_{t3} = 1 y_{ij} = 1)$	0.04	0.12	0.13	0.14
$\Pr(y_{t4} = 1 y_{ij} = 1)$	0.05	0.15	0.16	0.20
Suma	1.00	1.00	1.00	1.00

Źródło: obliczenia własne.

Rysunek 1.

Wartości statystyki CuSum dla parametrów α i β .



Źródło: obliczenia własne.