

Jerzy Marzec

BAYESOWSKI MODEL WIELOMIANOWY Z ROZKŁADEM t STUDENTA DLA KATEGORII UPORZĄDKOWANYCH¹

1. Wprowadzenie

W literaturze ekonometrycznej modele 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.

Powyższy model jest oryginalną propozycją Aitchisona i Silveya z 1957 roku i (jak wiele innych modeli dla zmiennych jakościowych) zastosowano go po raz pierwszy w biostatystyce. Dopiero w 1975 roku został przeniesiony na grunt nauk społecznych przez McKelveya i Zavoina. Wykorzystali go do analizy zachowań kongresmenów amerykańskich podczas głosowania nad ustawą z 1965 roku, dotyczącą rządowego programu pomagającego osobom o niskich dochodach pokryć koszty leczenia. W tym przypadku, jako zmienne objaśniające wykorzystano: przynależność partyjną (demokraci, republikanie), region (południe, północ), poziom bezrobocia, odsetek ludności powyżej 65 roku życia i gęstość zaludnienia w regionie, który reprezentował kongresmen. Autorzy rozważali trzy stanowiska: głosujących za przyjęciem najbardziej „socjalnej” wersji ustawy, głosujących za wersją z poprawkami zgłoszonymi przez umiarkowanych przeciwników tego programu oraz głosujących za odrzuceniem ustawy w całości. Postawy kongresmanów są podobne do skali odpowiedzi respondentów w badaniach ankietowych, tzw. skali Likerta: zdecydowanie tak, raczej tak, nie mam zdania, raczej nie i zdecydowanie nie. Innym przykładem zmiennych mierzonych na skali porządkowej jest np. informacja o zatrudnieniu (bezrobotny, zatrudniony w niepełnym wymiarze godzin, na pełny etat). Klasyfikację modeli dyskretnych ze względu na rodzaj wartości, jakie przyjmuje zmienna endogeniczna wraz z zastosowaniami w ekonomii, prezentuje m.in. Maddala [1983].

Głównym celem niniejszego artykułu jest prezentacja i wykorzystanie modelu wielomianowego z rozkładem t Studenta w analizie niespłacalności kredytów detalicznych. W szczególności przedstawiona zostanie bayesowska specyfikacja i estymacja tego modelu. Przedstawione w artykule wyniki stanowią kontynuację podjętej problematyki, zaprezentowanej wcześniej w pracach Marzec [2003b,c].

2. Model wielomianowy dla kategorii uporządkowanych

W niniejszym artykule przedmiotem analizy jest wyłącznie wielomianowy model dla kategorii uporządkowanych przy założeniu jednakowej liczby alternatyw i posiadaniu danych charakteryzujących jedynie podmiot dokonujący wybór. Wprowadzając ciągłe, nieobserwowalne (ukryte) zmienne z_i , których wartości determinują obserwowaną kategorię zmiennej y_i , model ma następującą postać (zob. McKelvey i Zavoina [1975])

¹ Artykuł powstał w ramach badań statutowych finansowanych przez Akademię Ekonomiczną w Krakowie w 2005 r.

$$\begin{cases} z_t = \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \\ y_{ij} = I_{[\alpha_{j-1}, \alpha_j)}(z_t) \text{ dla } t = 1, \dots, T \quad j = 1, \dots, J \end{cases} \quad (1)$$

gdzie $I_{\mathcal{A}}(\omega)$ jest funkcją charakterystyczną zbioru, zaś α_j są tzw. punktami granicznymi. Ponadto \mathbf{x}_t jest wektorem zmiennych egzogenicznych (lub ich znanych funkcji) charakteryzujących jednostkę podejmującą wybór. Zmienna y_t jest zmienną skalarną przyjmującą wartości odpowiadające numerom kategorii, czyli $1, 2, \dots, J$, zatem zmienna pomocnicza y_{ij} przyjmuje wartość jeden, gdy $y_t = j$ albo zero w pozostałych przypadkach. O składnikach losowych ε_t zakłada się, że posiadają identyczne i niezależne rozkłady o wartości oczekiwanej równej zero i nieznanym parametrze skali. Z uwagi na identyfikowalność parametrów przyjmuje się, analogicznie jak w modelu dychotomicznym, że parametr skali rozkładu zmiennej ε_t jest równy jedności. Ponadto identyfikowalność wymaga, aby $\alpha_0 = -\infty$ i $\alpha_J = +\infty$ oraz $\alpha_j \neq 0$, jeżeli w równaniu dla zmiennej z_t występuje wyraz wolny, co zakładamy w tej pracy. Zatem wymiar wektora $\boldsymbol{\alpha}$ wynosi $J-2$.

W dalszej części zakładamy, że zmienna ε_t posiada rozkład t Studenta z nieznaną liczbą stopni swobody (ν), z zerową modalną ($\mu = 0$) i jednostkową precyzją ($\tau = 1$), o funkcji gęstości

$$f_S(\varepsilon_t | \mu, \nu, \tau) = B\left(\frac{\nu}{2}, \frac{1}{2}\right)^{-1} \cdot \left(\frac{\tau}{\nu}\right)^{0.5} \cdot \left(1 + \frac{\tau}{\nu}(\varepsilon_t - \mu)^2\right)^{-0.5(\nu+1)} \quad (2)$$

gdzie $B\left(\frac{1}{2}\nu, \frac{1}{2}\right)$ jest funkcją beta. Wówczas funkcja wiarygodności modelu (1) ma postać

$$p(\mathbf{y} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \nu) = \prod_{t=1}^T \prod_{j=1}^J (p_{ij})^{y_{ij}}, \quad (3)$$

gdzie prawdopodobieństwo zaobserwowania kategorii j wynosi

$$p_{ij} \equiv \Pr(y_{ij} = 1) = \Pr(\alpha_{j-1} < z_t < \alpha_j) = F_S(\alpha_j - \mathbf{x}_t \boldsymbol{\beta}) - F_S(\alpha_{j-1} - \mathbf{x}_t \boldsymbol{\beta}), \quad (4)$$

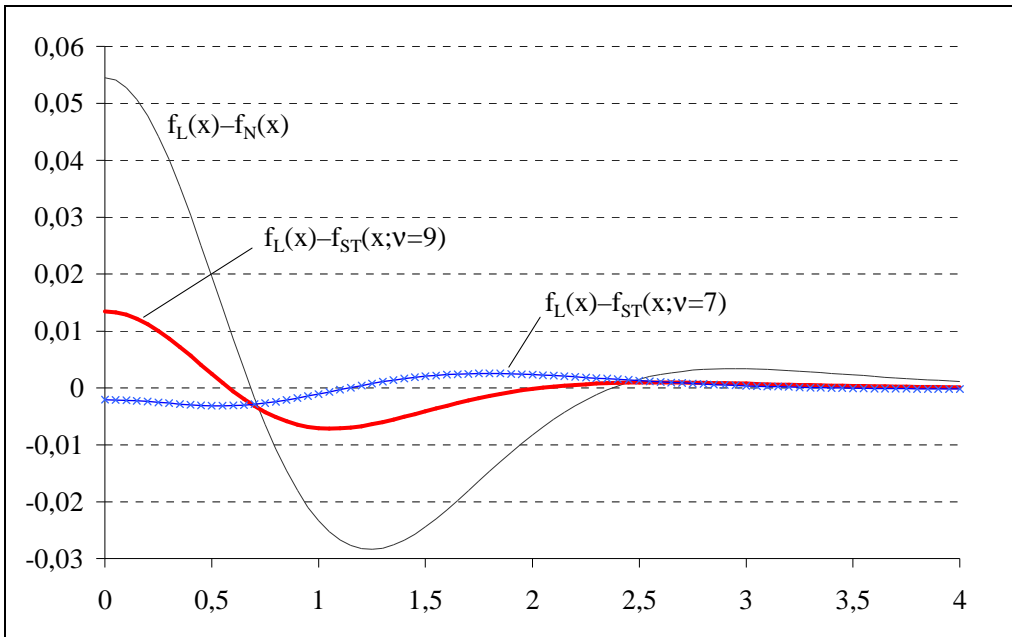
przy czym $\sum_{j=1}^J p_{ij} = 1$, zaś $F_S(a)$ jest wartością dystrybuanty zmiennej losowej ε_t o rozkładzie t Studenta z ν stopniami swobody w punkcie a . Uzyskaliśmy przez to uogólnienie standardowego modelu probitowego (z rozkładem normalnym). Wprowadzenie dodatkowego parametru ν umożliwia naturalne uwzględnienie modelu standardowego jako przypadku granicznego, gdy $\nu \rightarrow +\infty$.

Zastosowanie rodziny rozkładów t Studenta jest tym bardziej uzasadnione, że jak zauważyli Mudholkar i George [1978], kształt dystrybuanty rozkładu logistycznego jest zbliżony raczej do gęstości rozkładu t Studenta niż rozkładu normalnego. Rozkład t Studenta i logistyczny charakteryzują się grubszymi ogonami niż rozkład normalny, dla którego kurtoza (moment centralny czwartego rzędu podzielony przez kwadrat wariancji) wynosi 3. Współczynnik ekscesu (kurtoza pomniejszona o 3), mierzący grubość ogonów w stosunku do rozkładu normalnego, dla rozkładu logistycznego wynosi $6/5$, zaś w przypadku t Studenta $6/(\nu-4)$. Zatem dla $\nu=9$ współczynnik ekscesu dla obu rozkładów przyjmuje identyczną wartość. Mudholkar i George pokazują graficznie, że dla rozkładu logistycznego zmiennej standaryzowanej mniejsze różnice między dystrybuantami obserwuje się w stosunku do rozkładu t Studenta (o $\nu=9$ i jednostkowej wariancji) niż rozkładu normalnego. Albert i Chib [1993] przedstawiają wyniki własnych badań, które to potwierdzają. W uzupełnieniu dołączamy Rysunek 1 przedstawiający różnice między funkcją gęstości standaryzowanego rozkładu logistycznego, a standaryzowanym rozkładem normalnym oraz t Studenta o jednostkowej wariancji i różnej liczbie stopni swobody. Zauważmy, że najmniejsze różnice dla wartości zmiennej z przedziału od -4 do 4 występują w stosunku do rozkładu t Studenta o $\nu=7$. Wartość $\nu=7,3$ minimalizuje miarę podobieństwa między rozkładami, opartą na metryce odległości Kulbacka-Leiblera, zdefiniowaną jako $\int_{-\infty}^{+\infty} f_L(x) \cdot |\ln[f_L(x)/f_{ST}(x; \nu)]| dx$. Zatem przyjmijmy, że rozkład logistyczny można aproksymować rozkładem t Studenta o około 7–9 stopniach swobody.

W ramach modelu t Studenta istnieje prosta możliwość testowania empirycznej adekwatności dwóch specyfikacji, które najczęściej stosuje się w przypadku danych jakościowych, tj. modelu probitowego i logitowego. Dla tych modeli podstawową metodą estymacji jest metoda największej wiarygodności (MNW); zob. Aitchison i Silvey [1957] oraz McKelvey i Zavoina [1975]. Natomiast zastosowanie rozkładu t Studenta zmusza do korzystania z innej metody estymacji niż MNW, ponieważ jej własności, nawet w modelu regresji

liniowej ze składnikiem losowym o rozkładzie t Studenta z nieznanym parametrem stopni swobody, nie są poznane. Stanowi to motywację do wykorzystania małopróbkowego podejścia bayesowskiego.

Rysunek 1. Różnice między funkcją gęstości rozkładu logistycznego $f_L(x)$ a rozkładu normalnego $f_N(x)$ i t Studenta $f_{ST}(x; \nu)$ o ν stopniach swobody (rozkłady o jednostkowej wariancji)



Źródło: obliczenia własne.

W przypadku modelu (1) zakładamy, że równanie regresji z_t jest liniowe względem β . W literaturze najczęściej zakłada się, że to równanie jest także liniowe względem oryginalnych zmiennych egzogenicznych. W artykule wykorzystujemy zależność w formie wielomianu stopnia drugiego względem oryginalnych zmiennych egzogenicznych w_{th}

$$\mathbf{x}_t \boldsymbol{\beta} = \beta_1 + \sum_h \beta_h w_{th} + \sum_h \sum_{i \geq h} \beta_{hi} w_{th} w_{ti} = G(\mathbf{w}_t, \boldsymbol{\beta}). \quad (5)$$

Osiewalski i Marzec [2004] proponują więc nazywać taki przypadek modelem II rzędu, jako proste uogólnienie przypadku liniowego $\mathbf{x}_t \boldsymbol{\beta} = \beta_1 + \sum_h \beta_h w_{th}$, czyli modelu I rzędu. Z punktu widzenia estymacji, modele I i II rzędu różnią się jedynie liczbą parametrów, czyli wymiarem wektora β . W pierwszym modelu liczba parametrów wynosi $k = m+1$, zaś w modelu II rzędu wymiar wektora β jest nie większy niż $(1+m)(1+m/2)$ uwzględniając wyraz wolny, gdzie m oznacza liczbę oryginalnych zmiennych egzogenicznych. Zasadnicza przewaga drugiego modelu nad pierwszym przejawia się w charakterze efektów krańcowych, co było główną motywacją zastosowania modelu II rzędu w pracy Marzec [2003a].

Wnioskowanie o sile i kierunku wpływu zmiennych egzogenicznych \mathbf{w}_t może opierać się na efektach krańcowych zdefiniowanych dla zmiennych ciągłych jako

$$\eta_{ijh} = \frac{\partial p_{ij}}{\partial w_{th}} = -(f_s(\alpha_j - \mathbf{x}_t \boldsymbol{\beta}) - f_s(\alpha_{j-1} - \mathbf{x}_t \boldsymbol{\beta})) \frac{\partial G(\mathbf{w}_t, \boldsymbol{\beta})}{\partial w_{th}}, \quad (6)$$

gdzie $f_s(a)$ jest wartością funkcji gęstości zmiennej losowej ε_t w punkcie a . W przypadku modelu I rzędu otrzymujemy $\partial G(\mathbf{w}_t, \boldsymbol{\beta}) / \partial w_{th} = \beta_h$. Natomiast dla zmiennych zero-jedynkowych efekt krańcowy oblicza się jako różnicę prawdopodobieństw

$$\eta_{ijh} = \Pr(y_{ij} = 1; w_{th} = 1) - \Pr(y_{ij} = 1; w_{th} = 0). \quad (7)$$

Własności efektów krańcowych w obu modelach są istotnie różne. W modelu I rzędu znak efektu krańcowego dla prawdopodobieństwa zaobserwowania jednej ze skrajnych kategorii względem ustalonej zmiennej w_{th} jest identyczny dla wszystkich obserwacji i zależy od znaku parametru β_h . Ponadto iloraz

efektów krańcowych względem pary zmiennych w_{th} i w_{ti} dla ustalonej kategorii i obserwacji jest równy ilorazowi parametrów, tj. β_h/β_i . Zatem nie zależy on od wartości tych zmiennych i jest identyczny dla wszystkich obserwacji. Możemy uwolnić się od tych dwóch ograniczeń przyjmując, że z_t związane jest ze zmiennymi egzogenicznymi w_t poprzez funkcję $G(\mathbf{w}_t, \boldsymbol{\beta})$, daną wzorem (5).

Ponadto niezależną od postaci $G(\mathbf{w}_t, \boldsymbol{\beta})$ własnością efektów krańcowych jest to, że dla ustalonej obserwacji i zmiennej objaśniającej w_{th} suma efektów krańcowych η_{ijh} po j wynosi zero. W konsekwencji efekty krańcowe dla dwóch skrajnych alternatyw, tj. η_{t1h} i η_{tJh} , charakteryzują się przeciwnymi znakami; zob. Greene [2003]. O pozostałych możemy powiedzieć tylko tyle, że istnieje taka kategoria j ($j > 1$), że efekty krańcowe dla poprzedzających ją kategorii posiadają znak identyczny jak η_{t1h} i jednocześnie znak efektów dla pozostałych kategorii $j+1, j+2$ itd. jest zgodny ze znakiem efektu krańcowego dla $\Pr(y_{it} = 1)$, czyli η_{tjh} . Ta własność zachodzi, jeżeli zmienna ε_t ma rozkład jednomodalny, a takim jest rozkład t Studenta o funkcji gęstości określonej przez wzór (2).

3. Bayesowska specyfikacja i estymacja modelu

Na gruncie bayesowskim model statystyczny jest zdefiniowany poprzez łączny rozkład prawdopodobieństwa dla obserwacji i parametrów lub innych wielkości nieobserwowalnych. W tym przypadku jest on określony przez uogólnioną łączną funkcję gęstości

$$p(\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \nu) = p(\mathbf{y} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \nu) \cdot p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \nu) = \prod_{t=1}^T \prod_{j=1}^J (p_{ij})^{y_{ij}} \cdot p(\boldsymbol{\alpha}) \cdot p(\boldsymbol{\beta}) \cdot p(\nu), \quad (8)$$

jako iloczyn brzegowego rozkładu a priori $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \nu)$ i funkcji wiarygodności (3). Niezależność rozkładu a priori dla poszczególnych parametrów jest naturalnym założeniem. Ponadto wielkość p_{ij} , dana wzorem (4) jest funkcją $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ i ν .

Podstawą estymacji bayesowskiej modelu parametrycznego jest wyznaczenie z gęstości (8) dla interesującej wielkości (np. dla wybranego parametru bądź znanej funkcji parametru) gęstości rozkładu warunkowego względem obserwacji \mathbf{y} i jednocześnie brzegowego względem pozostałych składowych, czyli tzw. funkcji gęstości rozkładu a posteriori. Zagadnieniu wnioskowania bayesowskiego są poświęcone takie fundamentalne prace jak np. Zellner [1971] i O'Hagan [1994], zaś w języku polskim monografie Osiewalski [1991] i [2001].

Podjęcie bayesowskie wymaga od badacza specyfikacji rozkładu a priori na przestrzeni nieznanymi parametrów. Analizowany model jest zbyt skomplikowany, aby próbować wyznaczyć analitycznie wzorcowe rozkłady a priori w sensie Jeffreysa. W modelu z rozkładem t Studenta, w przypadku parametru ν przyjmuje się wyłącznie właściwe rozkłady a priori, określone na dziedzinie $(0, +\infty)$. Jest to konieczne, aby istniał rozkład a posteriori. Ponadto, w praktyce rozkład t Studenta o $\nu > 30$ może być przybliżany rozkładem normalnym. Przyjmując zatem niewłaściwy rozkład o priori postaci $p(\nu) \propto c \cdot I_{(0, +\infty)}(\nu)$, gdzie c jest dowolną dodatnią stałą, otrzymujemy, iż $\Pr(\nu \leq 30) / \Pr(\nu > 30) = 0$. Oznacza to, że taki rozkład jest tylko na pozór nieinformacyjny. W rzeczywistości jest rozkładem silnie informacyjnym, gdyż z prawdopodobieństwem równym jeden dopuszcza wyłącznie normalność. Właściwy rozkład a priori należy przyjąć także dla pozostałych parametrów. Albert i Chib [1993] proponując model z rozkładem t Studenta dla binarnej zmiennej endogenicznej zastosowali nieinformacyjny rozkład a priori dla $\boldsymbol{\beta}$. Przeprowadzone własne badania symulacyjne wskazały na problemy numeryczne z istnieniem właściwych rozkładów a posteriori, gdy stosuje się nieinformacyjne rozkłady a priori dla $\boldsymbol{\alpha}$ lub $\boldsymbol{\beta}$ w modelu wielomianowym z rozkładem t Studenta. O problemach z istnieniem rozkład a posteriori dla β w modelu regresji pisał Geweke [1993]. Rozważał on liniowy model regresji, przy założeniu, że składniki losowe mają niezależne rozkłady t Studenta o ustalonej

liczbie stopni swobody. Pokazał on, że jeżeli przyjmuje się nieinformacyjny rozkład a priori dla β , to wartość oczekiwana a posteriori dla β istnieje i jest skończona, gdy $\nu > 2$ oraz odchylenie standardowe a posteriori istnieje i jest skończone, gdy $\nu > 4$. Zatem, w celu zapewnienia istnienia rozkładów a posteriori i dogodnej symulacji w modelu t Studenta stosujemy informacyjną strukturę a priori postaci

$$p(\alpha, \beta, \nu) = f_N^k(\beta | \beta^*, H^{*-1}) \cdot f_{EXP}(\nu | r) \cdot \prod_{j=2}^{J-1} f_{EXP}(\alpha_j | \alpha_j^*) \cdot I_{(\alpha_{j-1}, \alpha_{j+1})}(\alpha_j). \quad (9)$$

Dla β przyjmujemy a priori k -wymiarowy rozkład normalny o wartości oczekiwanej β^* i macierzy kowariancji postaci $H^{*-1} = s_*^2 \cdot \mathbf{I}$, zaś dla α_j ($j=2, \dots, J-1$) rozkłady wykładnicze o wartości oczekiwanej i odchyleniu standardowym α_j^* . Z uwagi na restrykcję $\alpha_{j-1} < \alpha_j < \alpha_{j+1}$, stałe definiujące rozkład a priori, czyli α_j^* , dobieramy tak, aby $\alpha_{j-1}^* < \alpha_j^* < \alpha_{j+1}^*$, przy czym $\alpha_1^* = 0$. Alternatywą dla α_j jest rozkład normalny ucięty na lewo od zera. W przypadku parametru ν przyjęto, jak często czyni się to w literaturze, rozkład wykładniczy o wartości oczekiwanej i odchyleniu standardowym r . Stałe β^* , s_*^2 i α_j^* powinny być tak dobrane, aby implikowane rozkłady a priori efektów krańcowych (η_{ijk}) i prawdopodobieństw (p_{ij}), będących nieliniowymi funkcjami α , β i ν , były zgodne z posiadaną wiedzą ekonomiczną bądź odzwierciedlały jej brak. W artykule przyjęliśmy $r = 10$ i $\beta^* = \mathbf{0}$, zaś dla α_2^* i α_3^* różne zestawy wartości np. 1 i 2, 2 i 4 oraz 4 i 10. Dla s_*^2 ustaliliśmy wartość 4, jednocześnie badając wrażliwość uzyskanych wyników a posteriori na inne wartości, np. 1, 9 oraz 16. Przyjęte założenia o tych stałych uczyniły rozkłady a priori dla α i ν rozkładami słabo informacyjnymi. Porównanie rozkładów a priori i a posteriori dla η_{ijk} i p_{ij} prezentujemy w dalszej części tego artykułu.

W celu wyznaczenia rozkładów a posteriori lub ich charakterystyk całkowanie analityczne stosuje się tylko w najprostszych przypadkach; zob. Zellner [1971], Osiewalski [1991]. Skomplikowana postać funkcji wiarygodności (3) nie pozwala na to. W omawianym modelu konieczne jest całkowanie numeryczne, które realizuje się wykorzystując metody Monte Carlo. Pierwotnie Albert i Chib [1993] zaproponowali losowanie Gibbsa. Natomiast w niniejszym artykule zastosowano algorytm Metropolis i Hastingsa, który okazał się efektywną, a w stosunku do próbnika Gibbsa także szybszą metodą uzyskiwania próbek z rozkładu a posteriori. Zastosowanie algorytmu Gibbsa wymagało niesłychanie długiego czasu obliczeń, ponieważ w tym schemacie korzysta się ze zmiennych ukrytych z_t , których ilość jest równa liczbie obserwacji, a tych w prezentowanych badaniach jest ich ponad 39 tys.

Stosowanie algorytmu Metropolis i Hastingsa opiera się na jądrze uogólnionej funkcji gęstości (8). Wygodnie jest przyjąć taką parametryzację, aby zbiory dopuszczalnych wartości wszystkich parametrów były zbiorami liczb rzeczywistych, np. $\theta = [\beta' \theta_{k+1} \theta_{k+2} \dots \theta_{k+J-1}]'$, gdzie $\theta_{k+1} = \ln(\nu/r)$, $\theta_{k+h} = \ln(\alpha_h / \alpha_h^*)$ dla $h=2, \dots, J-1$, podobnie jak w pracy Osiewalski i Marzec [2004]. Wówczas jądro funkcji gęstości rozkładu a posteriori ma postać

$$p(\theta | y) \propto f_N^k(\beta | \beta^*, H^{*-1}) \cdot \exp(\theta_{k+1} - \exp(\theta_{k+1})) \times \\ \times \prod_{h=2}^{J-1} \exp(\theta_{k+h} - \exp(\theta_{k+h})) \cdot I(\theta_{k+h}) \cdot \prod_{t=1}^T \prod_{j=1}^J (p_{tj})^{y_{tj}}, \quad (10)$$

gdzie funkcja $I(\theta)$ uwzględnia te restrykcje jak $I(\alpha)$ we wzorze (9), tzn.

$$I(\theta) = \begin{cases} I_{(-\infty, \theta_{k+3} + \ln(\alpha_3^* / \alpha_2^*))}(\theta_{k+2}) \\ I_{(\theta_{k+2} + \ln(\alpha_2^* / \alpha_3^*); \theta_{k+4} + \ln(\alpha_4^* / \alpha_3^*))}(\theta_{k+3}) \\ I_{(\theta_{k+3} + \ln(\alpha_3^* / \alpha_4^*); \theta_{k+5} + \ln(\alpha_5^* / \alpha_4^*))}(\theta_{k+4}) \\ \dots \\ I_{(\theta_{k+J-2} + \ln(\alpha_{J-2}^* / \alpha_{J-1}^*); +\infty)}(\theta_{k+J-1}) \end{cases}$$

Otrzymywanie próbek $\theta_1, \theta_2, \dots, \theta_n$ z rozkładu a posteriori $p(\theta | y)$ uzyskuje się za pomocą mechanizmu losowania opisanego m.in. w pracy Osiewalski i Marzec [2004]. Szczegółowe omówienie

zastosowanych procedur Monte Carlo można znaleźć m.in. w pracach: O'Hagan [1994], Tierney [1994], Geweke [1996], Gamerman [1997] lub Koop [2003].

4. Wyniki empiryczne

4.1. Definicja zmiennych

Omówiony powyżej bayesowski model wielomianowy z rozkładem t Studenta wykorzystano do badania spłacalności kredytów detalicznych udzielonych przez polski bank komercyjny. Wcześniej te dane, pochodzące z lat 2000-2001, zostały wykorzystane w pracach Marzec [2003a,b,c], [2005] i Osiewalski, Marzec [2004]. Zbiór danych obejmował 39034 rachunków kredytowych.

Konstrukcja zmiennej wielomianowej opiera się na klasyfikacji należności pozostających do spłaty przez kredytobiorcę, określonej przez Komisję Nadzoru Bankowego. Wobec powyższego przyjęliśmy, iż zmienna objaśniana y , przyjmuje cztery wartości ($J=4$), które jednocześnie oznaczają kategorie należności lub równoważnie okres opóźnienia ze spłatą 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%

Jako potencjalne zmienne wyjaśniające ryzyko pojedynczej umowy kredytowej wprowadziliśmy (jak we wcześniejszych pracach):

- płeć (zmienna przyjmuje wartość 1, jeżeli klientem jest mężczyzna, 0 w przypadku kobiety),
- wiek kredytobiorcy (w setkach lat),
- wpływy, tzn. wielkość kwartalnych wpływów w latach 2000-2001 (w setkach tys. zł) na rachunki typu ROR kredytobiorcy w badanym banku,
- posiadanie rachunku ROR w analizowanym banku (1 – posiada, 0 – nie posiada),
- informację o tym, czy kredytobiorca posiada karty płatnicze lub kredytowe wydane przez ten bank (1 – posiada choć jedną kartę płatniczą, 0 – nie posiada),
- sposób udzielenia kredytu (1 – poprzez pośrednika kredytowego, 0 – bezpośrednio przez rozważany bank),
- typ kredytu (1 – kredyt konsumpcyjny, 0 – kredyt hipoteczny),
- okres trwania umowy kredytowej (w dziesiątkach lat),
- kwota przyznanego kredytu (w setkach tys. zł),
- waluta kredytu (1 – EUR, DEM lub USD, 0 – PLN),
- 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, przy czym za kategorię referencyjną przyjęliśmy umowę o pracę ($zrdoch1 = zrdoch2 = zrdoch3 = 0$); 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.

4.2. Porównywanie mocy wyjaśniającej modeli

W niniejszym artykule wykorzystano uogólnienie probitowego modelu wielomianowego. Zatem rodzi się pytanie: czy w świetle posiadanego materiału empirycznego to uogólnienie jest potrzebne? Który z modeli lepiej opisuje badane zjawisko?

Rozważyliśmy cztery specyfikacje, tj. model najogólniejszy z rozkładem t Studenta z nieznaną liczbą stopni swobody (ν) i z częścią regresyjną postaci wielomianu II stopnia (M_1), następnie model I rzędu z rozkładem t Studenta (M_2), model probitowy ze specyfikacją II rzędu (M_3) oraz standardowy model probitowy (M_4). Modele M_2 , M_3 i M_4 uzyskujemy poprzez warunkowanie względem β lub ν . Narzucenie odpowiednich restrykcji na parametry ν lub β w rozkładzie a priori implikuje rozkłady a priori dla poszczególnych modeli. W bayesowskim porównywaniu modeli wymaga się, aby rozkłady a priori dla parametrów swoistych w każdym modelu (a takimi są β i ν) były rozkładami właściwymi. Jest to dodatkowy argument za stosowaniem właściwych rozkładów a priori. Bayesowskie porównywanie modeli opiera się na obliczeniu prawdopodobieństw a posteriori każdego modelu wg wzoru

$$p(M_i | \mathbf{y}) = \frac{p(\mathbf{y} | M_i) \cdot p(M_i)}{\sum_{h=1}^4 p(\mathbf{y} | M_h) \cdot p(M_h)} \quad \text{dla } i \in \{1, 2, 3, 4\}, \quad (11)$$

gdzie $p(\mathbf{y} | M_i)$ i $p(M_i)$ są odpowiednio brzegową gęstością wektora obserwacji i określonym przez badacza prawdopodobieństwem a priori modelu M_i ; zob. np. Zellner 1971], O'Hagan A. [1994], Osiewalski [2001], Koop[2003].

Wyniki wskazały, iż model najogólniejszy (M_1) jest najbardziej prawdopodobny a posteriori, bez względu czy przyjmiemy jednakowe szanse a priori każdego modelu czy faworyzujemy te najprostsze, czyli najoszczędniej sparametryzowane. Prawdopodobieństwo a posteriori pozostałych modeli (M_2 , M_3 i M_4) w odniesieniu do M_1 wynosi praktycznie zero. Ponadto warto wspomnieć, że spośród dwóch możliwych rozszerzeń standardowego modelu probitowego (M_4), dane liczbowe zdecydowanie wskazują na aproksymację II rzędu (M_3) niż na wprowadzenie rozkładu t Studenta (M_2). Szczegółowe wyniki prezentuje Tabela 1. Wszystkie prezentowane dalej wyniki dotyczą modelu najogólniejszego, tzn. modelu II rzędu z rozkładem t Studenta.

Tabela 1. Brzegowe gęstości wektora obserwacji i prawdopodobieństwa a posteriori badanych modeli

Model	M_1	M_2	M_3	M_4
Liczba parametrów (k')	82	17	81	16
$\ln p(\mathbf{y} M_i)$	-21776	-22124	-21883	-22260
$p(M_i)$	0,25	0,25	0,25	0,25
$P(M_i \mathbf{y})$	1	4×10^{-152}	$2,6 \times 10^{-47}$	4×10^{-211}
$p(M_i) \propto 2^{-k_i}$	9×10^{-21}	0,3333	$1,8 \times 10^{-20}$	0,6666
$P(M_i \mathbf{y})$	1	$1,5 \times 10^{-132}$	$5,2 \times 10^{-47}$	$2,7 \times 10^{-191}$

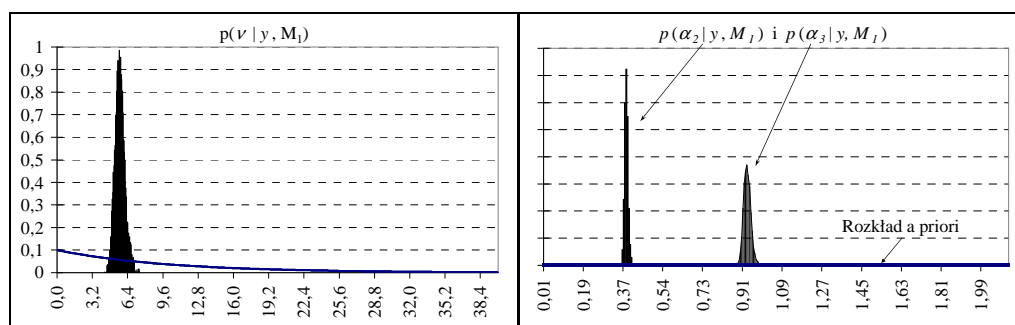
Źródło: obliczenia własne.

4.3. Wnioskowanie o wybranych parametrach modelu

Rysunek 2 przedstawia histogramy rozkładu a priori i a posteriori dla ν oraz α . Rozkłady a priori, w przeciwieństwie do rozkładów a posteriori, są bardzo rozproszone, zaś dane wnoszą nową informację w stosunku do tej zawartej w rozkładzie a priori. Rozkład a posteriori dla ν jest skupiony w wąskim paśmie 4,5-7, przy czym prawdopodobieństwo a priori, że $\nu \in (4,5; 7)$ wynosi jedynie 0,14 dla $r = 10$. Wartość oczekiwana a posteriori dla stopni swobody ν wynosi 5,7, zaś odchylenie standardowe 0,44 i te wyniki nie są wrażliwe na zmiany stałej r . Jeśliby przyjąć, że dla $\nu > 30$ rozkład t Studenta można przybliżać rozkładem

normalnym, to zaproponowana struktura a priori dopuszcza taki przypadek z prawdopodobieństwem $\Pr[\nu > 30] = 0,05$. Prawdopodobieństwo a posteriori takiego modelu wynosi praktycznie zero. Wyniki zdecydowanie odrzucają model probitowy, a nawet logitowy, gdyby przyjąć, że rozkład logistyczny może być aproksymowany przez rozkład logistyczny dla 7-9 stopni swobody. Z drugiej strony wartości $\nu \in (0; 2]$ są mało prawdopodobne a posteriori, więc rozkład składnika losowego ε_t w (1) posiada momenty pierwszego i drugiego rzędu, a nawet wyższych rzędów. Wyniki dotyczące parametru ν potwierdzają, w przypadku analizowanych danych, zasadność wprowadzonego rozkładu t Studenta. W przypadku parametrów α_2 i α_3 prawie cała masa prawdopodobieństwa a posteriori znajduje się w przedziałach $(0,35; 0,42)$ i $(0,87; 1,0)$. Wartości oczekiwane rozkładu a posteriori dla α_2 i α_3 wynoszą odpowiednio 0,377 i 0,924, zaś odchylenia standardowe 0,008 i 0,016. Tymczasem przyjmując np. $\alpha_2^* = 4$ i $\alpha_3^* = 10$ prawdopodobieństwo a priori, że $\alpha_2 \in (0,35; 0,42)$ wynosi 0,02 zaś $\Pr[\alpha_3 \in (0,87; 1,0)] = 0,01$. Wyniki a posteriori nie były wrażliwe na dobór stałych definiujących rozkład a priori badanych wielkości.

Rysunek 2. Brzegowe rozkłady a priori (linia ciągła) i a posteriori (słupki) parametru ν modelu M_1



Źródło: obliczenia własne.

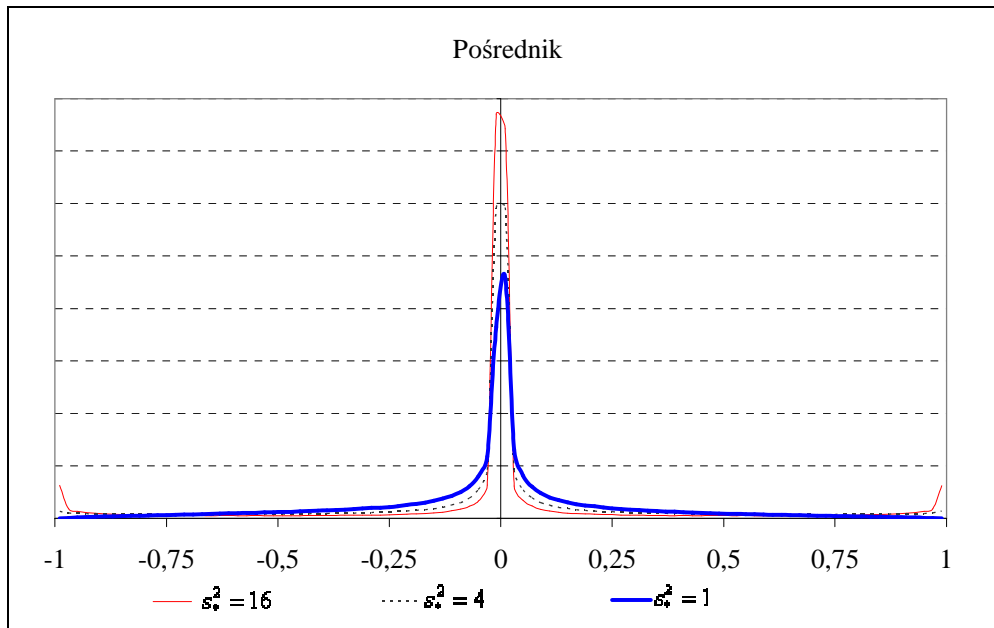
4.4. Wnioskowanie o niespłacalności kredytów

O niespłacalności kredytów detalicznych wnioskujemy w oparciu o efekty krańcowe (η_{ijk}) i prawdopodobieństwa zakwalifikowania danego kredytu do jednej z czterech kategorii należności (p_{ij}). Zagadnienie to omówimy na przykładzie dwóch wybranych klientów. Typowy kredytobiorca ma ponad 40 lat, jest mężczyzną zatrudnionym na umowę o pracę, jego wpływy na rachunek ROR wynoszą 8,5 tys. zł na kwartał, posiada karty płatnicze lub kredytowe, w banku udzielono mu złotowego kredytu konsumpcyjnego o wartości 11 tys. zł na okres 31 miesięcy. Natomiast drugi, hipotetyczny kredytobiorca – biznesmen - ma 20 lat i w odróżnieniu od typowego klienta nie posiada rachunku ROR i nie korzysta z kart płatniczych tego banku, prowadzi własną działalność gospodarczą, zaś kredyt został mu udzielony poprzez pośrednika, nie zaś bezpośrednio w banku.

Rysunki 4 i 5 przedstawiają rozkłady a priori i a posteriori efektów krańcowych w przypadku tych kredytobiorców. Porównanie obu rozkładów niesie informację o tym jak dane modyfikują wiedzę a priori o badanych wielkościach η_{ijk} . Zauważmy, że dla dwóch diametralnie różnych kredytobiorców kształt i położenie rozkładów a priori są prawie identyczne. Rozkłady a priori dla η_{ijk} są silnie skupione w otoczeniu zera, lecz charakteryzują się grubymi ogonami. Wraz ze wzrostem stałej s_*^2 rozkład a priori dla efektów krańcowych ma charakter coraz bardziej leptokurtyczny, por. Rysunek 3. Identyczne rezultaty otrzymamy przyjmując dla każdego β_h rozkład jednostajny na ustalonym przedziale, np. $(-10; 10)$, co skłania nas do przypuszczeń, że rozkład niewłaściwy indukuje trypunktowy rozkład a priori dla efektów krańcowych, tj. o największej masie prawdopodobieństwa wokół zera, a następnie w ogonach rozkładu. Dla zero-jedynkowych zmiennych objaśniających prawdopodobieństwo a priori, że $\eta_{ijk} \in (-0,25; 0,25)$ wynosi od 0,2 do 0,25 i nie zależy od s_*^2 . W przypadku zmiennych ciągłych prawdopodobieństwo, że $\eta_{ijk} \in (-0,75; 0,75)$

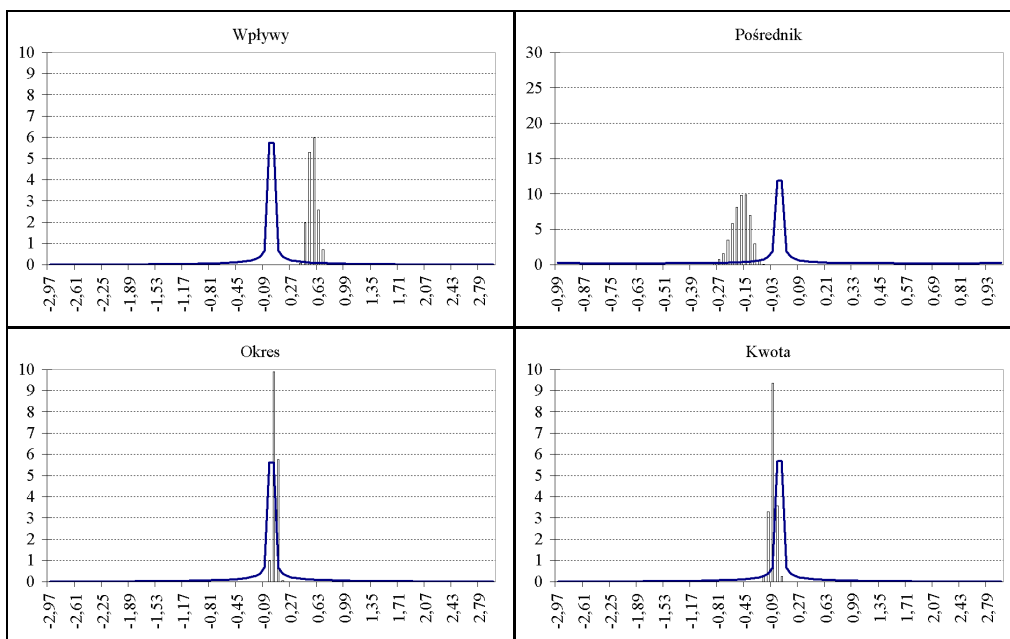
jest niższe i waha się między 0,03 a 0,05. Natomiast rozkłady a posteriori, w zależności od informacji zawartej w danych, mogą istotnie różnić się co do kształtu i położenia w stosunku do rozkładów a priori. Przykładowo, w przypadku typowego klienta położenie rozkładu a posteriori efektu krańcowego względem okresu kredytowania wskazuje, że dane nie wnoszą innej informacji niż rozkład a priori. Oznacza to brak wpływu a posteriori tej zmiennej na prawdopodobieństwo zakwalifikowania kredytu tego klienta do pierwszej kategorii należności (p_{11}). Natomiast zmienne *wpływy* i *pośrednik* silnie modyfikują informację a priori, więc rozkłady a posteriori są zlokalizowane kilka odchyleń standardowych na lewo lub prawo od zera. Następnie, zgodnie z intuicją wnioskowanie o typowym kredytobiorcy jest precyzyjniejsze niż o hipotetycznym (młodym biznesmenie), co przejawia się dużym rozproszeniem rozkładów a posteriori tego ostatniego klienta.

Rysunek 3. Rozkłady a priori efektu krańcowego względem zmiennej pośrednik w zależności od s_e^2

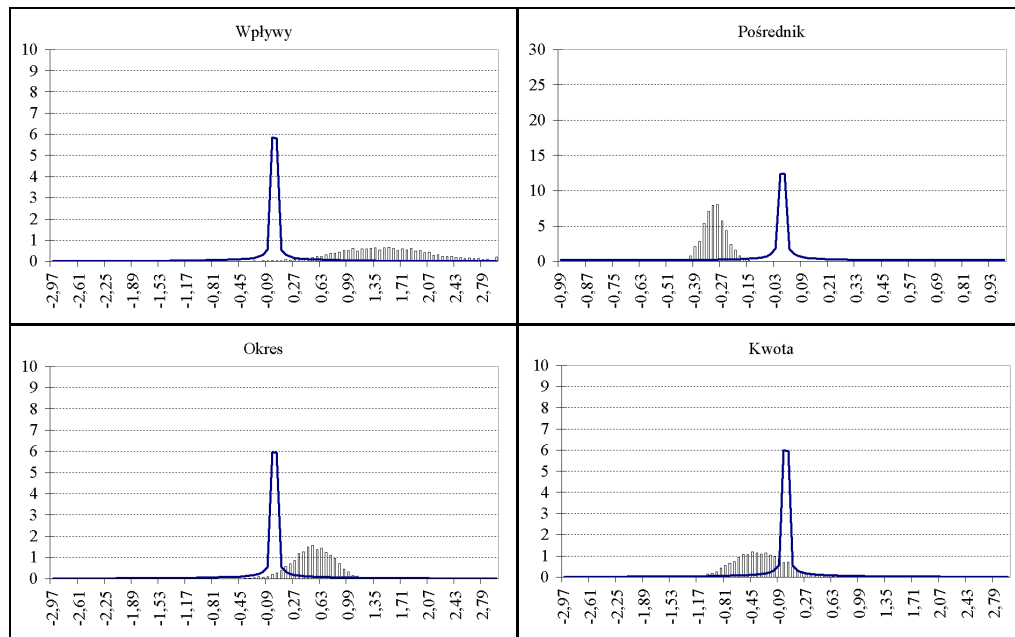


Źródło: obliczenia własne.

Rysunek 4. Brzegowe rozkłady a priori (linia ciągła) i a posteriori (słupki) efektów krańcowych η_j dla $j=1$ względem wybranych zmiennych dla typowego klienta



Źródło: obliczenia własne.

Rysunek 5. Brzegowe rozkłady a priori (linia ciągła) i a posteriori (słupki) efektów krańcowych η_j dla $j=1$ względem wybranych zmiennych dla młodego biznesmena

Źródło: obliczenia własne.

Szczegółowe charakterystyki rozkładów a posteriori efektów krańcowych w badanym modelu przedstawiają tabele 2, 3 i 4. Udzielenie kredytu konsumpcyjnego bezpośrednio przez bank ($Pośrednik = 0$) powoduje u typowego klienta wzrost o 0,17 (z błędem $\pm 0,04$) prawdopodobieństwa zakwalifikowania kredytu do pierwszej kategorii (p_{11}) oraz spadek prawdopodobieństwa średnio o 0,06 w przypadku pozostałych kategorii należności *ceteris paribus*. Wzrost jego wpływów o 1 tys. zł na kwartał powoduje wzrost p_{11} o 0,006. Udzielenie kredytu konsumpcyjnego ($typ = 1$) zamiast hipotecznego zmniejsza p_{11} o prawie 0,03. Rola pozostałych zmiennych, z wyłączeniem wieku, posiadania kart, okresu i waluty kredytu, jest istotna, ale niewielka z punktu widzenia ryzyka kredytowego.

W przypadku młodego biznesmena wpływ zmiennych *pośrednik* i *wpływy* na p_{11} jest silniejszy niż u typowego klienta. Przyznanie kredytu przez pośrednika zmniejsza p_{11} o 0,30 ($\pm 0,05$) i jednocześnie zwiększa o 0,16 ($\pm 0,03$) prawdopodobieństwo zakwalifikowania go do należności straconych (p_{14}). Szanse przynależności kredytu od pośrednika do drugiej i trzeciej kategorii wzrastają o 0,05 i 0,09. Wzrost kwartalnych wpływów o 1 tys. zł zwiększa p_{11} o 0,015 ($\pm 0,006$) i jednocześnie zmniejsza p_{14} o 0,01 ($\pm 0,005$). Dla 20-letniego klienta wiek jest dodatkowym czynnikiem ryzyka. Z roku na rok szanse uznania jego kredytu jako należności normalnych rosną o 0,006, zaś jako należności stracone maleją o 0,004. Wysoka wartość efektu względem zmiennej *karty* odzwierciedla brak zaufania banku do klienta, któremu nie udzielono kart płatniczych i kredytowych ($karty=0$) z powodu zbyt niskich lub nieregularnych wpływów bądź braku rachunku ROR. Duże znaczenie ma też waluta kredytu, tzn. kredyt w walucie obcej zwiększa p_{11} o 0,44 ($\pm 0,16$), zaś zmniejsza p_{12} o 0,10 ($\pm 0,04$) i p_{13} o 0,14 ($\pm 0,05$) oraz p_{14} o 0,20 ($\pm 0,08$). Wpływ zmiennej *typ* kredytu jest nieistotny, ale wskazuje, iż udzielenie pożyczki bądź kredytu hipotecznego, które są najczęściej w walucie obcej, raczej zwiększa prawdopodobieństwo regularnej spłaty rat i odsetek.

W uzupełnieniu Tabela 4 zawiera uśrednione po wszystkich obserwacjach wartości efektów krańcowych. Zauważmy, że jeżeli kwartalne wpływy na rachunek ROR wzrosną o 1 tys. zł (*ceteris paribus*), to prawdopodobieństwo terminowej spłaty kredytu konsumpcyjnego (p_{11}) wzrośnie średnio o prawie 0,023 ($\pm 0,003$), a prawdopodobieństwo opóźnienia w spłacie od 1 do 3 miesięcy (p_{12}) obniży się o 0,004, opóźnienia od 3 do 6 miesięcy (p_{13}) spadnie prawie o 0,007, zaś prawdopodobieństwo opóźnienia dłuższego niż 6 miesięcy (p_{14}) obniży się o 0,012 ($\pm 0,001$). Ponadto wraz z wydłużeniem o jeden rok okresu kredytowania prawdopodobieństwo terminowej spłaty rat wzrośnie o 0,02 ($\pm 0,01$), zaś szansa całkowitego

zaniechania spłaty spadnie o 0,01. Wraz ze wzrostem o 1 tys. zł kwoty udzielonego kredytu szanse zakwalifikowania go do należności normalnych maleją o 0,004 ($\pm 0,001$), zaś do należności straconych rosną o prawie 0,002 ($\pm 0,001$). Korzystanie przez kredytobiorcę z kart płatniczych i kredytowych jest przejawem większego zaufania banku, p_{11} wzrasta o 0,1 i jednocześnie p_{14} maleje o 0,05. Udzielenie kredytu poprzez pośrednika także istotnie zwiększa ryzyko kredytowe, tzn. p_{11} maleje o 0,27 ($\pm 0,01$), zaś p_{14} rośnie o 0,11 ($\pm 0,01$). Przyznanie kredytu konsumpcyjnego zamiast hipotecznego zmniejsza p_{11} o 0,18 ($\pm 0,02$) oraz zwiększa p_{12} i p_{13} o około 0,05, zaś p_{14} o 0,07. 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 typów klientów rozważanych ze względu na źródło ich dochodów. Efekty krańcowe względem zmiennych ROR i waluta kredytu charakteryzują się relatywnie dużymi odchyleniami standardowymi a posteriori szacunku, co wskazuje na brak wpływu tych zmiennych na wielkość p_{ij} .

Reasumując, w analizowanych trzech przypadkach efekty krańcowe względem tej samej zmiennej posiadają identyczne znaki (z wyłączeniem nieistotnej zmiennej ROR dla biznesmena). Znaki efektów krańcowych względem ustalonej zmiennej dla p_{12} , p_{13} i p_{14} są identyczne. Głównymi czynnikami determinującymi spłatę kredytu są typ, kwota, waluta i sposób jego przyznania oraz wpływy na rachunek ROR klienta.

Tabela 2. Wartości oczekiwane i odchylenia standardowe a posteriori efektów krańcowych w przypadku typowego klienta

Kategoria	$j = 1$ normalne		$j = 2$ poniżej standardu		$j = 3$ wątpliwe		$j = 4$ stracone	
	E(y)	D(y)	E(y)	D(y)	E(y)	D(y)	E(y)	D(y)
Płeć	-0,007	(0,003)	0,003	(0,001)	0,002	(0,001)	0,002	(0,001)
Wiek	0,013	(0,013)	-0,005	(0,006)	-0,004	(0,004)	-0,004	(0,004)
Wpływy	0,551	(0,06)	-0,226	(0,026)	-0,172	(0,02)	-0,154	(0,016)
ROR	-0,010	(0,003)	0,004	(0,001)	0,003	(0,001)	0,003	(0,001)
Karty	0,007	(0,004)	-0,003	(0,002)	-0,002	(0,001)	-0,002	(0,001)
Pośrednik	-0,173	(0,038)	0,064	(0,012)	0,056	(0,013)	0,052	(0,013)
Typ kredytu	-0,030	(0,004)	0,012	(0,002)	0,009	(0,001)	0,009	(0,001)
Okres kredytu	0,047	(0,029)	-0,019	(0,012)	-0,015	(0,009)	-0,013	(0,008)
Kwota	-0,089	(0,040)	0,036	(0,016)	0,028	(0,013)	0,025	(0,011)
Waluta	-0,087	(0,110)	0,031	(0,035)	0,028	(0,036)	0,028	(0,041)

Źródło: obliczenia własne.

Tabela 3. Wartości oczekiwane i odchylenia standardowe a posteriori efektów krańcowych w przypadku biznesmena

Kategoria	$j = 1$ normalne		$j = 2$ poniżej standardu		$j = 3$ wątpliwe		$j = 4$ stracone	
	E(y)	D(y)	E(y)	D(y)	E(y)	D(y)	E(y)	D(y)
Płeć	0,019	(0,040)	0,001	(0,002)	-0,004	(0,010)	-0,015	(0,031)
Wiek	0,557	(0,201)	0,001	(0,028)	-0,135	(0,045)	-0,423	(0,179)
Wpływy	1,478	(0,633)	-0,011	(0,076)	-0,368	(0,169)	-1,099	(0,486)
ROR	0,012	(0,077)	-0,003	(0,005)	-0,005	(0,02)	-0,004	(0,057)
Karty	0,341	(0,170)	-0,070	(0,043)	-0,109	(0,053)	-0,163	(0,082)
Pośrednik	-0,303	(0,051)	0,048	(0,014)	0,095	(0,016)	0,159	(0,029)
Typ kredytu	-0,188	(0,193)	0,037	(0,035)	0,062	(0,056)	0,088	(0,116)
Okres kredytu	0,517	(0,265)	-0,005	(0,028)	-0,129	(0,070)	-0,384	(0,201)
Kwota	-0,353	(0,335)	0,003	(0,023)	0,088	(0,086)	0,262	(0,250)
Waluta	0,440	(0,163)	-0,102	(0,041)	-0,142	(0,049)	-0,197	(0,081)

Źródło: obliczenia własne.

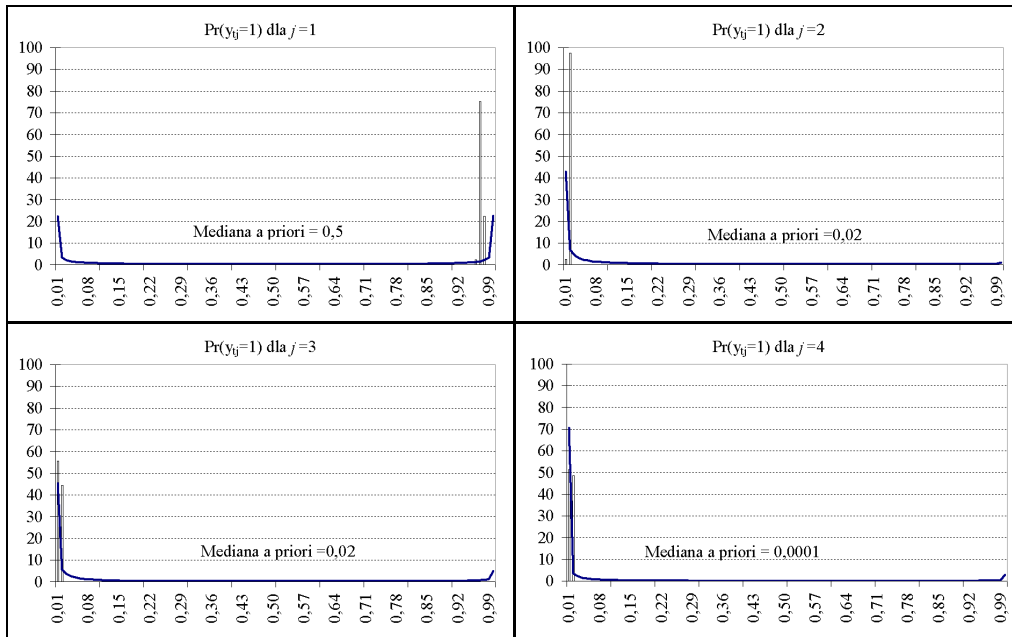
Tabela 4. Wartości oczekiwane i odchylenia standardowe a posteriori uśrednionych efektów krańcowych, $T^{-1} \sum_t \eta_{ijk}$

Kategoria	$j = 1$ normalne		$j = 2$ poniżej standardu		$j = 3$ wątpliwe		$j = 4$ stracone	
	E(y)	D(y)	E(y)	D(y)	E(y)	D(y)	E(y)	D(y)
Płeć	-0,005	(0,003)	0,001	(0,001)	0,001	(0,001)	0,003	(0,002)
Wiek	0,135	(0,025)	-0,020	(0,005)	-0,040	(0,008)	-0,075	(0,014)
Wpływy	2,291	(0,253)	-0,362	(0,036)	-0,685	(0,077)	-1,244	(0,145)
ROR	-0,008	(0,014)	0,001	(0,002)	0,001	(0,004)	0,006	(0,009)
Karty	0,108	(0,055)	-0,028	(0,015)	-0,035	(0,018)	-0,045	(0,023)
Pośrednik	-0,273	(0,012)	0,074	(0,003)	0,088	(0,004)	0,110	(0,006)
Typ kredytu	-0,178	(0,019)	0,049	(0,007)	0,056	(0,006)	0,073	(0,006)
Okres kredytu	0,205	(0,113)	-0,032	(0,018)	-0,059	(0,034)	-0,115	(0,061)
Kwota	-0,425	(0,146)	0,072	(0,024)	0,128	(0,044)	0,226	(0,079)
Waluta	0,077	(0,085)	-0,024	(0,019)	-0,027	(0,024)	-0,026	(0,044)
Zrdoch1	0,025	(0,01)	-0,007	(0,002)	-0,008	(0,003)	-0,010	(0,005)
Zrdoch2	-0,022	(0,006)	0,012	(0,002)	0,007	(0,002)	0,003	(0,003)
Zrdoch3	0,119	(0,021)	-0,039	(0,004)	-0,039	(0,006)	-0,041	(0,012)

Źródło: obliczenia własne.

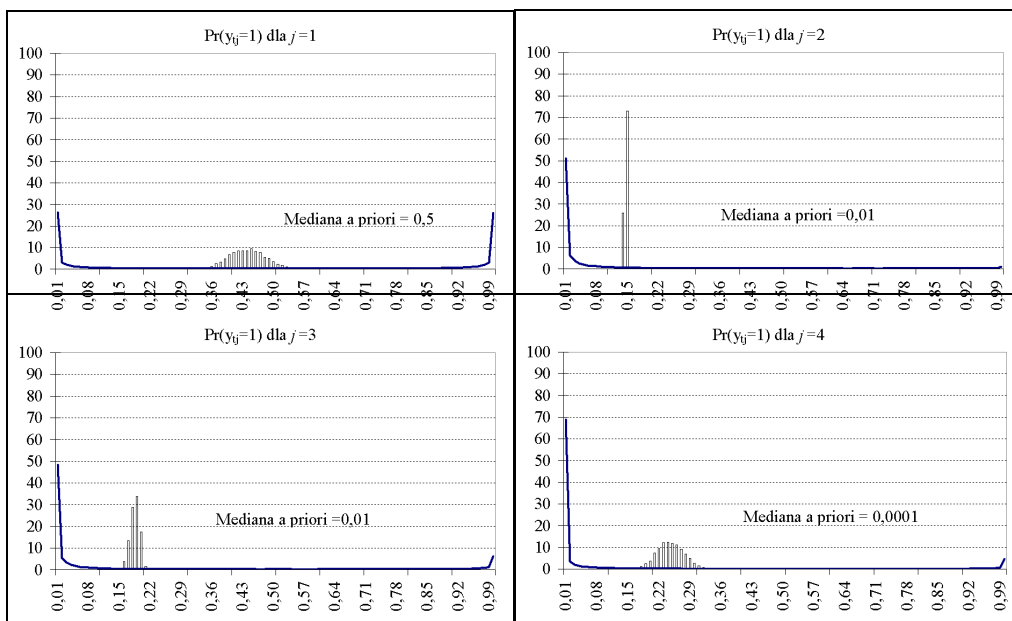
Oszacowany model możemy wykorzystać do celów prognostycznych, czyli prognozowania kategorii należności kredytu udzielonego wybranemu klientowi. Rysunki 6 i 7 przedstawiają rozkłady a priori i a posteriori prawdopodobieństw zakwalifikowania kredytu konsumpcyjnego wybranych klientów do jednej z czterech kategorii (p_{ij}). Rozkłady a priori p_{ij} dla poszczególnych klientów są do siebie bardzo zbliżone. Przyjmując $s_*^2 = 4$ (lub większe) otrzymujemy równe szanse a priori, że prawdopodobieństwo zakwalifikowania kredytu do pierwszej kategorii należności jest większe niż 0,5 albo mniejsze niż 0,5. Mediany a priori dla pozostałych p_{ij} są rzędu 0,02 lub mniej i wskazują, że pozostałe kategorie należności są mało prawdopodobne a priori. Może to pozornie wskazywać na silnie informacyjny charakter rozkładu a priori dla p_{ij} , gdy $j=2, 3, 4$. Z drugiej strony, te rozkłady są 1-kształtne, więc mediany tych rozkładów są mniejsze niż wartości oczekiwane. Rozkłady te są jednak rozproszone, na co wskazują wartości oczekiwane i odchylenia standardowe (w nawiasie): $E[p_{i1}] = 0,50 (\pm 0,44)$, $E[p_{i2}] = 0,21 (\pm 0,31)$, $E[p_{i3}] = 0,12 (\pm 0,24)$, $E[p_{i4}] = 0,17 (\pm 0,32)$. Zatem rozkłady a priori dla p_{ij} ($j=2,3,4$) mogą być akceptowalne z punktu widzenia posiadanej wiedzy o niespłacalności kredytów. Mediana rozkładu a posteriori dla p_{i1} typowego klienta wynosi 0,97. W przypadku biznesmena mediany a posteriori p_{ij} wynoszą odpowiednio 0,44, 0,14, 0,18 i 0,24 dla $j = 1, \dots, 4$. Dla obu kredytobiorców dane empiryczne silnie zmodyfikowały informację a priori o p_{ij} . Precyzja rozkładów a posteriori w stosunku do rozkładów a priori jest większa u typowego klienta niż u hipotetycznego biznesmena.

Rysunek 6. Brzegowe rozkłady a priori (linia ciągła) i a posteriori (słupki) prawdopodobieństwa zakwalifikowania kredytu typowego klienta do poszczególnych kategorii należności



Źródło: obliczenia własne.

Rysunek 7. Brzegowe rozkłady a priori (linia ciągła) i a posteriori (słupki) prawdopodobieństwa zakwalifikowania kredytu młodego biznesmena do poszczególnych kategorii należności



Źródło: obliczenia własne.

Spośród czterech rozważanych poniżej kredytobiorców najmniejsze ryzyko kredytowe związane jest ze 60-letnią klientką utrzymującą się z emerytury w kwocie 1 tysiąca złotych netto, której udzielono kredytu hipotecznego. Prawdopodobieństwo terminowej spłaty przez nią rat kapitałowo-odsetkowych jest praktycznie równe jedności, zob. Tabela 5. Największe ryzyko kredytowe związane jest z kredytem konsumpcyjnym, który został udzielony poprzez pośrednika młodemu biznesmenowi. Prawdopodobieństwo, że będzie on terminowo spłacał kredyt wynosi tylko $0,44 (\pm 0,04)$, a prawdopodobieństwo opóźnienia spłaty od jednego do 3 miesięcy wynosi 0,14, zaś opóźnienia od 3 do 6 miesięcy $0,18 (\pm 0,01)$. Natomiast prawdopodobieństwo opóźnienia dłuższego niż 6 miesięcy (czwarta kategoria należności), które powoduje obowiązek tworzenia 100% rezerw celowych, wynosi aż $0,24 (\pm 0,03)$. Istotny, negatywny wpływ korzystania przez bank z usług pośredników kredytowych na ryzyko kredytowe potwierdza analiza

typowego klienta. Jeżeli udzielono mu kredytu bezpośrednio przez bank, a zatem jego zdolność kredytowa została szczegółowo zweryfikowana przez pracownika banku, wówczas prawdopodobieństwo dotrzymania przez niego umowy jest bardzo wysokie i wynosi ponad 0,97, natomiast prawdopodobieństwo opóźnienia w spłacie dłuższego niż 6 miesięcy jest znikome – 0,01. Gdyby udzielono mu kredytu poprzez pośrednika, wówczas prawdopodobieństwo zakwalifikowania tego kredytu do pierwszej kategorii obniżyłoby się do poziomu 0,80 ($\pm 0,04$), zakwalifikowania do drugiej wynosiłoby 0,08 ($\pm 0,01$), do trzeciej prawie 0,07 ($\pm 0,01$), a prawdopodobieństwo całkowitego zaniechania spłaty kształtowałoby się na poziomie 0,06 ($\pm 0,01$). Zauważmy, że oszacowane wielkości p_{ij} dla typowego kredytobiorcy odpowiadają w przybliżeniu empirycznym udziałom ilości poszczególnych kategorii należności w badanym portfelu kredytów detalicznych.

Tabela 5. Prognozy prawdopodobieństwa zakwalifikowania kredytu (p_{ij}) wybranego klienta do poszczególnych kategorii ryzyka

Klient	Kategoria	$j = 1$	$j = 2$	$j = 3$	$j = 4$
Typowy (pośrednik=1)	E(y)	0,795	0,077	0,066	0,062
	D(y)	(0,039)	(0,013)	(0,013)	(0,013)
Typowy (pośrednik=0)	E(y)	0,967	0,013	0,010	0,010
	D(y)	(0,004)	(0,002)	(0,001)	(0,001)
Młody Biznesmen	E(y)	0,437	0,142	0,180	0,241
	D(y)	(0,042)	(0,003)	(0,011)	(0,031)
Starsza Pani	E(y)	0,991	0,003	0,002	0,003
	D(y)	(0,006)	(0,003)	(0,002)	(0,002)

Źródło: obliczenia własne.

5. Podsumowanie

Celem artykułu było przedstawienie wielomianowego modelu z rozkładem t Studenta dla kategorii uporządkowanych. W szczególności zaprezentowano bayesowską specyfikację modelu i jego estymację przy użyciu metod typu Monte Carlo. W artykule wykorzystano pewne uogólnienie modeli najczęściej stosowanych w praktyce, tj. modelu probitowego i logitowego. Uogólnienie polegało na wprowadzeniu rozkładu z szerszej klasy, tj. rozkładu t Studenta z nieznaną liczbą stopni swobody, a także na zastosowaniu w równaniu dla zmiennej ukrytej z_i wielomianu stopnia drugiego względem zmiennych egzogenicznych zamiast zależności liniowej. W przypadku posiadanego materiału empirycznego uzyskane wyniki preferują model najogólniejszy t Studenta z około 6 stopniami swobody i uzasadniają potrzebę wprowadzonego uogólnienia. Dokonana analiza wrażliwości wskazuje, iż dane empiryczne niosą dużo informacji, które modyfikują wiedzę a priori zarówno o parametrach, jak efektach krańcowych, a dobór stałych definiujących rozkłady a priori nie ma wpływu na wyniki a posteriori.

Z punktu widzenia zarządzania ryzykiem pojedynczego kredytu detalicznego wyróżniono determinanty kształtujące prawdopodobieństwo zakwalifikowania kredytu do jednej z czterech kategorii należności. Głównymi czynnikami determinującymi spłatę kredytu były typ, kwota i waluta kredytu oraz sposób jego przyznania (przez pośrednika albo bezpośrednio przez bank) i wpływy klienta.

Bibliografia

- Aitchison J., S. Silvey [1957], The Generalization of Probit Analysis to the Case of Multiple Responses, *Biometrika*, 44, s. 131-140
- Albert J., S. Chib [1993], Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, 88, s. 669-679

- Gamerman D. [1997], Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference, Chapman and Hall, London
- Geweke J. [1993], Bayesian Treatment of the Independent Student-t Linear Model, Journal of Applied Econometrics, vol. 8, s. 19-40
- Geweke J. [1996], Monte Carlo Simulation and Numerical Integration: in H. Amman, D. Kendrick and J. Rust (eds.), Handbook of Computational Economics, North-Holland, Amsterdam
- Koop G. [2003], Bayesian Econometrics, Wiley, Chichester.
- Maddala G.S. [1983], Limited Dependent and Qualitative Variables in Econometrics, Cambridge University Press, Cambridge
- 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
- Marzec J. [2003a], Bayesowska analiza modeli dyskretnego wyboru (dwumianowych), Przegląd Statystyczny, t. 50, s. 129-146
- Marzec J. [2003b], Modele wielomianowe dla kategorii uporządkowanych w badaniu niespłacalności kredytów konsumpcyjnych, w: Prognozowanie w zarządzaniu firmą (red. P. Dittmann), Prace Naukowe AE we Wrocławiu nr 1001, s.143-152
- Marzec J. [2003c], Bayesowska analiza wielomianowego modelu probitowego dla kategorii uporządkowanych, Folia Oeconomica Cracoviensia, vol. 43-44. s. 63-80
- Marzec J. [2005], Bayesowski model tobitowy z rozkładem t Studenta w analizie niespłacalności kredytów, Metody ilościowe w naukach ekonomicznych (red. A. Welfe), Wydawnictwo SGH w Warszawie, s. 147-164
- Mudholkar G., E. George [1978], A Remark on the shape of the logistic distribution, Biometrika, nr 65, s. 667-668.
- O'Hagan A. [1994], Bayesian Inference, J. Wiley, New York
- Osiewalski J. [1991], Bayesowska estymacja i predykcja dla jednorównaniowych modeli ekonometrycznych, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, Seria specjalna: Monografie, Kraków, nr 100
- Osiewalski J. [2001], Ekonometria bayesowska w zastosowaniach, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków
- 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
- Tierney L. [1994], Markov Chains for Exploring Posterior Distributions (with discussion), Annals of Statistics, 22, s. 1701-1762
- Zellner A. [1971], An Introduction to Bayesian Inference in Econometrics, J.Wiley, New York