

Jerzy Marzec

Akademia Ekonomiczna w Krakowie

ZASTOSOWANIE STANDARDOWEGO MODELU TOBITOWEGO W PROGNOZOWANIU OPÓŹNIENIA W SPŁACIE KREDYTU *

1. Wstęp

Model tobitowy jest rozszerzeniem modelu dla dychotomicznej zmiennej endogenicznej. Jego nazwa, wprowadzona przez A. Goldbergera w 1964 r., pochodzi od J. Tobina, który w 1958 r. po raz pierwszy w ekonomii zastosował model dla cenzurowanej zmiennej endogenicznej. Przegląd szerokich zastosowań modeli tobitowych przedstawia m.in. [Amemiya 1984; Greene 1993; McDonald i Moffitt 1980].

Celem niniejszego opracowania jest zastosowanie modelu tobitowego w prognozowaniu wielkości opóźnienia ze spłatą kredytów detalicznych. Omówiono jego konstrukcję i estymację oraz sposób wnioskowania i prognozowania o badanym zjawisku. W pracy są przedstawione mierniki służące ocenie dopasowania modelu tobitowego do danych. W części empirycznej zaprezentowano podstawowe wyniki badań, dotyczące m.in. prognoz okresu ewentualnej zaległości ze spłatą rat kapitałowo-odsetkowych w przypadku wybranych kredytobiorców.

2. Standardowy model tobitowy – definicja

Podstawowy model tobitowy dla dyskretno-ciągłej zmiennej y_i , której wartości są ograniczone od dołu przez β_0 , ma postać:

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

$$\begin{cases} y_t = \beta_0 & \text{gdy } z_t \leq \beta_0 \quad \text{dla } t = 1, \dots, T_0, \\ y_t = z_t & \text{gdy } z_t > \beta_0 \quad \text{dla } t = T_0 + 1, \dots, T, \\ z_t = x_t \beta + \varepsilon_t, \end{cases} \quad (1)$$

gdzie: $(\beta, \sigma^2) \in R^k \times (0, +\infty)$ – parametry modelu,

x_t – wektor-wiersz k zmiennych egzogenicznych (lub ich znane funkcje),

ε_t – zmienne losowe posiadające niezależne rozkłady normalne o zerowej wartości oczekiwanej i nieznanej wariancji σ^2

W powyższej specyfikacji, podobnie jak w pozostałych modelach dla danych jakościowych, z_t jest zmienną ukrytą, której realizacje decydują o tym, jakie obserwujemy wartości zmiennej zależnej y_t . Bez utraty ogólności, w celu zapewnienia identyfikacji parametrów przyjmuje się, że punkt ucięcia (stała wartość progowa) jest równy zeru ($\beta_0 = 0$), gdy w równaniu regresji dla z_t występuje wyraz wolny (np. $x_{t1} \equiv 1$), co zakładamy w niniejszej pracy. W literaturze powyższy model nosi nazwę standardowego modelu tobitowego. W 1984 r. T. Amemiya szacował, że ten model był wykorzystywany aż w 95% badań empirycznych opartych na modelach zmiennych cenzurowanych.

3. Estymacja modelu

Z punktu widzenia metodologii warto zauważyć, że w literaturze zwykle rozważa się model tobitowy i jego uogólnienia przy założeniu wyłącznie rozkładu normalnego dla składnika losowego ε_t . Wówczas naturalną metodą estymacji jest metoda największej wiarygodności (MNV). Łączny rozkład wektora obserwacji y i zmiennych ukrytych $z_0 = \{z_t \leq 0: t = 1, \dots, T_0\}$, warunkowy względem parametrów, ma postać

$$p(y, z_0 | \beta, \sigma^2) = \prod_{t: y_t > 0} f_N(y_t | x_t \beta, \sigma^2) \cdot \prod_{t: y_t = 0} (1 - y_t) \cdot I_{(-\infty, 0]} \cdot f_N(z_t | x_t \beta, \sigma^2), \quad (2)$$

gdzie: $f_N(\zeta | a, b)$ – funkcja gęstości rozkładu normalnego zmiennej ζ o wartości oczekiwanej a i wariancji b oraz $I_\Omega(\omega) = 1$, gdy $\omega \in \Omega$ i $I_\Omega(\omega) = 0$, jeżeli $\omega \notin \Omega$.

Uogólnioną funkcję gęstości rozkładu próbkowego dla dyskretno-ciągłej zmiennej endogenicznej y_t otrzymuje się, całkując gęstość (2) względem z_0 i otrzymując

$$p(y, z_0 | \beta, \sigma^2) = \begin{cases} \Pr(z_t \leq 0) = \Pr(\varepsilon_t \leq -x_t \beta) = \Phi(-x_t \beta / \sigma) & \text{dla } y_t = 0 \\ f_N(y_t | x_t \beta, \sigma^2) = \phi((y_t - x_t \beta) / \sigma) & \text{dla } y_t > 0 \end{cases} \quad (3)$$

gdzie: $\phi(a)$ i $\Phi(a)$ – odpowiednio wartość funkcji gęstości i dystrybucyjny standaryzowanej zmiennej o rozkładzie normalnym w punkcie a .

W efekcie funkcja wiarygodności ma postać

$$L(\beta, \sigma^2; y) = \prod_{t: y_t=0} \Phi\left(\frac{-x_t \beta}{\sigma}\right) \cdot \prod_{t: y_t>0} \phi\left(\frac{y_t - x_t \beta}{\sigma}\right). \quad (4)$$

Estymator MNW, $\hat{\theta}$, uzyskuje się poprzez numeryczne rozwiązanie, ze względu na β i σ^2 , układu następujących równań

$$\begin{aligned} \sigma^{-1} \sum_{t: y_t=0} \lambda(-x_t \beta / \sigma) \cdot x'_t - \sigma^{-2} \sum_{t: y_t>0} (y_t - x_t \beta) \cdot x'_t &= 0, \\ \sigma^{-3} \sum_{t: y_t=0} \lambda(-x_t \beta / \sigma) \cdot x_t \beta + \sigma^{-4} \sum_{t: y_t>0} (y_t - x_t \beta)^2 - \sigma^{-2} (T - T_0) &= 0, \end{aligned} \quad (5)$$

gdzie: T_0 – liczebność zbioru obserwacji cenzurowanych ($y_t = 0$),

$\lambda(a) = \phi(a)/\Phi(a)$ – funkcja hazardu lub odwrotność ilorazu Millsa, przy czym dla $a \rightarrow -\infty$ stosuje się aproksymację $\lambda(a) \approx -a$.

Rozwiązanie powyższego układu równań nieliniowych otrzymuje się, stosując iteracyjne metody optymalizacji, np. Newtona-Raphsona lub Berndta-Halla-Halla-Hausmana; zob. [Greene 1993]. Istotny wkład w rozwój tej klasy modeli miał T. Amemiya [1973], który dowiódł m.in., iż estymator $\hat{\theta} = [\hat{\beta} \ \hat{\sigma}^2]$ jest estymatorem zgodnym, o asymptotycznym rozkładzie normalnym przy założeniu homoscedastyczności i braku autokorelacji składnika losowego. Robinson [1982] udowodnił, że $\hat{\theta}_{NM}$ zachowuje te własności nawet wówczas, gdy występuje autokorelacja składników losowych. W literaturze stosuje się częściej następującą parametryzację $c = \beta/\sigma$ i $h = 1/\sigma$, która prowadzi do logarytmu funkcji wiarygodności o postaci

$$\ln L(c, h; y) = \text{Const} + \sum_{t: y_t=0} \ln \Phi(-x_t c) + T_1 \ln h - 0,5 \sum_{t: y_t>0} (hy_t - x_t c)^2 \quad (6)$$

Wówczas układ równań – warunków koniecznych na istnienie maksimum funkcji wiarygodności – ma postać

$$\begin{aligned} \left(- \sum_{t: y_t=0} \lambda(x_t c) + \sum_{t: y_t>0} (hy_t - x_t c) \right) x'_t &= 0, \\ T_1 / h - \sum_{t: y_t>0} (hy_t - x_t c) y_t &= 0, \end{aligned} \quad (7)$$

gdzie $\lambda(a)$ jest zdefiniowane jak we wzorze (5), zaś $T_1 = T - T_0$. Dla powyższej specyfikacji Olsen [1978] pokazał, że macierz drugich pochodnych cząstkowych (hesjan) logarytmu wiarygodności jest globalnie (pół)określona ujemnie. Oznacza to w praktyce, że jeżeli znajdziemy maksimum, to jest to maksimum globalne, więc estymator $\hat{\theta}$ jest określony jednoznacznie. Szczegółowe wzory dotyczące postaci analitycznej hesjanu i macierzy informacyjnej Fishera dla obu rozważanych specyfikacji można znaleźć m.in. w

pracach [Amemiya 1973; Gourieroux 2000]¹. O innych metodach estymacji modelu tobitowego i jego uogólnieniach przy standardowych założeniach piszą m.in. [Amemiya 1985 i 1984; Gourieroux 2000; Greene 1993; Maddala 1983]. Najczęściej wymieniane są algorytm Faira czy inne metody dwustopniowe, które, choć prostsze w użyciu, nie zawsze gwarantują zbieżność lub są obciążone w małych próbach. Wydaje się, że w erze szybkiego rozwoju techniki obliczeniowej MNW jest podstawową i skuteczną metodą estymacji modelu (1).

4. Wnioskowanie i predykcja na podstawie modelu

W modelu tobitowym punktem wyjścia interpretacji wyników liczbowych, czyli wnioskowania o badanym zjawisku empirycznym, a następnie predykcji, jest obliczenie bezwarunkowej i warunkowej wartości oczekiwanych zmiennej endogenicznej y_i . W standardowym modelu tobitowym warunkowa wartość oczekiwana y_i względem obserwacji powyżej progu ($y_i > 0$) wyraża się wzorem

$$E(y_i | y_i > 0) = \lambda(c_i) \cdot \sigma + x_i \beta, \quad (8)$$

natomiast bezwarunkowa wartość oczekiwana ma postać

$$E(y_i) = \Pr(y_i > 0) \cdot E(y_i | y_i > 0) = \Phi(c_i) \cdot E(y_i | y_i > 0), \quad (9)$$

gdzie $c_i = x_i \beta / \sigma$, przy czym $E(y_i | y_i > 0) > E(y_i)$. Wówczas wariancja warunkowa i bezwarunkowa zmiennej y_i wynosi

$$\begin{aligned} V(y_i | y_i > 0) &= E(y_i^2 | y_i > 0) - E(y_i | y_i > 0)^2 = \\ &= \sigma^2 (1 + c_i^2 + c_i \lambda(c_i)) - E(y_i | y_i > 0)^2 = \sigma^2 (1 - c_i \lambda(c_i) - \lambda(c_i)^2) \quad (10) \\ V(y_i) &= E(y_i^2) - E(y_i)^2 = \Phi(c_i) \cdot E(y_i^2 | y_i > 0) - \Phi(c_i)^2 \cdot E(y_i | y_i > 0)^2 \end{aligned}$$

Efekty krańcowe, które określają siłę i kierunek wpływu jednostkowej zmiany h -tej ciągłej zmiennej egzogenicznej na wartość zmiennej objaśnianej w przypadku, gdy z_i jest liniową funkcją oryginalnych zmiennych egzogenicznych, oblicza się według wzorów:

$$\frac{\partial E(y_i | y_i > 0)}{\partial x_{ih}} = \beta_h \cdot \left[1 - \frac{\lambda(c_i)}{\sigma} E(y_i | y_i > 0) \right] = \beta_h \cdot [1 - c_i \lambda(c_i) - \lambda(c_i)^2], \quad (11)$$

$$\frac{\partial E(y_i)}{\partial x_{ih}} = \beta_h \cdot \Phi(c_i). \quad (12)$$

¹ W pracy [Gourieroux 2000] występują błędy we wzorach na stronie 176-179, m.in. w formułach 7.5.1.1 i 7.5.1.2.

Korzystając z własności funkcji hazardu, można pokazać, że $[1 - \lambda(c_i) \cdot (c_i + \lambda(c_i))] \in (0,1)$, więc oba efekty krańcowe przyjmują wartości mniejsze niż wartość bezwzględna β_h . W przypadku skokowej zmiennej egzogenicznej x_{ih} efekty krańcowe liczy się jako różnicę między wartością oczekiwaną zmiennej y_i , daną wzorem (8) albo (9), obliczoną dla $x_{ih} = 1$ i $x_{ih} = 0$ przy ustalonych wartościach pozostałych zmiennych egzogenicznych.

Najprostszą modyfikacją modelu (1) jest przyjęcie innej postaci funkcyjnej dla zmiennej z_i , np. w formie wielomianu stopnia drugiego, który jest lepszą aproksymacją nieznaną zależność między z_i a zmiennymi egzogenicznymi niż funkcja liniowa². Osiewalski i Marzec [2004] proponują nazywać taki przypadek modelem II rzędu, jako proste uogólnienie przypadku liniowego, czyli modelu I rzędu (model M_1). W modelu II rzędu (M_2) zmienna z_i jest wciąż liniową funkcją β , więc nie ma jakichkolwiek komplikacji na etapie estymacji tego modelu, oprócz ewentualnej współliniowości składowych wektora x_i . Oba modele różnią się liczbą parametrów, czyli wymiarem wektora β , który w modelu M_2 wynosi maksymalnie $(1+m)(1+m/2)$, z uwzględnieniem wyrazu wolnego, gdzie m oznacza liczbę oryginalnych zmiennych egzogenicznych w_{ih} . Przyjmijmy więc, że

$$x_i \beta = \beta_1 + \sum_h \beta_h w_{ih} + \sum_h \sum_{i \geq h} \beta_{hi} w_{ih} w_{ii} = G(w_i, \beta). \quad (13)$$

Zatem w formułach (11) i (12), opisujących efekty krańcowe względem zmiennych egzogenicznych w_{ih} , pojawia się zamiast parametru β_h wyrażenie

$$\frac{\partial G(w_i, \beta)}{\partial w_{ih}} = \left(\beta_h + 2\beta_{hh} w_{ih} + \sum_{i < h} \beta_{ih} w_{ii} + \sum_{i > h} \beta_{hi} w_{ii} \right). \quad (14)$$

Wówczas efekty krańcowe względem tej samej zmiennej w_{ih} mogą mieć przeciwne znaki dla różnych obserwacji, w przeciwieństwie do modelu M_1 , w którym to mają identyczny znak jak β_h . Model M_1 jest szczególnym przypadkiem specyfikacji M_2 , więc testowanie względem siebie obu modeli sprowadza się do weryfikacji hipotezy $\beta_{hi} = 0$ dla każdego h oraz $i \geq h$. W części empirycznej przedstawiamy wyniki uzyskane na podstawie obu modeli.

5. Mierniki dopasowania modelu

Bogata literatura dotycząca modeli tobitowych jest poświęcona wielu zastosowaniom tych modeli w praktyce. Jednakże trudno znaleźć tam szerszą dyskusję na temat mierników dopasowania modelu do danych empirycznych albo choćby kryteriów, któ-

² Innym kierunkiem rozszerzenia standardowego modelu jest wprowadzenie rozkładu t -Studenta, naturalnego uogólnienia rozkładu normalnego, co zaprezentowane zostało w pracy [Marzec 2005].

rymi powinniśmy się kierować przy konstrukcji tych miar. Dyskretno-ciągły charakter zmiennej endogenicznej y_i powoduje, iż konstrukcja takich mierników jest trudna. Veall i Zimmermann [1994] dokonali przeglądu tej literatury, zwracając uwagę na tę nierozstrzygniętą kwestię. Zaprezentowali oni dziesięć mierników, zaproponowanych wcześniej w klasycznej literaturze przedmiotu, opartych m.in. na modyfikacji klasycznego współczynnika determinacji R^2 bądź ilorazu wiarygodności modelu bez restrykcji i z restrykcjami. Przeprowadzając badania symulacyjne, zbadali ich własności w zależności od udziału obserwacji cenzurowanych ($y_i = 0$) i liczebności próby. W efekcie proponują stosowanie modyfikacji miernika, zaproponowanego wcześniej przez McKelveya i Zavoina [1975] w modelu dla dychotomicznej zmiennej endogenicznej, o postaci

$$R_{MZ}^2 = \frac{\sum_{i=1}^T (\hat{z}_i - \bar{\hat{z}})^2}{\sum_{i=1}^T (\hat{z}_i - \bar{\hat{z}})^2 + T \cdot \hat{\sigma}^2}, \quad (15)$$

gdzie: $\hat{z}_i = x_i \hat{\beta}$, $\bar{\hat{z}}$ – średnia arytmetyczna dla $\hat{z}_i$.

Natomiast poddają oni krytyce mierniki wykorzystujące jedynie obserwacje powyżej progu ($y_i > 0$), jak np. współczynnik korelacji między warunkową wartością oczekiwaną zmiennej endogenicznej a jej rzeczywistą wartością powyżej progu (propozycja Dhrymesa z 1986 r.):

$$R_D^2 = \left[\text{korelacja}(\hat{y}_i^E, y_i) \right]_{i: y_i > 0}^2, \text{ gdzie } \hat{y}_i^E = \lambda(x_i \hat{\beta} / \hat{\sigma}) \cdot \hat{\sigma} + x_i \hat{\beta}. \quad (16)$$

W przypadku danych jakościowych często stosuje się współczynnik determinacji McFaddena o postaci

$$R_{MF}^2 = 1 - \frac{\ln L(\hat{\theta} | y)}{\ln L(\hat{\theta}_0 | y)}, \quad (17)$$

gdzie: $\ln L(\hat{\theta}_0 | y)$ – wartość logarytmu wiarygodności w modelu (1) z zerowymi restrykcjami dotyczącymi wszystkich parametrów oprócz β_1 ; zob. np. [Amemiya 1981]. Miernik R_{MF}^2 w przypadku modelu tobitowego, w odróżnieniu od modelu dychotomicznego czy wielomianowego dla jakościowej zmiennej endogenicznej, może przyjmować niepoprawne wartości. Może się zdarzyć, że R_{MF}^2 przyjmuje wartości większe niż jeden, gdy $\ln L(\hat{\theta} | y)$ jest dodatnie i jednocześnie $\ln L(\hat{\theta}_0 | y)$ ujemne. Ponadto badania Vealla i Zimmermanna pokazują małą przydatność współczynnika determinacji McFaddena jako miernika dopasowania oszacowanego modelu tobitowego do danych. Z uwagi na te kłopoty z konstrukcją odpowiednich miar porównawczych interesującą alternatywą jest

użycie podejścia bayesowskiego. Wówczas możliwe jest obliczenie prawdopodobieństwa *a posteriori* dla każdego z konkurencyjnych modeli próbkowych, co w przypadku tych danych, w modelu wielomianowego dla kategorii uporządkowanych, przedstawia Marzec [2006].

6. Ilustracja empiryczna – badanie opóźnienia w spłacie kredytu

Standardowy model tobitowy został wykorzystany do prognozowania opóźnienia w spłacie kredytów detalicznych, udzielonych klientom detalicznym w polskim banku komercyjnym. Zbiór danych obejmował 39 034 rachunków udzielonych w okresie od 01.01.2000 r. do 30.09.2001 r. Dane te zostały wykorzystane w pracy [Marzec 2005, 2006], a także częściowo w pracach [Marzec 2003a,b,c,d,e] oraz [Osiewalski, Marzec 2004] w przypadku modeli dwumianowych i modelu wielomianowego dla kategorii uporządkowanych.

Przyjęto, iż zmienna endogeniczna reprezentuje wielkość opóźnienia ze spłatą rat kapitałowo-odsetkowych przez kredytobiorców, jaki zaobserwowano na 30.09.2002 r. Opóźnienie wyrażone w dniach definiujemy jako różnicę między datą 30.09.2002 r. a ustaloną w harmonogramie spłaty kredytu datą ostatniej niespłaconej w całości raty kapitałowo-odsetkowej. W przypadku kredytów spłaconych w całości przyjęliśmy, że opóźnienie wynosi zero. Zatem wartość progowa dla zmiennej endogenicznej równa jest zero, przy czym udział obserwacji $y_i = 0$ wynosi 73%.

Zbiór potencjalnych zmiennych egzogenicznych wyjaśniających ryzyko pojedynczej umowy kredytowej zawierał (jak we wcześniejszych pracach):

- płeć (zmienna przyjmuje wartość 1, jeżeli klientem jest mężczyzna, 0 w przypadku kobiety),
- wiek kredytobiorcy (w setkach lat³),
- 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),

³ Jednostki ciągłych zmiennych egzogenicznych (wiek, wpływy, okres i kwota kredytu) zostały tak dobrane, gdyż pozwoliło to na etapie estymacji modelu M_2 uniknąć problemów numerycznych.

- 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 egzogeniczna, odzwierciedlająca źródło dochodu kredytobiorcy, ma charakter cechy nominalnej i odzwierciedla cztery różne przypadki. Chcąc ją uwzględnić w równaniu regresji z wyrazem wolnym, wprowadzono trzy zmienne zero-jedynkowe, przy czym kategorią referencyjną była umowa o pracę ($zrdoch1 = zrdoch2 = zrdoch3 = 0$). Jeżeli $zrdoch1 = 1$, to źródłem dochodu kredytobiorcy była renta lub emerytura, gdy $zrdoch2 = 1$, to źródłem dochodu kredytobiorcy jest własna działalność, natomiast $zrdoch3 = 1$ oznacza inne źródło dochodu. Średni wiek kredytobiorcy i kwartalne wpływy na rachunek ROR wynoszą odpowiednio 40 lat i 8,5 tys. zł. Przeciętna wartość udzielonego kredytu kształtuje się na poziomie 10,8 tys. zł, okres kredytowania zaś wynosi ponad 2,5 roku.

Podstawowe wyniki estymacji w postaci ocen MNW i błędy średnie szacunku dla parametrów podstawowego modelu M_2 przedstawia tab. 1. Na poziomie istotności 0,05 i większym przynajmniej 40% wszystkich ocen parametrów jest statystycznie istotna, w tym przy kwadratach zmiennych ciągłych, jak okres i kwota kredytu. Zastosowane w pracy rozszerzenie modelu M_1 , polegające na wprowadzeniu wielomianu stopnia drugiego dla z_t , jest przedmiotem testowania. Wartość logarytmu wiarygodności w modelu M_2 wynosi -84356,8, w modelu M_1 z restrykcjami ($\beta_{hi} = 0$) kształtuje się na poziomie -84731,4, a w modelu jedynie z wyrazem wolnym -91606,5. Wartość klasycznej statystyki opartej na ilorazie wiarygodności (*likelihood ratio test*; LR) dla testu redukcji modelu M_1 do M_2 jest wysoka i wynosi ponad 749⁴. Wartość statystyki χ^2 dla 65 stopni swobody i poziomu istotności 0,01 wynosi 94,4. Zatem wynik testu LR faworyzuje model M_2 , więc zastosowane rozszerzenie w świetle posiadanych danych było zasadne. Tabela 1 zawiera wartości mierników dopasowania obu modeli. Współczynnik determinacji McFaddena (R_{MF}^2) i kwadrat współczynnika korelacji (R_D^2) przyjmują bardzo niskie i prawie identyczne wartości. Natomiast zmodyfikowany współczynnik McKelveya i Zavoina różnicuje oba modele. W modelu M_2 R_{MZ}^2 jest wysoki i wynosi 0,99, natomiast w modelu M_1 kształtuje się na poziomie niższym 0,51. Ocena MNW dla σ^2 w modelu M_2 wynosi 111 977, zaś w M_1 przyjmuje wartość większą, tj. 114 331. Te wyniki także potwierdzają zasadność modelu M_2 .

⁴ $LR = -2[\ln L(\theta_R|y) - \ln L(\theta|y)]$, gdzie $\ln L(\theta_R|y)$ jest wartością logarytmu wiarygodności dla modelu z restrykcjami. Jeżeli $T \rightarrow \infty$, statystyka LR ma rozkład χ^2 o liczbie stopni swobody równej liczbie restrykcji, zob. np. [Greene 1993].

Tabela 1. Oceny MNW parametrów modelu M_2

Zmienna	Parametr	Ocena	Błąd szacunku	Zmienna	Parametr	Ocena	Błąd szacunku
1	β_1	-383,527	69,093	$w_3 \cdot w_{10}$	β_{41}	-22,459	81,193
Plec (w_1)	β_2	-0,135	59,399	$w_3 \cdot w_{11}$	β_{42}	-881,856	402,270
Wiek (w_2)	β_3	-367,745	160,189	$w_3 \cdot w_{12}$	β_{43}	225,912	62,941
Wpływy (w_3)	β_4	-540,028	87,072	$w_3 \cdot w_{13}$	β_{44}	-249,303	287,528
ROR (w_4)	β_5	-75,341	22,514	$w_4 \cdot w_5$	β_{45}	52,010	121,183
Karty (w_5)	β_6	-178,080	126,647	$w_4 \cdot w_6$	β_{46}	110,656	33,174
Pośrednik (w_6)	β_7	512,492	32,131	$w_4 \cdot w_8$	β_{47}	-147,567	55,764
Typ kredytu (w_7)	β_8	-3,213	53,865	$w_4 \cdot w_9$	β_{48}	100,928	72,236
Okres (w_8)	β_9	207,265	124,519	$w_4 \cdot w_{10}$	β_{49}	302,214	103,391
Kwota (w_9)	β_{10}	49,727	124,861	$w_4 \cdot w_{11}$	β_{50}	-4,882	35,171
Waluta (w_{10})	β_{11}	-295,863	172,131	$w_4 \cdot w_{12}$	β_{51}	-20,714	40,291
Zrdoch1 (w_{11})	β_{12}	142,477	54,812	$w_4 \cdot w_{13}$	β_{52}	-154,890	86,591
Zrdoch2 (w_{12})	β_{13}	276,637	105,999	$w_5 \cdot w_6$	β_{53}	66,327	35,889
Zrdoch3 (w_{13})	β_{14}	54,721	116,837	$w_5 \cdot w_8$	β_{54}	34,579	38,965
$w_1 \cdot w_2$	β_{15}	-75,061	48,616	$w_5 \cdot w_9$	β_{55}	-22,660	40,870
$w_1 \cdot w_3$	β_{16}	-59,610	59,141	$w_5 \cdot w_{10}$	β_{56}	4,717	59,825
$w_1 \cdot w_4$	β_{17}	32,269	21,280	$w_5 \cdot w_{11}$	β_{57}	35,045	36,383
$w_1 \cdot w_5$	β_{18}	22,460	18,422	$w_5 \cdot w_{12}$	β_{58}	35,710	28,204
$w_1 \cdot w_6$	β_{19}	-24,265	18,009	$w_5 \cdot w_{13}$	β_{59}	87,364	56,476
$w_1 \cdot w_7$	β_{20}	73,770	51,791	$w_6 \cdot w_8$	β_{60}	-506,891	58,300
$w_1 \cdot w_8$	β_{21}	-68,070	30,493	$w_6 \cdot w_9$	β_{61}	29,088	114,074
$w_1 \cdot w_9$	β_{22}	37,248	42,449	$w_6 \cdot w_{10}$	β_{62}	-29,795	27,431
$w_1 \cdot w_{10}$	β_{23}	80,261	63,428	$w_6 \cdot w_{12}$	β_{63}	-134,184	38,404
$w_1 \cdot w_{11}$	β_{24}	-37,077	16,395	$w_6 \cdot w_{13}$	β_{64}	-113,067	88,582
$w_1 \cdot w_{12}$	β_{25}	-56,586	23,497	$w_7 \cdot w_8$	β_{65}	344,216	83,267
$w_1 \cdot w_{13}$	β_{26}	-71,739	43,135	$w_7 \cdot w_9$	β_{66}	82,437	77,098
$(w_2)^2$	β_{27}	24,604	189,668	$w_7 \cdot w_{12}$	β_{67}	-107,115	79,708
$w_2 \cdot w_5$	β_{28}	164,632	79,896	$(w_8)^2$	β_{68}	-121,577	37,813
$w_2 \cdot w_6$	β_{29}	57,618	63,665	$w_8 \cdot w_9$	β_{69}	-159,772	41,079
$w_2 \cdot w_8$	β_{30}	576,276	147,160	$w_8 \cdot w_{10}$	β_{70}	145,094	80,313
$w_2 \cdot w_9$	β_{31}	182,321	192,412	$w_8 \cdot w_{11}$	β_{71}	27,115	52,101
$w_2 \cdot w_{10}$	β_{32}	-790,395	260,712	$w_8 \cdot w_{12}$	β_{72}	19,522	49,552
$w_2 \cdot w_{11}$	β_{33}	-254,595	90,133	$w_8 \cdot w_{13}$	β_{73}	-591,273	89,448

Zmienna	Parametr	Ocena	Błąd szacunku	Zmienna	Parametr	Ocena	Błąd szacunku
$w_2 \cdot w_{12}$	β_{34}	-68,051	113,501	$(w_9)^2$	β_{74}	-55,039	7,075
$w_2 \cdot w_{13}$	β_{35}	354,210	206,866	$w_9 \cdot w_{10}$	β_{75}	289,140	68,429
$(w_3)^2$	β_{36}	9,314	9,197	$w_9 \cdot w_{11}$	β_{76}	142,808	78,554
$w_3 \cdot w_5$	β_{37}	68,887	70,681	$w_9 \cdot w_{12}$	β_{77}	-11,084	41,789
$w_3 \cdot w_6$	β_{38}	-2021,780	499,242	$w_9 \cdot w_{13}$	β_{78}	799,839	115,016
$w_3 \cdot w_8$	β_{39}	254,182	56,287	$w_{10} \cdot w_{12}$	β_{79}	-239,901	78,840
$w_3 \cdot w_9$	β_{40}	-4,631	26,328	—	σ^2	111977	1793

Źródło: obliczenia własne.

Tabela 2. Wartości wybranych wskaźników dopasowania modelu

Miernik / Model	M_1	M_2
R_{MZ}^2	0,505	0,996
R_D^2	0,00395	0,00398
R_{MF}^2	0,075	0,079

Źródło: obliczenia własne.

Praktycznym sposobem wykorzystania modelu tobitowego jest prognozowanie opóźnienia spłaty kredytu dla wybranych klientów. Rozważmy cztery hipotetyczne sylwetki kredytobiorców: dwóch o cechach najczęstszych (dotyczy zmiennych jakościowych) lub przeciętnych (dla zmiennych ciągłych) w badanej zbiorowości, różniących się jedynie sposobem udzielenia kredytu konsumpcyjnego, oraz dwóch innych o cechach tak przeciwstawnie skonstruowanych, że charakteryzują się oni najmniejszym i największym ryzykiem kredytowym⁵.

W modelu M_2 , w przypadku typowego kredytobiorcy udzielenie mu kredytu poprzez pośrednika zamiast bezpośrednio przez bank skutkuje wzrostem oczekiwanego opóźnienia z 19 do 96 dni, czyli o trzy miesiące. Jeżeli pojawiłyby się jakiegokolwiek problemy ze spłatą kredytu ($y_i > 0$), a kredyt został udzielony bezpośrednio w banku, to opóźnienie wyniesie aż 163 (5,5 miesiąca). Opóźnienie wydłuży się o kolejne 66 dni (2 miesiące) i przekroczy 7,5 miesiąca, gdy kredyt zostanie sprzedany z pomocą pośrednika. W przypadku młodych, dwudziestoletnich klientów prowadzących własną działalność gospodarczą, którzy zaciągając kredyt konsumpcyjny, nie posiadają kart płatniczych ani rachunku ROR w badanym banku, spodziewane opóźnienie spłaty kredytu wynosi 190 dni, a w sytuacji, gdy pojawiają się problemy z jego spłatą ($y_i > 0$), wzrasta do 307 dni (ponad 10 miesięcy).

⁵ Szczegółowa charakterystyka hipotetycznych kredytobiorców przedstawiona została w innych pracach, zob. [Marzec 2003c,e; Osiewalski, Marzec 2004].

Najmniejszym ryzykiem charakteryzuje się sześćdziesięcioletnia klientka banku, utrzymująca się z emerytury i korzystająca z szerokiej oferty produktów banku, która spłaca kredyt hipoteczny. Aczkolwiek w sytuacji, gdy zarówno ona, jak i typowy kredytobiorca spłacają nieregularnie raty kredytowo-odsetkowe, spodziewamy się średnio 5-miesięcznego opóźnienia ze spłatą, co powoduje, że bank jest zobowiązany do tworzenia 50% rezerw od kwoty należności przeterminowanych pomniejszonych o zabezpieczenia. Przy opóźnieniu powyżej 6 miesięcy rezerwa celowa wynosi 100%. Oba modele zgodnie prognozują skalę ryzyka kredytowego wybranych czterech klientów. Możliwość prognozowania opóźnienia spłaty kredytu, a tym samym określenia ewentualnych kosztów utraconych korzyści (związanych z rezerwami celowymi), daje przewagę modelom tobitowym nad modelami dla dychotomicznej zmiennej endogenicznej, które są często stosowane w tego typu analizach.

Tabela 3. $E(y_i | y_i > 0)$ i $E(y_i)$ w przypadku wybranych kredytobiorców (błędy średnie szacunku w nawiasach)

Model	M_1		M_2	
	$E(y_i y_i > 0)$	$E(y_i)$	$E(y_i y_i > 0)$	$E(y_i)$
Kredytobiorca				
Typowy (pośrednik = 1)	256	116	239	96
Typowy (pośrednik = 0)	153	12	163	19
Młody biznesmen	363	266	307	190
Starsza pani	125	3	141	7

Źródło: obliczenia własne.

Tabela 3 prezentuje oszacowane dla czterech kredytobiorców wartości efektów krańcowych. W przypadku typowego kredytobiorcy wpływ potencjalnych czynników na opóźnienie ze spłatą kredytu, z wyjątkiem zmiennej *pośrednik*, jest niewielki. Natomiast w przypadku młodego przedsiębiorcy zaproponowane determinanty mają o wiele większy wpływ. Wzrost jego kwartalnych wpływów na ROR o tysiąc złotych zmniejsza opóźnienie średnio o 10-14 dni. W hipotetycznym przypadku, gdyby udzielono mu kredytu w walucie obcej (np. kredytu hipotecznego zamiast konsumpcyjnego), to ewentualne opóźnienie w spłacie obniży się o co najmniej 153 dni (5 miesięcy). Udzielenie kredytu na okres o 1 rok dłuższy zmniejsza potencjalne opóźnienie o 11-16 dni. Podsumowując, spośród potencjalnych czynników wpływających na wzrost wielkości opóźnienia w spłacie kredytu decydujący wpływ mają przede wszystkim: udzielenie kredytu poprzez pośrednika zamiast bezpośrednio w banku, niewielki poziom wpływów na rachunek ROR, udzielenie kredytu konsumpcyjnego zamiast hipotecznego, w mniejszym stopniu zaś młody wiek kredytobiorcy bądź krótki okres kredytowania.

Tabela 4. Efekty krańcowe dla $E(y_i | y_i > 0)$ i $E(y_i)$ w przypadku wybranych kredytobiorców w modelu M_2

Kredytobiorca	Typowy (pośrednik = 1)		Typowy (pośrednik = 0)		Młody biznesmen		Starsza pani	
	$E(y_i)$	$E(y_i y_i > 0)$	$E(y_i)$	$E(y_i y_i > 0)$	$E(y_i)$	$E(y_i y_i > 0)$	$E(y_i)$	$E(y_i y_i > 0)$
Płeć	13	10	6	10	-24	-16	-2	-6
Wiek w setkach lat	-48	-38	-20	-31	-122	-86	-14	-36
Wpływy w setkach tys. zł	-1004	-781	-54	-83	-1405	-987	-68	-180
ROR	15	12	-10	-13	12	9	-3	-8
Karty	18	13	-3	-4	-8	-5	1	2
Pośrednik	78	76	78	76	114	84	128	127
Typ Kredytu	53	46	13	26	36	26	7	14
Okres kredytu w dziesiątkach lat	153	119	43	67	165	116	27	70
Kwota kredytu w setkach tys. zł	28	22	8	12	14	10	13	35
Waluta	-51	-44	-12	-25	-176	-153	-7	-41

Źródło: obliczenia własne.

7. Podsumowanie

W niniejszym opracowaniu przedstawiono specyfikację, estymację, wnioskowanie i prognozowanie na podstawie standardowego modelu tobitowego. Klasyczny test ilorazu wiarygodności wyraźnie wskazuje, iż proste uogólnienie, poprzez aproksymację w postaci wielomianu drugiego stopnia dla z_i , w przypadku posiadanych danych empirycznych było uzasadnione. Z punktu widzenia zarządzania ryzykiem pojedynczego kredytu przedstawiono prognozy opóźnienia spłaty kredytu oraz informacje o spodziewanej sile i kierunku oddziaływania wybranych czynników egzogenicznych na wielkość opóźnienia w przypadku wybranych kredytobiorców. Uzyskane wyniki niosą nową informację o ryzyku kredytowym w stosunku do rezultatów uzyskanych na podstawie modeli dla dychotomicznej lub wielomianowej zmiennej endogenicznej, które autor rozważał we wcześniejszych pracach.

Literatura

- Amemiya T., *Regression Analysis when the Dependent Variable is Truncated Normal*, „Econometrica” 1973, vol. 41, nr 6, s. 997-1016.
- Amemiya T., *Qualitative Response Models: A Survey*, „Journal of Economic Literature” 1981, vol. 19, s. 1483-1536.
- Amemiya T., *Tobit Models: A Survey*, „Journal of Econometrics” 1984, vol. 24, 3-61.
- Amemiya T., *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts 1985.
- Arabmazar A., P. Schmidt, *Further Evidence on the Robustness of the Tobit Estimator to Heteroscedasticity*, „Journal of Econometrics” 1981, 17, s. 253-258.
- Arabmazar A., P. Schmidt, *An Investigation of the Robustness of the Tobit Estimator to Non-Normality*, „Econometrica” 1982, vol. 50, nr 4, s. 1055-1063.
- Goldberger A.S., *Abnormal Selection Bias*, [in:] S. Karlin, T. Amemiya, L. Goodman, *Studies in Econometrics, Time Series, and Multivariate Statistics*, s. 67-84, New York, Academic Press 1983.
- Gourieroux C., *Econometrics of Qualitative Dependent Variables*, Cambridge University Press, Cambridge 2000.
- Greene W.H., *Econometric Analysis*, Macmillan Publishing Company, New York 1993.
- Hurd M., *Estimation in Truncated Samples When There is Heteroscedasticity*, „Journal of Econometrics” 1979, 11, s. 247-258.
- Maddala G.S., *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge 1983.
- Marzec J., *Badanie niewypłacalności kredytobiorcy na podstawie modeli logitowych i probitowych*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, nr 628, Kraków 2003a, s. 103-117.
- Marzec J., *Badanie niespłacalności kredytów za pomocą bayesowskich modeli dychotomicznych - założenia i wyniki*, [w:] *Metody ilościowe w naukach ekonomicznych*, red. A. Welfe, Wydawnictwo SGH, Warszawa 2003b, s. 73-86.
- Marzec J., *Bayesowska analiza modeli dyskretnego wyboru (dwumianowych)*, „Przegląd Statystyczny” 2003c, t. 50, s. 129-146.
- Marzec J., *Modele wielomianowe dla kategorii uporządkowanych w badaniu niespłacalności kredytów konsumpcyjnych*, [w:] *Prognostowanie w zarządzaniu firmą*, red. P. Dittmann, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu* nr 1001, AE, Wrocław 2003d, s. 143-152.
- Marzec J., *Bayesowska analiza wielomianowego modelu probitowego dla kategorii uporządkowanych*, *Folia Oeconomica Cracoviensia* 2003e, vol. 43-44, s. 63-80.
- Marzec J., *Bayesowski model tobitowy z rozkładem t-Studenta w analizie niespłacalności kredytów*, [w:] *Metody ilościowe w naukach ekonomicznych*, red. A. Welfe, Wydawnictwo SGH, Warszawa 2005, s. 147-164.
- Marzec J., *Bayesowski model wielomianowy z rozkładem t-Studenta dla kategorii uporządkowanych*, [w:] *Metody ilościowe w naukach ekonomicznych*, red. A. Welfe, Wydawnictwo SGH, Warszawa (w druku).
- McKelvey R.D., W. Zavoina, *A Statistical Model for the Analysis of Ordinary Level Dependent Variables*, „Journal of Mathematical Sociology” 1975, 4, s. 103-120.
- McDonald J.F., R. A. Moffitt, *The Uses of Tobit Analysis*, „The Review of Economics and Statistics” 1980, vol. 62, s. 318-321.
- Olsen R.J., *Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model*, „Econometrica” 1978, vol. 46, nr 5, s. 1211-1209.
- Osiewalski J., J. Marzec, *Uogólnienie dychotomicznego modelu probitowego z wykorzystaniem skośnego rozkładu Studenta*, „Przegląd Statystyczny” 2004, t. 51, s. 13-24.
- Tobin J., *Estimation of Relationships for Limited Dependent Variables*, „Econometrica” 1958, vol. 26, nr 1, s. 24-36.
- Veall M.R., K.F. Zimmerman, *Goodness of Fit Measures in the Tobit Model*, „Oxford Bulletin of Economics and Statistics” 1994, vol. 56, nr 4, s. 485-499.

APPLICATION OF STANDARD TOBIT MODEL TO FORECAST OF DELAY IN PAYING OFF LOANS

Summary

The aim of this study is to present a use of the tobit model to examine how debtors pay off debt consumer loans. A explained variables represents a length of a delay in paying off loan and thus the dependent variable is zero for a significant fraction of the observations. Therefore the use tobit model appeared to be relevant for purposes of this study. We show interesting results of the empirical research using the data from a large Polish commercial bank. In a case of various clients we predict the delay in credit loan payments which depends on eleven explanatory variables such as: sex, age, size and source of client's income, loan amount and period, type and currency 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).