PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

JADWIGA KOSTRZEWSKA

PRZEGLĄD SEMIPARAMETRYCZNYCH METOD ESTYMACJI MODELI TOBITOWYCH TYPU I–III

1. WPROWADZENIE

Modele tobitowe rozwijają się od 1958 roku, kiedy to po raz pierwszy James Tobin w pracy [45] rozważył model regresji dla zmiennej zależnej cenzurowanej. Model ten nazwano modelem tobitowym od nazwiska jego twórcy oraz ze względu na pewne pokrewieństwo z modelem probitowym. W kolejnych latach następował rozwój modeli zmiennej zależnej ograniczonej (*limited dependent variable models*), tzn. modeli zmiennej zależnej uciętej (*truncated models*), cenzurowanej (*censored models*) lub podlegającej nieprzypadkowej selekcji[1] (*sample-selection models*, *selection models*). Takeshi Amemiya w pracy [1] dokonał przeglądu istniejących modeli i usystematyzował je ze względu na postać funkcji wiarygodności. W rezultacie wyodrębnił pięć głównych typów modeli tobitowych[2].

W celu estymacji parametrów w modelach tobitowych typu I–III (lub wyższego typu) na ogół stosuje się metodę największej wiarygodności. Metoda ta jednak wymaga założenia o normalności rozkładu składnika losowego. Metodę można stosować także w modelach tobitowych przy założeniu heteroskedastyczności składników losowych. W modelach z rodziny tobitowych o funkcji wiarygodności posiadającej więcej niż jedno ekstremum można stosować także dwustopniową metodę zaproponowaną przez J. Heckmana w pracy [11], która wymaga słabszych założeń. Metoda ta w pierwszym kroku wykorzystuje analizę probitową, zaś w drugim – korzysta z klasycznej metody najmniejszych kwadratów. Obecnie stosuje się ją zazwyczaj do wyliczenia wartości początkowych estymatorów zgodnych w procedurze iteracyjnej wyznaczania estymatorów największej wiarygodności.

[1] Zmienne ucięte (ang. *truncated variables*) to zmienne, których wartości są obserwowalne tylko w pewnym przedziale, a nie są brane pod uwagę wartości spoza niego. Zmienne cenzurowane (ang. *censored variables*) to takie zmienne, których dokładne wartości są znane, gdy występują w pewnym przedziale, a nie są znane te wartości, które są spoza niego i wówczas zmiennej zostaje przypisana umowna wartość z brzegu przedziału. Jeżeli to, czy wartości badanej zmiennej są obserwowane czy nie, zależy od wartości innej skorelowanej z nią zmiennej, wówczas występuje tzw. nieprzypadkowa selekcja (selekcja, samoselekcja, ang. *self-selection, sample selection*), będąca uogólnieniem ucięcia lub cenzurowania. Pojęcie zmiennej uciętej, cenzurowanej i podlegającej nieprzypadkowej selekcji omówiono np. w pracach [21], [22].

[2] Przegląd modeli tobitowych można odnaleźć we wspomnianej pracy [1] oraz [9], [22], [24].

Gdy nie ma żadnych przesłanek odnośnie rozkładu składnika losowego modelu zmiennej zależnej cenzurowanej lub, gdy w wyniku przeprowadzenia odpowiedniego testu założenie o rozkładzie normalnym zostanie odrzucone, odpowiednim narzędziem do oszacowania modelu mogą okazać się metody semiparametryczne. Metody te wymagają słabszych założeń o rozkładzie, przy których można uzyskać estymatory o dobrych własnościach, tj. zgodne, asymptotycznie normalne. Niektóre z metod nie wymagają założenia o homoskedastyczności składnika losowego.

W literaturze anglojęzycznej metody semiparametryczne dla modeli zmiennej zależnej cenzurowanej (modeli tobitowych) omawiane są od szeregu lat. W artykule zaprezentowano niektóre metody estymacji semiparametrycznej oraz wskazano na ich przydatność przy weryfikacji założeń narzuconych na rozkład składnika losowego. Skupiono uwagę na metodach odnoszących się do modeli tobitowych typu I, II oraz III zgodnie z klasyfikacją T. Amemiyi (por. [1]).

2. MODELE TOBITOWE TYPU I–III

W celu prezentacji poszczególnych metod wprowadzono następujące oznaczenia[3]. Niech Y^* oraz Z^* będą zmiennymi ukrytymi (*latent variable*), tzn. zmiennymi, których wartości zazwyczaj nie są obserwowalne lub są obserwowalne w zależności od pewnych warunków (kryterium obserwowalności). Przez n oznaczono liczbę elementów w próbie, przy czym dla wygody zapisu przyjęto, że pierwszych n_1 elementów ($i = 1, \ldots, n_1$) w próbie odnosi się do dokładnie zaobserwowanych wartości zmiennej Y^*, natomiast następne $(n - n_1)$ elementów ($i = (n_1 + 1), \ldots, n$) – do przypisanych wartości umownych (o ile istnieją).

Model tobitowy typu I jest postaci[4] (por. [24]):

$$y_i = \begin{cases} y_i^* & \text{gdy } y_i^* > 0, \quad i = 1, \ldots, n_1 \\ c & \text{gdy } y_i^* \leqslant 0, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{1}$$

gdzie:

$y_i^* = \mathbf{X}_i \boldsymbol{\beta} + u_i$ – równanie regresji,

$\mathbf{X}_i$ – $(k+1)$-elementowy wektor (poziomy) wartości zmiennych objaśniających ze zbioru $X = \{X_1, \ldots, X_k\}$, odnotowanych dla i–tej obserwacji (jednostki), przy czym pierwszy element wektora $\mathbf{X}_i$ jest równy jedności,

$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ – wektor (pionowy) parametrów równania regresji,

u_i – składniki losowe równania, niezależne i pochodzące z tego samego rozkładu o wartości oczekiwanej równej 0 oraz stałej wariancji σ_u. Najczęściej (w celu estymacji metodą największej wiarygodności) zakłada się, że składniki losowe podlegają rozkładowi normalnemu.

---

[3] Zob. pracę [24].

[4] W artykule rozważono modele z wartością progową równą zero. Nie wiąże się to ze stratą ogólności.

Model tobitowy typu II jest postaci (por. [24]):

$$y_i = \begin{cases} \mathbf{X}_i \boldsymbol{\beta} + u_i & \text{gdy} \quad I_i = 1, \quad i = 1, \ldots, n_1 \\ 0 \text{ lub brak} & \text{gdy} \quad I_i = 0, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{2}$$

gdzie:

$I_i = \begin{cases} 1 & \text{gdy } z_i^* > 0 \\ 0 & \text{gdy } z_i^* \leqslant 0 \end{cases}$ – zmienna binarna opisująca kryterium obserwowalności,

$z_i^* = \mathbf{W}_i \boldsymbol{\gamma} + v_i$ – równanie selekcji,

$\mathbf{W}_i$ – $(m+1)$-elementowy wektor (poziomy) wartości zmiennych objaśniających ze zbioru $W = \{W_1, \ldots, W_m\}$, odnotowanych dla i–tej obserwacji (jednostki), przy czym pierwszy element wektora $\mathbf{W}_i$ jest równy jedności,

$\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_m)'$– wektor (pionowy) parametrów równania selekcji,

v_i – składniki losowe równania selekcji,

$\mathbf{X}_i, \mathbf{W}_i, \beta,\ u_i,$ – określone jak wcześniej, przy czym najczęściej (w celu estymacji metodą największej wiarygodności) zakłada się, że:

$(u_i, v_i) \sim N\left(0,\ 0; \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)$, gdzie ρ – współczynnik korelacji między u_i i v_i.

W modelu tobitowym typu II kryterium obserwowalności opisane jest przez zmienną binarną I. Z tego względu przyjmuje się $\sigma_v = 1$, podobnie jak w modelu probitowym.

Model tobitowy typu III jest postaci (por. [24]):

$$y_i = \begin{cases} \mathbf{X}_i \boldsymbol{\beta} + u_i & \text{gdy} \quad z_i > 0, \quad i = 1, \ldots, n_1 \\ 0 \text{ lub brak} & \text{gdy} \quad z_i = 0, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{3}$$

gdzie tym razem zmienna opisująca kryterium obserwowalności jest postaci:

$$z_i = \begin{cases} z_i^* & \text{gdy} \quad z_i^* > 0 \\ 0 & \text{gdy} \quad z_i^* \leqslant 0 \end{cases},$$

przy pozostałych oznaczeniach jak wcześniej.

Przegląd modeli tobitowych (modeli zmiennej zależnej cenzurowanej) można znaleźć m.in. w pracach: [1], [9], [27], [28], a także [22], [24].

3. ESTYMACJA SEMIPARAMETRYCZNA W MODELU TOBITOWYM TYPU I

Wśród wielu metod semiparametrycznych stosowanych w wypadku modelu tobitowego typu I najbardziej popularne są metody wprowadzone przez J.L. Powella (zob. [38], [39], [40], [41]) oraz B.E. Honoré'a i J.L. Powella ([16]). Są to m.in.

metody: CLAD, STLS/SCLS, ICLAD oraz ICLS. W metodach tych określa się postać parametryczną równania regresji, ale nie narzuca się postaci parametrycznej rozkładu składnika losowego. Zaprezentowane metody semiparametryczne dają estymatory o dobrych własnościach dopuszczając dowolny rozkład składnika losowego i/lub jego heteroskedastyczność, przy pewnych słabszych założeniach. Poniżej metody te omówiono dla modelu tobitowego typu I z cenzurowaniem w zerze.

Metoda najmniejszych odchyleń bezwzględnych z cenzurowaniem (*censored least absolute deviations*, *censored* LAD, CLAD). Metoda CLAD, wprowadzona przez J.L. Powella w pracy [38], jest metodą estymacji semiparametrycznej stosowaną w modelu tobitowym typu I. Przy dość ogólnych założeniach estymatory otrzymane metodą CLAD są zgodne i asymptotycznie normalne. Spośród wymaganych założeń należy wspomnieć następujące[5]:

- składniki losowe u_i są niezależne, o tym samym rozkładzie, przy czym ich rozkład może być dowolnym rozkładem, w którym mediana równa jest zero (warunkowo od $\mathbf{X}_i$),
- procent obserwacji „zerowych" (ocenzurowanych) nie jest zbyt wysoki,
- dla obserwacji „zerowych" odpowiadające im zmienne $\mathbf{X}_i$ nie są współliniowe.

Jeżeli w próbie jest znaczący procent obserwacji „zerowych", można zastosować estymator oparty na cenzurowanej regresji kwantylowej (ang. *censored quantile regression*) bazując na kwantylu wyższego rzędu niż mediana[6] (por. [39]). Główną zaletą metody jest fakt, że dopuszcza dowolny, w tym również asymetryczny (por. [5], s. 34), rozkład składnika losowego oraz heteroskedastyczność[7]. Metoda CLAD charakteryzuje się dużą złożonością numeryczną.

Celem metody CLAD jest uzyskanie estymatora wektora parametrów $\boldsymbol{\beta}$. Estymator ten otrzymuje się wyznaczając wektor $\hat{\boldsymbol{\beta}}_{CLAD}$ minimalizujący wyrażenie:

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \max\{0; \mathbf{X}_i\boldsymbol{\beta}\}|, \tag{4}$$

przy czym mediana zmiennej zależnej wynosi: $Me(y_i) = \max\{0; \mathbf{X}_i\boldsymbol{\beta}\}$.

Sposób wyznaczenia zgodnego estymatora asymptotycznej macierzy kowariancji można znaleźć w pracach J.L. Powella (zob. [38], [41]). W wypadku estymatorów otrzymanych metodą CLAD w modelu tobitowym typu I można stosować test Walda[8] odnośnie ewentualnych restrykcji nałożonych na parametry.

Metoda CLAD oraz inne metody semiparametryczne oparte na kwantylach nie są odpowiednie do estymacji modeli regresji uciętej.

[5] Szczegóły można odnaleźć w pracy J.L. Powella [38], s. 307.

[6] Estymator CLAD jest szczególnym przypadkiem estymatora opartego na cenzurowanej regresji kwantylowej.

[7] Przy heteroskedastycznym składniku losowym estymatory otrzymane metodą CLAD nadal pozostaną zgodne i asymptotycznie normalne (por. [38], s. 308).

[8] Postać testu Walda można odnaleźć np. w [28].

Metoda najmniejszych kwadratów z symetrycznym przycinaniem lub cenzurowaniem (ang. *symmetrically trimmed least squares*, STLS; *symmetrically censored least squares*, SCLS). Metoda ta, również zaproponowana przez J.L. Powella w pracy [40], znajduje zastosowanie zarówno w modelu regresji uciętej (metoda STLS) jak i cenzurowanej (metoda SCLS). W niniejszym artykule omówiono tylko wersję metody dla cenzurowania. Metoda SCLS jest prostsza w zastosowaniach niż metoda CLAD, ale daje estymatory asymptotycznie mniej efektywne (por. [17], s. S65).

J.L. Powell w pracy [40] zauważył, że cenzurowanie lub ucięcie zmiennej zależnej sprawia, że rozkład składnika losowego jest asymetryczny, co jest główną przyczyną niezgodności estymatorów otrzymanych metodą najmniejszych kwadratów. W metodzie SCLS zakłada się[9], że u_i są niezależne o tym samym symetrycznym rozkładzie warunkowo od $\mathbf{X}_i$ (metoda dopuszcza dowolny symetryczny rozkład, niekoniecznie normalny). Korzystając z tego założenia sztucznie cenzuruje się dane: obserwacjom y_i takim, że $y_i > 2\mathbf{X}_i\boldsymbol{\beta}$ przypisuje się wartość $2\mathbf{X}_i\boldsymbol{\beta}$.

Estymator $\hat{\beta}_{SCLS}$ otrzymuje się wyznaczając argument minimalizujący wyrażenie:

$$\sum_{i=1}^{n}(y_i - \max\{\tfrac{1}{2}y_i; \mathbf{X}_i\boldsymbol{\beta}\})^2 + \sum_{i=1}^{n} 1\{y_i > 2\mathbf{X}_i\boldsymbol{\beta}\}\cdot(\tfrac{1}{4}y_i^2 - \max\{0; \mathbf{X}_i\boldsymbol{\beta}\}^2), \tag{5}$$

gdzie $1\{A\}$ jest funkcją charakterystyczną zbioru A.

Także ta metoda prowadzi do estymatorów zgodnych i asymptotycznie normalnych[10], nawet przy wystąpieniu heteroskedastyczności (o nieznanej formie). J.L. Powell (zob. [40], s. 1444–1445) wskazał zgodny estymator asymptotycznej macierzy kowariancji (por. także prace [18] oraz [33], s. 129).

Metoda oparta na różnicy par (ang. *pairwise difference method*). Metoda ta została zaproponowana przez B.E. Honoré'a i J.L. Powella w pracy [16] dla modeli regresji uciętej lub cenzurowanej, przy głównym założeniu, że u_i są niezależne o tym samym rozkładzie, a przy tym niezależne od zmiennych objaśniających $\mathbf{X}_i$. Dopuszczalny jest rozkład asymetryczny składnika losowego u_i, natomiast wykluczona jest jego heteroskedastyczność (por. także [5], s. 34). Idea metody polega na symetrycznym sztucznym cenzurowaniu par danych w taki sposób, aby po cenzurowaniu pary te miały identyczny rozkład[11]. Estymator parametrów modelu otrzymuje się wyznaczając argument minimalizujący wyrażenie (por. [4], s. 18):

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} s(y_i, y_j, (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}), \tag{6}$$

[9] W pracy [33], s. 127, opisano sposoby weryfikacji założeń gwarantujących zgodność estymatorów otrzymanych metodą SCLS dla modelu tobitowego.

[10] Por. [40], gdzie zamieszczono szerszą dyskusję odnośnie istnienia globalnego minimum oraz omówiono własności otrzymanych estymatorów przy pewnych warunkach.

[11] Opis metody można znaleźć w [16], a także [4], s. 16, lub [15], s. 112.

gdzie:

$$s(y_1, y_2, \delta) = \begin{cases} \Xi(y_1) - (y_2 + \delta)\xi(y_1) & \text{gdy } \delta \leqslant -y_2 \\ \Xi(y_1 - y_2 - \delta) & \text{gdy } -y_2 < \delta < y_1 \\ \Xi(-y_2) + (y_1 - \delta)\xi(-y_2) & \text{gdy } y_1 \leqslant \delta \end{cases}, \tag{7}$$

$\Xi(.)$ – pewna parzysta i wypukła funkcja,

$\xi(.)$ – pewna nieparzysta funkcja, dla której powyższe wyrażenia mają sens.

Estymatory uzyskane tą metodą są zgodne i asymptotycznie normalne. Jednak za jej pomocą nie można wyznaczyć estymatora wyrazu wolnego bez nałożenia dodatkowych warunków (założenie o niezależności u_i od $\mathbf{X}_i$ nie nakłada restrykcji na położenie rozkładu składnika losowego)[12].

B.E. Honoré i J.L. Powell w pracy [16] przyjęli $\Xi(d) = |d|$ otrzymując estymatory o dobrych własnościach nawet dla małej próby. W literaturze przedmiotu (por. [5], s. 32) w tym wypadku metodę nazywa się metodą najmniejszych odchyleń bezwzględnych z identycznym cenzurowaniem (ang. *identically censored least absolute deviations*, ICLAD). W przypadku, gdy $\Xi(d) = d^2$ dla metody stosuje się nazwę metoda najmniejszych kwadratów z identycznym cenzurowaniem (ang. *identically censored least squares*, ICLS). Stosowanie estymatora ICLAD jest bardziej wskazane niż estymatora ICLS (por. [5], [16]).

Estymator $\hat{\boldsymbol{\beta}}_{ICLAD}$ otrzymuje się wyznaczając argument minimalizujący wyrażenie (por. [15], s. 113):

$$\sum_{i<j} \left(1 - 1\left\{y_i \leqslant \max\{0, (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\}; y_j \leqslant \max\{0, (\mathbf{X}_j - \mathbf{X}_i)\boldsymbol{\beta}\}\right\}\right) \cdot \left|y_i - y_j - (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\right|, \tag{8}$$

natomiast estymator $\hat{\boldsymbol{\beta}}_{ICLS}$ – minimalizujący wyrażenie (por. [15], s. 113):

$$\begin{aligned} &\sum_{i<j} \left(\max\{y_i, (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\} - \max\{y_j, (\mathbf{X}_j - \mathbf{X}_i)\boldsymbol{\beta}\} - (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\right)^2 + \\ &+2 \cdot 1\{y_i < (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\} \cdot ((\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta} - y_i)y_j + \\ &+2 \cdot 1\{y_j < (\mathbf{X}_j - \mathbf{X}_i)\boldsymbol{\beta}\} \cdot ((\mathbf{X}_j - \mathbf{X}_i)\boldsymbol{\beta} - y_j)y_i \end{aligned} . \tag{9}$$

Na zakończenie rozważań o metodach semiparametrycznych dla modelu tobitowego typu I warto wspomnieć o pewnej własności przydatnej przy wnioskowaniu o poprawności specyfikacji modelu, w tym poprawności założeń. Mianowicie wartości estymatorów otrzymanych metodą CLAD można traktować jako pewnego rodzaju nieformalne kryterium weryfikujące poprawność założeń modelu, gdyż estymator ten jest zgodny przy:

- założeniu normalności składnika losowego u_i – wymaganym do zgodności estymatorów otrzymanych metodą największej wiarygodności,

[12] Por. np. [4], s. 16.

- założeniu niezależności u_i – wymaganym do zgodności estymatorów otrzymanych metodą ICLAD lub ICLS,
- założeniu warunkowej symetryczności rozkładu u_i – wymaganym do zgodności estymatorów otrzymanych metodą SCLS.

Łączne porównanie[13] wartości estymatorów otrzymanych różnymi metodami pozwala wnioskować o ewentualnym naruszeniu założeń (por. [4], s. 21–24).

4. ESTYMACJA SEMIPARAMETRYCZNA W MODELU TOBITOWYM TYPU II

Metody semiparametryczne dla modeli, w których równanie selekcji jest opisane za pomocą zmiennej binarnej I, w tym dla modelu tobitowego typu II danego wzorem (2), na ogół wymagają, aby co najmniej jedna zmienna należąca do zbioru zmiennych W opisującego równanie selekcji, nie należała do zbioru zmiennych X objaśniającego równanie regresji (por. [5], s. 35, [30], s. 24, [46],[47], s. 219). W zastosowaniach założenie to nie jest wygodne i prawdopodobnie z tego względu semiparametryczne modele tobitowe typu II są znacznie rzadziej stosowane niż np. modele tobitowe typu I lub III. Ponadto całkowity brak znajomości wartości zmiennej Z^* w modelu tobitowym typu II sprawia, że nie jest możliwe w szczególności zastosowanie metod analogicznych do tych, odpowiednich dla modelu tobitowego typu I. Częściej natomiast stosuje się metody nieparametryczne.

Tak jak poprzednio metody semiparametryczne pozwalają na osłabienie założenia o rozkładzie normalnym składnika losowego.

Większość z metod dla modelu tobitowego typu II przebiega w dwóch etapach, podobnie jak dwustopniowa metoda Heckmana[14]. W każdym z etapów stosuje się metody semiparametryczne lub nieparametryczne. Podstawą estymacji jest zapisanie warunkowej wartości oczekiwanej zmiennej zależnej w następujący sposób:

$$E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i) = \mathbf{X}_i\boldsymbol{\beta} + E(u_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i). \tag{10}$$

Przy tym zazwyczaj zakłada się, że warunkowa wartość oczekiwana składnika losowego u_i jest funkcją $\mathbf{W}_i\boldsymbol{\gamma}$ i tylko w ten sposób zależy od $\mathbf{W}_i$, tzn. (por. [46], s. 139 i n.):

$$E(u_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i) = g(\mathbf{W}_i\boldsymbol{\gamma}), \tag{11}$$

gdzie g jest nieznaną funkcją[15].

W pierwszym kroku estymuje się wektor parametrów $\boldsymbol{\gamma}$ w równaniu selekcji bez narzucania konkretnego rozkładu na składnik losowy v_i równania selekcji. Można za-

[13] Porównania te przeprowadza się w sposób podobny do testu Hausmana.

[14] Por. [11], [12].

[15] Warunkową wartość oczekiwaną można zapisać w ogólny sposób $E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i) = g_1(\mathbf{X}_i) + g_2(\mathbf{W}_i\boldsymbol{\gamma})$, gdzie g_1 oraz g_2 – nieznane funkcje, przy czym zazwyczaj określa się postać funkcji g_1 np. liniową. Szerszą dyskusję odnośnie odpowiednich założeń i postaci funkcji można znaleźć np. w pracy [29], s. 353–358.

stosować jedną z metod semiparametrycznych lub nieparametrycznych dla modeli binarnych np. metodę R.W. Kleina i R.H. Spady (zob. [19])[16] prowadzącą do zgodnych i asymptotycznie normalnych estymatorów.

Również w drugim kroku estymacji semiparametrycznej modelu tobitowego typu II można postąpić na wiele sposobów[17]. Jedna z metod przebiega następująco (por. [31]). Korzystając z estymatora $\mathbf{W}_i\hat{\boldsymbol{\gamma}}$ oraz przybliżając funkcję g za pomocą szeregu aproksymacji opartego na odwrotności współczynnika Millsa równanie regresji modelu tobitowego typu II należy zapisać w postaci[18]:

$$y_i = \mathbf{X}_i\boldsymbol{\beta} + \zeta_0 + \zeta_1\lambda(\mathbf{W}_i\hat{\boldsymbol{\gamma}}) + \zeta_2\lambda^2(\mathbf{W}_i\hat{\boldsymbol{\gamma}}) + \ldots + \zeta_L\lambda^L(\mathbf{W}_i\hat{\boldsymbol{\gamma}}) + \varepsilon_i, \qquad (12)$$

gdzie:

$\lambda(\alpha) = \dfrac{\varphi(\alpha)}{1-\Phi(\alpha)}$ – odwrotność współczynnika Millsa dla cenzurowania lewostronnego,

$\beta_0 + \zeta_0, \quad \beta_1, \ldots, \beta_k, \quad \zeta_1, \ldots, \zeta_L$ – nieznane parametry.

Ten sposób aproksymacji ma tę zaletę, że przy $L = 1$ model (12) jest tożsamy z modelem tobitowym typu II z założeniem rozkładu normalnego. Jeżeli oszacowane parametry $\zeta_2, \ldots, \zeta_L$ modelu (12) nie są łącznie statystycznie nieistotne, można przypuszczać, że założenie o rozkładzie normalnym nie jest poprawne (por. [25], s. 333), a także [37]). Postać asymptotycznej wariancji podano w pracy [34] autorstwa W.K. Neweya.

Estymację parametrów równania (12) przeprowadza się na podstawie obserwacji, dla których $I_i = 1$. Należy zastosować jedną z metod odpornych na heteroskedastyczność składnika losowego ε_i (np. metodę P.M. Robinsona [43], por. [31], s. 70).

Przy zastosowaniu aproksymacji (12) nie ma jednoznacznej reguły określającej wartość L. Jedna z propozycji ustala kryterium doboru L w taki sposób, by dodanie następnego elementu w szeregu aproksymacji nie zmieniało zasadniczo wartości wyrazu wolnego (por. [25]).

Wyraz wolny modelu (12) jest sumą dwóch wyrazów wolnych: $\beta_0 + \zeta_0$. W celu estymacji oddzielnie β_0 i ζ_0 można skorzystać z propozycji J. Heckmana w pracy [13] lub D. Andrewsa i M. Schafgans w pracy [2][19].

5. ESTYMACJA SEMIPARAMETRYCZNA W MODELU TOBITOWYM TYPU III

Sposoby estymacji semiparametrycznej w wypadku modelu tobitowego typu III danego wzorem (3) są odmienne od tych mających zastosowanie w modelu tobitowym

[16] Opis metody można znaleźć np. w [30], s. 25, [36], s. 283. Inną metodę zaproponowano np. w pracy [42]. W pracy [30], Appendix A, przedstawiono także testy umożliwiające porównanie równania selekcji opisanego za pomocą modelu probitowego z tym otrzymanym metodą semi- lub nieparametryczną.

[17] W artykule zaprezentowano tylko jedno z wielu możliwych podejść. Szeroki opis stosowanych metod można znaleźć np. w [41] lub [46].

[18] Inne postaci aproksymacji zaprezentowano np. w pracy [8], [25], [34], [37].

[19] Por. prace [25] lub [44] odnośnie dyskusji na temat własności tego estymatora.

typu II. Różnice wynikają przede wszystkim z faktu, że obecnie dostępnych jest więcej informacji: znane są wartości zmiennej Z. Oczywiście w celu estymacji modelu tobitowego typu III można zastosować również metody semiparametryczne odpowiednie do estymacji modelu tobitowego typu II wykorzystując tylko informację o znaku zmiennej Z^*. Jednakże ze względu na utratę informacji (nie korzysta się ze znanych dokładnych wartości zmiennej cenzurowanej Z), otrzymane estymatory zazwyczaj będą mniej efektywne.

W literaturze przedmiotu metody semiparametryczne dla modelu tobitowego typu III rozważano m.in. w pracach: [15], a także [6], [26]. W artykule zaprezentowano metody opisane w pierwszej z tych prac.

Pierwsza metoda zaproponowana w pracy[20] [15] opiera się głównie na założeniu, że dwuwymiarowy rozkład składników losowych (u_i, v_i) warunkowo od $(\mathbf{X}_i, \mathbf{W}_i)$ jest rozkładem symetrycznym w takim sensie, że (u_i, v_i) pochodzą z tego samego rozkładu co $(-u_i, -v_i)$. Ponadto, aby otrzymane estymatory parametrów $\boldsymbol{\beta}$ oraz $\boldsymbol{\gamma}$ były zgodne i asymptotycznie normalne, muszą być spełnione pewne dodatkowe warunki regularności[21]. Warto podkreślić, że metoda dopuszcza heteroskedastyczność składnika losowego.

W pierwszym kroku omawianej metody należy zastosować metodę CLAD lub SCLS do równania selekcji (będącego modelem tobitowym typu I) w celu wyznaczenia estymatora wektora parametrów $\boldsymbol{\gamma}$ (por. wzór (4) lub, odpowiednio, (5)). W drugim kroku wartości $\mathbf{W}_i\hat{\boldsymbol{\gamma}}_{CLAD}$ (odpowiednio $\mathbf{W}_i\hat{\boldsymbol{\gamma}}_{SCLS}$) wykorzystuje się do symetrycznego przycięcia (cenzurowania) rozkładu składników losowych u_i. Następnie wystarczy zastosować np. metodę LAD lub, odpowiednio, metodę najmniejszych kwadratów do równania regresji po przycięciu, czyli na nowej, przyciętej próbie. W rezultacie estymator wektora parametrów $\boldsymbol{\beta}$ otrzymuje się wyznaczając argument $\tilde{\tilde{\boldsymbol{\beta}}}_{CLAD}$ minimalizujący wyrażenie:

$$\frac{1}{n}\sum_{i=1}^{n} 1\{0 < z_i < 2\mathbf{W}_i\hat{\boldsymbol{\gamma}}_{CLAD}\} \cdot |y_i - \mathbf{X}_i\boldsymbol{\beta}|, \tag{13}$$

przy zastosowaniu metody LAD lub argument $\tilde{\tilde{\boldsymbol{\beta}}}_{SCLS}$ minimalizujący wyrażenie:

$$\frac{1}{n}\sum_{i=1}^{n} 1\{0 < z_i < 2\mathbf{W}_i\hat{\boldsymbol{\gamma}}\} \cdot (y_i - \mathbf{X}_i\boldsymbol{\beta})^2, \tag{14}$$

przy zastosowaniu metody najmniejszych kwadratów.

Druga z proponowanych w pracy [15] metod bazuje głównie na założeniu, że zmienne objaśniające X oraz W są niezależne od składników losowych u oraz v, przy czym nie narzuca się założenia odnośnie symetryczności rozkładu składników

[20] W pracy [15] podano także sposób wyznaczania estymatorów w wypadku modeli selekcji uciętej.

[21] Szczegółowe założenia można znaleźć w [15], s. 115.

losowych[22]. W celu uzyskania estymatorów zgodnych i asymptotycznie normalnych wymagane są pewne dodatkowe warunki regularności[23].

Tym razem w pierwszym kroku metody następuje estymacja równania selekcji metodą semiparametryczną służącą do estymacji modelu tobitowego typu I przy założeniu niezależności zmiennych objaśniających i składników losowych. W pracy [15] stosowano metody ICLAD lub ICLS zaproponowane przez B.E. Honoré'a i J.L. Powella (zob. [16]), tzn. wyznaczono estymator $\hat{\gamma}_{ICLAD}$ (por. wzór (8)) lub odpowiednio $\hat{\gamma}_{ICLS}$ (por. wzór (9)) wektora parametrów $\boldsymbol{\gamma}$, traktując równanie selekcji jako model tobitowy typu I.

W drugim kroku w celu eliminacji efektu selekcji zastosowano podejście bazujące na rozkładzie różnicy par[24], analogicznie jak w opisanej wcześniej semiparametrycznej metodzie B.E. Honoré'a i J.L. Powella (por. [16]) dla modelu tobitowego typu I (por. wzory (8) i (9)). Estymatory wektora parametrów otrzymuje się wyznaczając argument $\tilde{\tilde{\boldsymbol{\beta}}}_{ICLAD}$ minimalizujący wyrażenie:

$$\sum_{i<j} 1\left\{z_i > \max\{0, (\mathbf{W}_i - \mathbf{W}_j)\hat{\boldsymbol{\gamma}}_{ICLAD}\}; z_j > \max\{0, (\mathbf{W}_j - \mathbf{W}_i)\hat{\boldsymbol{\gamma}}_{ICLAD}\}\right\} \cdot \left|y_i - y_j - (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\right|$$

lub odpowiednio $\tilde{\tilde{\boldsymbol{\beta}}}_{ICLS}$ minimalizujący wyrażenie:

$$\sum_{i<j} 1\left\{z_i > \max\{0, (\mathbf{W}_i - \mathbf{W}_j)\hat{\boldsymbol{\gamma}}_{ICLS}\}; z_j > \max\{0, (\mathbf{W}_j - \mathbf{W}_i)\hat{\boldsymbol{\gamma}}_{ICLS}\}\right\} \cdot \left(y_i - y_j - (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\right)^2.$$

Przy tym, tak jak poprzednio, bardziej wskazane jest stosowanie metody bazującej na ICLAD.

6. WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Przydatna jest możliwość porównania różnych specyfikacji modeli tobitowych np. postaci parametrycznej z modelem semiparametrycznym. W literaturze przedmiotu stosuje się np. test J.A. Hausmana (zob. [10]). Idea tego testu jest następująca[25]. Niech $M-$elementowe wektory $\hat{\boldsymbol{\pi}}_0$ oraz $\hat{\boldsymbol{\pi}}_1$ będą dwoma estymatorami wektora parametrów $\boldsymbol{\pi}$, które przy założeniu hipotezy zerowej są zgodne i asymptotycznie normalne o asymptotycznych macierzach wariancji odpowiednio: V_0 oraz V_1. Ponadto przy założeniu hipotezy zerowej $\hat{\boldsymbol{\pi}}_0$ jest estymatorem asymptotycznie efektywnym oraz $V_1 - V_0$ jest nieujemnie określona. Przy założeniu hipotezy alternatywnej estymator $\hat{\boldsymbol{\pi}}_1$ pozostaje

[22] W przypadku przyjęcia wspomnianego założenia o niezależności należy rozpatrywać model bez wyrazu wolnego w równaniu regresji i w równaniu selekcji (por. [15], s. 112).

[23] Por. [15], s. 117.

[24] Idea metody bazuje na fakcie, że rozkład różnicy niezależnych zmiennych losowych pochodzących z tego samego rozkładu jest symetryczny wokół zera.

[25] Por. np. [28], s. 558–559, [32], s. 1321, [33].

zgodny, ale estymator $\hat{\boldsymbol{\pi}}_0$ traci tę własność (nie jest zgodny). Statystyka testowa jest postaci:

$$N \cdot (\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_0)'(\hat{V}_1 - \hat{V}_0)^{-1}(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_0) \sim \chi^2[M], \tag{15}$$

gdzie N jest liczbą elementów próby.

W.K. Newey (zob. [33]) podał sposób zastosowania testu Hausmana w celu porównania estymatorów parametrów modelu tobitowego typu I otrzymanych metodą największej wiarygodności przy założeniu rozkładu normalnego i homoskedastyczności składnika losowego z estymatorami otrzymanymi metodą semiparametryczną SCLS odporną na naruszenie tych założeń. Dzięki odporności tego ostatniego estymatora test Hausmana można zastosować jako narzędzie sprawdzające podstawowe założenia parametrycznego modelu tobitowego typu I. W przypadku odrzucenia hipotezy zerowej należy uznać, że co najmniej jedno z założeń (o rozkładzie normalnym lub o homoskedastyczności) nie jest spełnione. Do przeprowadzenia dalszego wnioskowania należy posłużyć się estymatorami otrzymanymi metodą semiparametryczną.

Na podobnej zasadzie można skonstruować test porównujący estymatory otrzymane metodą największej wiarygodności z estymatorami otrzymanymi dowolną inną metodą semiparametryczną (por. [17], s. S67, [39]). Przy tym należy zwrócić uwagę, że macierze wariancji estymatorów $\hat{\boldsymbol{\pi}}_0$ oraz $\hat{\boldsymbol{\pi}}_1$ wyznacza się przy założeniu hipotezy zerowej. Jest to szczególnie przydatne, gdyż macierze wariancji estymatorów otrzymanych metodami semiparametrycznymi są zazwyczaj proste do wyznaczenie przy założeniu rozkładu normalnego (por. [4]).

B. Melenberg i A. van Soest w pracy [31] zastosowali test Hausmana w celu porównania modelu tobitowego z rozkładem normalnym z modelem otrzymanym metodą CLAD, jak również przy porównaniu semiparametrycznych specyfikacji modelu. Także w pracy [4] przeprowadzono analizę różnic wartości semiparametrycznych estymatorów powołując się na podobieństwo do testu Hausmana. Głównym celem takiej analizy jest wskazanie, które z założeń modelu zostało naruszone, gdyż każdy z omówionych estymatorów semiparametrycznych modelu tobitowego wymaga innych założeń.

Test Hausmana w odpowiedniej postaci można również zastosować do porównania specyfikacji parametrycznych (model probitowy) i semiparametrycznych równania selekcji (por. np. [35]).

Także analiza reszt modelu w semiparametrycznych modelach tobitowych może pomóc w określeniu, czy któreś z założeń o rozkładzie składników losowych jest naruszone (por. np. [4]). W literaturze przedmiotu zazwyczaj stosuje się mniej formalne testy, w tym w dużej mierze graficzne (por. [7], [17]).

7. UWAGI KOŃCOWE

Zaprezentowane metody estymacji semiparametrycznej znajdują zastosowanie w modelach tobitowych typu I–III, a także innych modelach zmiennej zależnej ograniczonej. Metody te pozwalają w szczególności na rozluźnienie założenia o rozkładzie normalnym składników losowych, narzucając jednocześnie inne, na ogół mniej krępujące warunki. Ponadto sięgnięcie po metody semiparametryczne umożliwia weryfikację, które z założeń o rozkładzie składnika losowego modelu zostało naruszone. Dzięki temu można zweryfikować, czy można zastosować np. metodę największej wiarygodności, umożliwiającą m.in. poprawną interpretację wyników.

Omawiane metody nie są jednak pozbawione pewnych wad. Jedną z nich jest występujący w niektórych metodach problem niejednoznacznej identyfikacji wyrazu wolnego modelu. Ponadto niektóre metody, np. CLAD, charakteryzują się dużą złożonością obliczeniową. Kolejną niedogodnością jest brak metod pozwalających na poprawną interpretację uzyskanych wyników. Specyfika modeli tobitowych typu I–III, oraz innych modeli zmiennej zależnej ograniczonej, sprawia, że wpływy krańcowe zmiennych objaśniających nie są równe parametrom modelu[26]. Ze względu na brak określenia parametrycznej postaci rozkładu składnika losowego występują trudności z wyznaczeniem wpływów krańcowych zmiennych objaśniających na analizowane zjawisko. Brak możliwości wyznaczenia wpływów krańcowych równoznaczny jest z brakiem poprawnej interpretacji otrzymanych wyników na podstawie modeli tobitowych. Jest to duże utrudnienie w stosowaniu metod semiparametrycznych w wypadku rozważanych modeli tobitowych.

Uniwersytet Ekonomiczny w Krakowie

LITERATURA

[1] Amemiya T. (1984), *Tobit Models: A survey*, Journal of Econometrics, vol. 24, s. 3-61.
[2] Andrews D., Schafgans M. (1998), *Semiparametric Estimation of a Sample Selection Model*, Review of Economic Studies, vol. 65, nr 224, s. 497-518.
[3] Bloomfield P., Steiger W.L. (1983), *Least Absolute Deviations: Theory, Applications, and Algorithms*, Progress in Probability and Statistics, vol. 6, Birkhäuser Boston, Boston, Mass, USA.
[4] Chay K.Y., Honoré B.E. (1998), *Estimation of Semiparametric Regression Models*, Journal of Human Resources, vol. 33, nr 1, s. 4-38.
[5] Chay K.Y., Powell J.L. (2001), *Semiparametric Censored Regression Models*, Journal of Economic Perspectives, vol. 15, nr 4, s. 29-42.
[6] Chen S. (1997), *Semiparametric Estimation of the Type-3 Tobit Model*, Journal of Econometrics, vol. 80, s. 1-34
[7] Chesher A., Irish M. (1987), *Residual Analysis on the Grouped and Censored Normal Linear Model*, Journal of Econometrics, vol. 34, s. 33-61.
[8] Gallant R., Nychka G. (1987), *Semi-Nonparametric Maximum Likelihood Estimatnion*, Econometrica, vol. 55, s. 363-390.
[9] Gruszczyński M. (2002), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, wyd. 2, Monografie i Opracowania 490, Oficyna Wydawnicza SGH, Warszawa.
[10] Hausman J.A. (1978), *Specification Tests in Econometrics*, Ecoonmetrica, vol. 46, s. 1251-1272.

[26] Zob. szerszą dyskusję na temat interpretacji w modelach tobitowych w pracy [24].

[11] Heckman J. (1976), *The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models*, Annals of Economic and Social Measurement, nr 5/4, 1976, s. 475-492.

[12] Heckman J. (1979), *Sample Selection Bias as a Specification Error*, Econometrica, vol. 47, nr 1, s. 153-161.

[13] Heckman J. (1990), *Varieties of Selection Bias*, American Economic Review, 80, s. 313-318.

[14] Heckman J. (2003), *Microdata, Heterogeneity and the Evaluation of Public Policy* [w:] *Nobel Lectures in Economic Sciences 1996 – 2000*, T. Persson (ed.), World Scientific Publishing Co., Singapore, s. 255-322.

[15] Honoré B.E., Kyriazidou E., Udry C. (1997), *Estimation of Type 3 Tobit Models Using Symmetric Trimming and Pairwise Comperisions*, Journal of Econometrics, vol. 76, s. 107-128.

[16] Honoré B.E., Powell J.L. (1994), *Pairwise Difference Estimators of Censored and Truncated Regression Models*, Journal of Econometrics, vol. 64, nr 1–2, s. 241-278.

[17] Horowitz J.L., Neumann G.R. (1989), *Specification Testing in Censored Regression Models: Parametric and Semiparametric Methods*, Journal of Applied Econometrics, vol. 4, s. S61-S86.

[18] Johnston J. i DiNardo J. (1997), *Econometric Methods*, McGraw-Hill Companies, Inc., New York.

[19] Klein R.W., Spady R.H. (1993), *An Efficient Semiparametric Estimator of Binary Response Models*, Econometrica, vol. 61, nr 2, s. 387-421.

[20] Kostrzewska J. (2003), *Model tobitowy jako szczególny przypadek cenzurowanego modelu regresji* [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, red. A. Zeliaś, Wyd. AE w Krakowie, Kraków, s. 397-404.

[21] Kostrzewska J. (2008), *Zastosowanie wybranych modeli tobitowych do opisu tygodniowej liczby godzin pracy* [w:] *Taksonomia 15. Klasyfikacja i analiza danych – teoria i zastosowania*, pod red. K. Jajugi i M. Walesiaka, Wyd. AE we Wrocławiu, Wrocław.

[22] Kostrzewska J. (2009), *Truncated and Censored Dependent Variables and Their Models*, referat wygłoszony na 15th Polish-Slovak-Ukrainian Scientific Seminar "Statistical Analysis of Socio-Economic Consequences of Transition Processes in Central-East European Countries", Kraków, 21–24 października 2008 r.

[23] Kostrzewska J. (2011), *Analiza porównawcza tygodniowego czasu pracy zamężnych kobiet pracujących na własny rachunek lub najemnie*, [w:] *Osiągnięcia i perspektywy modelowania i prognozowania zjawisk społeczno-gospodarczych*, pod red. B. Pawełek, Wyd. UEK, Kraków, w druku.

[24] Kostrzewska J. (2011), *Interpretacja w modelach tobitowych*, Przegląd Statystyczny, tom 58, z. 3-4, s. 256-280.

[25] Lanot G., Walker I. (1998), *The union/non-union wage differential: An application of semiparametric methods*, Journal of Econometrics, vol. 84, nr 2, s. 327-349.

[26] Lee L.-F. (1994), *Semiparametric Two-Stage Estimation of Sample-Selection Models Subject to Tobit-Type Selection Rules*, Journal of Econometrics, vol. 61, s. 305-344.

[27] Maddala G.S. (1994), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

[28] Maddala G.S. (2006), *Ekonometria*, Wyd. Nauk. PWN, Warszawa.

[29] Manski C.F. (1989), *Anatomy of the Selection Problem*, The Journal of Human Resources, vol. 24, nr 3, s. 343-360.

[30] Martins M.F. (2001), *Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal*, Journal of Applied Econometrics, vol. 16, nr 1, s. 23-39.

[31] Melenberg B., van Soest A. (1996), *Parametric and Semi-Parametric Modelling on Vacation Expenditures*, Journal of Applied Econometrics, vol. 11, s. 59-76.

[32] Nelson F.D. (1981), *A Test for Misspecification in the Censored Normal Model*, Econometrica, vol. 49, s. 1317-1329.

[33] Newey W.K. (1987), *Specification Tests for Distributional Assumption in the Tobit Model*, Journal of Econometrics, vol. 34, s. 125-145.
[34] Newey W.K. (1991), *Efficient Estimation of Tobit Models under Conditional Symmetry* [w:] *Nonparametric and Semiparametric Estimation Methods in Econometrics and Statistics*, pod red. W.A. Barnetta, J. Powella, G.E. Tauchena, Cambridge University Press, Cambridge.
[35] Newey W.K., Powell J.L., Walker J.R. (1990), *Semiparametric Estimation of Selection Models: Some Empirical Results*, The American Economic Review, vol.80, nr 2, s. 324-328.
[36] Pagan A.R., Ullah A. (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
[37] Pagan A.R., Vella F. (1989), *Diagnostic Tests for Models Based on Individual Data: A Survey*, Journal of Applied Econometrics, vol. 4, s. S29-S59.
[38] Powell J.L. (1984), *Least Absolute Deviations Estimation for the Censored Regression Model*, Journal of Econometrics, vol. 25, nr 3, s. 303-325.
[39] Powell J.L. (1986), *Censored Regression Quantiles*, Journal of Econometrics, vol. 32, s. 143-155.
[40] Powell J.L. (1986), *Symmetrically Trimmed Least Squares Estimation for Tobit Models*, Econometrica, vol. 54, nr 6, s. 1435-1460.
[41] Powell J.L. (1994), *Estimation of Semiparametric Models* [w:] *Handbook of Econometrics*, vol. IV, red. R.F. Engle, D.L. McFadden, Elsevier Science Publishers B.V., s. 2444-2514.
[42] Powell J.L., Stock J.H., Stoker T.M. (1989), *Semiparametric Estimation of Index Coefficients*, Econometrica, vol. 57, nr 6, s. 1403-1430.
[43] Robinson P.M. (1987), *Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form*, Econometrica, vol. 55, s. 875-891.
[44] Schafgans M. (2004), *Finite Sample Properties for the Semiparametric Estimation of the Intercept of a Censored Regression Model*, Statistica Neerlandica, vol. 58, nr. 1, s. 35-56.
[45] Tobin J. (1958), *Estimation of Relationships for Limited Dependent Variables*, Econometrica, vol. 26, s. 24-36.
[46] Vella F. (1998), *Estimating Models with Sample Selection Bias: A Survey*, Journal of Human Resources, vol. 33, nr 1, s. 127-169.
[47] Verbeek M. (2001), *A Guide to Modern Econometrics*, John Wiley & Sons, Chichester.

PRZEGLĄD SEMIPARAMETRYCZNYCH METOD ESTYMACJI MODELI TOBITOWYCH TYPU I–III

Streszczenie

Estymacja parametrów modeli tobitowych, tzn. modeli opisujących zmienną zależną cenzurowaną, za pomocą metody największej wiarygodności wymaga m.in. założenia o normalności rozkładu składników losowych. Sięgnięcie po metody semiparametryczne pozwalające na osłabienie założeń jest szczególnie przydatne. W kolejnych punktach krótko przedstawiono modele tobitowe typu I, II oraz III. Następnie zaprezentowano wybrane metody estymacji semiparametrycznej tych modeli oraz wskazano przydatność przedstawionych metod przy weryfikacji założeń narzuconych na rozkład składnika losowego.

THE REVIEW OF THE SEMI-PARAMETRIC METHODS FOR THE TOBIT TYPE I–III MODELS

S u m m a r y

The estimation of parameters of the tobit type models, that is models for a censored dependent variable, by the MLE method requires an assumption of the normal distribution of disturbances. Semi-parametric methods are very useful, because allow to weaken these assumptions. In the paper, at first the tobit type I, II and III models are introduced, next some known in the literature semi-parametric methods for these models are presented. At the end of the paper there is shown the usefulness of semi-parametric methods to a verification of assumptions imposed on the distribution of disturbances.