

Drucker, P. F. (1998). The Coming of the New Organization, Harvard Business Review on Knowledge Management, Harvard Business School Press, USA.

Egbu, C.O. (2002). *Information Technologies for Knowledge Management: Their Usage and Effectiveness*, Electronic Journal of Information Technology in Construction.

Egbu, C. O., & BOTTERILL, K. (2002). *Information Technologies for Knowledge Management: Their Usage and Effectiveness*, Electronic Journal of Information Technology in Construction.

Garvin, D. A. (1998). Building a Learning Organization, Harvard Business Review on Knowledge Management, Harvard Business School Press, USA.

Kleiner, A., & ROTH, G. (1998). How to Make Experience Your Company's Best Teacher, Harvard Business Review on Knowledge Management, Harvard Business School Press, USA.

Koivu, T. J. (2002). *Future of Product Modeling and Knowledge Sharing in the FM/AEC Industry*, Electronic Journal of Information Technology in Construction.

Koenders, E. A. B., at al (2004). *The (Re) Use of Knowledge for Academics and Build-*

*ing Industry*, Delft University of Technology, Delft, The Netherlands.

Leonard, D., & STRAUS, S. (1998). Putting Your Company's Whole Brain to Work, Harvard Business Review on Knowledge Management, Harvard Business School Press, USA.

Lima, C., et al (2005). *Ontology-Based Optimization of in e-Construction*, Electronic Journal of Information Technology in Construction.

Lima, C., et al (2003). The BCXML: Supporting Ecommerce and Knowledge Management in the construction industry, Electronic Journal of Information Technology in Construction.

Nonaka, I. (1998). The Knowledge-Creating Company, Harvard Business Review on Knowledge Management, Harvard Business School Press, USA.

Quinn, J. B., at al (1998). Managing Professional Intellect: Making the Most of the Best, Harvard Business Review on Knowledge Management, Harvard Business School Press, USA.

Pakanen, J. E., at al (2001). *A web-based Information System for Diagnosing, Servicing and Operating Heating Systems*, Electronic Journal of Information Technology in Construction.

Slak, M. (2002). *Ko bi le vedeli, kaj vemo*, Sistem, Infomediji, d. o. o., Slovenija.

Stojanovic, N., HANDSCHUH, S. (2002). *A Framework for Knowledge Management on the Semantic Web*, University of Karlsruhe, Karlsruhe, Germany.

Staab, S., et al (2001). *Knowledge Processes and Ontologies*, IEEE Intelligent Systems 16(1), January/February 2001, Special Issue on Knowledge Management.

Wilson, T. D. (2002). The Nonsense of 'Knowledge Management', Information Research, 8(1), paper no. 144 (<http://InformationR.net/ir/8-1/paper144.html>).

Zack, M. H. (1999). Managing Codified Knowledge, Sloan Management Review, 40(4) (<http://web.cba.neu.edu/~mzack/articles/kmarch/kmarch.htm>).

#### Povezave na spletu

Microsoft SharePoint  
[<http://www.microsoft.com/windowsserver2003/technologies/sharepoint/default.aspx>],

[<http://www.microsoft.com/resources/documentation/wss/2/all/adminguide/en-us/stsa02.msp>].

Urbanistični inštitut Slovenije [www.uirs.si].

Boštjan KERBLER – Kefo

## Modeli diskretne izbire

V članku je sistematično predstavljena posebna oblika regresijskih metod – modelov diskretne izbire –, imenovanih tudi verjetnostni modeli. Poleg njihovega pomena so opisane še metodološke značilnosti pri njihovi izvedbi, natančneje pa so predstavljeni modeli binarne izbire in tisti z omejeno odvisno spremenljivko, logistični model ter modela probit in tobit kot izhodiščni metodološki pristopi k izvedbi modelov.

The paper systematically describes special regression methods – discrete choice models – known as probability models. The meaning of models and their methodological characteristics are described, as well as different types of models, especially binary-choice models and censored regression models. We considered three most commonly used approaches to estimating such models – logit, probit and tobit model.

*Logistični model*  
**Modeli diskretne izbire**  
**Modeli binarne izbire**  
**Modeli z omejeno odvisno spremenljivko**  
**Model probit**  
**Regresijska analiza**  
**Statistične metode**  
**Model tobit**

**Binary-choice models**  
**Censored regression models**  
**Discrete choice models**  
**Logit model**  
**Probit model**  
**Regression analysis**  
**Statistical methods**  
**Tobit model**

### 1. Poimenovanja in pomen modelov diskretne izbire

Za modele diskretne izbire (angl. *discrete choice models*) avtorji upo-

rabljajo različna poimenovanja. Izraz povzemamo po Greenu (2003), ki jih imenuje tudi modeli kvalitativnega odziva (angl. qualitative response (QR) models), Pindyck in Rubinfeld (1991) pa jih imenujeta modeli kvalitativne izbire (angl.

*qualitative choice models*). Maddala (1999) uporablja termin diskretni regresijski modeli (angl. discrete regression models), Liao (1994) pa govori o verjetnostnih modelih (angl. *probability models*). Modeli diskretne izbire sodijo med regresijske

modele – Fox (1997) jih uvršča med posplošene linearne (regresijske) modele (angl. *generalized linear (regression) models* – GLM), Wooldridge (2002) pa med nelinearne regresijske modele (angl. *nonlinear regression models*). Podobno kot pri klasičnih linearnih regresijskih modelih analiziramo tudi pri modelih diskretne izbire kavzalne zveze med pojasnjeno, odvisno spremenljivko ( $Y$ ) in eno ali več pojasnjevalnimi, neodvisnimi spremenljivkami ( $X_1 \dots X_k$ ). Glavni pomen modelov diskretne izbire je, da v nasprotju s klasičnimi linearnimi regresijskimi modeli omogočajo vpogled v kavzalne zveze, če ne poznamo zveznih, kvantitativnih vrednosti odvisne spremenljivke, temveč ločimo le končno število izidov, ki zavzemajo diskretne, kvalitativne vrednosti (Juvančič, 2002, povzeto po Maddala, 1999, Wooldridge, 2002). Takšno odvisno spremenljivko imenujemo omejena odvisna spremenljivka (angl. *limited dependent variable*), uporaba konvencionalnih regresijskih metod pa v takih primerih ni ustrezna (Greene, 2003).

Ker je po Foxu (1997) za regresijske modele značilno, da lahko na podlagi sprejetega modela in ocen njegovih parametrov iz vrednosti pojasnjevalnih spremenljivk napovemo vrednost odvisne spremenljivke, omogočajo modeli diskretne izbire – kot verjetnostni modeli – napovedovanje verjetnosti odziva oziroma izbire (angl. *forecasting response/choice probability*) (Liao, 1994, Wooldridge, 2002).

Prob (dogodek  $j$  se zgodi) =  
= Prob ( $Y = j$ ) =  $P$ [relevantne posledice, parametri]

Pri tem pomeni Prob ( $Y = j$ ) verjetnost, da se dogodek  $j$  zgodi, dogodek (angl. *event*) pa posameznikovo izbiro med alternativami – izidi, ki jih zavzema odvisna spremenljivka  $Y$ .

Opis modelov diskretne izbire je tako prvič objavljen v slovenščini.<sup>[1]</sup>

## 2. Metodološke značilnosti pri izvedbi modelov diskretne izbire

Glavne metodološke značilnosti modelov diskretne izbire so: Osnovna oblika zapisa modelov diskretne izbire temelji na regresijskem modelu, ki ga z enačbo zapišemo kot

$$Y_i = \beta X_i + \varepsilon_i$$

Pri tem je  $Y_i$  vektor vrednosti odvisne spremenljivke,  $X_i$  matrika vrednosti pojasnjevalnih spremenljivk,  $\beta$  vektor koeficientov, ki jih ocenjujemo,  $\varepsilon_i$  pa vektor slučajnih napak (ostankov) oziroma vplivov (Košmelj, 2001).

Ocenjevanje postavljenega modela oziroma njegovih parametrov temelji na metodi največjega verjetja (angl. *maximum likelihood method*).

Za preizkušanje hipotez in ugotavljanje statistične značilnosti postavljenega modela uporabljamo test z razmerjem verjetij (angl. *likelihood ratio test*).

Skladnost postavljenega modela izražamo z različnimi oblikami determinacijskega koeficienta psevdor<sup>2</sup> (angl. *pseudo-R<sup>2</sup>*), npr. z indeksom razmerja verjetij (angl. *likelihood ratio index*), različnimi oblikami informacijskih kriterijev (angl. *information criteria*) in deležem pravilno napovedanih odzivov – cenitev R<sup>2</sup> (angl. *count R<sup>2</sup>*). Postavljeni model interpretiramo s cenilkami mejnih učinkov (angl. *estimates of marginal effects*) pojasnjevalnih spremenljivk.

## 3. Vrste modelov diskretne izbire

Modele diskretne izbire razvrščajo avtorji, kot npr. Greene (2003), Pindyck in Rubinfeld (1991), Maddala (1999), Fox (1997), Liao (1994) Wooldridge (2002), glede na lastnosti odvisne spremenljivke  $Y_i$ . Čeprav se razvrstitve v podrobnostih med seboj razlikujejo, je njihova

skupna značilnost ta, da so modeli diskretne izbire v osnovi razdeljeni glede na število izidov oziroma diskretnih vrednosti, ki jih zavzema odvisna spremenljivka:

Modeli diskretne izbire, pri katerih zavzema odvisna spremenljivka dva izida – modeli binarne izbire (angl. *binary-choice/response models, dichotomous choice models*, itd.):

$$Y_i = 0, 1$$

Pri tem  $Y_i = 1$  pomeni, da se dogodek zgodi,  $Y_i = 0$  pa, da se dogodek ne zgodi.

Spremenljivko, ki je prirejena vsaki vrednosti nominalne spremenljivke in dobi vrednost 1, če ima enota izbrano vrednost nominalne spremenljivke, ter vrednost 0, če enota nima izbrane vrednosti nominalne spremenljivke, imenujemo slamnata (umetna) spremenljivka (angl. *dummy variable*) (Košmelj in drugi, 2001).

Izhodiščna metodološka pristopa pri izvedbi modelov binarne izbire sta logistični model, imenovan tudi logistična regresija (angl. *logit model/logistic regression*), in model probit, imenovan tudi model normit (angl. *probit/normit model*).

Pri obeh metodoloških pristopih običajno ocenjujemo vpliv ene pojasnjevalne spremenljivke ali več na eno odvisno spremenljivko z dvema izidoma. Po Greenu (2003) lahko pri modelu probit opazujemo tudi vpliv ene pojasnjevalne spremenljivke ali več na več kot eno odvisno spremenljivko, ki zavzema dva izida. Kadar imamo dve odvisni spremenljivki, govorimo o bivariatnem modelu probit (angl. *bivariate probit model*), kadar so v analizo vključeni več kot dve odvisni spremenljivki, pa o multivariatnem modelu probit (angl. *multivariate probit model*). Modeli diskretne izbire, pri katerih zavzema odvisna spremenljivka več kot dva izida – modeli multiple izbire (angl. *multiple-choice/response models, polytomous (polychotomous) choice models*, itd.):

$$Y_i = 0, 1, 2, 3, \dots j$$

Pri tem so izidi lahko razvrščeni na urejenostni merski lestvici (angl. *ordered outcomes*) ali pa niso rangirani (angl. *unordered outcomes*).

Tako kot pri modelih binarne izbire sta tudi pri izvedbi modelov multiple izbire izhodiščna metodološka pristopa logistični model in model probit<sup>[2]</sup>, ki sta izpeljana v več različicah. Najpogosteje sta v rabi multinomski logistični model (angl. *multinomial logit model* – MNL) in multinomski model probit (angl. *multinomial probit model* – MNP).

Nekateri avtorji, kot npr. Gujarati (1995) in Pindyck, Rubinfeld (1991), uvrščajo med modele diskretne izbire tudi tiste, pri katerih ima odvisna spremenljivka omejene (angl. *censored*)<sup>[3]</sup> vrednosti:

Modeli diskretne izbire, pri katerih zavzema odvisna spremenljivka eno diskretno vrednost in eno ali več zveznih vrednosti – po Gujarati (1995) modeli z omejeno odvisno spremenljivko (angl. *censored regression models/limited dependent variable models*):

$$Y_i = 0$$

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

če je desna stran enačbe (*DSE*)  $> 0$  Pri tem je  $Y_i = 0$  diskretna vrednost, *DSE* pa zavzema zvezne vrednosti.

Metodološki pristop pri izvedbi modelov diskretne izbire te vrste je model tobit (angl. *tobit model*).

### 3.1 Modeli binarne izbire

Regresijska oblika zapisa modelov binarne izbire je

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Pri tem predpostavljamo, da je pričakovana vrednost slučajne napake  $E(\varepsilon_i)$  enaka 0, saj s tem dosežemo, da so cenilke nepristranske (Gujarati, 1995).

Skupna značilnost modelov je, da zavzema odvisna spremenljivka  $Y_i$  le dve vrednosti (0 in 1), zato zapišemo po Pindycku in Rubinfeldu (1991) verjetnostno porazdelitev  $Y_i$  kot

$$P_i = \text{Prob}(Y_i = 1) \text{ in} \\ 1 - P_i = \text{Prob}(Y_i = 0)$$

Pri tem je verjetnost  $P_i$  – verjetnost, da se bo dogodek pri danih  $X_i$  zgodil – enaka razmerju med številom za dogodek  $Y_i = 1$  ugodnih možnosti in številom vseh možnosti, ob pogoju, da imajo vse možnosti enako priložnosti za nastop (Košmelj in drugi, 2001). Verjetnost  $P_i$  zavzema vrednosti na intervalu med 0 in 1 (Pindyck, Rubinfeld, 1991).

Ker zavzame  $Y_i$  le dve vrednosti, interpretiramo po Gujarati (1995) pri modelih binarne izbire verjetnost  $P_i$  tudi kot pogojno pričakovano vrednost odvisne spremenljivke  $Y_i$

$$E(Y_i) = 1(P_i) + 0(1 - P_i) = P_i$$

Pri logističnem modelu in modelu probit binarne izbire je po Gujarati (1995) odnos med verjetnostjo  $P_i$  in pojasnjevalnimi spremenljivkami  $X_i$  nelinearen. Porazdelitvena funkcija ima obliko črke S in je po Foxu (1997) asimptota – vrednostma 0 in 1 se približuje v neskončnost (njen razpon je med  $-\infty$  in  $+\infty$ ). Pri obeh metodoloških pristopih je funkcija porazdeljena kumulativno (angl. *cumulative distribution function* – CDF oziroma *F*). Nagib *F* je največji pri  $P_i = 0,5$ , ko je  $X_i = 0$ . To pomeni, da imajo spremembe vrednosti pojasnjevalnih spremenljivk  $X_i$  največji vpliv na verjetnost izbire v razpolovišču porazdelitve, zaradi majhne nagnjenost porazdelitvene krivulje na njenih koncih pa so spremembe v verjetnosti majhne, tudi če so spremembe v vrednostih pojasnjevalnih spremenljivk  $X_i$  velike (Pindyck, Rubinfeld, 1991).

Kumulativna porazdelitvena funkcija *F* se pri obeh izhodiščnih metodoloških pristopih izvedbe modelov binarne izbire – logističnem in modelu probit – razlikuje v tem, da prvi predpostavlja logistično, drugi pa normalno kumulativno porazdelitev.

Logistično kumulativno verjetnostno porazdelitveno funkcijo zapišemo kot

$$P_i = F(\alpha + \beta X_i) = F(Z_i) = \\ = \frac{1}{1 + e^{-(\alpha + \beta X_i)}} = \Lambda(Z_i)$$

Pri tem je  $\Lambda$  dogovorjeni simbol za logistično porazdelitev (Greene, 2003),  $Z_i$  je po Pindycku in Rubinfeldu (1991) teoretični zvezni indeks (angl. *theoretical continuous index* –  $Z_i = \alpha + \beta X_i$ ),  $F(Z_i)$  kumulativna porazdelitvena funkcija zveznega indeksa, *e* pa osnova za naravni logaritem – matematična konstanta, ki znaša približno 2,718 oziroma po Foxu (1997) eksponenta negativne vrednosti zveznega indeksa ( $-Z_i$ ).

Normalno standardizirano kumulativno verjetnostno porazdelitveno funkcijo zapišemo kot

$$P_i = F(\alpha + \beta X_i) = F(Z_i) = \\ = \int_{-\infty}^{Z_i} \phi(t) dt = \Phi(Z_i)$$

Pri tem je  $\Phi$  dogovorjeni simbol za normalno standardno porazdelitev (Greene, 2003), *t* pa je normalno porazdeljena slučajna, naključna spremenljivka (angl. *random variable*), s povprečno vrednostjo 0 in enoto variance  $\sigma^2$  (Pindyck, Rubinfeld, 1991). Po Košmeljevi in drugih (2001) zavzame slučajna spremenljivka vrednosti v kateremkoli intervalu z znano verjetnostjo. Vrednosti so odvisne od slučajja, naključja (Košmelj, 2001).

Porazdelitveni funkciji sta si zelo podobni, izrazitejša je le razlika v repih porazdelitev – pri logistični je rep rahlo sploščen, pri normalni pa strmejši, zato se krivulja hitreje približuje vrednostma 0 in 1. Izбира med logističnim modelom in modelom probit binarne izbire je odvisna predvsem od računalniške programske opreme, ki je raziskovalcu na voljo za analizo (Gujarati, 1995). Po Maddalaju (1999) se namreč ob predpostavki, da je vzorec preučevanja dovolj velik, rezultati obeh metodoloških pristopov izvedbe modelov binarne izbire ne razlikujejo bistveno.

Tudi pri postopku ocenjevanja postavljenih modelov z metodo največjega verjetja med logističnim modelom in modelom probit binarne izbire ni bistvenih razlik. Po Pindycku in Rubinfeldu (1991) pomeni pri logističnem modelu  $P_i$

verjetnost, ki je povezana s kumulativno logistično, pri modelih probit pa s kumulativno normalno porazdelitveno funkcijo  $F$ . Funkcija verjetja  $L$  ima pri obeh metodoloških pristopih izvedbe modelov binarne izbire enako obliko

$$L = P_1 \dots (1 - P_N) = \prod_{Y_i=1} P_i \prod_{Y_i=0} (1 - P_i) = \prod_{i=1}^N P_i^{Y_i} (1 - P_i)^{(1 - Y_i)}$$

Pri tem pomeni  $N$  število vseh ugodnih možnosti za dogodke – za dogodek  $Y_i = 1$  in dogodek  $Y_i = 0$  –,  $\prod$  pa produkt ugodnih možnosti za dogodke.

Za maksimiranja funkcije verjetja je ustreznejše, da izvedemo računski postopek z logaritemsko transformacijo funkcije verjetja – logaritemska funkcija verjetja (angl. *log-likelihood function* –  $\log L$ ), pri čemer je  $\log L$  naravni logaritem funkcije verjetja ( $\ln L$ ). Logaritemska funkcija namreč ohranja ureditve, vrednost funkcije pa ni nikoli negativna. Po Pindycku in Rubinfeldu (1991) ter Maddalaju (1999) zapišemo za logistični model in model probit binarne izbire logaritemsko funkcijo verjetja kot

$$\ln L = \sum_{Y_i=1} \ln P_i + \sum_{Y_i=0} \ln(1 - P_i)$$

Čeprav cenilke obeh metodoloških pristopov izvedbe modelov binarne izbire zaradi razlik v variancah porazdelitev niso neposredno primerljive med seboj, Gujarati (1995) in Maddala (1999) po Amemiyanu (1981) predlagata, da lahko primerljivost med cenilkami logističnega modela in modela probit povečamo, če cenilke, pridobljene z logističnim modelom, pomnožimo z vrednostjo 0,625.

Pri logističnem modelu in modelu probit binarne izbire temelji na skupni enačbi tudi izračun mejnih učinkov pojasnjevalnih spremenljivk na verjetnost dogodka, in sicer

$$\frac{\partial \text{Prob}(Y_i = 1)}{\partial X_i} = \frac{\partial E(Y_i)}{\partial X_i} = \frac{\partial P_i}{\partial X_i} = f(\alpha + \beta X_i) \beta$$

Pri tem je konstanta  $\alpha = 1$ ,  $f$  je funkcija verjetnosti gostote, ki se ujema s kumulativno verjetnostno porazdelitveno funkcijo  $F$ , pojasnjevalne spremenljivke  $X_i$  pa zavzemajo zvezne vrednosti (Liao, 1994, Greene, 2003, Anderson, Newell, 2003, Cornelissen, 2005).

Kadar zavzemajo pojasnjevalne spremenljivke  $X_i$  binarne vrednosti, Greene (2003) predlaga izračun mejnih učinkov po enačbi

$$\begin{aligned} \text{Mejni } u &= \text{Prob}[Y_i = 1 | \bar{X}_{(d)}] \\ d &= \lfloor 1 \rfloor - \text{Prob}[Y_i = 1 | \bar{X}_{(d)}] \\ d &= 0 \end{aligned}$$

Pri tem pomeni  $d$  pojasnjevalno spremenljivko z binarno vrednostjo,  $\bar{X}_{(d)}$  pa povprečno vrednost preostalih pojasnjevalnih spremenljivk v modelu. Vendar so, kljub binarni vrednosti pojasnjevalnih spremenljivk, vrednosti mejnih učinkov, izračunane po enačbi z delnim odvajanjem, pogosto prav tako natančne.

### 3.2 Modeli z omejeno odvisno spremenljivko

Modeli z omejeno odvisno spremenljivko so podobni konvencionalnim regresijskim modelom, vendar razpolagamo pri njih le s podatki tistih opazovanih enot, pri katerih se dogodek zgodi. V takih primerih zavzema odvisna spremenljivka zvezne vrednosti, vendar enote v analizo vključimo le, če so zvezne vrednosti odvisnih spremenljivk pozitivne. Opazovanih enot, za katere nimamo podatkov o vrednostih, ki jih zavzemajo odvisne spremenljivke (gre za opazovane enote, pri katerih se dogodek ne zgodi), razpolagamo pa z vrednostmi, ki jih zavzemajo pojasnjevalne spremenljivke, ne izključimo iz analize, ampak takšne odvisne spremenljivke omejimo, najpogosteje z vrednostjo 0 ( $Y_i = 0$ ), ki pa je diskretna in ne zvezna vrednost (Pindyck in Rubinfeld, 1991, Gujarati, 1995, Greene, 2003).

Modele z omejeno odvisno spremenljivko imenujemo po Gujarati (1995) tudi modeli tobit, saj je ta metodološki pristop pri izvedbi modelov te vrste najpogostejši.

Ime je skovanka med priimkom avtorja tega modela J. Tobinom in poimenovanjem model probit, modelom, na katerem temelji model tobit (Bierens 2004). Model tobit predpostavlja, da opazovana odvisna spremenljivka  $Y_i$  ne more zavzeti vrednosti, manjše od 0, in mora za  $i = 1 \dots, N$  opazovanih enot ustrezati kriteriju

$$Y_i = \max(Y_i^*, 0)$$

Pri tem je  $Y_i^*$  latentna spremenljivka, ki ni opazovana in je izpeljana iz klasičnega linearnega regresijskega modela

$$Y_i^* = \alpha + \beta X_i + \varepsilon_i$$

Pri tem modelu predpostavljamo, da je konstanta  $\alpha = 1$ , slučajna napaka  $\varepsilon_i$  pa neodvisno in normalno porazdeljena s povprečno vrednostjo 0 in varianco  $\sigma^2$  –  $N(0, \sigma^2)$  ter pogojena z neodvisnimi spremenljivkami  $X_i$  (Bierens, 2004, Oladele, 2005, vir 1).

Odnos med opazovano odvisno spremenljivko  $Y_i$  in latentno spremenljivko  $Y_i^*$  matematično zapišemo v obliki enačbe

$$Y_i = \begin{cases} Y_i^* & \text{če je } Y_i^* > 0 \\ 0 & \text{če je } Y_i^* \leq 0 \end{cases}$$

Ker so podatki omejeni, je pri modelu tobit porazdelitvena funkcija mešanica med diskretno in zvezno porazdelitvijo (Greene, 2003), vendar temelji po Bierensu (2004) in Oladeleju (2005) na normalni kumulativni porazdelitveni funkciji, ki jo ob predpostavki, da se dogodek zgodi ( $Y_i > 0$ ), zapišemo kot

$$F(Z_i) = \int_{-\infty}^{Z_i} \phi(t) dt = \Phi(Z_i)$$

$$\text{Pri tem je } Z_i = \frac{\beta X_i}{\sigma},$$

-simbol  $\sigma$  pa standardni odklon.

Pri modelu tobit kumulativna porazdelitvena funkcija in pričakovana vrednost odvisne spremenljivke  $Y_i E(E_i)$  nista izenačeni. Ob predpostavki, da je  $Y_i > 0$ , izračunamo to po formuli

$$E(Y_i) = (\beta X_i) F(Z_i) + \sigma f(Z_i)$$

Pri tem je  $f(Z_i)$  funkcija verjetnostne gostote za normalno porazdeli-

tev (Pindyck in Rubinfeld, 1991, Bierens, 2004, Oladele, 2005).

Ker je pri modelu tobit porazdelitev delno diskretna in delno zvezna, zapišemo po Greenu (2003) logaritemsko funkcijo verjetja za model tobit v oblik enačbe kot

$$\ln L = \sum_{Y_i > 0} - \frac{1}{2} \left[ \log(2\pi) + \ln \sigma^2 + \frac{(Y_i - \beta X_i)^2}{\sigma^2} \right] + \sum_{Y_i = 0} \ln[1 - \Phi(Z_i)]$$

Pri tem je enačba razdeljena na dva dela, in sicer prvi del pripada opazovanim enotam, ki nimajo omejitev oziroma zavzemajo odvisne spremenljivke zvezne vrednosti ( $Y_i > 0$ ), drugi del pa opazovanim enotam, katerih odvisne spremenljivke so omejene ( $Y_i > 0$ ).

Zaradi omejitve odvisne spremenljivke ocenjujemo model tobit, kljub podobnostim s konvencionalnimi regresijskim modeli, po metodi največjega verjetja, pred tem pa po Greenu (2003) in Bierensu (2004) logaritemsko funkcijo verjetja poenostavimo s t. i. Olsnovo reparametrizacijo, po kateri sta

$$\gamma = \frac{\beta}{\sigma} \quad \text{in} \quad \theta = \frac{1}{\sigma},$$

$$\ln L = \sum_{Y_i > 0} - \frac{1}{2} [\ln(2\pi) - \ln \theta^2 + (\theta Y_i - X_i \gamma)^2] + \sum_{Y_i = 0} \ln[1 - \Phi(X_i \gamma)]$$

Pri modelu tobit izračunamo determinacijski koeficient  $R^2$  enako kot pri konvencionalnih regresijskih modelih, in sicer po enačbi

$$R^2 = 1 - \frac{\sum \hat{\epsilon}_i^2}{\sum (Y_i - \bar{Y})^2} =$$

Pri tem pomeni  $\bar{Y}$  povprečje vrednosti odvisnih spremenljivk  $Y_i$ . V nasprotju s konvencionalnimi regresijskimi modeli je pri modelu tobit izračun mejnih učinkov po enačbi

$$\frac{\partial E(Y_i^*)}{\partial X_i} = \beta$$

omejen, saj je  $Y_i^*$  latentna spremenljivka. Greene (2003) zato predlaga, da moramo namesto  $Y_i^*$  upoštevati opazovano odvisno spremenljivko  $Y_i$ .

$$\frac{\partial E(Y_i)}{\partial X_i} = \beta \Phi\left(\frac{\alpha + \beta X_i}{\sigma}\right)$$

Pri tem predpostavljamo, da je konstanta  $\alpha = 1$ , slučajna napaka je normalno porazdeljena, opazovana odvisna spremenljivka  $Y_i$  pa je omejena z vrednostjo 0.

Greene (2003) povzema po McDonaldu in Mofittu (1980) razčlenbo enačbe na dva dela

$$\frac{\partial E(Y_i)}{\partial X_i} = \text{Prob}(Y_i > 0) \frac{\partial E(Y_i | Y_i > 0)}{\partial X_i} + E(Y_i | Y_i > 0) \frac{\partial \text{Prob}(Y_i > 0)}{\partial X_i}$$

Pri tem ima sprememba pojasnjevalne spremenljivke  $X_i$  dvojni pomen: v pozitivnem delu porazdelitve vpliva na pogojno povprečno vrednost latentne spremenljivke  $Y_i^*$  in na verjetnost, da bo v tem delu porazdelitve tudi opazovana enota.

## 4. Sklep

Čeprav članek temelji na pregledu strokovne literature s področja statistike in ekonometrije, ima za slovensko znanost velik pomen, saj zelo sistematično in podrobno opiše metode diskretne izbire, kar v slovenskem jeziku doslej še ni bilo narejeno. To je vzrok, da smo številne izraze zapisali tudi v izvirnem (angleškem) jeziku.

Modeli diskretne izbire so kot statistično metodo pri nas redko uporabljali tudi v empiričnih raziskavah, najpogosteje sna področju ekonomije in medicine. V doktorski disertaciji smo prvi v Sloveniji uporabili modele na področju socialne geografije in z njimi ugotavljali povezanost med socialnogeografsko strukturo na slovenskih hribovskih kmetijah in nasledstvom na njih. V članku predstavljena teoretična metodološka podlaga je pripomogla k ustrezni uporabi modelov in uspešni izvedbi raziskave.

Boštjan Kerbler - Kefo, univ. dipl. geogr., Urbanistični inštitut RS, Ljubljana  
E-pošta: kefo@uirs.si

### Pojasnilo

[1] Podrobnejši opis metod in njihovo aplikacijo lahko najdete v avtorjevi doktorski disertaciji.

### Opombe

[2] Čeprav je v Sloveniji raba izraza probit model že ustaljena, v skladu s pravili Slovenskega pravopisa (2001) določilo probit zapisujemo v vlogi desnega prilastka. Pravilen zapis se torej glasi model probit. Enako velja tudi za izraz model tobit.

[3] Juvančič (2002) izraz »censored« preva kot okrnjen, čeprav je angleški prevod izraza okrnjen »truncated«.

### Viri in literatura

Anderson, S., Newell, R. (2003) Simplified marginal effects in discrete choice models. Dostopno na: <http://www.rff.org/Documents/RFF-DP-03-38.pdf> (7. 6. 2006).

Bierens, H. J. (2004) The tobit model. Dostopno na: [http://econ.la.psu.edu/~hbierens/EasyRegTours/TOBIT\\_Tourfiles/TOBIT.PDF](http://econ.la.psu.edu/~hbierens/EasyRegTours/TOBIT_Tourfiles/TOBIT.PDF) (20. 6. 2006).

Cornelissen, T. (2005) Standard errors of marginal effects in the heteroskedastic probit model. Dostopno na: <http://www.wiwi.uni-hannover.de/Forschung/Diskussionspapiere/dp-320.pdf> (8. 6. 2006).

Fox, J. (1997) Applied regression analysis, linear models, and related methods. Thousand Oaks, Sage Publications, London, New Delhi.

Greene, W. H. (2003) Econometric analysis. Prentice Hall: Pearson Education International, Upper Saddle River (New Jersey).

Gujarati, D. N. (1995) Basic econometrics. McGraw-Hill Inc., New York.

Juvančič, L. (2002) Model odločanja o zaposlovanju na kmečkih gospodarstvih v Sloveniji. Biotehniška fakulteta, Oddelek za zootehniko, Ljubljana.

Košmelj, B., Arh, F., Doberšek Urbanc, A., Ferligoj, A., Omladič, M. (2001) Statistični terminološki slovar. Statistično društvo Slovenije, Statistični urad Republike Slovenije, Ljubljana.

Košmelj, K. (2001) Osnove logistične regresije (1. del). Zbornik Biotehniške fakultete Univerze v Ljubljani – Kmetijstvo, št. 2/2001, str. 227–238.

Liao, T. F. (1994) Interpreting probability models: logit, probit, and other generalized linear models. Thousand Oaks, Sage Publications, London, New Delhi.

Maddala, G. S. (1999) Limited-dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge, New York, Melbourne.

Oladele, O. I. (2005) A tobit analysis of propensity to discontinue adoption of agricultural technology among farmers in Southwestern Nigeria. Journal of Central European agriculture, št. 3/2005, str. 249–254.

Pindyck, R. S., Rubinfeld, D. L. (1991) Econometric models and econometric forecasts. McGraw-Hill Inc., New York.

Vir 1. [http://en.wikipedia.org/wiki/Tobit\\_model](http://en.wikipedia.org/wiki/Tobit_model) (20. 6. 2006).

Wooldbridge, J. M. (2002) Econometric analysis of cross section and panel data. The MIT Press, Cambridge, London.