PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 7 (1207) 2008
TAKSONOMIA 15
Klasyfikacja i analiza danych – teoria i zastosowania

Joanna Małgorzata Landmesser
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

AKTYWNOŚĆ EKONOMICZNA LUDNOŚCI: KLASYFIKACJA OSÓB ZA POMOCĄ WIELOMIANOWYCH MODELI LOGITOWYCH ORAZ JEJ ZWIĄZEK Z MODELAMI HAZARDU DLA CZASÓW TRWANIA*

1. Wstęp

Badając procesy za pomocą metod z zakresu analizy historii zdarzeń, przyjmuje się zazwyczaj, że rozkład zdarzeń oparty jest na ciągłej zmiennej czasowej T. W rzeczywistości jednak natura dostępnych danych empirycznych sugeruje konieczność zastosowania modeli o czasie dyskretnym. Sytuacja taka występuje wówczas, gdy w analizach chcemy wykorzystać dane dotyczące czasu zatrudnienia, których źródłem jest Badanie Aktywności Ekonomicznej Ludności w Polsce (BAEL). Z tego powodu w niniejszym artykule długość czasu trwania aktywności zawodowej osób będziemy szacować, wykorzystując modele hazardu o czasie dyskretnym. Specjalny typ modeli hazardu dla ryzyka konkurencyjnego pozwoli opisać warunkowe prawdopodobieństwa przejścia między takimi stanami, jak: zatrudnienie, bezrobocie, brak aktywności zawodowej.

Estymacji modeli hazardu dla ryzyka konkurencyjnego o czasie dyskretnym można dokonać m.in. za pomocą wielomianowych modeli logitowych (MLM). Model MLM stanowi dobrą aproksymantę modelu hazardu opartego na danych grupowanych w przedziałach o stałych stopach hazardu. Dysponując oszacowanym modelem MLM, określimy przynależność osób do klas (pracujący, bezrobotni, nieaktywni zawodowo). Ocena jakości uzyskanej klasyfikacji może być pomocna we wnioskowaniu na temat jakości wyjściowego modelu hazardu.

2. Modele hazardu z czasem ciągłym i dyskretnym

Rozkład długości czasu trwania zjawiska można przedstawić za pomocą modelu hazardu. Prezentacja rozkładu ciągłego czasu T wystąpienia zdarzenia jest moż-

* Pracę wykonano w ramach projektu badawczego z puli JM Rektora SGGW „Badanie aktywności zawodowej ludności wiejskiej przy wykorzystaniu metod analizy czasu trwania" (nr 504-08270017).

liwa za pomocą następujących form [Kalbfleisch, Prentice 1980; Kiefer 1988; Frątczak i in. 2005]:

- funkcji rozkładu $F(t)$: $F(t) = \Pr[T \le t] = \int_0^t f(u)du$, (1)
- funkcji przeżycia $S(t)$: $S(t) = \Pr[T > t] = 1 - F(t)$, (2)
- funkcji hazardu $h(t)$: $h(t) = \frac{f(t)}{S(t)} = \lim_{dt \to 0} \frac{\Pr[t \le T < t + dt \mid T \ge t]}{dt}$. (3)

Funkcja hazardu szacuje bezpośrednie ryzyko tego, że wydarzenie nastąpi w przedziale czasowym $[t, t + dt)$ pod warunkiem że do danej chwili t wydarzenie to nie nastąpiło.

Spośród modeli hazardu o czasie ciągłym na uwagę zasługują parametryczne modele proporcjonalnego hazardu (PH). W modelach tych funkcja zmiennych objaśniających (funkcja $g_0(X)$) działa w sposób multiplikatywny na hazard bazowy $h_0(t)$. Funkcję stopy hazardu modelujemy w wypadku PH następująco:

$$h(t, X) = g_0(X)h_0(t) = \exp(\beta' X)h_0(t), \tag{4}$$

co jest równoważne zapisowi

$$\log h(t, X) = \beta' X + \log h_0(t). \tag{5}$$

Gdy dane dotyczące ciągłych czasów trwania są pogrupowane w postaci przedziałów $(t_{j-1}, t_j]$, wówczas do opisu proporcjonalnego hazardu można się posłużyć następującą dyskretną reprezentacją [Jenkins 2004]:

$$\begin{aligned} \theta(t_j, X) &= \frac{S(t_{j-1}, X) - S(t_j, X)}{S(t_{j-1}, X)} = 1 - \frac{S(t_j, X)}{S(t_{j-1}, X)} = \\ &= 1 - \exp\left[-\exp(\beta' X)\left(\int_0^{t_j} h_0(u, X)du - \int_0^{t_{j-1}} h_0(u, X)du \right) \right]. \end{aligned} \tag{6}$$

Na drodze transformacji komplementarnej log-log otrzymujemy z powyższego

$$\log(-\log(1 - \theta(t_j, X))) = \beta' X + \log\left(\int_{t_{j-1}}^{t_j} h_0(u, X)du \right) = \beta' X + \gamma_j. \tag{7}$$

Estymatory dla parametrów β w modelu (7) odpowiadają estymatorom parametrów β w modelu PH z czasem ciągłym (5).

Dla procesów składających się ze zdarzeń występujących tylko w określonych punktach czasu (wtedy dyskretna zmienna czasowa $T = t_j$, dla $j = 1, 2, 3 \ldots$) stosuje się modele z czasem wewnętrznie dyskretnym. Do klasy tej należy model proporcjonalnych ilorazów szans, w którym zakłada się, że iloraz szans dla przejścia w

momencie t_j, pod warunkiem przetrwania do momentu poprzedzającego t_{j-1}, opisuje formuła:

$$\frac{\theta(t_j, X)}{1-\theta(t_j, X)} = \exp(\beta' X)\left[\frac{\theta_0(t_j, X)}{1-\theta_0(t_j, X)}\right], \tag{8}$$

gdzie: $\theta(t_j, X)$ – stopa hazardu dla czasu dyskretnego w chwili t_j,

$\theta_0(t_j, X)$ – korespondujący hazard bazowy [Yamaguchi 1991].

Z powyższego zapisu wynika logitowy model hazardu:

$$\text{logit}[\theta(t_j, X)] = \log\left[\frac{\theta(t_j, X)}{1-\theta(t_j, X)}\right] = \beta' X + \text{logit}[\theta_0(t_j)] = \beta' X + \alpha_j. \tag{9}$$

W praktyce dla relatywnie niskich stóp hazardu komplementarny log-log (7) oraz logitowy model hazardu (9), przy założeniu takiej samej zależności od czasu trwania procesu (co wyraża się poprzez przyjęcie jednakowych parametryzacji dla α_j i γ_j) oraz dla tego samego zestawu zmiennych objaśniających X, mają po oszacowaniu podobne oceny parametrów β [Jenkins 2004; Landmesser 2007]. Dla niskich stóp hazardu logitowy model hazardu (9) jest więc aproksymantą modelu (5).

3. Modele dla ryzyka konkurencyjnego

Jeśli zakończenie badanego procesu następuje poprzez przejście do jednego z kilku niezależnych od siebie stanów, np. wyjście do stanu A lub do stanu B (w sytuacji braku zdarzenia wyjścia mówimy o cenzurowaniu C), to zjawisko takie możemy modelować za pomocą modelu ryzyka konkurencyjnego [Narendranathan, Stewart 1993]. Na przykład formuła dla stopy hazardu wiążącej się z przejściem do stanu A ma postać:

$$h_A(t) = \lim_{dt \to 0} \frac{P(t \le T < t+dt, \delta^A = 1 \mid T \ge t)}{dt}, \tag{10}$$

gdzie δ^A – zmienna 0-1 wskazująca, czy zostało wybrane przejście do stanu A.

Dla modeli hazardu o czasie ciągłym logarytm funkcji wiarygodności dla modelu z wieloma stanami wyjścia może być rozłożony na sumę logarytmów funkcji wiarygodności dla poszczególnych modeli opisujących pojedyncze wyjścia. W przypadku modelu ryzyka konkurencyjnego z czasem wewnętrznie dyskretnym

$$\begin{aligned} L &= (L^A)^{\delta^A}(L^B)^{\delta^B}(L^C)^{\delta^{1-\delta^A-\delta^B}} = \\ &= \left[\frac{h_A(j)}{1-h_A(j)-h_B(j)}\right]^{\delta^A}\left[\frac{h_B(j)}{1-h_A(j)-h_B(j)}\right]^{\delta^B}\prod_{k=1}^{j}[1-h_A(k)-h_B(k)] \end{aligned}. \tag{11}$$

Allison [1982] wykazał, że o ile

$$h_A(k) = \frac{\exp(\beta_A' X)}{1 + \exp(\beta_A' X) + \exp(\beta_B' X)} \text{ oraz } h_B(k) = \frac{\exp(\beta_B' X)}{1 + \exp(\beta_A' X) + \exp(\beta_B' X)}, \quad (12)$$

to funkcja wiarygodności (11) ma taką samą formę jak funkcja wiarygodności dla standardowego wielomianowego modelu logitowego MLM (po uprzednim przeorganizowaniu danych do ich rozszerzonej postaci typu jedna osoba – jeden okres dla każdej obserwacji). Mając na uwadze treści z poprzedniego rozdziału, można twierdzić, że w wypadku niskich stóp hazardu model MLM będzie stanowił również dobrą aproksymantę dla modelu ryzyka konkurencyjnego PH opartego na danych pogrupowanych w przedziałach o stałych stopach hazardu [Tutz 1995; Jenkins 2004].

4. Modelowanie długości czasu zatrudnienia

Na podstawie próby 10 387 pełnoletnich osób pochodzącej z badania BAEL będziemy estymować modele hazardu dla długości czasu zatrudnienia. Badane osoby wykazywały w okresie 1994-2002 przynajmniej jeden epizod zatrudnienia (praca przynajmniej przez rok) oraz na koniec okresu objętego badaniem pracowały (7347 osób) albo były zaszeregowane jako bezrobotne (1076) lub bierne zawodowo, jak np. emeryci, renciści (1964). Długości trwania zatrudnienia w ostatnich miejscach pracy w latach dla poszczególnych osób tworzą zmienną „czas_pracy". Listę zmiennych objaśniających zawiera tab. 1.

Tabela 1. Lista zmiennych objaśniających

Nazwa	Opis zmiennej
pl	zmienna 0-1 przyjmująca wartość 1, gdy respondent jest mężczyzną, 0 w p.p
wiek1	zmienna 0-1 przyjmująca wartość 1, gdy wiek respondenta wynosił 18-24 lata, 0 w p.p.
wiek2	zmienna 0-1 przyjmująca wartość 1, gdy wiek respondenta wynosił 25-34 lata, 0 w p.p.
wiek3	zmienna 0-1 przyjmująca wartość 1, gdy wiek respondenta wynosił 35-44 lata, 0 w p.p.
wiek4	zmienna 0-1 przyjmująca wartość 1, gdy wiek respondenta wynosił 45-54 lata, 0 w p.p.
wiek5	zmienna 0-1 przyjmująca wartość 1, gdy wiek respondenta wynosił 55 i więcej lat, 0 w p.p.
zwiazek	zmienna 0-1; 1, gdy respondent pozostawał w związku małżeńskim, 0 w p.p.
wyksz1	zmienna 0-1; 1, gdy respondent posiadał wykształcenie wyższe, 0 w p.p.
wyksz2	zmienna 0-1; 1, gdy respondent posiadał wykształcenie policealne, średnie zawodowe lub średnie ogólnokształcące, 0 w p.p.
wyksz3	zmienna 0-1; 1, gdy respondent posiadał wykształcenie gimnazjalne, zasadnicze zawodowe lub podstawowe, 0 w p.p.
wyksz4	zmienna 0-1; 1, gdy respondent był bez wykształcenia lub posiadał wykształcenie niepełne podstawowe, 0 w p.p.
wlasrach	zmienna 0-1; 1, gdy respondent pracował na własny rachunek lub jako pomagający członek rodziny, 0 – jako pracownik najemny
wlaspryw	zmienna 0-1; 1, gdy respondent pracował w instytucji o własności prywatnej, 0 w p.p.
wies	zmienna 0-1; 1, gdy miejscem zamieszkania respondenta była wieś, 0 w p.p.

Źródło: opracowanie własne.

Mimo że dane empiryczne są danymi pogrupowanymi w przedziałach o długości 1 roku, to na podstawie wyjściowych 10 387 obserwacji dokonano wstępnie estymacji modelu hazardu PH Weibulla o czasie ciągłym (5) z jednym stanem wyjścia (= zaprzestanie pracy). Po przeorganizowaniu danych uzyskano 137 644 obserwacji i na ich podstawie oszacowano komplementarny log-log model hazardu (7). Przyjęto parametryzację $\gamma_j = \ln(j)$. Wyniki estymacji parametrów β w obu modelach są zbliżone (zob. tab. 2).

Tabela 2. Modele hazardu Weibulla oraz komplementarny log-log dla jednego wyjścia

	Model Weibulla				Model komplementarny log-log		
	β	p-value	exp(β)		β	p-value	exp(β)
pl	–0,279	0,000	0,757	pl	–0,281	0,000	0,755
wiek1	1,789	0,000	5,984	wiek1	1,599	0,000	4,950
wiek2	0,903	0,000	2,468	wiek2	0,822	0,000	2,276
wiek4	–0,152	0,010	0,859	wiek4	–0,096	0,104	0,909
wiek5	0,282	0,000	1,326	wiek5	0,387	0,000	1,473
zwiazek	–0,133	0,002	0,876	zwiazek	–0,128	0,003	0,880
wyksz1	–0,810	0,000	0,445	wyksz1	–0,880	0,000	0,415
wyksz2	–0,236	0,128	0,790	wyksz2	–0,286	0,065	0,751
wyksz3	0,221	0,142	1,247	wyksz3	0,181	0,228	1,198
wlasrach	–1,666	0,000	0,189	wlasrach	–1,625	0,000	0,197
wlaspryw	0,555	0,000	1,741	wlaspryw	0,524	0,000	1,688
wies	–0,167	0,000	0,846	wies	–0,155	0,000	0,856
cons	–4,269	0,000		logt	0,088	0,000	
p	1,230	0,000		cons	–3,762	0,000	
Number of obs = 10 387 Log likelihood = –7598,75 LR chi2(12) = 1916,47 Prob > chi2 = 0,000				Number of obs = 137 644 Log likelihood = –13 689,07 LR chi2(13) = 1816,11 Prob > chi2 = 0,000			

Źródło: obliczenia własne.

Sądząc po ocenach parametrów w modelu komplementarnym log-log, można stwierdzić np., że: ryzyko zaprzestania aktualnego zatrudnienia dla mężczyzny jest o (100 – 75,5)% = 24,5% niższe niż dla kobiety; ryzyko przerwania pracy u osób w wieku 18-24 lata jest o 395% wyższe niż wśród osób w wieku 35-44 lata; to samo ryzyko w wypadku osoby z wyższym wykształceniem jest o 58,5% niższe niż u osoby bez wykształcenia lub z niepełnym podstawowym; ryzyko zaprzestania zatrudnienia dla osoby z firmy prywatnej jest o 68,8% wyższe niż w dla osoby z firmy o innej formie własności.

Następnie przystąpiono do oszacowania modelu ryzyka konkurencyjnego dla dwóch możliwych wyjść: w stan bezrobocia i w stan bierności zawodowej (tab. 3). Można zaobserwować różnice między procesami przejścia. Mężczyźni, w porównaniu z kobietami, wykazują niższe ryzyko przejścia w stan bierności zawodowej. Z wiekiem maleje ryzyko bezrobocia, ale dla osób starszych (55 i więcej lat) notujemy zwiększone ryzyko bierności zawodowej. Poziom wykształcenia respondenta wywiera jedynie statystycznie istotny

wpływ na prawdopodobieństwo przejścia w stan bierności zawodowej, obniżając je w wypadku wyższego wykształcenia. W wypadku osoby z firmy prywatnej ryzyko stania się bezrobotnym jest wyższe niż ryzyko opuszczenia zasobów siły roboczej. Im dłuższy czas trwania aktualnego zatrudnienia, tym mniejsze ryzyko stania się bezrobotnym, ale za to wyższe ryzyko bierności zawodowej.

Tabela 3. Komplementarny log-log model hazardu dla dwóch wyjść

	Dla wyjścia = bezrobocie			Dla wyjścia = bierność zawodowa		
	β	p-value	exp(β)	β	p-value	exp(β)
pl	0,082	0,195	1,085	–0,505	0,000	0,603
wiek1	0,690	0,000	1,994	2,112	0,000	8,264
wiek2	0,307	0,000	1,359	1,238	0,000	3,448
wiek4	–0,479	0,000	0,620	0,354	0,000	1,424
wiek5	–1,784	0,000	0,168	1,094	0,000	2,985
zwiazek	–0,408	0,000	0,665	0,081	0,152	1,085
wyksz1	–0,457	0,654	0,633	–0,659	0,000	0,517
wyksz2	0,408	0,686	1,504	–0,167	0,299	0,846
wyksz3	0,902	0,371	2,464	0,286	0,062	1,331
wlasrach	–1,901	0,000	0,149	–1,511	0,000	0,221
wlaspryw	0,819	0,000	2,267	0,306	0,000	1,358
wies	–0,160	0,017	0,852	–0,163	0,003	0,850
logt	–0,579	0,000		0,515	0,000	
cons	–3,987	0,000		–5,746	0,000	
	Number of obs = 137 644 Log likelihood = –5172,26 LR chi2(13) = 2239,31 Prob > chi2 = 0,000			Number of obs = 137 644 Log likelihood = –9627,74 LR chi2(13) = 1337,14 Prob > chi2 = 0,000		

Źródło: obliczenia własne.

Tabela 4. Wielomianowy model logitowy

Równanie 1			Równanie 2		
	β	p-value		β	p-value
pl	0,0713	0,268	pl	–0,5107	0,000
wiek1	0,7659	0,000	wiek1	2,2035	0,000
wiek2	0,3300	0,000	wiek2	1,2595	0,000
wiek4	–0,4816	0,000	wiek4	0,3500	0,000
wiek5	–1,7770	0,000	wiek5	1,0967	0,000
zwiazek	–0,4128	0,000	zwiazek	0,0786	0,170
wyksz1	–0,4845	0,636	wyksz1	–0,6752	0,000
wyksz2	0,3915	0,699	wyksz2	–0,1771	0,276
wyksz3	0,9047	0,370	wyksz3	0,2874	0,063
wlasrach	–1,9322	0,000	wlasrach	–1,5385	0,000
wlaspryw	0,8353	0,000	wlaspryw	0,3230	0,000
wies	–0,1722	0,012	wies	–0,1691	0,003
logt	–0,5786	0,000	logt	0,5127	0,000
cons	–3,9561	0,000	cons	–5,7175	0,000
Number of obs = 137 644 LR chi2(26) = 3578,89 Log likelihood = –14 783,25 Prob > chi2 = 0,000					

Źródło: obliczenia własne.

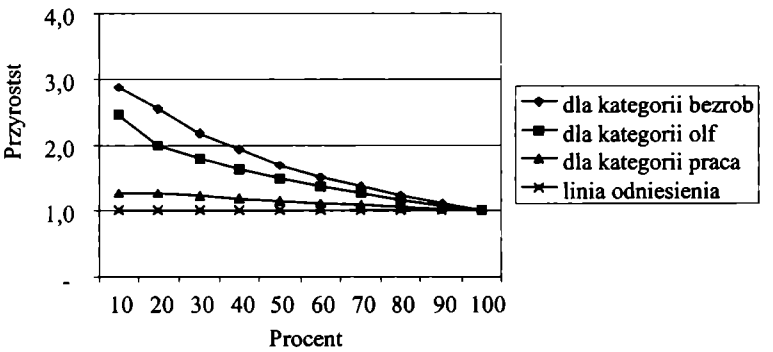
W tab. 4 zaprezentowano wyniki estymacji wielomianowego modelu logitowego MLM. Oszacowania parametrów w komplementarnym log-log modelu hazardu dla dwóch typów ryzyka oraz szacunki parametrów w modelu MLM są podobne, czego oczekiwaliśmy.

Po oszacowaniu modelu każdej obserwacji można przyporządkować najbardziej prawdopodobną kategorię zmiennej zależnej. Utworzenie tabeli krzyżowej prognozy i zmiennej zależnej pozwala stwierdzić, jaki odsetek obserwacji został poprawnie zaklasyfikowany przez model. Miara ta może być jednak myląca w wypadku dokonanych analiz opartych na próbie BAEL ze względu na niezrównoważoną zmienną zależną. Stąd w celu predykcji posłużono się specjalnie wyznaczonymi punktami odcięcia.

Tabela 5. Tabela krzyżowa prognozy i zmiennej zależnej

	Prog_status = praca	Prog_status = bezrob	Prog_status = olf	Razem
Status = praca	$n_{00} = 3097$	$n_{01} = 2561$	$n_{02} = 1689$	$n_{0.} = 7347$
Status = bezrob	$n_{10} = 135$	$n_{11} = 797$	$n_{12} = 144$	$n_{1.} = 1076$
Status = olf	$n_{20} = 367$	$n_{21} = 427$	$n_{22} = 1170$	$n_{2.} = 1964$
Razem	$n_{.0} = 3599$	$n_{.1} = 3785$	$n_{.2} = 3003$	$n = 10\,387$

Źródło: obliczenia własne.



Rys. 1. Skumulowane wykresy przyrostu

Źródło: obliczenia własne.

Z danych tab. 5 wynika, że wartość współczynnika poprawnie rozpoznanych osób w grupie bezrobotnych ($n_{11}/n_{1.}$) wynosi 74,07%, współczynnika poprawnie rozpoznanych osób w grupie nieaktywnych zawodowo (ozn. olf) ($n_{22}/n_{2.}$) 59,57%, ale współczynnika poprawnie rozpoznanych w grupie pracujących ($n_{00}/n_{0.}$) tylko 42,15%. Jednak z punktu widzenia wnioskowania na podstawie modeli hazardu wydaje się mieć znaczenie tylko poprawne klasyfikowanie osób do dwóch grup: bezrobotnych i nieaktywnych (każdy pracujący przecież kiedyś przerwie zatrudnienie). Ponownie wykonane przyporządkowanie, polegające na kierowaniu osób tylko do dwóch grup, cechowało się współczynnikiem poprawnie rozpoznanych

wśród bezrobotnych na wysokim poziomie 81,32% oraz współczynnikiem poprawnie rozpoznanych pośród nieaktywnych na poziomie 74,34%.

Graficznym obrazem użyteczności modeli logitowych do przewidywania wartości zmiennej zależnej skategoryzowanej jest wykres przyrostu (*lift chart*). Na rys. 1 widać, o ile częściej w stosunku do całego zbioru danych przypadki należące do badanej klasy występują w podzbiorach danych zawierających frakcje przypadków (10%, 20% itd.) o największym, wynikającym z modelu prawdopodobieństwie przynależności do tej klasy.

5. Wnioski

Gdy dane empiryczne są zgrupowane w dyskretnych momentach czasu, wówczas szacowanie poszczególnych równań modelu hazardu dla ryzyka konkurencyjnego może być wykonane za pomocą wielomianowego modelu logitowego MLM. Zadowalająca ocena klasyfikacji osób do klasy bezrobotnych oraz klasy nieaktywnych zawodowo, dokonanej na podstawie oszacowanego modelu MLM, pozwala pozytywnie wnioskować na temat jakości wyjściowego modelu hazardu.

Literatura

Allison P. (1982), *Discrete-Time Methods for the Analysis of Event Histories*, „Sociological Methodology", vol. 13, s. 61-98.

Frątczak E., Gach-Ciepiela U., Babiker H. (2005), *Analiza historii zdarzeń., Elementy teorii, wybrane przykłady zastosowań*, Oficyna Wydawnicza SGH, Warszawa.

Landmesser J.M. (2007), *Estymacja modeli ryzyk konkurencyjnych o czasie dyskretnym za pomocą wielomianowych modeli logitowych*, [w:] *Metody ilościowe w badaniach ekonomicznych – VIII*, red. B. Borkowski, Wydawnictwo SGGW, Warszawa, s. 183-192.

Jenkins S.P. (2004), *Survival Analysis*, manuskrypt, University of Essex, Colchester.

Kalbfleisch J., Prentice R. (1980), *The Statistical Analysis of Failure Time Data*, John Wiley and Sons, New York.

Kiefer N. (1988), *Economic Duration Data and Hazard Functions*, „Journal of Economic Literature", 26, s. 646-679.

Narendranathan W., Stewart M.B. (1993), *Modelling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Baseline Hazards*, „Journal of the Royal Statistical Society, Applied Statistics", 42, s. 63-83.

Tutz G. (1995), *Competing Risks Models in Discrete Time with Nominal or Ordinal Categories of Response*, „Quality & Quantity", 29, s. 405-420.

Yamaguchi K. (1991), *Event History Analysis*, „Applied Social Research Methods Series", vol. 28, Sage Publ., London-New Delhi.

ECONOMIC ACTIVITY OF PEOPLE: CLASSIFICATION USING MULTINOMIAL LOGIT MODELS AND THEIR CONNECTION WITH HAZARD MODELS

Summary

In this paper the usage of a multinomial logit model MLM for the estimation of a discrete time competing risks hazard models is described. The MLM model provides a close approximation to a hazard model for interval-censored data for which the hazard rate is constant within each interval. The usefulness of the information provided by the MLM model for predicting a categorical outcome variable reflects a quality of the prior hazard model. To illustrate this statement we use employment duration data from the Labour Force Survey in Poland (BAEL).