



DOI:10.18276/sip.2016.45/2-01

**Beata Bieszk-Stolorz\***

Uniwersytet Szczeciński

## MODELE HAZARDU A MODEL LOGITOWY

### Streszczenie

Celem artykułu jest porównanie dwóch grup modeli stosowanych w analizie historii zdarzeń. Pierwsza z nich to modele z czasem ciągłym opisujące intensywność zachodzenia zjawiska (hazard) w dowolnym momencie czasu. W badaniu wykorzystano model proporcjonalnego hazardu Coxa. Druga grupa to modele z czasem dyskretnym. Analizę oparto na modelu regresji logistycznej (prawdopodobieństwo zajścia zdarzenia w czasie dyskretnym) i modelu hazardu empirycznego (dla danych pogrupowanych). Materiał badawczy stanowiły dane indywidualne osób bezrobotnych zarejestrowanych w PUP w Szczecinie w 2012 roku obserwowane do końca 2013 roku. Wyznaczono intensywność względną wychodzenia z bezrobocia i szansę względną na podjęcie zatrudnienia w zależności od płci, wieku, wykształcenia i stażu pracy osoby bezrobotnej.

**Słowa kluczowe:** model regresji Coxa, hazard empiryczny, regresja logistyczna, bezrobocie

### Wstęp

W większości metod analizy historii zdarzeń zakłada się, że czas jest zmienną ciągłą, to znaczy może przyjmować dowolne wartości nieujemne. Jednak w praktyce obserwacje są prowadzone co pewien czas i jeżeli nie są zbyt częste, bardziej odpowiednie jest wykorzystanie w analizie modeli dyskretnych. Sytuacja taka ma miejsce, gdy analizowane zjawisko występuje tylko w czasie dyskretnym (czas trwania kadencji prezydenta) lub gdy zjawisko występuje w czasie ciągłym, ale do-

---

\* Adres e-mail: [beatus@wneiz.pl](mailto:beatus@wneiz.pl).

stępane są tylko dane o charakterze dyskretnym. Tak jest w przypadku wielu badań ankietowych, kiedy respondenci są pytani na przykład o czas trwania małżeństwa. Dlatego też wśród metod analizy historii zdarzeń omawiane są modele z czasem ciągłym i czasem dyskretnym (Allison, 1984; Yamaguchi, 1991).

## 1. Modele czasu ciągłego

W przypadku modelu czasu ciągłego każdą z badanych  $m$  jednostek ( $j = 1, 2, \dots, m$ ) obserwuje się, począwszy od momentu czasu  $t = 0$  aż do wystąpienia zdarzenia. Badanie może się zakończyć, zanim wystąpi zdarzenie lub gdy jednostka znika z pola obserwacji. W takich przypadkach obserwację uznaje się za cenzurowaną.

Funkcję hazardu definiuje się w następujący sposób (Allison, 1982):

$$h(t) = \lim_{\Delta t \rightarrow 0} P(t \leq T < t + \Delta t | T \geq t) / \Delta t \quad (1)$$

Wartości  $h(t)$  nie interpretuje się jako prawdopodobieństwa, ponieważ może być większa od 1 (Allison, 1982). Najczęściej przyjmuje się, że jest to intensywność zajścia zdarzenia w momencie czasu. Można pokazać, że:

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (2)$$

gdzie:

$f(t)$  – gęstość rozkładu zmiennej losowej  $T$ ,

$F(t)$  – dystrybuanta zmiennej losowej  $T$ .

Następnym krokiem jest wyznaczenie postaci analitycznej funkcji hazardu. Najczęściej stosowaną formą jest model proporcjonalnego hazardu:

$$h(t, x) = \exp\left(\alpha_0(t) + \sum_{i=1}^n \alpha_i x_i\right) \quad (3)$$

gdzie:

$x$  – wektor zmiennych niezależnych  $x_i$ ,

$n$  – liczba zmiennych niezależnych.

Postać funkcji  $\alpha_0(t)$  w modelu (3) implikowana jest przez rozkład zmiennej losowej  $T$ :

- jeżeli  $\alpha_0(t) = \alpha$ , to  $T$  ma rozkład wykładniczy,
- jeżeli  $\alpha_0(t) = \alpha_0 + \alpha_1 \ln t$ , to  $T$  ma rozkład Weibulla,
- jeżeli  $\alpha_0(t) = \alpha_0 + \alpha_1 t$ , to  $T$  ma rozkład Gompertza.

Model (3) można zapisać w postaci zlogarytmowanej:

$$\ln(h(t, x)) = \alpha_0(t) + \sum_{i=1}^n \alpha_i x_i \quad (4)$$

W przypadku, gdy  $\alpha_0(t)$  jest niewyspecyfikowaną funkcją czasu, model (3) ma postać:

$$h(t, x) = h_0(t) \exp\left(\sum_{i=1}^n \alpha_i x_i\right) \quad (5)$$

i nosi nazwę modelu proporcjonalnego hazardu Coxa. Funkcja  $h_0(t)$  w tym modelu nazywana jest hazardem bazowym. Nazwa *proporcjonalny* wynika z założenia stałości ilorazów hazardu w każdym punkcie czasu.

## 2. Modele czasu dyskretnego

W przypadku czasu dyskretnego obserwacja przebiega podobnie jak dla czasu ciągłego. Każdą z badanych  $n$  jednostek ( $j = 1, 2, \dots, m$ ) zaczyna się obserwować w momencie czasu  $t = 0$ . Jednak czas obserwacji może przyjmować tylko wartości naturalne, aż do momentu końcowego  $t_k$  ( $t = 1, 2, \dots, t_k$ ). W praktyce gromadzone dane mają charakter dyskretny. Jeżeli częstość występowania zdarzeń jest duża, najczęściej ignoruje się nieciągłość i czas traktuje się jako zmienną ciągłą. Pojawia się problem, gdy odstęp czasu między kolejnymi zdarzeniami jest stosunkowo duży. Istnieją dwa główne podejścia do tego problemu. Najprostszym jest traktowanie czasu tak, jakby był naprawdę dyskretny (Myers, Hankey, Mantel, 1973; Byar, Mantel, 1975; Brown, 1975; Mantel, Hankey, 1978). Alternatywne jest, aby najpierw zastosować model ciągły, zwykle model proporcjonalnego hazardu (3), w celu uzyskania estymatorów, które są odpowiednie dla danych pogrupowanych w odstępach czasu (Holford, 1976, 1980; Thompson, 1977; Prentice, Gloeckler, 1978). Niezależnie od wyboru kierunku wyniki są bardzo podobne. W przypadku modelu dyskretnego definiuje się prawdopodobieństwo warunkowe zajścia zdarzenia w chwili  $T$  pod warunkiem, że nie zaszło ono do momentu  $T$ :

$$P_t = [t_t = T | t_t \geq T, x_t] \quad (6)$$

Najbardziej popularną postacią analityczną modelu (Cox, 1972; Myers, Hankey, Mantel, 1973; Byar, Mantel, 1975; Brown, 1975; Thompson, 1977; Mantel, Hankey, 1978) jest funkcja regresji logistycznej<sup>1</sup>:

$$P_i = 1 / \left[ 1 + \exp \left( -\beta_i - \sum_{i=1}^n \beta_i x_{ii} \right) \right] \quad (7)$$

Wzór (7) można przekształcić do postaci:

$$P_i / (1 - P_i) = \exp \left( \beta_i + \sum_{i=1}^n \beta_i x_{ii} \right) \quad (8)$$

która po zlogarytmowaniu przyjmuje postać logitu:

$$\text{logit}(P_i) = \ln [P_i / (1 - P_i)] = \beta_i + \sum_{i=1}^n \beta_i x_{ii} \quad (9)$$

Innym sposobem wyznaczenia hazardu w czasie dyskretnym jest wykorzystanie modelu hazardu empirycznego dla danych pogrupowanych w przedziały czasu:

$$h_i = \frac{d_i}{n_i} \quad (10)$$

gdzie:

$d_i$  – liczba zdarzeń zachodzących w  $i$ -tym przedziale,

$n_i$  – liczba jednostek objętych obserwacją w  $i$ -tym przedziale.

Należy tutaj zaznaczyć, że hazard wyznaczony na podstawie wzoru (10) nie przyjmuje wartości większych od 1, gdyż  $d_i \leq n_i$ .

### 3. Modele hazardu i model logitowy – podobieństwa i różnice

Podobieństwa między modelem proporcjonalnego hazardu Coxa a modelem regresji logistycznej dotyczą przede wszystkim postaci analitycznej obu modeli. Gdy oba modele są przekształcone do postaci (5) i (8), widoczne są analogie między parametrami  $\alpha_i$  i  $\beta_i$ . Analogie te są związane również z ich interpretacją, analizie podlegają bowiem nie same parametry, ale ich postać wykładnicza. W przypadku modelu regresji Coxa wyznacza się ilorazy hazardu (hazard względny, *hazard ratio*) określone wzorem  $HR = \exp \alpha_i$ . Natomiast w modelu regresji logistycznej analizuje się ilorazy szans (szansa względna, *odds ratio*)  $OR = \exp \beta_i$ . Inna jest jednak interpretacja tych

<sup>1</sup> Innym modelem stosowanym w przypadku czasu dyskretnego jest (Holford, 1976; Prentice, Gloeckler, 1978)  $P_i = 1 - \exp \left( -\exp \left( \beta_i + \sum_{i=1}^n \beta_i x_{ii} \right) \right)$ , który po przekształceniu polegającym na podwójnym zlogarytmowaniu przyjmuje postać (nazywaną funkcją log-log):  $\ln[-\ln(1 - P_i)] = \beta_i + \sum_{i=1}^n \beta_i x_{ii}$ .

ilorazów. Pierwszy z nich określa intensywność względną zajścia zdarzenia w chwili  $t$  (czas ciągły). Natomiast drugi pozwala na obliczenie szansy<sup>2</sup> względnej zajścia zdarzenia w momencie  $T$  (czas dyskretny) pod warunkiem, że nie zaszło ono do momentu  $T$ .

Tabela 1. Liczebność grup dla czasu obserwacji  $T = 1$

	Brak zdarzenia (0)	Zdarzenie (1)	Razem
Grupa 1	$l_{10}$	$l_{11}$	$l_{10} + l_{11}$
Grupa 2	$l_{20}$	$l_{21}$	$l_{20} + l_{21}$

Źródło: opracowanie własne.

Związek między modelem hazardu empirycznego a modelem logitowym jest widoczny w przypadku wartości hazardu i szansy wyznaczonej dla określonej grupy. W tabeli 1 podano teoretyczne liczebności dwóch grup 1 i 2 w momencie obserwacji  $T = 1$ . Zaznaczono, dla ilu jednostek zaszło i dla ilu nie zaszło zdarzenie. Hazard dla grupy 1 wyznaczony ze wzoru (10) jest równy:

$${}_1h_1 = i_{11} / (l_{10} + l_{11}) \quad (11)$$

Natomiast szansa  ${}_1s_1$  zajścia zdarzenia dla grupy 1 ma wartość:

$${}_1s_1 = p / (1 - p) = i_{11} / l_{10} \quad (12)$$

Po przekształceniu wzorów (11) i (12) otrzymuje się związek:

$$1 / {}_1h_1 = 1 / {}_1s_1 + 1 \quad (13)$$

#### 4. Metodyka badania i wykorzystane dane statystyczne

Przeprowadzone badanie polegało na wyznaczeniu intensywności względnej wychodzenia z bezrobocia na podstawie modeli (5), (10) i szansy względnej podjęcia zatrudnienia na podstawie modelu (9) w zależności od płci, wieku, wykształcenia i stażu pracy osoby bezrobotnej. Analizę przeprowadzono w trzech etapach. Pierwszy etap polegał na oszacowaniu parametrów modelu proporcjonalnego hazardu Coxa, drugi – na estymacji parametrów modelu regresji logistycznej, trzeci – na wyznaczeniu ilorazów hazardu na podstawie modelu hazardu empirycznego.

<sup>2</sup> W regresji logistycznej wyrażenie  $p/(1-p)$ , gdzie  $p$  jest prawdopodobieństwem, jest nazywane szansą.

Badanie przeprowadzono na podstawie anonimowych danych indywidualnych 21 398 osób bezrobotnych zarejestrowanych w Powiatowym Urzędzie Pracy w Szczecinie w 2012 roku. Obserwację prowadzono przez okres 24 miesięcy, czyli do końca 2013 roku. Dane gromadzone w systemie Syriusz stosowanym przez urzędy pracy zawierają dokładną datę rejestracji i wyrejestrowania, co pozwoliło na wyznaczenie indywidualnego czasu zarejestrowania w urzędzie każdej badanej jednostki. W przypadku, gdy obserwacja zakończyła się wyrejestrowaniem z powodu podjęcia szeroko rozumianego zatrudnienia: podjęcie pracy, prac interwencyjnych, prac społecznie użytecznych, podjęcie działalności gospodarczej, to uznano ją za pełną. Dane przyjęto za cenzurowane<sup>3</sup>, jeżeli wyrejestrowanie nie nastąpiło do końca 2013 roku lub nastąpiło przed końcem 2013 roku z przyczyny innej niż podjęcie zatrudnienia. Wyodrębniono grupy osób bezrobotnych ze względu na ich płeć, wiek, wykształcenie i zadeklarowany staż pracy (tabela 2).

Tabela 2. Liczebność grup obserwowanych osób bezrobotnych

Cecha		Obserwacje pełne	Obserwacje cenzurowane	Razem	
1		2	3	4	
Razem		8559	12 839	21398	100%
Płeć	mężczyźni ( $M$ )	4214	7316	11530	53,9%
	kobiety ( $K$ )	4345	5523	9868	46,1%
Wiek	18–25 ( $W_1$ )	703	1237	1940	9,1%
	25–35 ( $W_2$ )	3188	4029	7217	33,7%
	35–45 ( $W_3$ )	1938	2845	4783	22,4%
	45–55 ( $W_4$ )	1780	2634	4414	20,6%
	55–60 ( $W_5$ )	803	1667	2470	11,5%
	60–65 ( $W_6$ )	147	427	574	2,7%
Wykształcenie	co najwyżej gimnazjalne ( $S_1$ )	1303	3626	4929	23,0%
	zasadnicze zawodowe ( $S_2$ )	2009	3468	5477	25,6%
	średnie ogólnokształcące ( $S_3$ )	981	1371	2352	11,0%
	średnie zawodowe, pomaturalne/policealne ( $S_4$ )	1776	2418	4194	19,6%
	wyższe (w tym licencjat) ( $S_5$ )	2490	1956	4446	20,8%

<sup>3</sup> Szerzej na temat cenzurowania w: Frątczak, Gach-Ciepiela, Babiker (2005), s. 47–54.

		1	2	3	4	
Staż pracy	brak stażu ( $Z_1$ )		1009	2819	3828	17,9%
	0-1 ( $Z_2$ )		1029	2086	3115	14,6%
	1-5 ( $Z_3$ )		2114	2910	5024	23,5%
	5-10 ( $Z_4$ )		1520	1614	3134	14,6%
	10-20 ( $Z_5$ )		1196	1299	2495	11,7%
	20-30 ( $Z_6$ )		1294	1490	2784	13,0%
	powyżej 30 lat ( $Z_7$ )		397	621	1018	4,7%

Źródło: opracowanie własne.

W modelu regresji Coxa czas przyjęto jako zmienną ciągłą wyrażoną w miesiącach<sup>4</sup>. Parametry modelu regresji logistycznej i hazard empiryczny oszacowano dla czasu  $T = 2$  (lata).

## 5. Analiza intensywności i prawdopodobieństwa wychodzenia z bezrobocia

Zbadano zależność intensywności wychodzenia z bezrobocia i prawdopodobieństwo znalezienia pracy przez osoby bezrobotne zarejestrowane w PUP w Szczecinie w roku 2012 w zależności od płci ( $P$ ), wieku ( $W_i$ ), wykształcenia ( $S_i$ ) i stażu pracy ( $Z_i$ ). Jako grupy referencyjne przyjęto: mężczyzn ( $M$ ), osoby w wieku 18–24 ( $W_1$ ), osoby z wykształceniem co najwyżej gimnazjalnym ( $S_1$ ), bezrobotnych nieposiadających doświadczenia zawodowego ( $Z_1$ ).

Najpierw wyznaczono parametry modelu proporcjonalnego hazardu Coxa<sup>5</sup> (5). Otrzymano następujące modele:

$$h(t, x) = h_o(t) \exp(0,1166 \cdot P) \quad (14)$$

[0,0216]

$$h(t, x) = h_o(t) \exp\left(0,0020 \cdot W_2 - 0,2199 \cdot W_3 - 0,3037 \cdot W_4 - 0,5837 \cdot W_5 - 0,8946 \cdot W_6\right) \quad (15)$$

[0,0417]      [0,0441]      [0,0447]      [0,0518]      [0,0908]

$$h(t, x) = h_o(t) \exp\left(0,3622 \cdot S_2 + 0,6021 \cdot S_3 + 0,5340 \cdot S_4 + 0,9728 \cdot S_5\right) \quad (16)$$

[0,0356]      [0,0423]      [0,0365]      [0,0342]

<sup>4</sup> Wybór jednostki czasu nie ma wpływu na wartość hazardu wyznaczonego na podstawie modelu proporcjonalnego hazardu Coxa.

<sup>5</sup> Szerzej na temat zastosowania modeli proporcjonalnego i nieproporcjonalnego hazardu Coxa w badaniu bezrobocia w pracach: Bieszk-Stolorz, Markowicz (2012), Bieszk-Stolorz (2013).

$$h(t, x) = h_0(t) \exp \left( \begin{array}{l} 0,2361 \cdot Z_2 + 0,4690 \cdot Z_3 + 0,5297 \cdot Z_4 + 0,4704 \cdot Z_5 + \\ [0,0443] \quad [0,0383] \quad [0,0406] \quad [0,0428] \\ + 0,3585 \cdot Z_6 + 0,1687 \cdot Z_7 \\ [0,0420] \quad [0,0593] \end{array} \right) \quad (17)$$

W wyniku estymacji parametrów modeli logitowych (8) otrzymano następujące wyniki:

$$\text{logit}(P) = -0,5517 + 0,3118 \cdot P \quad (18)$$

[0,0193]      [0,0280]

$$\text{logit}(P) = -0,5651 + 0,3310 \cdot W_2 + 0,1812 \cdot W_3 + 0,1732 \cdot W_4 - 0,1653 \cdot W_5 - 0,5013 \cdot W_6 \quad (19)$$

[0,0472]      [0,0528]      [0,0557]      [0,0563]      [0,0638]      [0,1067]

$$\text{logit}(P) = -1,0235 + 0,4775 \cdot S_2 + 0,6887 \cdot S_3 + 0,7149 \cdot S_4 + 1,2648 \cdot S_5 \quad (20)$$

[0,0323]      [0,0428]      [0,0528]      [0,0449]      [0,0442]

$$\text{logit}(P) = -1,0274 + 0,3208 \cdot Z_2 + 0,7079 \cdot Z_3 + 0,9674 \cdot Z_4 + \quad (21)$$

[0,0367]      [0,0529]      [0,0465]      [0,0512]

$$+ 0,9448 \cdot Z_5 + 0,8864 \cdot Z_6 + 0,5800 \cdot Z_7$$

[0,0543]      [0,0528]      [0,0740]

W modelach (14)–(21) wszystkie parametry są istotne statystycznie z wyjątkiem parametru stojącego przy zmiennej  $W_2$  w modelu (15), co oznacza, że intensywności wychodzenia z bezrobocia przez osoby w wieku 25–35 lat i 18–25 lat (grupa referencyjna) były do siebie zbliżone. Następnie korzystając z modeli regresji Coxa i modeli logistycznych, wyznaczono ilorazy hazardu i ilorazy szans, a także wyznaczono ilorazy hazardu na podstawie modelu (10) (rysunki 1–4). Czarną linią zaznaczono poziom odniesienia, czyli grupę referencyjną.

W analizowanym okresie kobiety intensywniej niż mężczyźni wychodziły z bezrobocia i miały większe szanse na podjęcie zatrudnienia. Wskazują na to ilorazy hazardu i szans wyznaczone na podstawie parametrów wszystkich trzech modeli (rysunek 1).

Oszacowane parametry modelu proporcjonalnego hazardu Coxa prowadzą do wniosku, że największą intensywność wychodzenia z bezrobocia miały osoby młode z grup wieku 18–25 lat i 25–35 lat. Następnie intensywność ta malała dla kolejnych grup wieku. Nieco innych wniosków dostarczają ilorazy hazardu empirycznego. Największa intensywność cechowała osoby w wieku 25–35 lat, następnie osoby w wieku 35–45 lat i 45–55 lat. Podobnie było z prawdopodobieństwem podjęcia zatrudnienia (rysunek 2).

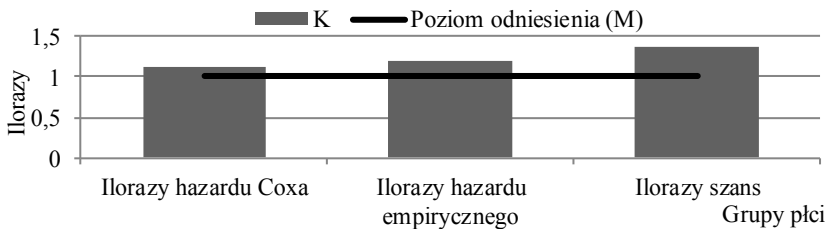
Ilorazy hazardu i szans przedstawione na rysunku 3 wskazują na to, że najintensywniej wychodziły z bezrobocia osoby z wykształceniem wyższym. Miały one również największe szanse na podjęcie zatrudnienia. W najgorszej sytuacji były osoby z wykształceniem co najwyżej gimnazjalnym (grupa porównawcza).



Wszystkie trzy modele w podobny sposób ukazują sytuację osób bezrobotnych w zależności od stażu pracy. Najmniej intensywnie wychodziły z bezrobocia i najmniejsze szanse na podjęcie zatrudnienia miały osoby deklarujące brak doświadczenia zawodowego. W najlepszej sytuacji były osoby posiadające staż pracy 5–10-letni, a także 10–20-letni i 20–30-letni (rysunek 4).

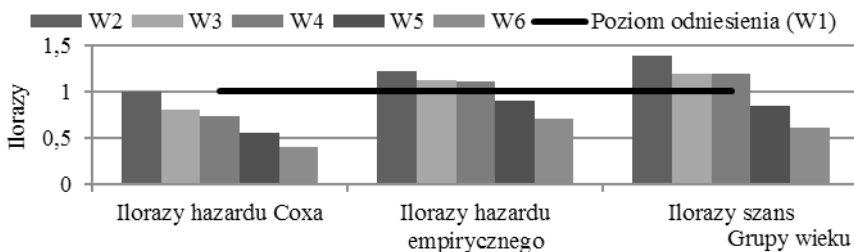
W przypadku ilorazów hazardu empirycznego i ilorazów szans zauważalny jest związek określony wzorem (13). Grupy tych ilorazów mają zbliżony kształt w sensie geometrycznym.

Rysunek 1. Intensywność względna wyjścia z bezrobocia i szansa względna podjęcia zatrudnienia w zależności od płci



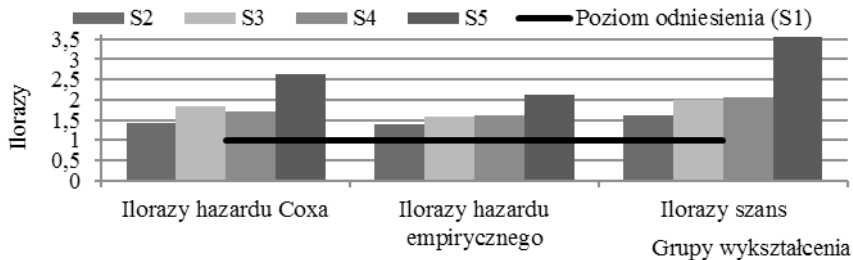
Źródło: opracowanie własne.

Rysunek 2. Intensywność względna wyjścia z bezrobocia i szansa względna podjęcia zatrudnienia w zależności od wieku



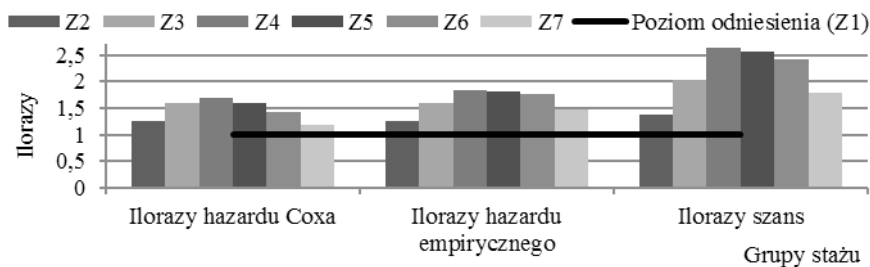
Źródło: opracowanie własne.

Rysunek 3. Intensywność względna wyjścia z bezrobocia i szansa względna podjęcia zatrudnienia w zależności od wykształcenia



Źródło: opracowanie własne.

Rysunek 4. Intensywność względna wyjścia z bezrobocia i szansa względna podjęcia zatrudnienia w zależności od stażu



Źródło: opracowanie własne.

## Podsumowanie

Zastosowane modele hazardu i logitowy dostarczyły podobnych wniosków, jednak wyznaczone wartości ilorazów hazardu i ilorazów szans różniły się w zależności od sposobu uwzględnienia czasu trwania bezrobocia. W przypadku przeprowadzonego badania czas można było potraktować dwojako. Dane z urzędu pracy pozwoliły na wyznaczenie czasu zarejestrowania z dokładnością do jednego dnia, nie było więc przeszkód w potraktowaniu go jako zmiennej ciągłej i zastosowaniu do niego modelu hazardu Coxa. W przypadku, gdy dokona się pogrupowania obserwacji w dłuższe przedziały (miesiące, kwartały itd.), wygodny w stosowaniu jest dyskretny model hazardu empirycznego. W niniejszym badaniu dane pogrupowano

w okres dwuletni. Oba modele hazardu pozwoliły uwzględnić dane cenzurowane. Model logitowy jako przykład modelu z czasem dyskretnym umożliwił oszacowanie szansy zajścia zdarzenia na koniec okresu obserwacji, w tym przypadku po dwóch latach. Co ważne, ilorazy hazardu i szans wyznaczone za pomocą tych modeli bez względu na sposób uwzględnia czasu (ciągły lub dyskretny) są stałe w czasie trwania obserwacji. Ilorazy otrzymane na podstawie modelu regresji logistycznej i hazardu empirycznego są do siebie zbliżone. Wynika to z zależności opisanej wzorem (13). Natomiast zauważalne są niewielkie różnice w ilorazach wyznaczonych na podstawie modelu regresji Coxa w porównaniu z pozostałymi modelami.

## Literatura

- Allison, P.D. (1982). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology*, 13, 61–98.
- Allison, P.D. (1984). *Event History Analysis: Regression for Longitudinal Event Data*. Beverly Hills CA: SAGE Publications.
- Bieszk-Stolorz, B. (2013). *Analiza historii zdarzeń w badaniu bezrobocia*. Szczecin: Volumina.pl Daniel Krzanowski.
- Bieszk-Stolorz, B., Markowicz, I. (2012). *Modele regresji Coxa w analizie bezrobocia*. Warszawa: CeDeWu.
- Brown, C.C. (1975). On the Use of Indicator Variables for Studying the Time Dependence of Parameters in a Response-Time Model. *Biometrics*, 31, 863–872.
- Byar, D.P., Mantel, N. (1975). Some Interrelationships among the Regression Coefficient Estimates Arising in a Class of Models Appropriate to Response-Time Data. *Biometrics*, 31, 943–947.
- Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B*, 34, 187–220.
- Frączzak, E., Gach-Ciepiela, U., Babiker, H. (2005). *Analiza historii zdarzeń. Elementy teorii, wybrane przykłady zastosowań*. Warszawa: Wyd. SGH.
- Holford, T.R. (1976). Life Tables with Concomitant Information. *Biometrics*, 32, 587–597.
- Holford, T.R. (1980). The Analysis of Rates and of Survivorship Using Log-Linear Models. *Biometrics*, 36, 299–305.
- Mantel, N., Hankey, B. (1978). A Logistic Regression Analysis of Response-Time Data Where the Hazard Function is Time Dependent. *Communications in Statistics – Theory and Methods*, A7, 333–347.
- Myers, M.H., Hankey, B.F., Mantel, N. (1973). A Logistic-Exponential Model for Use with Response-Time Data Involving Regressor Variables. *Biometrics*, 29, 257–269.
- Prentice, R.L., Gloeckler, L.A. (1978). Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data. *Biometrics*, 34, 57–67.

- Thompson, W.A., JR. (1977). On the Treatment of Grouped Observations in Life Studies. *Biometrics*, 33, 463–470.
- Yamaguchi, K. (1991). *Event History Analysis*. Newbury Park CA: SAGE Publications.

## HAZARD MODEL VERSUS LOGIT MODEL

### Abstract

The aim of the article is the comparison of two groups of models used in the event history analysis. The first one encompasses continuous-time models which describe event intensity (hazard) at any moment of time. The Cox proportional hazard model are used in the study. The second group consists of discrete-time models. The analysis is based on the logistic regression model (the probability of an event to occur at the discrete time) and the empirical hazard model (for the grouped data). The research material includes individual data of the unemployed beneficiaries registered in 2012 by the Powiat Labour Office in Szczecin as observed by the end of 2013. The authors determine the relative unemployment exit intensity and the relative employment odds by the unemployed person's gender, age, education and employment history.

*Translated by Anita Zdrojewska*

**Keywords:** Cox's regression model, the empirical hazard model, logistic regression, unemployment

**JEL Codes:** C51, J64