

Joanna Małgorzata Landmesser

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

WYKORZYSTANIE MODELI RYZYKA KONKURENCYJNEGO W CELU SZACOWANIA CZASU TRWANIA AKTYWNOŚCI ZAWODOWEJ

1. Wstęp

W pracy dokonano szacowania długości czasu trwania zatrudnienia w Polsce. Wielkość tę modelowano, wykorzystując metody analizy czasu przetrwania. Estymacja semi-parametrycznych modeli hazardu pozwala na wyznaczenie bezpośredniego ryzyka tego, że zatrudniony przestanie wykonywać swą pracę w danym przedziale czasowym. Modele hazardu uwzględniają nie tylko aktualny czas trwania zjawiska jako istotną determinantę prawdopodobieństwa zajścia zdarzenia, stąd też możliwość określenia wpływu takich wielkości, jak: płeć, wiek, poziom wykształcenia, wielkość miejscowości zamieszkania czy rodzaj zakładu pracy, na indywidualną długość czasu zatrudnienia.

Za pomocą modeli opisuje się warunkowe prawdopodobieństwa przejścia między stanem zatrudnienia a stanem nieaktywności zawodowej (stanem bezrobocia lub stanem bierności zawodowej, np. emerytura, renta). W analizie zastosowanie znajdują dwa typy modelowania: modele ryzyka niekonkurencyjnego (*noncompeting risk models*) oraz modele ryzyka konkurencyjnego (*competing risk models*). W estymacji wykorzystano dane pochodzące z reprezentacyjnego Badania Aktywności Ekonomicznej Ludności Polski (BAEL).

Celem przeprowadzonych badań było zweryfikowanie za pomocą modeli następujących hipotez:

1. Młody wiek obniża prawdopodobieństwo zaprzestania aktualnego zatrudnienia.
2. Wysoki poziom wykształcenia obniża prawdopodobieństwo utraty pracy.
3. Ryzyko utraty pracy jest w wypadku mężczyzny mniejsze niż w przypadku kobiety.
4. Ryzyko utraty pracy w wypadku osoby ze wsi jest niższe niż w przypadku osoby z miasta.

5. Ryzyko utraty pracy w wypadku osoby z firmy prywatnej jest wyższe, niż w przypadku osoby z firmy o innej formie własności.
6. Osoby starsze cechują się większym ryzykiem stania się biernymi zawodowo niż bezrobotnymi.
7. Kobiety w porównaniu z mężczyznami wykazują wyższe ryzyko stania się biernymi zawodowo.

2. Opis metody analizy

Celem analizy czasu przetrwania jest wyrażenie zaobserwowanego rozkładu długości czasu trwania zjawiska za pomocą tzw. funkcji hazardu. Modele hazardu są budowane tam, gdzie celem jest prognozowanie momentu, w którym dojdzie do pewnego wydarzenia. Prezentacja rozkładu indywidualnego czasu wystąpienia zdarzenia T (w naszym przypadku zdarzeniem tym jest zaprzestanie zatrudnienia) możliwa jest za pomocą funkcji rozkładu, gęstości i przeżycia [Kalbfleisch, Prentice 1980; Kiefer 1988; Frątczak, Gach-Ciepiela, Babiker 2005]:

Funkcja rozkładu $F(t)$ dostarcza prawdopodobieństwo tego, że moment wystąpienia zdarzenia leży przed wybranym punktem czasu t albo dokładnie w tym punkcie; $F(t) = \Pr[T \leq t]$. Symbolem $f(t)$ oznaczamy funkcję gęstości. Prawdopodobieństwo przeżycia czasu t wyraża funkcja przeżycia $S(t)$:

$$S(t) = \Pr[T > t] = 1 - F(t). \quad (1)$$

Funkcja ta przyjmuje wartości od 0 do 1; jej przebieg zależy od właściwości obserwowanego procesu, z reguły jest to jednak krzywa monotonicznie malejąca.

Najczęściej stosowanym przedstawieniem rozkładu czasu trwania jest funkcja hazardu $h(t)$. Szacuje ona „prawdopodobieństwo” lub raczej bezpośrednie ryzyko tego, że wydarzenie nastąpi w przedziale czasowym pomiędzy t i $t + dt$, pod warunkiem że do danego momentu zajście jego nie nastąpiło.

$$h(t) = \frac{f(t)}{S(t)} = \lim_{dt \rightarrow 0} \frac{\Pr[t < T < t + dt \mid T > t]}{dt}. \quad (2)$$

3. Modele hazardu

Pośród modeli czasu przeżycia wyróżnia się modele nieparametryczne (np. estymator Kaplana-Meiera), modele parametryczne (modele proporcjonalnego hazardu lub *accelerated failure time*) oraz modele semiparametryczne (np. regresja Coxa).

Estymator Kaplana-Meiera jest estymatorem dla funkcji przeżycia:

$$\hat{S}(t) = \prod_{j|t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right), \quad (3)$$

gdzie n_j wyraża liczbę obiektów, które przeżyły do chwili t_j , natomiast d_j to liczba obiektów „umierających” w chwili t_j .

Najczęściej stosowaną specyfikacją dla modeli regresji opartych na stopach hazardu są modele proporcjonalnego hazardu. W modelach tych funkcja wielowymiarowego wektora zmiennych objaśniających $\mathbf{X}$ (funkcja $g_0(\mathbf{X})$) działa w sposób multiplikatywny na tzw. hazard bazowy $h_0(t)$. Poszczególne modele z tej klasy wykazują istotne różnice, jeśli chodzi o założenia dotyczące rozkładu hazardu bazowego. Warunkową funkcję stopy modelujemy:

$$h(t|\mathbf{X}) = h_0(t)g_0(\mathbf{X}) = h_0(t)\exp(\mathbf{X}\beta). \quad (4)$$

W modelach semiparametrycznych funkcja hazardu bazowego nie jest ani parametryzowana, ani szacowana. Może ona mieć dowolny przebieg. Rezygnacja z przyjęcia konkretnej funkcji rozkładu pozwala uniknąć wyciągania błędnych wniosków, co ma miejsce przy nieodpowiednim modelowaniu hazardu bazowego.

U podstaw modelu Coxa [Cox 1972; Cox, Oakes 1984] leży teza o proporcjonalnym hazardzie. Rozpatrzmy następujący iloraz będący stosunkiem funkcji hazardu dla dwóch jednostek i oraz j z różnymi wektorami zmiennych objaśniających X_i i X_j :

$$HR = \frac{h_0(t)\exp(X_i\beta)}{h_0(t)\exp(X_j\beta)} = \exp[(X_i - X_j)]. \quad (5)$$

Założenie proporcjonalności mówi o tym, że dla dwóch jednostek o określonych wartościach zmiennych objaśniających stosunek estymowanych wartości hazardu w czasie będzie stały (stąd nazwa metody: model proporcjonalnego hazardu). Iloraz HR nie zależy od czasu t , a stopa hazardu j -tej jednostki jest w każdym momencie krotnością stopy hazardu i -tej jednostki. Szacowanie modeli hazardu Coxa opiera się na maksymalizowaniu częściowej wiarygodności. Podczas tej maksymalizacji ignorowany jest pierwszy czynnik, czyli hazard bazowy, we wzorze na stopę hazardu, a drugi czynnik jest traktowany tak, jakby chodziło o zwyczajną funkcję wiarygodności.

Model hazardu Coxa jest używany głównie w celu osiągnięcia wglądu w proces ryzyka oraz otrzymania odpowiedzi na pytanie, jak to ryzyko się zmienia wraz ze zmianami zmiennych objaśniających. Gdy zakończenie pewnego procesu następuje poprzez przejście do jednego z kilku niezależnych od siebie stanów, zjawisko takie możemy modelować za pomocą modelu ryzyka konkurencyjnego [Narendranathan, Stewart 1993]. Niech T_a i T_b oznaczają momenty zajścia zdarzeń przejścia do dwóch stanów a oraz b . Wtedy moment zakończenia procesu wyjściowego definiujemy jako $T = \min\{T_a, T_b\}$. Pomysł analizowania ryzyka konkurencyjnego ilustruje formuła:

$$h_a(t|\mathbf{X}) = h_{0a}(t)g_0(\mathbf{X}) = h_{0a}(t)\exp(\mathbf{X}\beta_a), \quad (6)$$

gdzie $h_{0a}(t)$ jest bazową stopą hazardu dla zdarzenia a . Przyjmując czas wystąpienia zdarzenia innego niż zdarzenie a jako czas cenzurowany, estymujemy osobno ryzyko zajścia zdarzenia a . W ten sposób specyfikacja modeli ryzyka konkurencyjnego następuje poprzez szacowanie poszczególnych modeli proporcjonalnego hazardu.

4. Zakres czasowy badania oraz obiekty badań

Do estymacji modeli hazardu dla długości czasu zatrudnienia wykorzystano dane pochodzące z Badania Aktywności Ekonomicznej Ludności (BAEL) przeprowadzonego w Polsce w czwartym kwartale 2002 r. Ponieważ próba BAEL jest panelem rotacyjnym, aby ustalić, jak długo trwa aktualne zatrudnienie respondenta lub jak długo trwała jego praca w ostatnim miejscu zatrudnienia, należy oprzeć się na zawartych w ankiecie pytaniach retrospektywnych.

Tabela 1. Lista zmiennych objaśniających

Nazwa	Opis zmiennej
Pl	zmienna dychotomiczna dotycząca płci przyjmująca wartość 1, gdy mężczyzna, 0, gdy kobieta
wiek1	zmienna dychotomiczna przyjmująca wartość 1, gdy wiek respondenta w chwili podjęcia pracy wynosił 15-17 lat, 0 w p.p.
wiek2	zmienna dychotomiczna przyjmująca wartość 1, gdy wiek respondenta w chwili podjęcia pracy wynosił 18-24 lat, 0 w p.p.
wiek3	zmienna dychotomiczna przyjmująca wartość 1, gdy wiek respondenta w chwili podjęcia pracy wynosił 25-34 lat, 0 w p.p.
wiek4	zmienna dychotomiczna przyjmująca wartość 1, gdy wiek respondenta w chwili podjęcia pracy wynosił 35-44 lat, 0 w p.p.
wiek5	zmienna dychotomiczna przyjmująca wartość 1, gdy wiek respondenta w chwili podjęcia pracy wynosił 45-59/64 lat, 0 w p.p.
wiek6	zmienna dychotomiczna przyjmująca wartość 1, gdy wiek respondenta w chwili podjęcia pracy wynosił więcej niż 60/65 lat, 0 w p.p.
wyksz1	zmienna dychotomiczna przyjmująca wartość 1, gdy respondent posiada wyższe wykształcenie, 0 w p.p.
wyksz2	zmienna dychotomiczna przyjmująca wartość 1, gdy respondent posiada średnie wykształcenie, 0 w p.p.
wyksz3	zmienna dychotomiczna przyjmująca wartość 1, gdy respondent posiada podstawowe wykształcenie, 0 w p.p.
klm0	zmienna dychotomiczna dotycząca klasy miejscowości zamieszkania respondenta przyjmująca wartość 1, gdy miasto ma powyżej 100 tys. mieszkańców, 0 w p.p.
klm1	zmienna dychotomiczna dotycząca klasy miejscowości zamieszkania przyjmująca wartość 1, gdy miasto ma 50-100 tys. mieszkańców, 0 w p.p.
klm2	zmienna dychotomiczna dotycząca klasy miejscowości zamieszkania przyjmująca wartość 1, gdy miasto ma 20-50 tys. mieszkańców, 0 w p.p.
klm3	zmienna dychotomiczna dotycząca klasy miejscowości zamieszkania przyjmująca wartość 1, gdy miasto ma do 20 tys. mieszkańców, 0 w p.p.
klm9	zmienna dychotomiczna dotycząca klasy miejscowości zamieszkania przyjmująca wartość 1, gdy wieś, 0 w p.p.
włas1	zmienna dychotomiczna przyjmująca wartość 1, gdy respondent pracował w instytucji, której własność jest prywatna, 0 w p.p.
włas2	zmienna dychotomiczna przyjmująca wartość 1, gdy respondent pracował w instytucji, której własność nie jest prywatna, 0 w p.p.

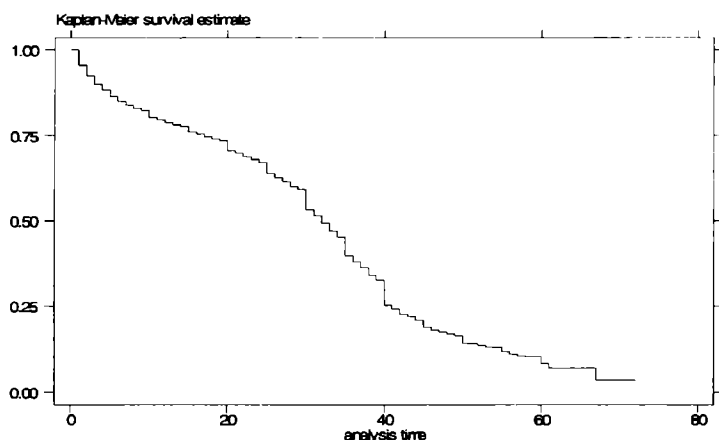
Źródło: obliczenia własne.

Pełną próbę BAEL (47 806 obserwacji) ograniczono w badaniu do podpróby 25 918 aktywnych zawodowo osób w wieku od 15 lat wzwyż, które pracowały w

pełnym lub niepełnym wymiarze czasu pracy (17 969 osób) albo były zatrudnione zanim nie trafiły do grupy osób bezrobotnych (2628) i biernych zawodowo (5321). Uwzględniono osoby, które w trakcie badania wykazały co najmniej jeden epizod oznaczający zatrudnienie (pracowały przynajmniej przez rok). Z powodu konstrukcji badania BAEL z analizy wyłączono dane dotyczące osób, które zakończyły swoje zatrudnienie przed 1 stycznia 1996 r. (cenzurowanie lewostronne). Gdy w rzeczywistości proces zatrudnienia był dłuższy niż ten obserwowany w czasie trwania badania, uznano, że mamy do czynienia z cenzurowaniem prawostronnym. Uszeregowane długości trwania zatrudnienia w ostatnich miejscach pracy w latach dla poszczególnych osób tworzą zmienną „latapracy” (por. [Devine, Kieler 1991; Lancaster 1979]). Średni czas zatrudnienia w jednym miejscu pracy wyniósł 13,33 roku. Jako zmienne objaśniające do modelu zostały włączone cechy przedstawione w tab. 1. Grupa referencyjna, względem której analizowane są wyniki estymacji modeli hazardu, składa się z osób o następujących cechach: kobiety, z wykształceniem podstawowym, w wieku 35-44 lata, mieszkające w miastach powyżej 100 tys. mieszkańców, pracujące w instytucji niebędącej własnością prywatną.

5. Konstrukcja i estymacja modeli hazardu

Korzystając ze wzoru na estymator Kaplana-Meiera, wyznaczono w pierwszym kroku empiryczne prawdopodobieństwo przeżycia, czyli prawdopodobieństwo tego, że zatrudniony będzie pracował w dalszym ciągu w swoim aktualnym miejscu pracy.



Rys. 1. Empiryczne prawdopodobieństwa kontynuacji zatrudnienia
Źródło: obliczenia własne.

Z rysunku 1 wynika, że im więcej lat upłynie od rozpoczęcia aktualnego zatrudnienia, tym mniejsze jest prawdopodobieństwo jego kontynuacji.

Wyniki estymacji zwykłego modelu Coxa ryzyka niekonkurencyjnego prezentuje tab. 2.

Tabela 2. Wyniki estymacji modelu Coxa ryzyka niekonkurencyjnego

Lata pracy	Model ryzyka niekonkurencyjnego					
	regresja Coxa					
	współcz.	błąd std.	wartość p	stopa hazardu	błąd std.	wartość p
pl	-0,286***	0,023	0,000	0,751***	0,017	0,000
wiek1	-1,773***	0,050	0,000	0,170***	0,005	0,000
wiek2	-1,105***	0,035	0,000	0,331***	0,012	0,000
wiek3	-0,645***	0,035	0,000	0,525***	0,019	0,000
wiek5	0,514***	0,042	0,000	1,672***	0,070	0,000
wiek6	0,535***	0,171	0,002	1,708***	0,293	0,002
wyksz1	-0,759***	0,048	0,000	0,468***	0,022	0,000
wyksz2	-0,251***	0,027	0,000	0,778***	0,021	0,000
klm1	0,204***	0,042	0,000	1,226***	0,052	0,000
klm2	0,108***	0,040	0,007	1,114***	0,045	0,007
klm3	0,115***	0,037	0,002	1,121***	0,041	0,002
klm9	-0,304***	0,031	0,000	0,738***	0,023	0,000
włas1	0,927***	0,025	0,000	2,528***	0,063	0,000
Ln L	-69886,27			-69886,7		
LR chi2	4908,52			4908,52		
Liczba obserwacji	25918			25918		

*** – parametr jest istotny na poziomie $\alpha = 0,01$.

Źródło: obliczenia własne.

Oto przykładowe interpretacje ocen parametrów:

- dla oceny parametru $-0,286$ przy zmiennej „pl”: relatywne ryzyko utraty pracy jest w wypadku mężczyzny mniejsze o 24,9% niż w przypadku kobiety ($(\exp(-0,286)-1) 100\% = -24,9\%$),
- dla oceny parametru $-0,645$ przy zmiennej „wiek3”: relatywne ryzyko utraty pracy w wypadku osoby w wieku 25-34 lata jest o 47,53% niższe niż w przypadku osoby w wieku 35-44 lat,
- relatywne ryzyko utraty pracy w wypadku osoby w wieku 45-59/64 lata jest o 67,2% wyższe niż w przypadku osoby w wieku 35-44 lata,
- relatywne ryzyko utraty pracy w wypadku osoby z wyższym wykształceniem jest o 53,19% niższe niż w przypadku osoby z wykształceniem podstawowym,
- relatywne ryzyko utraty pracy w wypadku osoby ze wsi jest o 26,21% niższe niż w przypadku osoby z miasta mającego powyżej 100 tys. mieszkańców,
- relatywne ryzyko utraty pracy w wypadku osoby z firmy prywatnej jest o 152,69% wyższe niż w przypadku osoby z firmy o innej formie własności.

Oceny „dobroci” modelu dokonuje się za pomocą oceny wiarygodności uzyskanych wyników, której miarą jest ujemny dwukrotny logarytm funkcji wiarygodności ($-2\ln L$). Poprawność modelu ocenia się również w tzw. *likelihood-ratio-test*.

Rozważamy sytuację, w której mamy do czynienia z dwoma rodzajami zdarzeń mogących nastąpić po zakończeniu zatrudnienia, są to przejście w stan bezrobocia lub przejście w stan bierności zawodowej (osoby, które nie mają pracy i jej nie poszukują). Oba te zdarzenia są od siebie niezależne, a suma prawdopodobieństw ich zajścia wynosi 1. Ryzyko ich zajścia określa się jako konkurencyjne. Analiza funkcji skojarzonych z każdym z możliwych zdarzeń może być przeprowadzona osobno, z uwzględnieniem tego, że wszystkie obserwacje odnoszące się do zdarzeń innych niż analizowane, powinny być traktowane jako cenzurowane. Specyfikacja modeli ryzyka konkurencyjnego następuje więc poprzez szacowanie poszczególnych modeli proporcjonalnego hazardu. Wyniki ekonometrycznej analizy czasu zajścia zdarzeń „stanie się bezrobotnym” oraz „przejście do grupy biernych zawodowo” przedstawiono w tab.3.

Tabela 3. Wyniki estymacji modelu Coxa ryzyka konkurencyjnego

Model ryzyka konkurencyjnego						
Lata pracy	regresja Coxa					
	bezrobotni			bierni zawodowo		
	stopa hazardu	błąd std.	wartość p	stopa hazardu	błąd std.	wartość p
pl	1,088***	0,043	0,034	0,620***	0,018	0,000
wiek1	0,309***	0,031	0,000	0,114***	0,007	0,000
wiek2	0,629***	0,034	0,000	0,205***	0,009	0,000
wiek3	0,763***	0,043	0,000	0,376***	0,017	0,000
wiek5	0,895	0,069	0,152	2,527***	0,127	0,000
wiek6	0,248***	0,176	0,049	3,155***	0,561	0,000
wyksz1	0,379***	0,035	0,000	0,504***	0,028	0,000
wyksz2	0,774***	0,034	0,000	0,787***	0,026	0,000
klm1	1,245***	0,089	0,002	1,224***	0,064	0,000
klm2	1,062	0,074	0,391	1,155***	0,057	0,003
klm3	1,202***	0,074	0,003	1,074	0,049	0,118
klm9	0,864***	0,046	0,006	0,653***	0,025	0,000
włas1	3,986***	0,170	0,000	1,965***	0,063	0,003
ln L	-24852,88			-44478,22		
LR chi2	1790,32			4228,53		
liczba obserwacji	25 918			25 918		

*** – parametr jest istotny na poziomie $\alpha = 0,01$.

Źródło: obliczenia własne.

Porównując wyniki, znajdujemy różnice we wpływie zmiennych objaśniających na czasy zajścia obu analizowanych zdarzeń. Analizując wartości ocen parametrów, stwierdzamy, iż:

- mężczyźni w porównaniu z kobietami wykazują większe o 8,8% relatywne ryzyko przejścia w stan bezrobocia oraz o 38% niższe ryzyko stania się biernymi zawodowo,

- ryzyko przejścia w stan bezrobocia w wypadku osoby w wieku 25-34 lata jest o 23,7% niższe niż w przypadku osoby w wieku 35-44 lata, natomiast ryzyko stania się biernym zawodowo jest niższe aż o 62,4%,

- ryzyko przejścia w stan bezrobocia w wypadku osoby w wieku 60/65 i więcej lat jest o 75,2% niższe niż w przypadku osoby w wieku 35-44 lata, ale już ryzyko stania się biernym zawodowo jest wyższe o 215,5%,
- ryzyko przejścia w stan bezrobocia w wypadku osoby z wyższym wykształceniem jest o 62,1% niższe niż w przypadku osoby z wykształceniem podstawowym; również ryzyko stania się biernym zawodowo jest niższe o 49,6%,
- ryzyko bycia bezrobotnym w wypadku osoby z firmy prywatnej jest o 298,6% wyższe niż w przypadku osoby z firmy o innej formie własności; również ryzyko stania się biernym zawodowo jest w tym wypadku wyższe o 96,5%.

6. Podsumowanie

Celem dokonanej w pracy estymacji modeli ryzyka było opisanie warunkowego prawdopodobieństwa przejścia między stanem zatrudnienia a stanem nieaktywności zawodowej oraz określenie wpływu wybranych wielkości na indywidualny czas trwania zatrudnienia. Wyniki wskazują na to, że młody wiek i wysoki poziom wykształcenia istotnie obniżają stopę hazardu określającą prawdopodobieństwo zaprzestania aktualnego zatrudnienia. Osoby starsze cechują się większym ryzykiem stania się biernymi zawodowo niż ryzykiem pozostania zarejestrowanymi jako bezrobotne. Kobiety wykazują wyższe ryzyko stania się biernymi zawodowo. Ryzyko bycia bezrobotnym w wypadku osoby z firmy prywatnej jest wyższe, niż w przypadku osoby z firmy o innej formie własności; również ryzyko stania się biernym zawodowo jest w tym wypadku wyższe. W świetle badań można stwierdzić, że modele hazardu mogą być odpowiednimi narzędziami do analizy czasu zatrudnienia.

Literatura

- Cox D.R. (1972), *Regression Models and Life Tables (with Discussion)*, „Journal of the Royal Statistical Society”, Series B 34, s. 187-220.
- Cox D.R., Oakes D. (1984), *Analysis of Survival Data*, Chapman and Hall, London.
- Devine T., Kiefer N. (1991), *Empirical Labor Economics*, New York, Oxford.
- Frątczak E., Gach-Ciepiela U., Babiker H. (2005), *Analiza historii zdarzeń. Elementy teorii, wybrane przykłady zastosowań*, Wydawnictwo SGH, Warszawa.
- Kalbfleisch J., Prentice R. (1980), *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Kiefer N. (1988), *Economic Duration Data and Hazard Functions*, „Journal of Economic Literature” 26, s. 646-679.
- Lancaster T. (1979), *Econometric Methods for the Duration of Unemployment*, „Econometrica” 47, s. 939-956.

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.

APPLICATION OF COMPETING RISK MODELS TO ESTIMATION OF EMPLOYMENT DURATION

Summary

The aim of this paper is to model the duration of employment period in Poland using the survival analysis. Estimating the noncompeting risk model and competing risk model we calculate the conditional probabilities of transition from employment state to the unemployment state (exits to unemployment and job search or exits to out of the labour force, e.g. retirement, disability). We estimate hazard rate models to assess the effect of gender, age, education level and place of residence on the individual's employment duration. For the estimation, we use data from the Labour Force Survey in Poland (Badanie Aktywności Ekonomicznej Ludności Polski BAEL).